



УНИВЕРЗИТЕТ „СВ. КИРИЛ И МЕТОДИЈ“ ВО СКОПЈЕ
ФАКУЛТЕТ ЗА ЕЛЕКТРОТЕХНИКА И ИНФОРМАЦИСКИ ТЕХНОЛОГИИ
ИНСТИТУТ ЗА АВТОМАТИКА И СИСТЕМСКО ИНЖЕНЕРСТВО



м-р Стефан Златинов

**ПРОАКТИВНА БЕЗБЕДНОСТ НА ПЕШАЦИТЕ
ПРИ АВТОНОМНО УПРАВУВАЊЕ НА ВОЗИЛА
ПРЕКУ УЧЕЊЕ СО ПОТТИКНУВАЊЕ
БАЗИРАНО НА МОДЕЛ**

Докторски труд

Скопје, 2026

Кандидат: **м-р Стефан Златинов**
магистер по електротехника и информациски технологии
број на индекс: 7/2018

Ментор: **проф. д-р Миле Станковски**
Универзитет „Св. Кирил и Методиј“ во Скопје
Факултет за електротехника и информациски технологии
Институт за автоматика и системско инженерство

Комисија:

Претседател: **проф. д-р Весна Ојлеска Латкоска**
Универзитет „Св. Кирил и Методиј“ во Скопје
Факултет за електротехника и информациски технологии
Институт за автоматика и системско инженерство

Ментор: **проф. д-р Миле Станковски**
Универзитет „Св. Кирил и Методиј“ во Скопје
Факултет за електротехника и информациски технологии
Институт за автоматика и системско инженерство

Член: **вонр. проф. д-р Горјан Наџински**
Универзитет „Св. Кирил и Методиј“ во Скопје
Факултет за електротехника и информациски технологии
Институт за автоматика и системско инженерство

Член: **проф. д-р Зоран Ивановски**
Универзитет „Св. Кирил и Методиј“ во Скопје
Факултет за електротехника и информациски технологии
Институт за електроника

Член: **проф. д-р Даме Димитровски**
Универзитет „Св. Кирил и Методиј“ во Скопје
Машински факултет
Институт за термичко инженерство

Научна област: Електротехника и информациски технологии

Лектор: **Јулијана Јосифовска**

Датум на одбрана: _____

Апстракт

Безбедноста на пешаците е најтешкиот преостанат проблем во автономното градско возење. Модуларните системи ја разделуваат задачата на перцепција, предвидување и планирање, но секој модул се оптимизира за посредна цел, а не за самата безбедна интеракција. Токму реткото конфликтно пешачко однесување е местото каде овие каскадни цели најмногу потфрлаат. Пристапите со учење од крај до крај и со поттикнувано учење се предложени како алтернатива, но нивните внатрешни процеси на одлучување се нејасни, па тешко е да се утврди дали научената полиса е навистина проактивна или само реагира во последен момент. Оваа теза истражува дали агент со поттикнувано учење со модел развива проактивно однесување кон пешаци како произлезено својство на планирање преку претставувачки модел, кога е трениран со минимална награда за возачка компетентност без експлицитни безбедносни членови.

Истражувањето е изградено врз агент DREAMERV3 трениран во CARLA врз неконфликтни распределби на пешаци и врз *Библиотеката на конфликтни пешаци*, воведена како еден од оригиналните придонеси на овој труд. Таа е контролирана група сценарија со три конфликтни пешачки архетипови: *Најрајник* (доцно преминување преку рутата), *Нерешителен преминувач* (замрзнување–продолжување сред преминување) и *Пројеситант* (упорно заземање на лентата). Секој архетип испитува квалитативно различен режим на безбедносен неуспех. Тренираниот агент се евалуира на библиотеката со замрзнати тежини, без безбедносно обликувана награда во кој било стадиум од тренингот. Втор методолошки придонес што го овозможува ова истражување: проекциите од претставувачкиот модел произведени при секој реален чекор на одлучување се претвораат во интерпретабилна дијагностичка површина преку детерминистичка таксономија од шест класи на замислено однесување, преку метрики за проактивност на ниво на епизода (време на претходење на конфликтната свесност, време на претходење на намерата за избегнување, стапка на небезбедна упорност, времиња на престој за стабилност на намерата и индикатор за дефанзивна блокада во чекање) и преку поле на имагинација условено од состојба, кое ја проектира замислената намера во реализираниот простор растојание-до-пешак/брзина-на-возилото.

Основниот агент постигнува висока стапка на завршување рута во отсуство на конфликтни пешаци и не е нарушен од пасивни случајни минувачи, со што се утврдува дека последователните неуспеси не се општи возачки неуспеси. Под конфликтна евалуација, судирите со пешак се значајни и квалитативно се разликуваат меѓу архетиповите, при што типичните епизоди со средба достигнуваат вредности на ТТС блиски до судир. Агентот ја задржува физичката способност силно да кочи штом конфликтот стане непосреден, па неуспехот не е недостиг на управувачки авторитет. Доказите од имагинацијата го лоцираат неуспехот во претставувачкиот модел: замислените иднини навистина го достигнуваат пешачкиот регион пред најблиско-

то приближување, но свесноста не се претвора во посветеност. Голем дел од проекциите што го достигнуваат конфликтот остануваат право низ конфликтот, замислените планови се краткотрајни и нестабилни, а значаен дел од блиските подвижни состојби создаваат замислени иднини без давање првенство и без свртување. Избегнувањето што се појавува е силно доминирано од давање првенство, додека комбинираните планови со кочење и странично избегнување се ретки. Резултатот за Протестантот е најдијагностички: високата стапка на судир со локално ограничена блокада ја исклучува неизвесноста на пешачкото движење како доминантно објаснување и го лоцира неуспехот во научено априорно верување на претставувачкиот модел дека пешаците се расчистуваат од рутата. Посебен режим на неуспех се појавува кај Протестантот: дефанзивен застој, во кој агентот станува непосредно безбеден со запирање, но во својот краток хоризонт на имагинација не може да процени дали чекањето на крај ќе ја расчисти пречката.

Тезата заклучува дека проактивната безбедност на пешаци е архитектонски достапна, но не произлегува автоматски во поттикнувано учење со модел под минимална награда за возачка компетентност. Теснецот кај агентот изучуван тука не е ниту перцепцијата ниту управувањето, туку квалитетот на имагинацијата на претставувачкиот модел под распределбено поместување. Надвор од конкретниот наод, работата придонесува пренослив дијагностички протокол за секој агент од класата Dreamer на кој било бенчмарк со аналогна судирна точка.

Клучни зборови: поттикнувано учење со модел, DreamerV3, претставувачки модели, автономно возење, безбедност на пешаци, CARLA.

Abstract

Pedestrian safety is the hardest residual problem in autonomous urban driving. Modular pipelines decompose the task into perception, prediction, and planning, but each module is optimised for an intermediate objective rather than for the safe interaction itself, and the long tail of non-compliant pedestrian behaviour is precisely where these cascading objectives fail most consequentially. End-to-end and reinforcement learning approaches have been proposed as alternatives, but their internal decision processes are opaque, making it difficult to determine whether a learned policy is genuinely proactive or merely reactive at the last moment. This thesis investigates whether a model-based reinforcement learning agent develops proactive pedestrian-handling behaviour as an emergent property of world-model planning, when trained with a minimal driving-competence reward that contains no explicit safety terms.

The investigation is built on a DreamerV3 agent trained in CARLA on non-adversarial pedestrian distributions, and on the *Jaywalkers Library*, introduced as one of the original contributions of this work. The library consists of a controlled scenario suite with three adversarial pedestrian archetypes: the *Intruder*, who enters the route late at full speed; the *Indecisive Crosser*, who freezes and resumes mid-crossing; and the *Protester*, who persistently occupies the lane. Each archetype is designed to probe a qualitatively distinct safety failure mode. The trained agent is evaluated on the library at frozen weights, with no safety-shaped reward at any stage of training. A second methodological contribution makes the investigation possible: the world-model rollouts produced at every real decision step are turned into an interpretable diagnostic surface through a deterministic six-class taxonomy of imagined behaviour, a set of episode-scale proactivity metrics (conflict-awareness lead time, avoidance-intent lead time, unsafe-persistence rate, intent-stability dwell times, defensive-waiting-lock indicator), and a state-conditioned imagination field that projects imagined intent into the realised pedestrian-distance/ego-speed state space.

The baseline agent achieves high route-completion rates in the absence of adversarial pedestrians and is unaffected by passive bystanders, establishing that subsequent failures are not generic driving failures. Under adversarial evaluation, pedestrian-collision rates are substantial and differ qualitatively across archetypes, with typical encounter episodes reaching near-collision time-to-collision values. The agent retains the physical capacity to brake hard once the conflict is imminent, so the failure is not one of control authority. The imagination evidence locates the failure inside the world model: imagined futures do reach the pedestrian region ahead of closest approach, but awareness does not translate into commitment — a large fraction of conflict-reaching rollouts remain straight-through, imagined plans are short-lived and unstable, and a meaningful share of close, moving states produce imagined futures with neither yielding nor steering. Whatever avoidance appears is heavily yield-dominant, with combined brake-and-lateral plans rare. The Protester result is the most diagnostic: a high collision rate against a stationary obstacle rules out pedestrian motion uncertainty as the dominant explanation and locates the failure in a learned

world-model prior that pedestrians clear the route. A separate failure mode emerges in the Protester condition: defensive deadlock, in which the agent becomes immediately safe by stopping but cannot evaluate, within its short imagination horizon, whether waiting will eventually clear the obstruction.

The thesis concludes that proactive pedestrian safety is architecturally accessible but not automatically emergent in model-based reinforcement learning under a minimal driving-competence reward. The bottleneck for the agent studied here is neither perception nor control but world-model imagination quality under distribution shift. Beyond the specific finding, the work contributes a transferable diagnostic protocol for any Dreamer-class agent on any conflict-anchored benchmark.

Keywords: model-based reinforcement learning, DreamerV3, world models, autonomous driving, pedestrian safety, CARLA.

Благодарност

Најискрена благодарност му изразувам на мојот ментор, проф. д-р Миле Станковски, за трпението, разбирањето и посветеноста во текот на изработката на оваа докторска дисертација. Во текот на повеќегодишниот процес на истражување и работа, неговата доверба, советите и подготвеноста да даде насока имаа значајна улога во успешното завршување на оваа работа, особено во периодите кога истражувањето не се одвиваше со посакуваната динамика. Му благодарам за вложеното време, охрабрувањето и континуираниот ангажман во текот на моето научно и професионално усовршување.

Посебна благодарност упатувам и до вонр. проф. д-р Горјан Наџински за поддршката, достапноста и бројните корисни дискусии во текот на изработката на оваа дисертација. Неговите совети, сугестии и подготвеноста секогаш да помогне при различни истражувачки и технички предизвици претставуваа значајна помош во текот на работата.

Благодарност и до колегите од Институтот за автоматика и системско инженерство за поддршката и помошта што секогаш беа пружени кога беа потребни. Посебно им благодарам на проф. д-р Весна Ојлеска Латкоска за посветеноста и искрениот ангажман со кој придонесе процесот на докторирање да се одвива непречено, на вонр. проф. д-р Душко Ставров за деталните забелешки и корисните совети при подготовката на научните трудови, како и на моите колеги доц. д-р Филип Дончевски, м-р Благој Христов и Илија Мижимакоски за позитивната атмосфера и бројните разговори и размена на идеи во текот на изработката на оваа дисертација.

Благодарност упатувам и до Билјана Апостоловска и Марија Стоилкова за нивната љубезност и достапност при административниот дел од докторските студии. Во систем исполнет со сложени и често неинтуитивни процедури, нивните јасни насоки и практични решенија значително го олеснија процесот на докторирање.

Особена благодарност до проф. д-р Даме Димитровски за иницирањето и овозможувањето на проектната соработка во чии рамки беше поддржано ова истражување.

Од срце им благодарам на моите родители Мирче и Весна, брат ми Дамјан и снаа ми Анџела, како и целото пошироко семејство, за љубовта, разбирањето и трпението во текот на целиот докторски пат. Посебна благодарност до мојата сопруга Стефи за љубовта, неизмерната блискост и спокојот што го носеше во периодот на завршување на оваа дисертација. На крај, благодарност и до мојот син Петар, чие присуство беше постојан потсетник за тоа што е навистина важно.

Стефан

Изјава

Ова истражување беше финансирано преку програмата COMET K2—Competence Centers for Excellent Technologies, од Австриското сојузно министерство за транспорт, иновации и технологија (bmvit), Австриското сојузно министерство за дигитализација, деловни работи и претприемништво (bmdw), Австриската агенција за промоција на истражувањето (FFG), Покраината Штаерска и Штаерската агенција за промоција на стопанството (SFG).

This research was funded by the COMET K2—Competence Centers for Excellent Technologies Programme of the Austrian Federal Ministry for Transport, Innovation and Technology (bmvit), the Austrian Federal Ministry for Digital, Business and Enterprise (bmdw), the Austrian Research Promotion Agency (FFG), the Province of Styria, and the Styrian Business Promotion Agency (SFG).

Изјава

Изјавувам дека докторскиот труд го изработев самостојно, дека уредно ги цитирам сите користени извори и литература и дека трудот не е користен во рамките на други универзитетски студии или за стекнување на друго звање.

м-р Стефан Златинов, _____

Содржина

Апстракт	3
Abstract	5
Благодарност	7
Изјава	8
Изјава	9
Список на слики	12
Список на табели	13
Список на кратенки	14
1 Вовед	15
1.1 Проблемот на безбедноста на пешаците во автономното возење	15
1.2 Кон проактивна безбедност преку научени претставувачки модели	16
1.3 Истражувачки хипотези	17
1.4 Опфат и пристап	19
1.5 Придонеси	19
1.6 Организација на тезата	20
2 Тековни сознанија од поттикнуваното учење и безбедноста на пешаците во автономното возење	21
2.1 Системи за автономно возење: архитектури и безбедност	21
2.2 Однесување на пешаците и интеракција меѓу пешак и автономно возило	23
2.3 Безбедност на пешаците во применети системи за автономно возење	24
2.4 Перцепција заснована на машински вид за автономно возење	26
2.5 Поттикнувано учење за автономно возење	27
2.6 Претставувачки модели и латентна имажинација	29
2.7 Безбедно поттикнувано учење	30
2.8 Конфликтно тестирање и евалуација заснована на сценарија	31
2.9 Позиционирање на истражувањето	32
3 Теоретски и технички основи	34
3.1 Основи на длабоко учење	34
3.1.1 Конволуциски невронски мрежи	34
3.1.2 Семантичка сегментација	35
3.1.3 Проценка на длабочина и инверзна длабочина	35
3.1.4 Рекурентни невронски мрежи	36
3.1.5 Варијадиски автоенкодери	38
3.2 Поттикнувано учење	40
3.2.1 Поттикнувано учење без модел	42
3.2.2 Поттикнувано учење со модел	43
3.3 DreamerV3	44
3.3.1 Рекурентниот модел на простор на состојби	44

3.3.2	Тренирање на претставувачкиот модел	45
3.3.3	Учење на однесувањето	47
3.3.4	Тренинг-процес	49
3.4	Симулаторот CARLA	50
4	Експериментална платформа и конфигурација на агентот	53
4.1	Библиотека на конфликтни пешаци	53
4.1.1	Мотивација и принципи на дизајн	53
4.1.2	Конструкција на сценарија	54
4.1.3	Надрапник: доцно влегување на рутата	55
4.1.4	Нерешителен преминувач: однесувачка неизвесност	56
4.1.5	Протестант: постојано заземање на лента	57
4.1.6	Параметризација	58
4.2	Од симулатор до околина	58
4.2.1	Конфигурација на CARLA	59
4.2.2	Животен циклус на епизода	60
4.2.3	Создавање рути	61
4.2.4	Услови за терминација	61
4.3	Дизајн на набљудувања и акции	62
4.3.1	Сензори, во симулација и во реален свет	62
4.3.2	Канал за семантичка сегментација подредена по ризик	64
4.3.3	Канал за инверзна длабочина	64
4.3.4	Канал на карта на водилки	65
4.3.5	Вектор на состојба на возилото	66
4.3.6	Простор на акции	66
4.4	Архитектура на награда	67
4.4.1	Регулација на брзина	69
4.4.2	Терминални награди	69
4.4.3	Зошто наградата завршува тука: отфрлени алтернативи	70
4.5	Конфигурација на DreamerV3	71
4.6	Поставки за тренирање	72
4.6.1	Инфраструктура за тренирање	73
5	Анализа на безбедноста: однесување и имагинација	75
5.1	Протокол за евалуација	75
5.2	Основна способност за возење (Фаза 0)	77
5.3	Агрегирани безбедносни исходи во Фаза 1	78
5.4	Однесување по рута	79
5.5	Конфликтна динамика на ниво на чекор	83
5.6	Анатомија на судири	84
5.7	Технички вовед во проекциите и реконструкциите на Dreamer	87
5.8	Истражувачки траги на имагинација во единечна епизода	88
5.9	Таксономија на проекции	89
5.10	Агрегатни метрики за проактивност	92
5.11	Дискусија на резултатите	98
6	Заклучок	101
6.1	Сумирање на истражувањето	101
6.2	Наоди во однос на истражувачките хипотези	102
6.3	Придонеси	103
6.4	Ограничувања	103
6.5	Идна работа	104
6.6	Завршни забелешки	105

Список на слики

3.1	Просторна компресија на сцена преку CNN енкодер.	35
3.2	Семантичка сегментација и кодирање според ризик.	36
3.3	Нелинеарна предност на блиското поле со инверзна длабочина.	37
3.4	Механизам на порти и временска меморија кај GRU.	38
3.5	Веројатносно латентно претставување кај VAE.	39
3.6	Циклус на интеракција агент–околина во поттикнуваното учење.	40
3.7	Учење на претставувачкиот модел на DreamerV3.	46
3.8	Учење на однесувањето преку замислени проекции во DreamerV3.	49
3.9	Тренинг-циклус на DreamerV3: дејствување, учење и имагинација.	50
3.10	Пример на сцена рендерирана во симулаторот CARLA.	52
4.1	Просторна конфигурација на трите архетипа на конфликтни пешаци.	55
4.2	Карти за тренирање и евалуација: Town02 и Town01.	60
4.3	Тренинг-рута и набљудувачки водилки во Town02.	61
4.4	Сликовни канали на набљудување на агентот.	63
4.5	Дизајниран систем набљудување-до-акција.	68
4.6	Поврат по епизода низ текот на тренирањето.	73
5.1	Геометрија на пресметка на TTC со модел на дискови.	76
5.2	Геометрија на напредокот по рута и порамнување во Фаза 1.	77
5.3	Распределби на квалитетот на возење во Фаза 0.	79
5.4	Распределба на минимално TTC по архетип во Фаза 1.	80
5.5	Брзина на возилото по напредок по рута за секој архетип.	81
5.6	Напредок по рута низ време за Протестантот.	82
5.7	Надолжна наредба по напредок по рута за секој архетип.	83
5.8	Грешка во насоката по напредок по рута за секој архетип.	84
5.9	Реализирани акции наспроти TTC по архетип.	85
5.10	Брзина при најблиско приближување и при удар за судири.	86
5.11	Разложување на режими на неуспех по архетип.	86
5.12	Динамика на избегнување: странично поместување и пад на брзина.	87
5.13	Декодирана реконструкција на проекција при најблиско приближување.	88
5.14	Замислена брзина низ епизода за Нерешителен преминувач.	89
5.15	Замислени траектории наспроти реализирана патека.	90
5.16	Временска линија на класи на проекции по епизода.	93
5.17	Замислена намера условена од состојба во просторот растојание–брзина.	96

Список на табели

4.1	Дијагностички улоги на трите архетипа на конфликтни пешаци.	54
4.2	Параметризација на архетиповите на конфликтни пешаци.	58
4.3	Услови за терминација на епизода.	62
4.4	Кодирање на семантичка сегментација подредена по ризик.	64
4.5	Компоненти на векторот на состојба на возилото.	66
4.6	Активни компоненти на функцијата за награда.	68
5.1	Способност за возење во Фаза 0 низ четири услови.	78
5.2	Исходи по архетип во Фаза 1.	80
5.3	Метрики за проактивност засновани на имагинација во Фаза 1.	94
5.4	Бенчмарк метрики за имагинација изведени од полето на намера.	97

Список на кратенки

AEB	Автономно итно сопирање (анг. Autonomous Emergency Braking)
API	Апликациски програмски интерфејс (анг. Application Programming Interface)
CNN	Конволуциска невронска мрежа (анг. Convolutional Neural Network)
GRU	Рекурентна мрежа со контролни порти (анг. Gated Recurrent Unit)
IQR	Интерквартилен опсег (анг. Interquartile Range)
MBRL	Поттикнувано учење со модел (анг. Model-Based Reinforcement Learning)
MDP	Марков процес на одлучување (анг. Markov Decision Process)
MLP	Повеќеслојна мрежа (анг. Multilayer Perceptron)
RL	Поттикнувано учење (анг. Reinforcement Learning)
RNN	Рекурентна невронска мрежа (анг. Recurrent Neural Network)
RSSM	Рекурентен просторен модел на состојби (анг. Recurrent State Space Model)
TTC	Време до судир (анг. Time-To-Collision)
VAE	Варијациски автоенкодер (анг. Variational Autoencoder)

Глава 1

Вовед

1.1 Проблемот на безбедноста на пешаците во автономното возење

Урбаното возење е најкомплицираниот домен за автономните возила. За разлика од автопатиштата или руралните средини, каде што сообраќајот се одвива по јасно одвоени ленти и со мал број вкрстувања на траектории, урбаните средини се густе, неструктурирани и споделени со најранливата категорија учесници во сообраќајот — пешаците. Пешак удрен од возило што се движи со 50 km/h се соочува со приближно 80% ризик од смртен исход, додека при 30 km/h тој ризик паѓа на околу 10% [87]. Границата меѓу безбедна и фатална интеракција често ја определуваат мали разлики во брзината и во времето — разлики што автономното возило мора да ги совлада во услови на неизвесност за тоа како пешакот ќе постапи во следниот момент.

Размерот на проблемот е голем. На глобално ниво, пешаците сочинуваат повеќе од една четвртина од сите смртни случаи во патниот сообраќај, а нивниот удел дополнително расте во урбаните средини каде што изложеноста е најголема [17, 87]. Во Европа, пешаците претставуваат околу 20% од жртвите во сообраќајот, а во земјите со понизок доход тој удел надминува 40% [24]. Овие статистики не укажуваат само на ограничувањата на технологијата за детекција, туку и на подлабок предизвик: однесувањето на пешаците е слабо ограничено, многу променливо и често неусогласено со сообраќајните правила [30, 66].

Емпириските студии јасно ја документираат оваа варијабилност. Пешаците ги користат процепите меѓу возилата врз основа на проценки за оддалеченост и брзина што често се неточни [58]. Тие преминуваат надвор од обележани пешачки премини, минуваат на црвено и ја менуваат траекторијата среде преминување како одговор на она што го проценуваат како намера на возилото [56, 94]. Расеаноста — поради мобилен телефон, разговор или едноставно невнимание — дополнително ја намалува предвидливоста на движењата на пешаците [13]. Таквото неусогласено однесување не е исклучок — студиите покажуваат дека значителен дел од урбаните преминувања на пешаци прекршуваат барем едно сообраќајно правило [58, 65]. За автономното возило ова значи дека претпоставките засновани на правилата — „пешакот ќе чека на тротоарот“, „пешакот ќе преминува на обележан пешачки премин“ — често не важат, а системот мора безбедно да се справи токму со тие отстапувања.

Интеракцијата меѓу автономното возило и пешакот е дополнително отежната од отсуството на комуникацискиот канал што го регулира човечкото возење. Човечките возачи и пешаците преговараат преку контакт со очи, гестови и суптилни промени во кинематиката на возилото — богат, двонасочен систем на сигнализација што пешаците учат да го читаат уште од детството [19, 29]. Автономното возило, по дефиниција, го отстранува човечкиот возач од таа преговарачка постапка. Истражувањата за надворешните интерфејси меѓу човекот и машината (eHMI, англ. external Human-Machine Interface) — светлосни ленти, проектирани

сигнали, дисплеи — покажуваат извесен потенцијал, но емпириските докази укажуваат дека пешаците пред сè се потпираат на кинематското однесување на возилото, односно на неговата брзина и профилот на забавување, за да ја проценат неговата намера, без оглед на експлицитната сигнализација [68, 99]. Импликацијата е јасна: автономното возило ги соопштува своите намери преку начинот на кој вози, а не преку тоа што го прикажува. Возило што не забавува при приближување кон пешак што преминува всушност сигнализира дека нема намера да даде првенство, без оглед на која било друга индикација.

Традиционалните системи за автономно возење го решаваат проблемот на безбедноста на пешаците преку низа специјализирани модули: систем за перцепција го детектира пешакот, модул за предвидување ја прогнозира неговата идна траекторија, а планерот генерира патека без судири [11, 72]. Оваа модуларна декомпозиција е интуитивна и со години претставува доминантна архитектура во комерцијалниот развој на автономни возила [88]. Но истовремено внесува и структурна кривкост. Секој модул е оптимизиран за сопствената меѓуцел — детекторот за средна просечна прецизност, предвидувачот за грешка на поместување, планерот за изводливост на рутата — наместо за крајната цел, а тоа е безбедна интеракција [96]. Грешка во кој било модул се пренесува понатаму по синџирот. Пропуштена детекција произведува план што целосно го игнорира пешакот. Пристрасно предвидување на траекторијата произведува план што претпоставува дека пешакот ќе даде првенство тогаш кога тој нема да го стори тоа. Токму кај екстремните случаи на средби со пешаци — ретките, сложени и неусогласени интеракции што доминираат во статистиката на несреќи — овие каскадни неуспеси се најверојатни и носат најтешки последици [17].

Оваа ранливост го поттикна преминот кон пристапи засновани на учење, кои во принцип можат да го оптимизираат целиот синџир од перцепција до акција за задачата на возење, наместо да оптимизираат поединечни меѓуцели. Меѓу нив, поттикнуваното учење (RL, англ. Reinforcement Learning) се издвојува како особено ветувачка парадигма. RL агентот учи полиса — пресликување од набљудувања во акции — преку проба и грешка во симулирана околина, примајќи сигнали за награда што ја кодираат задачата на возење [14, 101]. За разлика од имитативното учење, кое е ограничено од квалитетот на човечките демонстрации, поттикнуваното учење може да открива нови стратегии и уште поважно, може да се изложи на сценарија што се премногу опасни или премногу ретки за систематски да се соберат во реалниот свет. Прашањето не е дали се потребни пристапи засновани на учење — ограничувањата на модуларните системи тоа доволно убедливо го покажуваат — туку која рамка за учење е најсоодветна за проблемот на безбедноста на пешаците и што таа рамка може, а што не може да постигне.

1.2 Кон проактивна безбедност преку научени претставувачки модели

Реактивен систем за безбедност открива непосредно претстојан судир и изведува итен маневар, на пример, автономно итно сопирање (AEB, англ. Autonomous Emergency Braking) применува максимално забавување кога се проценува дека судирот ќе се случи во краток временски хоризонт. Реактивните системи спасуваат животи, но тие интервенираат дури во последен момент, кога просторот на безбедни акции веќе е стеснет на една единствена можност. Проактивниот систем за безбедност, напротив, ги предвидува опасностите пред тие да станат итни случаи и го прилагодува однесувањето доволно рано за да зачува поширок простор на можни безбедни реакции. Разликата е иста како разликата меѓу нагло кочење во последна секунда и умерено намалување на брзината две секунди порано.

Во оваа теза, *проактивната безбедност на пешаците* се дефинира како способност да се предвидат опасности предизвикани од пешаци преку научен модел на околината и да се приспособи возењето пред судирот да стане непосреден. Оваа дефиниција има три составни

елементи. Прво, *антиципација*: системот мора да расудува за идни состојби, а не само да реагира на тековното набљудување. Второ, *учење*: способноста за предвидување мора да се стекне од искуство, а не да биде рачно програмирана, за да може да генерализира на нови ситуации. Трето, *рано прилагодување*: промената во однесувањето мора да се случи со доволна временска резерва за да претставува осмислена стратегија, а не итна реакција.

Поттикнуваното учење со модел (MBRL, англ. Model-Based Reinforcement Learning) нуди природна рамка за ваква дефиниција. MBRL агентот учи *претставувачки модел* — невронска мрежа што предвидува како ќе се развива околината како одговор на акциите на агентот — и го користи тој модел за да „замислува“ идни траектории пред да се обврзе на конкретна акција [33, 51]. Во контекст на безбедноста на пешаците, претставувачкиот модел, во принцип, треба да ја опфати динамиката на движењето на пешакот: каде се наоѓа пешакот, како се движи и што ќе се случи ако возилото ја задржи брзината, забави или сврти. Агентот потоа може да планира низ овие замислени иднини и да избира акции што ги избегнуваат опасните исходи.

Кај пристапите засновани на машински вид, ова добива уште еден слој на сложеност. Разбирањето на светот кај агентот е целосно посредувано од она што го перципира преку камерите. Пешак што е присутен во сцената, но е слабо претставен во набљудувањето — премал, со низок контраст или премногу оддалечен од главната оска на камерата — можеби воопшто нема да биде вграден во латентната состојба на претставувачкиот модел. Дури и кога пешакот е перципиран, претставувачкиот модел може да не успее да го задржи неговото присуство во замислените иднини. Затоа, квалитетот на имагинацијата на претставувачкиот модел е критична алка во синџирот од перцепција до проактивна акција: точната перцепција е неопходна, но не и доволна — претставувачкиот модел мора да го одржи пешакот во своите предвидувања низ хоризонт на планирање доволно долг за да овозможи навремено приспособување.

Оттука произлегува конкретно и проверливо прашање. Ако MBRL агентот се тренира со минимална функција за награда — таква што ја наградува способноста за возење, но не ја кодира експлицитно безбедноста на пешаците — дали агентот сепак развива проактивно однесување во интеракција со пешаци? И ако не развие, каде точно настанува неуспехот во синџирот перцепција–имагинација–акција? Дали агентот не успева да го перципира пешакот? Дали не успева да го задржи пешакот во своите замислени иднини? Дали има точна имагинација, но не успева да ја претвори во безбедна акција? Или самата функција за награда е структурирана така што проактивното однесување станува нерационално? Секој од овие начини на неуспех упатува на различна интервенција, а секоја интервенција открива нешто различно за тоа што е потребно за проактивна безбедност кај MBRL.

1.3 Истражувачки хипотези

Истражувањето е организирано околу четири работни хипотези. Секоја од нив е тврдeње за кое се очекува да важи, но се третира како отворено прашање — доказите може да ја потврдат, да ја прецизираат или да ја побијат, а секој од тие исходи е информативен.

X1. Агент со поттикнувано учење со модел има архитектонски капацитет да развие проактивна безбедност на пешаците како појавно својство на планирањето преку претставувачкиот модел, без експлицитни безбедносни термини во функцијата за награда.

- **X1.1:** Со тоа што замислува идни траектории пред да се обврзе на акција, агентот во принцип има можност однапред да ги согледа опасностите од пешаци пред тие да станат непосредни — не затоа што безбедноста е директно наградена, туку затоа што претставувачкиот модел ги прави видливи идните последици.

- **X1.2:** Дали овој капацитет ќе се појави во практика зависи од структурата на наградата, квалитетот на претставите за пешаците во претставувачкиот модел и распределбата на тренирање. Секоја од овие зависности може независно да се тестира преку насочени интервенции.

X2. Простор на набљудување центриран на камера — семантичка сегментација, инверзна длабочина, состојба на возилото и водилки по рутата — му обезбедува на претставувачкиот модел доволно информации за да ја претстави динамиката на пешаците на оддалечености релевантни за безбедноста.

- **X2.1:** Сегментацијата подредена според ризик и кодирањето со инверзна длабочина ги зачувуваат просторните информации и информациите за блискост што се најважни за безбедноста на пешаците, а притоа ја намалуваат димензионалноста што претставувачкиот модел треба да ја компресира во својата латентна состојба.
- **X2.2:** Резолуцијата на сликата и видното поле во избраната конфигурација може да бидат, а може и да не бидат доволни за детекција на пешаци на оддалеченостите што се потребни за проактивно избегнување.

X3. За да се опише безбедносното однесување на агентот е неопходна систематска евалуација врз неусогласени архетипови на пешаци. Евалуацијата во стандардни, усогласени сообраќајни услови ги остава критичните начини на неуспех незабележани.

- **X3.1:** Различните конфликтни архетипови испитуваат квалитативно различни начини на неуспех како временска компресија, неизвесност на траекторијата и постојана опструкција, кои една единствена метрика за стапка на судири или хомогена популација на пешаци би ги споила во едно.
- **X3.2:** Раздвојувањето на тренирањето (стандардни пешаци што преминуваат) од евалуацијата (конфликтни архетипови) овозможува да се разликува однесување што произлегува од планирањето преку претставувачкиот модел од однесување што е директно обликувано од конфликтни сигнали во наградата за време на тренирањето.

X4. Замислените идни траектории на претставувачкиот модел обезбедуваат мерлив прозорец во процесот на планирање на агентот, откривајќи што агентот очекува да се случи при средба со пешак и како тоа очекување е поврзано со неговите акции.

- **X4.1:** Декодираните проекции на имагинацијата можат да се споредат со вистинските иднини преку метрики за просторен преклоп и постојаност на пиксели, со што квалитетот на внатрешните предвидувања на агентот станува мерлив, а не само квалитативно опишлив.
- **X4.2:** Дали квалитетот на имагинацијата ги предвидува исходите, односно дали агентите што попрецизно ги претставуваат пешаците истовремено и побезбедно постапуваат со нив, останува отворено прашање.

Заедно, овие четири хипотези го обликуваат истражувачкиот правец на тезата. X1 е основа за експерименталниот дизајн со минимална награда. X2 го насочува изборот на просторот на набљудување. X3 ја оправдува употребата на Библиотеката на конфликтни пешаци и раздвојувањето меѓу тренирање и евалуација. X4 ја поставува методологијата за интроспекција на претставувачкиот модел. Ако X1 се потврди недвосмислено, тогаш MBRL овозможува проактивна безбедност под условите определени со X2. Ако не се потврди, X3 и X4 обезбедуваат алатки за да се утврди каде точно лежи ограничувањето.

1.4 Опфат и пристап

Истражувањето го користи DreamerV3 [33] како основа за поттикнувано учење со модел. DreamerV3 е избран не затоа што е единствениот алгоритам од овој вид, туку затоа што нуди три својства клучни за ова истражување. Прво, учи целосно преку имагинација: актерот се тренира врз траектории создадени од претставувачкиот модел, а не врз вистински транзиции од околината. Тоа значи дека квалитетот на претставувачкиот модел директно го одредува квалитетот на полисата, па затоа анализата на имагинацијата е навистина информативна за однесувањето на полисата. Второ, латентните состојби на претставувачкиот модел можат да се декодираат назад во просторот на набљудување, што овозможува прозорец разбирлив за човек во она што агентот „го замислува“ за време на планирањето. Трето, DreamerV3 покажува широка компетентност во различни домени без преголеми архитектурни измени специфични за задачата, со што се намалува ризикот дека евентуалните ограничувања што ќе се забележат се последица на лошо конфигуриран алгоритам, а не вистинско својство на MBRL применет на безбедноста на пешаците.

Сите експерименти се спроведуваат во симулаторот CARLA [21], симулатор со отворен код за урбано возење изграден врз Unreal Engine. CARLA овозможува контролирани перцепциски влезови со точни референтни податоци, како семантичка сегментација и длабочина, како и скриптабилно однесување на пешаци, што е неопходно за конструирање повторливи конфликтни сценарија.

Функцијата за награда на основниот агент намерно не содржи ниеден термин што директно ја кодира безбедноста на пешаците: нема казна за време до судир (TTC, англ. Time-To-Collision), нема казна за близина и нема поттик за давање првенство. Ова е истражувачка одлука, а не препорака за реална примена. Прашањето што го поставува тезата е какво однесување кон пешаците се појавува кога наградата ја кодира само способноста за возење. Ако безбедноста однапред се вгради во наградата, тогаш се одговара на поинакво прашање — колку добро обликувана награда може да произведе безбедно однесување, а улогата на претставувачкиот модел во антиципирањето на опасностите останува замаглена. Минималната награда овозможува придонесот на секоја компонента на системот (перцепција, претставувачки модел, награда, полиса) независно да се тестира преку насочени интервенции.

Експерименталната работа е организирана во две фази на евалуација. Основен DreamerV3 агент се тренира со минималната награда, најпрво се евалуира во неконфликтни услови за да се утврди возачката компетентност, а потоа се евалуира врз сите конфликтни сценарија со пешаци. Оваа евалуација го опишува однесувањето на агентот, ги анализира неговите замислени иднини и води до дијагноза на теснеците што ја ограничуваат проактивната безбедност.

1.5 Придонеси

Оваа теза дава четири главни придонеси.

- П1. Библиотека на конфликтни пешаци.** Систематски бенчмарк со конфликтни пешаци интегриран во CARLA. Библиотеката содржи три архетипови: Натрапник, Нерешителен преминувач и Протестант, при што секој е осмислен да испита различен начин на неуспех: временска компресија, неизвесност на траекторијата и постојана опструкција. Архетипите се дефинирани преку набљудливи параметри на движење, а не преку претпоставена психологија, што ги прави повторливи и преносливи во други симулатори.
- П2. Методологија за интроспекција на претставувачкиот модел.** Методологија за анализа на тоа што DreamerV3 агентот замислува за пешаците за време на планирањето.

Со зачувување на латентните состојби во моментот на средба со пешак и проектирање на априорниот модел 32 чекори напред, декодираните замислени иднини можат да се споредат со вистинските иднини преку метрики за просторен преклоп и постојаност на пиксели. На тој начин, внатрешниот процес на планирање на агентот станува видлив и мерлив.

- П3. Карактеризација на однесување под конфликтен притисок.** Детална карактеризација на тоа како MBRL агент со минимална награда се однесува низ систематски варијанти конфликтни средби со пешаци. Наместо да се пријавуваат само стапки на судири, однесувањето се разложува на динамика на избегнување (странично отстапување, намалување на брзината), безбедносни маргини (минимален TTC, брзина во конфликтниот прозорец) и крајни исходи, откривајќи стратегии чувствителни на архетипот.
- П4. Емпириски сознанија за MBRL и безбедноста на пешаците.** Заклучоци засновани на докази за тоа што е потребно за проактивна безбедност на пешаците во поттикнуваното учење со модел, изведени од заемното дејство меѓу дизајнот на наградата, распределбата на тренирање, надежноста на претставувачкиот модел и однесувањето кое произлегува од полисата.

1.6 Организација на тезата

Глава 1 (оваа глава) го воведува проблемот на безбедноста на пешаците, ја дефинира проактивната безбедност, ги поставува истражувачките хипотези и го опишува опфатот на истражувањето. Глава 2 ја прегледува релевантната литература — архитектури за автономно возење, однесување на пешаци, поттикнувано учење за возење, претставувачки модели, безбедно RL и конфликтно тестирање — и го идентификува истражувачкиот јаз што го адресира тезата. Глава 3 ги изложува техничките основи потребни за остатокот од работата: конволуциски мрежи, рекурентни мрежи, варијацијски автоенкодери, поттикнувано учење и архитектурата DreamerV3. Глава 4 ја опишува целата експериментална поставеност: симулациската околина CARLA, Библиотеката на конфликтни пешаци, дизајнот на набљудувањата, функцијата за награда и конфигурацијата на агентот. Глава 5 ги прикажува резултатите: способност за возење во номинални услови, однесување при средби со конфликтни пешаци, анализа по архетип, интроспекција на замислените иднини и дијагноза на теснеците. Глава 6 ги синтезира наодите, ги дискутира ограничувањата и ги посочува насоките за идни истражувања.

Глава 2

Тековни сознанија од поттикнуваното учење и безбедноста на пешаците во автономното возење

Истражувањето во оваа теза се потпира на неколку различни истражувачки правци, чии наоди заедно го дефинираат јазот што тезата го адресира. Прегледот е организиран по логиката на тој јаз. Прво се разгледуваат архитектурите на системите за автономно возење (2.1) и однесувањето на пешаците (2.2), со цел да се утврди зошто конфликтните интеракции остануваат нерешен проблем и за модуларните и за системите со учење. Потоа се испитува каде стојат применетите комерцијални системи (2.3) и какви перцепциски алатки се достапни за камера-центрични пристапи (2.4), за да се воспостави контекстот на дизајнерските избори во оваа теза. Следно, поттикнуваното учење за возење (2.5) и претставувачките модели со латентна имагинација (2.6) се анализираат од аспект на нивниот потенцијал и ограничувања за проактивна безбедност. Рамките за безбедно поттикнувано учење (2.7) и методите за конфликтно тестирање (2.8) го заокружуваат прегледот со поглед на расположливите алатки за вградување и верификација на безбедноста. Конечно, секцијата 2.9 ги поврзува сите правци и прецизно го лоцира истражувачкиот јаз.

2.1 Системи за автономно возење: архитектури и безбедност

Доминантната архитектура во комерцијалниот развој на автономни возила ја резделува задачата на возење на низа специјализирани модули: перцепција, локализација, предвидување, планирање и управување [11, 72]. Секој модул решава јасно определен потпроблем и има јасно определен интерфејс кон следниот модул, што го прави системот разбирлив и погоден за тестирање и верификација на ниво на модул. Инженерите можат да ги разгледуваат излезите на поединечните компоненти, да вметнуваат безбедносни ограничувања во фазата на планирање и да ги припишуваат неуспесите на конкретни модули кога тие ќе се појават [96].

Истата модуларност што ја прави оваа архитектура разбирлива ја воведува и нејзината главна слабост: пропагација на грешки. Интерфејсот меѓу модулите е граница на која дел од информацијата се губи. Систем за детекција што пропушта пешак не ја пренесува својата неизвесност до планерот — наместо тоа создава празна листа со детекции, а планерот постапува врз основа на тоа. Систем за предвидување што претпоставува дека пешакот ќе продолжи да преминува, иако тој всушност ќе застане, му предава на планерот траекторија што е само-

уверена, но погрешна. Бидејќи секој модул се оптимизира за сопствена меѓуцел, детекторот за средна просечна прецизност, а предвидувачот за грешка на поместување, а не за крајната цел — безбедно возење, системот може да биде многу успешен на секој поединечен чекор, а сепак во екстремни случаи да произведе опасен план [72, 88]. Модулите создаваат и информациски теснеци: контекстуалниот сигнал што е богат во суровите сензорски влезови, но не е опфатен во излезната претстава на модулот, неповратно се губи во системот.

Пристапите со учење од крај до крај го заменуваат модуларниот систем со една единствена невронска мрежа, научена да пресликува сурови сензорски влезови во наредби за управување [11, 60, 98]. Заедничкото тренирање ги отстранува теснеците на интерфејсите и ја оптимизира целата низа од перцепција до акција за задачата на возење, наместо за меѓучекорни цели. Системи како Learning by Cheating [10] и implicit affordance networks [84] покажаа дека е можно да се постигнат конкурентни резултати на бенчмарци без експлицитни меѓупретстави.

Придобивките од учењето од крај до крај доаѓаат со добро позната цена. Отсуството на меѓупретстави што може човек да ги толкува го отежнува разбирањето зошто е донесена конкретна одлука, вметнувањето безбедносни ограничувања на природни архитектонски граници и проверката дали системот се однесува исправно во услови што не биле видени при тренирањето [14, 96]. Овие системи бараат и многу податоци: потребни им се разновидни распределби за тренирање за да не станат кршливи, а нивното однесување во ретки или конфликтни сценарија што се недоволно застапени во податоците може значително да се влоши [11].

Истражувачката заедница постепено се насочи кон хибридни пристапи што го задржуваат духот на учењето од крај до крај, но воведуваат структурирани меѓупретстави како разбирлив мост меѓу перцепцијата и планирањето [72]. Наместо директно да создаваат наредби за управување на ниско ниво, многу современи системи предвидуваат богата меѓупретстава т.е. множество идни водилки, карта на трошоци или семантичка карта од птичја перспектива (BEV, англ. Bird's-Eye View), од која потоа надолен модул ја добива конечната траекторија [27, 92]. Таа меѓупретстава ја користи моќта на длабокото учење за сложеноста на фузијата на сензори и разбирањето на сцената, а притоа останува доволно структурирана за отстранување грешки и за примена на безбедносни правила.

Употребата на карти од птичја перспектива стана особено распространета. Системите што ги проектираат влезовите од камери, LiDAR и радар во претстава одозгора и расудуваат во тој простор покажаа силни резултати на навигациски бенчмарци [15, 26]. Сепак, ваквите пристапи најчесто бараат или фузија со LiDAR или проекција камера-до-BEV што зависи од силни претпоставки за геометријата на патот. Тезата во овој труд следи поинаков пат: таа користи строго камера-центричен простор на набљудување и го обработува од сопствената гледна точка на агентот, без глобална меѓупретстава одозгора.

Во распределбата на сообраќајните сценарија постојат ретки, но безбедносно критични екстремни ситуации што е тешко систематски да се опфатат [17]. Систем што е евалуиран на илјадници преминувања во согласност со правилата и натаму може да се однесува опасно кога ќе се соочи со конфликтен пешак што користи тесен временски прозорец, со пешак што ќе застане среде преминување или со протестант што намерно го блокира патот. Стандардната евалуација врз типични сценарија дава ограничени докази за однесувањето во ваквите екстремни случаи. Токму ова набљудување ги оправдува и литературата за конфликтно тестирање разгледана во секција 2.8 и Библиотеката на конфликтни пешаци воведена во оваа теза.

2.2 Однесување на пешаците и интеракција меѓу пешак и автономно возило

Однесувањето на пешаците во сообраќајот е значително попроменливо и помалку усогласено со правилата отколку што тоа го претпоставува сообраќајната регулатива. Емпириските студии во различни културни контексти бележат високи стапки на прекршување на правилата: пешаците често преминуваат на необележани места, ги игнорираат сигналите и динамички ги приспособуваат своите траектории во зависност од брзината и оддалеченоста на возилото [58, 94]. Таквото неусогласено однесување е стратешко: пешаците ги прифаќаат празнините во сообраќајот врз основа на груба проценка на брзината на возилото, сопствената способност за преминување и перципираната цена на чекањето [65]. Овие проценки често се неточни, особено кога станува збор за брзината [58], а изборот на такви празнини може да доведе до траектории што непредвидливо се вкрстуваат со динамиката на возилото.

Повеќе фактори влијаат врз зачестеноста и сериозноста на неусогласеното однесување. Возраста има добро утврден ефект: постарите пешаци имаат подолго почетно време на тргнување, пониски брзини на преминување и поинакви прагови за прифаќање празнини во споредба со помладите возрасни лица [59]. Одвлекувањето на вниманието од мобилни уреди значително ја намалува внимателноста кон сообраќајот и ја зголемува веројатноста за опасни преминувања [13, 56]. Влијаат и културни и демографски фактори: стапките на неусогласено однесување значително се разликуваат меѓу држави и урбани контексти, како одраз на разлики во инфраструктурата, нормите и спроведувањето [58, 94]. Практичната последица за автономното возење е јасна: однесувањето на пешаците не може да се моделира како една единствена распределба — тоа е силно хетерогено и условено од контекстот.

Кинематиката на пешаците воведува уште еден слој на променливост. Во теренски студии, брзините на преминување се движат од под 0.5 m/s до над 2 m/s, со распределби што зависат од возраста на пешакот, итноста и природата на самото преминување [38, 41]. Кога пешаците ќе забележат возило што се приближува и ќе се обидат да избегнат судир, нивните странични и надолжни прилагодувања се брзи, но ограничени: највисоките брзини на избегнување се пониски од слободната брзина на одење, а почетокот на избегнувачкото однесување доаѓа дури кога пешакот ќе го регистрира возилото како закана [47]. За автономното возило ова значи дека траекторијата на пешакот за време на конфликтот може суштински да се разликува од траекторијата пред конфликтот, што создава проблем на предвидување.

Со воведувањето на автономни возила, човечкиот возач исчезнува од преговарањето што вообичаено ја уредува интеракцијата меѓу возило и пешак. Човечките возачи и пешаците комуницираат преку контакт со очи, движења на главата, гестови со рацете и суптилни сигнали на забавување — богат двонасочен канал што пешаците го имаат усвоено уште од детството и што функционира под прагот на свесното разгледување [19, 29]. Автономното возило, со тоа што го отстранува видливиот човек, го нарушува токму овој канал.

Надворешните интерфејси меѓу човекот и машината (eHMI): светлосни ленти, проектирани патеки за преминување, текстуални прикази и звучни сигнали се предложени како начин повторно да се воспостави експлицитна комуникација [29]. Експерименталните евалуации покажуваат мешани резултати: eHMI може да ја намали неизвесноста кај пешаците во контролирани услови, но неговата ефикасност зависи од контекстот, а пешаците често го толкуваат погрешно [68]. Поконзистентни докази упатуваат дека кинематичките сигнали се примарниот канал преку кој пешаците ја заклучуваат намерата на возилото: возило што забавува се доживува како да дава првенство, без оглед на експлицитниот интерфејс. Возило што ја задржува брзината не се доживува така, без оглед на светлосниот сигнал што го прикажува [65, 99]. Ова има директни последици за дизајнот на полисата кај автономното возење: профилот на движење на возилото е неговиот главен комуникациски канал со пешаците.

Тезата дека кинематиката на возилото служи како главен комуникациски канал е поткрепена и од студии засновани на делумно набљудлив Марков процес на одлучување (POMDP,

анг. Partially Observable Markov Decision Process) за интеракцијата меѓу возило и пешак. Кога намерите на пешакот се третираат како скриена променлива, а акцијата на возилото влијае и врз веројатноста за судир и врз понатамошното однесување на пешакот преку забележаната кинематика, планирањето под неизвесност произведува квалитативно поинакво однесување отколку планирање наспроти фиксна пешачка траекторија [3, 34]. Пешакот ја ажурира својата намера според тоа што изгледа дека ќе направи возилото. Возилото, барем начелно, треба тоа да го земе предвид при планирањето. Дали агент со претставувачки модел, научен од искуство, имплицитно ја опфаќа оваа динамика на интеракција останува отворено прашање.

Точното предвидување каде ќе се наоѓа пешакот во следните неколку секунди е предуслов за проактивна безбедност. Полето на предвидување на пешачки траектории создаде обемна литература, од физички модели на социјални сили до современи пристапи со длабоко учење [9, 52, 97]. Еден наод што доследно се повторува е дека задачата на предвидување е природно мултимодална: пешак што стои на работ на тротоарот може да премине, да почека или да се оттурне од работ, па затоа стабилен систем мора да ја претстави таа распределба на можни иднини, наместо да се врзе за една единствена траекторија [54].

Рекурентните пристапи — Long Short-Term Memory (LSTM) и Gated Recurrent Unit (GRU) мрежи условени од набљудуваната траекторија на пешакот — беа доминантна парадигма неколку години [77]. Моделите со графовски невронски мрежи го проширија овој пристап за да ги опфатат социјалните интеракции меѓу повеќе пешаци и меѓу пешаци и возила. Поновите модели засновани на трансформери користат самовнимание за да ги фатат долгорочните зависимости меѓу сите учесници во сцената истовремено и постигнуваат најдобри резултати на стандардни бенчмарци [52, 80].

И покрај овој напредок, предвидувањето на траектории и натаму не е решен проблем кога се работи за конфликтни или неусогласени пешаци. Стандардни бенчмарци како ETH/UCY и nuScenes главно евалуираат преминувања во согласност со правилата. Методите научени на такви бенчмарци може да не се генерализираат на конфликтно преминување, запирање среде преминување или намерно попречување [9]. Оваа теза му пристапува на проблемот поинаку: наместо експлицитно да предвидува пешачки траектории, агентот учи имплицитен модел на пешачката динамика преку својот претставувачки модел. Квалитетот на тој имплицитен модел е централен предмет на анализата.

2.3 Безбедност на пешаците во применети системи за автономно возење

Разбирањето на тоа како зрелите програми за автономни возила ја третираат безбедноста на пешаците дава важен контекст за истражувачките избори во оваа теза. Следниот преглед ги опфаќа и комерцијалните системи засновани на фузија на сензори и камера-центричните пристапи.

Најспособните применети системи за автономни возила: Waymo, Cruise, Aurora, Zoox и слични програми, делат заедничка архитектонска определба кон повеќемодална фузија на сензори [11, 37]. LiDAR создава густ 3D облак од точки што ја разрешува формата и положбата на пешакот со сантиметарска прецизност на растојанија до 100 m и повеќе, без оглед на осветлувањето. Радарот обезбедува стабилни мерења на брзина и работи и во неповолни временски услови во кои камерите и LiDAR се влошуваат. Камерите, пак, додаваат семантичко богатство: текстура, боја и визуелни признаци што недостигаат во чисто геометриските претстави. Фузијата на овие три модалитети дава претстава што е истовремено прецизна во 3D просторот (LiDAR), робусна во неповолни услови (радар) и богата со семантичка содржина (камера) [37].

Во таков контекст, *детекцијата* на пешаци во голема мера е решен проблем внатре во

оперативниот домен на секој систем. Пешак што се наоѓа во работниот опсег на сензорите создава сигурна, локализирана 3D детекција со висока опфатеност. Потешките проблеми се поместени погоре по низата — кон *предвидувањето* и *планирањето*. Откако пешакот е детектиран, што ќе направи тој понатаму и како треба да одговори возилото? Индустриските системи за предвидување ги моделираат пешаците како агенти што создаваат траектории, при што нивните идни патеки зависат од сопствената историја, од положбите на блиските возила и инфраструктура и од заклучената намера. Архитектурите засновани на трансформери, кои заеднички ги моделираат сите учесници во сцената и нивните интеракции, станала доминантен пристап [52]. Модулите за планирање потоа ги користат овие предвидени траектории и расудуваат низ просторот на можни акции на возилото, избирајќи ја онаа што задржува прифатливи безбедносни маргини долж хоризонтот на предвидување.

Карактеристична одлика на овие системи е нивната конзервативност. Применетите системи за автономно возење обично држат поголеми маргини околу пешаците отколку што би држел човек-возач, даваат првенство однапред на премини и запираат кога траекторijата на пешакот е неизвесна. Оваа конзервативност е делумно директна последица на безбедносната архитектура — безбедносни слоеви наметнуваат минимални прагови на ТТС и минимални странични маргини без оглед на излезот на модулот за предвидување — а делумно е полиса обликувана од огромното оперативно искуство што овие програми го имаат акумулирано [69]. Цената на таквата конзервативност е ефикасноста: автономните возила често создаваат неочекувани ситуации за другите возачи со непотребно забавување, а нивно претпазливо прифаќање празнини на неконтролирани премини може да го наруши текот на сообраќајот.

Фузијата на сензори што ја овозможува оваа способност има и своја цена. LiDAR сензорите и натаму се скапи во споредба со камерите и додаваат од неколку илјади до десетици илјади долари на хардверската цена по возило, зависно од конфигурацијата [48]. Калибрацијата и синхронизацијата на повеќемодални сензорски состави внесуваат дополнителна инженерска сложеност. Уште поважно, зависноста од LiDAR го ограничува скалабилното распоредување на овие системи: безбедносна архитектура што се потпира на 3D детекција од облак од точки не може лесно да се преслика на хардвер што воопшто нема LiDAR.

Системот Full Self-Driving (FSD) на Tesla е најшироко распоредена програма за автономно возење што се потпира само на камери [11]. Tesla користи флота од сериски возила опремени со осум камери што обезбедуваат визуелна покриеност од 360°, без LiDAR и без високорезолуциски радар. Стекот FSD еволуираше од силно инженерски модуларен систем во претежно архитектура со учење од крај до крај, а најновите верзии користат една голема трансформерска мрежа што внесува видеострими од повеќе камери и создава наредби за возење. Ова движење кон невронско возење од крај до крај е во согласност со поширокиот истражувачки тренд разгледан во секција 2.1.

Пристапот на Tesla покажува дека системите само со камери можат да работат во голем обем низ широк спектар патни услови и типови на средби, вклучително и пешачки премини, конфликтни пешаци и велосипедисти во сложени урбани околина. Сепак, потпирањето само на камери воведува карактеристични слабости што системите опремени со LiDAR не ги делат: малите пешаци на поголемо растојание зафаќаат мал број пиксели, работата ноќе и во неповолни временски услови е ограничена од чувствителноста на камерите, и нема геометриска референтна вистина што би ја вкотила перцепцијата во 3D простор [17]. Прашањето дали системите само со камери можат да постигнат безбедносна рамноправност со системите опремени со LiDAR на рабовите на нивниот работен опсег и натаму останува отворено.

Mobileye, чии процесори за вид се интегрирани во напредни системи за помош на возачот (ADAS, англ. Advanced Driver Assistance Systems) системите на голем дел од светската флота на возила, зазема комплементарна позиција. Неговите системи за масовен пазар користат камери дополнети со радар, со што се избегнува цената на LiDAR, а сепак се задржува геометриската информација за брзина. Рамката Responsibility-Sensitive Safety (RSS) на Mobileye

формализира правила за безбедно возење и обезбедува проверлив безбедносен слој над стекот за перцепција и предвидување заснован на учење [89]. Со оваа архитектура, проверката на безбедноста се одвојува од научените компоненти и обезбедува формални гаранции што самиот систем на учење не може да ги даде.

Оваа теза користи дизајн на набљудување само со камери. Причината е истражувачка. Со исклучување на LiDAR, претставувачкиот модел мора да се потпре исклучиво на визуелниот влез, со што перцепциските ограничувања стануваат предмет на истражување, а не пречка за избегнување.

2.4 Перцепција заснована на машински вид за автономно возење

Комерцијалните автономни возила обично спојуваат податоци од повеќе сензорски модалитети: камери, LiDAR и радар. LiDAR обезбедува точна 3D геометрија, но е скап, троши многу енергија и е чувствителен на неповолни временски услови. Системите што работат само со камери се значително поевтини и директно добиваат од брзиот напредок во машинскиот вид, кој отвори можности за толкување на светот само од визуелни податоци што пред една деценија не беа достапни [37]. Сè поголем број истражувања покажуваат дека камера-центричните системи можат да се доближат до системите со LiDAR или да ги изедначат на клучни задачи од перцепцијата, што ја оправдува истражувачката програма што не зависи од скапа сензорска инфраструктура.

Сигурната детекција на пешаци е најосновниот перцепциски предуслов за безбедност. Развојот на детекцијата на пешаци во голема мера ја следи пошироката историја на машинскиот вид. Класичните пристапи засновани на рачно конструирани признаци: описникот Histogram of Oriented Gradients (HOG) во комбинација со Support Vector Machine (SVM), како и Deformable Part Model (DPM) за справување со артикулација и прекривање, поставија почетна основа, но беа ограничени од изразната моќ на инженерски зададените признаци [25, 102]. Длабокото учење решително ги потисна овие пристапи, а детекторите засновани на конволуциски невронски мрежи постигнаа значително подобри резултати на стандардни бенчмарци.

Подоцнежниот развој главно беше воден од две семејства на длабоки детектори. Двостепените детектори, најјасно претставени од семејството R-CNN што кулминира со Faster R-CNN [67], прво создаваат предлози за региони, а потоа ги класифицираат, при што добиваат висока точност по цена на времето за изведба. Едностепените детектори, пред сè семејството YOLO [83], ја третираат детекцијата како еден регресиски проблем и обезбедуваат работа во реално време, соодветна за применети системи за возење. Последователните верзии на YOLO постепено ја подобрија детекцијата на мали и делумно прекриени цели.

И покрај овој напредок, детекцијата на пешаци не е решен проблем. Перформансите опаѓаат при лошо осветлување, силен дожд и густа магла [17]. Силно прекриените или скратени пешаци и натаму се тешки за обработка. Особено малку внимание е посветено на проблемот со многу мали пешаци на поголемо растојание — случај што е директно релевантен за проактивната безбедност, бидејќи бара пешакот да се открие доволно рано за да има простор за навремено кочење.

Детекцијата со обиколни кутии дава груба локализација. Многу системи за возење бараат побогато просторно разбирање. Семантичката сегментација доделува класна ознака на секој пиксел, создавајќи густа карта на составот на сцената [70]. Целосно конволуциските енкодер-декодер архитектури, предводени од U-Net и DeepLab [12, 36], станаа стандарден пристап. Семантичката сегментација е вредна за автономното возење затоа што истовремено ги идентификува пикселите што припаѓаат на пешаци, ја издвојува проодната површина и ја открива структурата на сцената — информација што е распоредена низ целата слика, а не е сведена

на една обиколна кутија.

Проценката на длабочина од една камера ја враќа третата димензија изгубена при проекцијата [44]. Стереокамерите директно ја мерат длабочината преку дисперзија, додека монокуларните пристапи ја заклучуваат од научени визуелни сигнали. Во симулација, точната референтна длабочина е директно достапна од механизмот за рендерирање, со што се отстранува грешката на проценка и се добива чист сигнал што ги одразува само информациските ограничувања на гледната точка на камерата.

Во оваа теза се користат камерите за семантичка сегментација и длабочина во CARLA, кои обезбедуваат точни референтни податоци, компресирани и кодирани на начинот опишан во глава 4. Важно е што каналот за сегментација е уреден според ризикот — пикселите на пешаците добиваат највисока кодна вредност, пикселите на возилата средна вредност, а позадината најниска — со што агентот добива структурирана претстава за приоритетот на ризикот, а не само сурова карта на класни ознаки. Каналот за длабочина е кодиран како инверзна длабочина, со што поблиските објекти, како пешаци во блиско поле, се нагласуваат повеќе од далечните елементи во позадината.

Многу од најдобрите современи системи ги подигнуваат набљудувањата од камера во претстава од птичја перспектива (BEV), со што овозможуваат просторно расудување во рамнината гледана одозгора [27]. Таквите претстави се интуитивни за планирање и обезбедуваат природен простор за спојување на повеќе камери и повеќе сензорски модалитети. Сепак, проекцијата во BEV од слики бара или проценка на длабочина, што внесува грешки на растојание, или силни геометриски претпоставки за структурата на патот. Уште поважно, BEV претставите се глобални: тие ја кодираат целата сцена од перспектива што не му е достапна на ниту една поединечна камера.

Оваа теза намерно ги избегнува BEV претставите. Дизајнот на набљудувањето е строго камера-центричен: агентот ја перципира околината онака како што би ја перципирала камера поставена на неговото возило, без глобална перспектива одозгора. Овој избор го ограничува агентот на гледна точка што одговара на реалистична вградена сензорска поставка.

2.5 Поттикнувано учење за автономно возење

Поттикнуваното учење е широко применувано на задачи во автономното возење, од едноставно следење лента до целосна урбана навигација [6, 14, 101]. Пристапите без модел учат полиса или функција на вредност директно од интеракциите со околината, без да градат експлицитен модел на нејзината динамика. Deep Q-Network (DQN) покажа дека учењето засновано на вредност може да се прошири и на визуелни набљудувања [22]. Proximal Policy Optimisation (PPO) и Soft Actor-Critic (SAC) ги проширија методите со градиент на полиса и актер-критичар на континуирани простори на акции, со што се овозможи континуално управување со волан и гас наместо избор од дискретни маневри [49, 82]. Бенчмарците засновани на CARLA често се користат за евалуација на такви агенти, а повеќе од нив постигнаа конкурентни резултати на навигациски задачи [8, 18, 60].

Главното ограничување на пристапите без модел во оваа област е неефикасноста по примероци: за да научат сложени однесувања, им се потребни многу интеракции со околината, а интеракциите во CARLA се скапи во реално часовничко време поради рендерирањето и физичката симулација. Милиони чекори во околината можат да значат денови тренирање [86]. Освен прашањето на ефикасност, овие методи немаат експлицитен механизам за расудување за иднината: агентот постапува според научена функција на вредност, но не може да види или протолкува што токму таа функција предвидува за последиците од различни акции [51, 61, 79].

Поттикнуваното учење со модел ја адресира неефикасноста по примероци така што учи модел на динамиката на околината и потоа го користи за планирање или за проширување

на податоците [51, 61]. Во рамките од типот Дуна, научениот модел создава синтетички транзиции што го дополнуваат баферот со реални искуства, овозможувајќи повеќе ажурирања по чекор во околината. Кај пристапите со латентна имагинација, полисата се тренира целосно врз траектории симулирани од претставувачкиот модел, додека реалните интеракции со околината служат само за тренирање на самиот модел [33]. Клучната предност е што симулираните транзиции се евтини: научениот модел работи на графичкиот процесор (GPU, англ. Graphics Processing Unit) без товарот на симулаторот, што овозможува стотици ажурирања по секој вистински чекор во околината.

Во автономното возење, поттикнуваното учење со модел нуди уште една придобивка покрај ефикасноста по примероци: обезбедува експлицитен механизам за планирање нанапред. Агент со поттикнувано учење со модел не следи само научена функција на вредност — тој замислува идни траектории преку својот претставувачки модел и избира акции според предвидените последици. Претставувачките модели Think2Drive [46] и CarDreamer [26], засновани на DreamerV3, покажаа дека можат да научат сложени возачки маневри — менување лента, минување раскрсници и вклучување — преку тренирање со имагинација во CARLA. Raw2Drive [95] покажа дека усогласувањето на претставувачки модел со претходно научен јазичен модел може дополнително да го забрза учењето на сложени задачи за возење. Прегледот на претставувачки модели за автономно возење [23] ги издвојува моделите на латентна динамика како централна насока во тековното истражување и го нагласува нивниот потенцијал да ги обединат перцепцијата, предвидувањето и планирањето во една научена рамка.

Едно ограничување што го делат тековните системи за возење со поттикнувано учење со модел е тоа што речиси исклучиво се евалуирани во кооперативни или умерено стохастички сообраќајни услови. Постоечката литература не покажува дали капацитетот за планирање на претставувачкиот модел навистина се претвора во подобра безбедност кога интеракциите со пешаците се конфликтни, односно кога ризикот произлегува од намерни стратегии во однесувањето, а не од случајна варијација.

Дизајнот на функцијата за награда е еден од најзначајните избори во секоја примена на поттикнуваното учење, а возењето не е исклучок [1, 78]. Задачата на возење е природно повеќецелна: агентот мора да напредува кон дестинацијата, да ги почитува сообраќајните правила, да одржува удобност и да избегнува судири. Наградите за напредок најчесто се задаваат преку метрики за завршување рута — изминато растојание по планираната рута, достигнати водилки или поминати порти. Наградите за брзина поттикнуваат одржување целна брзина. Казните за удобност ги обесхрабруваат наглите вртења и кочења. Безбедноста обично се наметнува преку комбинација од казни за судир и, во многу системи, погусто безбедносни сигнали што даваат градиент уште пред да се случи судир.

Густото обликување на безбедносниот сигнал — казни засновани на ТТС што се активираат кога агентот е на траекторија што води кон судир, награди што растат со оддалеченоста од други учесници или експлицитни поттици за давање првенство — може значително да го стабилизира учењето во околини во кои има пешаци. Системите што ги вклучуваат овие елементи обично покажуваат пониски стапки на судири за време на тренирањето и побрзо доаѓаат до безбедно однесување [1, 22]. Но ваквото обликување внесува забуна: ако агентот се казнува само затоа што е близу до пешак, без разлика дали навистина би дошло до судир, тогаш на агентот директно му е кажано да избегнува пешаци, наместо тоа однесување да го научи од првите принципи преку искуство со судири. Во таков случај тешко е да се процени дали добиеното однесување навистина произлегува од научената динамика на интеракцијата или само од следење на сигналот од казната.

Основната функција за награда во оваа теза е минимална по дизајн: благо регулирање на брзината и терминална казна за успех и судир. Ова е истражувачки избор — без густо обликување на безбедносниот сигнал, она што ќе произлезе може да се припише директно на капацитетот за планирање на претставувачкиот модел.

2.6 Претставувачки модели и латентна имагинација

Идејата за користење научен модел на околината за оптимизација на полиса потекнува од раната литература за поттикнувано учење со модел, но современата форма на латентни претставувачки модели јасно се обликува со трудот World Models на Ха и Schmidhuber [?]. Во тој труд се учи компактна рекурентна претстава на визуелна околина, а едноставна полиса потоа се тренира во научениот латентен простор, покажувајќи дека тренирањето со имагинација може да доведе до ефикасно однесување.

Серијата Dreamer ја прошири оваа идеја кон задачи на континуирано управување од набљудувања од пиксели. DreamerV1 учеше Recurrent State-Space Model (RSSM) со континуирана стохастичка состојба и тренираше актер-критичар во имагинација, при што постигна силни резултати на DeepMind Control Suite [31]. DreamerV2 ја замени континуираната латентна состојба со дискретна категоријална распределба, со што истовремено ги подобри резултатите и на Atari игри и на задачи со континуирано управување [32]. DreamerV3 [33], алгоритмот што се користи во оваа теза, воведе повеќе техники за стабилизација: симетрично-логаритамски предвидувања, нормализација на наградата преку експоненцијални подвижни средини и скалирање на повратите според перцентили; што го прават алгоритмот робустен на различни размери на награда и домени без нагонување по задача. DreamerV3 покажа способност во широк опсег домени истовремено, со еден единствен сет хиперпараметри, со што се наметна како најопштоприменливиот достапен латентен претставувачки модел.

Примената на латентни претставувачки модели во автономното возење забрзано се развива. CarDreamer [26] обезбеди платформа со отворен код за возење засновано на претставувачки модел во CARLA и покажа дека DreamerV3 може да научи структурирани возачки маневри, вклучително и држење лента, движење низ раскрсници и менување лента, преку тренирање со имагинација. Think2Drive [46] покажа дека агентите со DreamerV3 можат да се генерализираат низ сложени урбани сценарија кога се тренираат на доволно разновидни рути. Raw2Drive [95] интегрира претходно научен претставувачки модел со основен модел за возење, спојувајќи го капацитетот за планирање на поттикнуваното учење со модел со разбирањето на сцената што го носат големи претходно научени мрежи.

Поширок преглед на претставувачки модели за автономно возење издвојува неколку истражувачки насоки: подобрување на веродостојноста на предвидувањата на претставувачкиот модел на долги хоризонти, скалирање на моделите за да опфатат посложена динамика на околината и користење на такви модели за планирање под неизвесност [23, 35]. Она што во голема мера недостасува во оваа литература е систематско испитување на тоа што претставувачките модели учат да претставуваат за конкретни динамички актери, а особено за пешаците; и дали квалитетот на таа претстава го одредува безбедносното однесување на агентот. Реконструираниите замислени иднини што ги создава декодерот на DreamerV3, барем начелно, може да се анализираат за да се процени што точно моделот предвидува за положбите, брзините и траекториите на пешаците. Оваа интроспективна анализа е еден од придонесите на оваа теза.

Квалитетот на замислените иднини создадени од еден претставувачки модел зависи од квалитетот на енкодерот, односно колку добро набљудувањата се компресираат во латентни состојби, од квалитетот на моделот на динамика, односно колку точно RSSM транзициите предвидуваат идни латентни состојби, и од квалитетот на декодерот, односно колку верно латентните состојби се реконструираат назад во просторот на набљудување. Откажувања може да се појават на секое ниво. Енкодерот можеби нема да ги фати признаците релевантни за пешаците. Моделот на динамика можеби нема точно да го пренесува нивното движење. Декодерот можеби нема верно да го прикаже пешакот иако релевантната информација постои во латентната состојба.

Натрופувањето на грешките на моделот низ долги хоризонти на имагинација е добро познат предизвик во поттикнуваното учење со модел [20, 39, 74]. Дури и мала грешка во пред-

видувањето по поединечен чекор се акумулира низ хоризонт од 32 чекори и може да доведе до замислени иднини што значително отстапуваат од реалноста. За чести и добро застапени сценарија, како возење по прав пат или присуство на статични пречки, моделот на динамика може да научи точни предвидувања од густ тренинг-искуства. За ретки, конфликтни структурирани интеракции, распределбата на тренирање обезбедува помалку примери, па квалитетот на предвидувањето може да биде понизок.

2.7 Безбедно поттикнувано учење

Стандардното поттикнувано учење ја максимизира наградата без тврди гаранции за безбедност. Кога безбедноста е задолжително барање, наместо спореден критериум, задачата се формулира со експлицитни ограничувања врз трошокот поврзан со судири, нарушување на безбедносни растојанија и слични ризици. Лагранжовите методи овозможуваат тие ограничувања да се вклучат директно во процесот на учење, така што полисата се насочува истовремено кон добра изведба и кон почитување на безбедносните граници. Во автономното возење, таквите пристапи се користат за истовремено одржување лента, избегнување судири и почитување сообраќајни правила [7].

Поттикнуваното учење со штит и пристапува на безбедноста поинаку: наместо да ја менува наградата или функцијата на вредност, меѓу полисата и околината вметнува безбедносен слој што го проверува или коригира излезот на полисата пред извршување [7]. Штитот спроведува тврди ограничувања така што ја надминува секоја акција што би нарушила определено безбедносно правило, обично пресметано од тековната состојба со формален модел, како пресметка на достижливост, проверка на време до судир или пресметка според Responsibility-Sensitive Safety (RSS). Полисата се тренира во присуство на таков штит и со текот на времето може да научи однапред да ги предвидува интервенциите на штитот и да избегнува состојби во кои тој би се активирал.

Клучното својство на поттикнуваното учење со штит е дека безбедноста е обезбедена по конструкција за време на извршувањето: штитот дава тврда гаранција без оглед на тоа што научила полисата. Цената е намалување на просторот на акции на полисата и можност за нарушување на сигналот за учење: ако штитот често ги надминува планираните акции на полисата, тогаш полисата се тренира во изменета околина што можеби не одговара добро на неограничената поставка при распоредување. Дали агентот навистина учи претходно да дејствува така што ја намалува зачестеноста на активациите на штитот, односно да ги избегнува токму ситуациите што би го активирале штитот, е емпириско прашање. Ова е емпириско прашање релевантно за разбирањето на границите на безбедното поттикнувано учење.

Стандардното поттикнувано учење ја оптимизира очекуваната награда, што значи дека е слепо за распределбата на исходите. Агент што создава ретки, но катастрофални исходи може да има иста очекувана награда како агент што создава постојано умерени исходи. Поттикнуваното учење чувствително на ризик го адресира ова така што оптимизира мерки на ризик што ја казнуваат варијансата на исходите или им даваат поголема тежина на неповолните исходи [?]. Условната вредност при ризик (CVaR) на ниво на доверба α ја мери очекуваната награда во најлошиот дел α од исходите и го поттикнува агентот да го избегнува екстремниот дел од распределбата. Поттикнуваното учење со распределба на поврати ја учи целата распределба на поврати, а не само нејзиното очекување, со што овозможува избор на акции чувствителен на ризик врз основа на целата распределба [50]. Овие пристапи се особено релевантни за безбедносно критични примени, каде што ретките, но скапи исходи, како судирите, не смеат да се третираат како занемарливи во просек.

Голема класа пристапи за безбедност на пешаците во поттикнуваното учење не користи формални рамки за безбедно учење, туку ја обликува функцијата за награда за да создаде погуст безбедносни сигнали. Казните засновани на ТТС се активираат кога предвиденото време до судир со пешак падне под зададен праг и даваат градиентен сигнал што го обес-

храбрува приближувањето кон пешак со висока брзина [1]. Казните за близина полека растат како што агентот се приближува до кој било пешак, создавајќи одбивно потенцијално поле во просторот на наградата. Експлицитните награди за давање првенство се активираат кога агентот забавува во присуство на пешак што има право на премин.

2.8 Конфликтно тестирање и евалуација заснована на сценарија

Бидејќи податоците од реалниот сообраќај ретко ги опфаќаат ретките, но безбедносно критични ситуации, проценката заснована на сценарија стана доминантен пристап за валидација на автономни возила [69, 100]. Сценариото е структуриран опис на една возачка ситуација т.е. геометрија на патот, учесници во сообраќајот, услови во околината и временски распоред, што системот треба правилно да ја разреши. Овој пристап овозможува контролирани и повторливи услови, насочени кон конкретни способности, и затоа е погоден за систематско покривање на однапред определени категории на опасност.

Каталозите на сценарија за тестирање на автономни возила се стандардизирани од тела како ASAM, Pegasus и протоколите EURO NCAP за проценка на АЕВ [69]. Овие каталози определуваат апстрактни сценарија, на пример “пешак преминува од десно”, што потоа се инстанцираат во конкретни тест-случаи со зададени вредности на параметрите. Анализата на покриеност прашува дали избраните конкретни сценарија доволно добро го покриваат релевантниот параметарски простор. Предизвикот е што тој простор е високодимензионален, а најголем дел од параметризациите водат до неинтересни ситуации. Безбедносно критичните случаи се концентрирани во мал дел од просторот што тешко се открива со рамномерно семплирање.

Семплирањето по важност и неговите варијанти го адресираат проблемот со ретките настани така што го насочуваат семплирањето кон безбедносно критичните области на просторот на сценарија. Adaptive stress testing [42] го формулира проблемот како секвенциски процес на одлучување, во кој конфликтен агент избира промени во околината, на пример положба и тајминг на пешакот, за да ја максимизира веројатноста да предизвика неуспех кај системот што се тестира. Овој пристап е ефикасен во пронаоѓање критични сценарија, но бара диференцијабилен, или барем оценлив, модел на однесувањето на системот. O’Kelly и соработниците [57] покажаа скалабилна симулација на ретки настани за системи со учење од крај до крај преку важностно семплирање над почетните услови.

Најголемиот дел од работата за конфликтно тестирање во автономното возење е насочен кон интеракции возило-возило, измени на сензорските влезови или намерно неповолни временски услови [43, 63]. ANTI-CARLA [63] вовеле рамка за конфликтно тестирање во CARLA што ги менува траекториите на други возила и сензорските влезови за да најде конфигурации што предизвикуваат неуспех. CRASH [43] тренираше конфликтен агент заснован на поттикнувано учење за да конструира предизвикувачки сценарија за целно автономно возило, исто така, главно насочени кон интеракции возило-возило.

Работата што конкретно се насочува кон конфликтно однесување на пешаци е многу поограничена. Priisalu и соработниците [62] создадоа разновидни сценарија на пешачко преминување преку варирање на параметри, но ги третираа пешаците како пасивни актери, а не како стратешки агенти со различни обрасци на однесување. Бенчмаркот HABIT [64] обезбеди структуриран сет сценарија на човечки дејства во CARLA, но беше насочен кон перцепциски задачи, а не кон евалуација на полиси во затворена јамка. Serrano и соработниците [76] користеа снимање на движење (motion capture) за да создадат реалистични пешачки анимации во CARLA, подобрувајќи ја природноста на изгледот и движењето на пешаците, но не конструираа систематски конфликтни сценарија на преминување.

Во досегашната литература не се појавил систематски бенчмарк на конфликтни пешач-

ки *архетипови* — однесувања дизајнирани да испитаат различни механизми на неуспех во интеракцијата меѓу полиса на автономно возило и неусогласен пешак. Библиотека на конфликтни пешаци воведена во оваа теза го пополнува токму овој јаз, преку три архетипови (Натрапник, Нерешителен преминувач и Протестант) што поставуваат квалитативно различни барања пред безбедносните способности на агентот.

Повторлива слабост и кај евалуацијата врз основа на реален сообраќај и кај конфликтното тестирање е тоа што резултатите обично се прикажуваат како стапки на судири или стапки на успешни епизоди, без неуспесите да се разложат според конкретните механизми што ги произвеле [69, 93]. Стапка на судири од 15% може да одговара на широк спектар квалитативно различни однесувања: агентот можеби постојано не успева навреме да го детектира пешакот. Можеби го детектира, но избира да ја задржи брзината и да сврти. Можеби се обидува да кочи, но премногу доцна. Или пак успешно кочи, а потоа повторно забрзува кон пешакот. Секој од овие механизми на неуспех упатува на различна насока за подобрување, а една единствена скаларна метрика не може да ги разликува.

Во оваа теза, рамката за евалуација го разложува однесувањето на крајни исходи, динамика на избегнување и распределби на безбедносни маргини, со што овозможува квалитативно да се опише *како* агентот одговара на секој архетип, а не само дали се судира. Токму оваа дијагностичка перспектива ги поврзува резултатите од евалуацијата со дизајнот на интервенциите.

2.9 Позиционирање на истражувањето

Изложениот преглед покажува јасно приближување на повеќе истражувачки правци. Истражувањето во автономното возење се помести од крути модуларни системи кон научени системи од крај до крај, при што најспособните понови пристапи користат структурирани меѓупретстави или експлицитни претставувачки модели за да ја спојат изразната моќ на длабокото учење со разбирливоста и погодноста за планирање што ја нудат структурирани претстави. Камера-центричните системи покажаа дека перцепцијата само со камери е конкурентна со системите засновани на LiDAR во многу задачи, а заедницата во машинскиот вид продолжува да ја турка границата кај детекцијата, сегментацијата и проценката на длабочина за пешаци. Поттикнуваното учење, особено поттикнуваното учење со модел и латентни претставувачки модели, се наметна како способен пристап за учење сложени возачки полиси од симулација, при што DreamerV3 и сродни алгоритми покажаа компетентност во разновидни домени. Безбедното поттикнувано учење обезбедува алатки за експлицитно вградување безбедносни цели, а конфликтното тестирање обезбедува средства за откривање случаи на неуспех што евалуацијата врз основа на реален сообраќај лесно би ги пропуштила.

Сепак, на пресекот на овие правци остануваат неколку недоволно адресирани проблеми.

Еден проблем е систематската евалуација на агенти со поттикнувано учење со модел во услови на конфликтна интеракција со пешаци. Постоечките системи за возење со поттикнувано учење со модел, како CarDreamer, Think2Drive и Raw2Drive, се евалуираат во кооперативен или слабо стохастички сообраќај. Дали нивниот капацитет за планирање преку претставувачки модел навистина се претвора во подобра безбедност кога пешаците се однесуваат конфликтно — го темпираат преминувањето за да го скратат временскиот прозорец за реакција на агентот, се двоумат и повторно тргнуваат среде преминување или упорно го погрчуваат патот — не е утврдено. За тоа е потребна евалуациска парадигма што систематски конструира конфликтни интеракции, наместо само да ги зема од распределбата на случаи во реалниот сообраќај.

Проблем е и разграничувањето меѓу безбедносното однесување што произлегува од капацитетот на претставувачкиот модел и однесувањето што директно е наметнато преку густо обликување на наградата. Повеќето системи за возење со поттикнувано учење што пријавуваат ниски стапки на судири со пешаци вклучуваат експлицитни безбедносни термини во

наградата — ТТС казни, награди за близина или поттици за давање првенство. Тие термини директно го вметнуваат избегнувањето пешаци во градиентниот сигнал, па не може да се одвои колкав дел од набљудуваното безбедносно однесување навистина би произлегол само од капацитетот за планирање на претставувачкиот модел, без инженерски насоки. Ако се сака да се проучи ова прашање, потребна е основна поставка што намерно го исклучува таквото обликување и потоа набљудува што ќе се развие.

Дијагностичката анализа на тоа на кој начин претставувачкиот модел не успева, а не само дали не успева, е исто така проблем. Самото пријавување стапка на судири дава ограничена насока за подобрување на системот. Од суштинско значење е да се разбере дали агентот не успева да го перципира пешакот, не успева да го задржи пешакот во своите замислени иднини, не успева точната имагинација да ја претвори во безбедна акција или рационално постапува според награда што кочењето го прави неисплатливо. Токму овие разлики се пресудни за точно да се лоцира теснецот и да се разбере природата на неуспехот. За тоа е потребна интроспекција на внатрешните претстави на агентот, заедно со набљудување на неговото надворешно однесување.

Оваа теза е поставена така што ги адресира сите три проблеми. Таа спојува контролирана конфликтна евалуациска рамка — Библиотеката на конфликтни пешаци — со агент DreamerV3 чија функција за награда намерно го изоставува безбедносното обликување, со што овозможува да се проучи однесувањето на агентот кон пешаците и ограничувањата на тоа однесување. Таа воведува методологија за интроспекција на претставувачкиот модел што ги прави замислените иднини на агентот видливи и мерливи, а детална карактеризација на однесувањето под конфликтен притисок ги поврзува надворешните резултати со внатрешната дијагноза. Заедно, овие придонеси го адресираат јазот што го откри претходниот преглед.

Глава 3

Теоретски и технички основи

Оваа глава ја поставува математичката и алгоритамската основа за поглавјата што следуваат. Материјалот е организиран така што го следи текот на информацијата низ агентот: суровите пиксели влегуваат во конволуциски енкодер, семантичките канали и каналите за длабочина се создаваат од сензорите на симулаторот, рекурентните и стохастичките модули пренесуваат компресирана состојба на околината низ времето, а парот актер-критичар учи полиса на однесување од замислени траектории. Секој дел завршува со кратка забелешка што го поврзува поимот со неговата конкретна улога во тезата.

Изложувањето го следи редоследот на зависностите: перцепциските компоненти — конволуциски мрежи, семантичка сегментација, инверзна длабочина, рекурентни мрежи и варијациски автоенкодери — се воведени прво (3.1), бидејќи RSSM ги претпоставува. Потоа следи поттикнуваното учење (3.2), кое го обезбедува математичкиот јазик за задачата на воzeње, а завршокот е DreamerV3 (3.3) и симулаторот CARLA (3.4), кои заедно ја формираат целосната експериментална поставка.

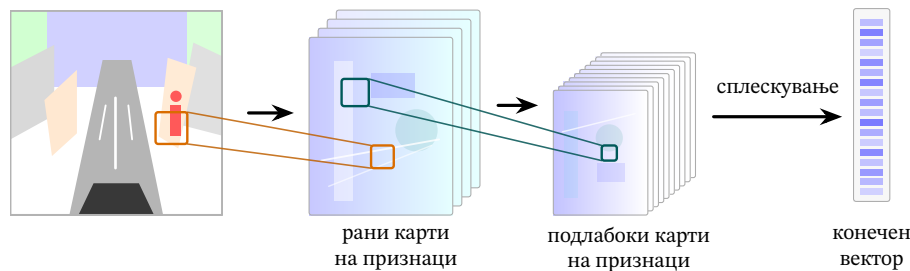
3.1 Основи на длабоко учење

Овој дел ги воведува компонентите на длабокото учење на кои се потпира агентот: конволуциски мрежи за просторна перцепција, семантичка сегментација и кодирање на длабочина за обработка на сензорскиот влез, рекурентни мрежи за временска меморија и варијациски автоенкодери за веројатносно латентно претставување.

3.1.1 Конволуциски невронски мрежи

Конволуциска невронска мрежа (CNN, англ. Convolutional Neural Network) извлекува просторно структурирани признаци од влезови како слики, кај кои податоците се организирани матрично [45]. Основната операција е конволуцијата: научлив конволуциски филтер $\mathbf{W} \in \mathbb{R}^{k \times k \times C_{\text{in}}}$ се поместува низ просторните димензии на влезот $\mathbf{x} \in \mathbb{R}^{H \times W \times C_{\text{in}}}$, создавајќи скаларен излез на секоја положба и така формирајќи карта од признаци $\mathbf{F} \in \mathbb{R}^{H' \times W' \times C_{\text{out}}}$. Повеќе филтри се применуваат паралелно, при што секој се специјализира за различен локален образец. По секоја конволуција следи нелинеарна активација — најчесто Rectified Linear Unit, $\text{ReLU}(x) = \max(0, x)$ — што ѝ овозможува на мрежата да претставува нелинеарни функции од влезот.

Два просторни хиперпараметри ја определуваат геометријата на секој слој. *Големината на јадро* k го определува просторниот опфат на рецептивното поле на секој филтер. *Прескоко* s определува колку се поместува филтерот меѓу две примени. Прескок од два ја преполовува секоја просторна димензија и така постигнува просторно потсемплирање. Кога се редат слоеви со прескок 2, просторната резолуција постепено се огрубува, а бројот на канали



Слика 3.1: Како CNN чита една сцена. Левиот панел прикажува слика од гледната точка на возилото како суров пикселски влез. Следниот куп прикажува рани карти на признаци. Обележаните прозорци ја илустрираат локалната природа на конволуцијата: мал исечок во една фаза придонесува за локален регион на активација во следната. Подлабокиот слој потоа создава погусто множество помали карти на признаци. Со сплескување на овие активации се добива векторот на признаци десно, каде што просторната слика е компресирана во компактна претстава за понатамошниот секвенциски модел.

расте, со што глобалната структура на сцената се компресирана во густ латентен вектор, како што е прикажано на слика 3.1. Ова просторно компресирање е намерен избор: ситниот просторен детал, кој е потребен за задачи на ниво на пиксел, се отфрла во полза на компактни претстави погодни за секвенциско моделирање.

CNN енкодерот е првата фаза од перцепцискиот систем на агентот. Во него заеднички влегуваат три просторни канали: карта на семантичка сегментација кодирана според ризикот, карта на инверзна длабочина и канал што ја прикажува планираната рута. Нивното спојување во компактен вектор му овозможува на агентот подоцна да гради единствена претстава на сцената во која се поврзани ризикот, геометријата и односот кон планираната патека.

3.1.2 Семантичка сегментација

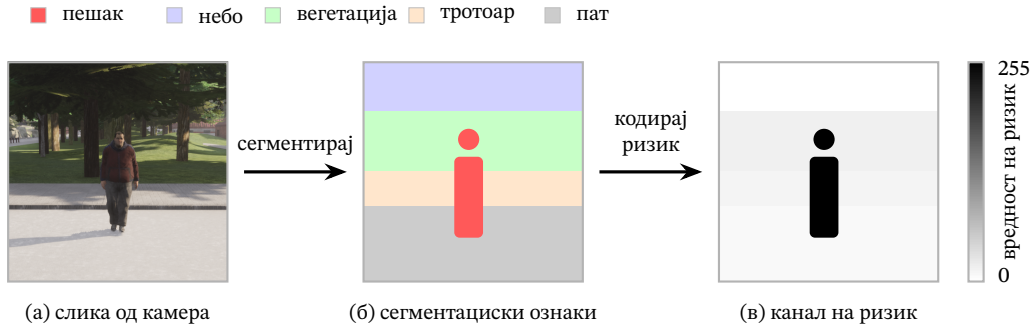
Семантичката сегментација е задача во која на секој пиксел од сликата му се доделува класна ознака [12]. За разлика од детекцијата на објекти, која ги локализира објектите со обиколни кутии, сегментацијата создава густа класна карта усогласена по пиксели, што го зачувува точниот просторен опфат на секој регион. За автономно возење, релевантни класи се: површината на патот, тротоарот, возилата, пешаците, зградите и вегетацијата.

Стандардната архитектура за семантичка сегментација е мрежа од типот енкодер-декодер. Енкодерот најчесто е основна CNN-мрежа што извлекува претстави на признаци додека ја намалува просторната резолуција. Декодерот потоа ги зголемува овие претстави назад до првобитната резолуција на сликата, враќајќи го финиот просторен детал преку врски за прескок што пренесуваат високорезолуциски признаци директно од слоевите на енкодерот до соодветните слоеви на декодерот [71]. Последниот слој применува softmax по пиксел низ речникот на класи за да создаде карта на веројатности.

Во оваа теза, CARLA обезбедува точни семантички ознаки по пиксел, па агентот добива чиста класна карта наместо излез од научен сегментатор. Таа карта се сведува на едноканално претставување кодирано според ризикот, прикажано на слика 3.2, во кое пешаците добиваат највисоки вредности, а позадинските класи се потиснати. Овој канал е главниот сигнал преку кој агентот ги перципира пешаците и нивната безбедносна релевантност.

3.1.3 Проценка на длабочина и инверзна длабочина

Проценката на длабочина го добива растојанието од камерата до секоја видлива точка на површината, создавајќи карта на длабочина по пиксел. Во реални системи, длабочината најчесто се проценува на еден од три начини. Стереопарови слики ја добиваат длабочината преку триангулација на дисперзијата меѓу соодветни пиксели во калибриран лево-десен пар



Слика 3.2: Семантичката сегментација доделува класна ознака на секој пиксел од сликата од камерата, создавајќи густа просторна карта на тоа што е присутно во сцената. Панелот (а) ја прикажува суровата слика од камерата. Панелот (б) го прикажува добиениот сегментациски преклоп, каде што секој пиксел е обоен според својата класа. Панелот (в) ја прикажува едноканалната карта кодирана според ризикот што се користи во оваа теза: секоја класа е заменета со скаларна вредност пропорционална на нејзината безбедносна релевантност, така што пешаците ја добиваат највисоката вредност, а позадинските класи се блиску до нула.

камери, користејќи епиполарна геометрија. Сензори со структурирана светлина, како LiDAR, ја мерат длабочината директно преку времето на прелет на ласерски пулс, создавајќи редок 3D облак од точки што потоа се проектира во просторот на сликата. Монокуларните невронски мрежи ја заклучуваат длабочината од една слика, користејќи заклучоци за перспектива, текстурни градиенти, прекривање и претходно научени претпоставки за сцената [44].

Без оглед на начинот на добивање, суровата линеарна длабочина не е секогаш најкорисната претстава. Наместо тоа, картата на длабочина може да се преобрази во *инверзна длабочина*:

$$d_{\text{inv}}(u, v) = \frac{1}{d(u, v) + \varepsilon} \quad (3.1)$$

каде што $d(u, v)$ е суровата длабочина во метри на пикселот (u, v) , а ε е мала константа што спречува делење со нула кај многу блиски површини. Инверзната длабочина има својство еднаквите прирастувања да одговараат на еднакви аголни промени во перспективата на камерата: објектите на блиско растојание, каде промената на визуелниот агол по единица растојание е голема, се претставуваат со висока прецизност, додека далечните објекти, каде оваа аголна промена е мала, се компресираат кон нула. Ова распределување е прикажано на слика 3.3.

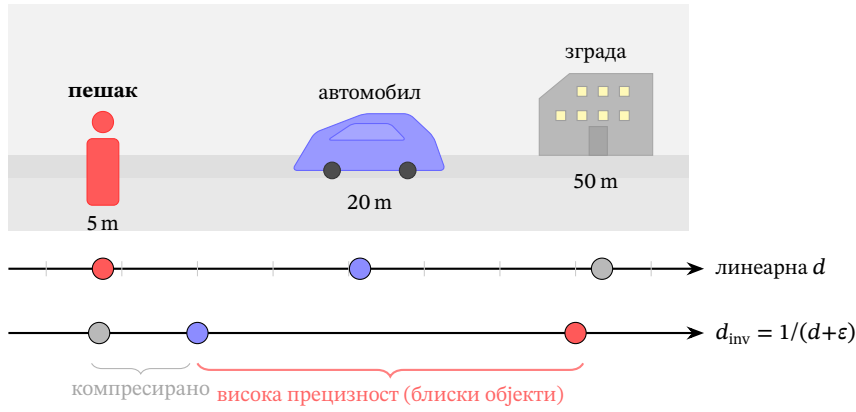
Во оваа теза, CARLA обезбедува точна карта на длабочина по пиксел, која се преобразува во инверзна длабочина и се нормализира на $[0, 1]$. Така агентот добива јасен просторен сигнал за тоа колку се блиску објектите во сцената. Во комбинација со сегментацијата кодирана според ризикот, овој канал му помага на агентот попрецизно да ги разликува блиските и критичните пешаци.

3.1.4 Рекурентни невронски мрежи

Рекурентна невронска мрежа (RNN, англ. Recurrent Neural Network) обработува секвенциски податоци така што одржува скриена состојба $\mathbf{h}_t \in \mathbb{R}^d$ што се ажурира на секој временски чекор како функција од тековниот влез $\mathbf{x}_t$ и претходната скриена состојба [73]:

$$\mathbf{h}_t = f(\mathbf{h}_{t-1}, \mathbf{x}_t; \theta) \quad (3.2)$$

каде што θ ги означува научливите параметри на функцијата за ажурирање f . Скриената состојба акумулира информација низ секвенцата, правејќи ја достапна на подоцнежните чекори без експлицитно чување на сите минати влезови. Ова својство е суштинско за делумно



Слика 3.3: Преобразбата во инверзна длабочина ја дава предноста на блиското поле. Горниот ред прикажува три објекти на различни растојанија онака како што би се појавиле во сцената. На оската на линеарна длабочина (средина), нивните положби се рамномерно распоредени. Под преобразбата на инверзна длабочина $d_{\text{inv}} = 1/(d+\epsilon)$, блискиот пешак (5 m) се пресликува во голема вредност и зафаќа широк, јасно разликувачки регион од каналот, додека далечната зграда (50 m) се компресирана близу до нула. Оваа нелинеарна претстава го распределува перцепцискиот капацитет таму каде што ризикот од судир е најголем.

набљудливи околина, каде што тековното набљудување не ја задава целосно состојбата на системот и агентот мора да ја обедини информацијата низ времето.

Канонската RNN користи едноставно нелинеарно ажурирање:

$$\mathbf{h}_t = \tanh(\mathbf{W}_h \mathbf{h}_{t-1} + \mathbf{W}_x \mathbf{x}_t + \mathbf{b}) \quad (3.3)$$

каде што $\mathbf{W}_h$ и $\mathbf{W}_x$ се научени матрици на тежини, а $\mathbf{b}$ е вектор на отстапување. Во практика, оваа формулација страда од *проблемот на исчезнувачки градиенти*: при враќање на градиентот низ времето, градиентите се множат со истата матрица на тежини $\mathbf{W}_h$ на секој чекор, поради што експоненцијално се намалуваат или растат [4]. Градиенти што исчезнуваат не можат да пренесат сигнал за учење низ долги секвенци. Градиенти што експлодираат, пак, создаваат нумеричка нестабилност.

Рекурентната мрежа со контролни порти (GRU, англ. Gated Recurrent Unit) го адресира овој проблем преку научени порти зависни од влезот, кои го контролираат текот на информацијата во и низ скриената состојба [16]. Две порти го управуваат ажурирањето:

$$\mathbf{r}_t = \sigma(\mathbf{W}_r[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_r) \quad (3.4)$$

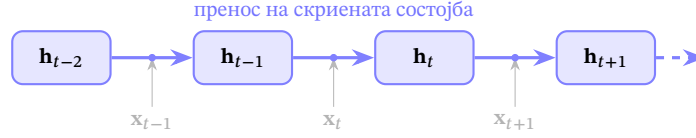
$$\mathbf{u}_t = \sigma(\mathbf{W}_u[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_u) \quad (3.5)$$

каде што $\sigma(\cdot)$ е сигмоидната функција што ги стеснува вредностите во $(0, 1)$, $[\mathbf{h}_{t-1}, \mathbf{x}_t]$ ја означува конкатенацијата на претходната скриена состојба и тековниот влез, а $\mathbf{W}_r$, $\mathbf{W}_u$, $\mathbf{b}_r$, $\mathbf{b}_u$ се научени матрици на тежини и вектори на отстапување.

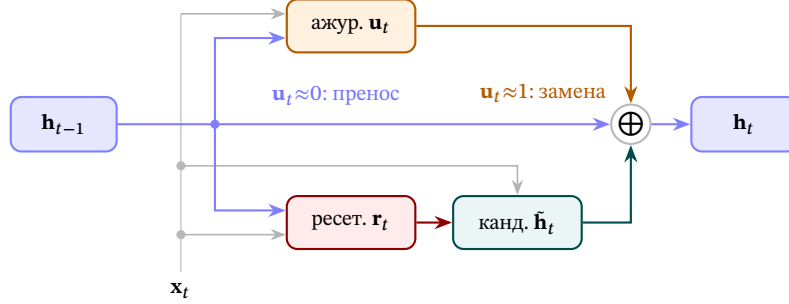
Портирање за ресетирање $\mathbf{r}_t \in [0, 1]^d$ одредува колку од претходната скриена состојба ќе придонесе во кандидатското ажурирање. Кога $\mathbf{r}_t \approx 0$, единицата ја игнорира претходната состојба и се однесува како нова активација. Кога $\mathbf{r}_t \approx 1$, претходната состојба придонесува целосно. *Портирање за ажурирање* $\mathbf{u}_t \in [0, 1]^d$ ја контролира мешавината меѓу претходната скриена состојба и новата кандидатска состојба:

$$\tilde{\mathbf{h}}_t = \tanh(\mathbf{W}_h[\mathbf{r}_t \odot \mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_h) \quad (3.6)$$

$$\mathbf{h}_t = (1 - \mathbf{u}_t) \odot \mathbf{h}_{t-1} + \mathbf{u}_t \odot \tilde{\mathbf{h}}_t \quad (3.7)$$



(а) Скриената состојба со текот на времето.



(б) Внатрешни порти на еден GRU-чекор.

Слика 3.4: Скриената состојба се пренесува низ секвенцата, акумулирајќи контекст од секој нов влез. На секој чекор, две порти одлучуваат што ќе се случи: портата за ресетирање одредува колку од минатото ќе влијае на кандидатското ажурирање, а портата за ажурирање одредува колкав дел од тој кандидат ќе ја замени постојната меморија. Кога портата за ажурирање е затворена, состојбата се пренесува непроменета — директен пат за градиентот што ѝ овозможува на мрежата да задржи информација низ произволно долги секвенци.

каде што $\odot$ означува множење по елементи, $\tilde{\mathbf{h}}_t$ е кандидатската нова состојба, а $\mathbf{W}_h, \mathbf{b}_h$ се матрицата на тежини и отстапувањето за пресметката на кандидатот.

Кога $\mathbf{u}_t \approx 0$, претходната скриена состојба се пренесува непроменета — речиси совршен пат на копирање што им овозможува на градиентите да течат назад низ времето без слабеење. Токму тоа е решението на исчезнувачките градиенти: мрежата учи кога да ја ажурира состојбата, а кога да ја зачува, како што е илустрирано на слика 3.4.

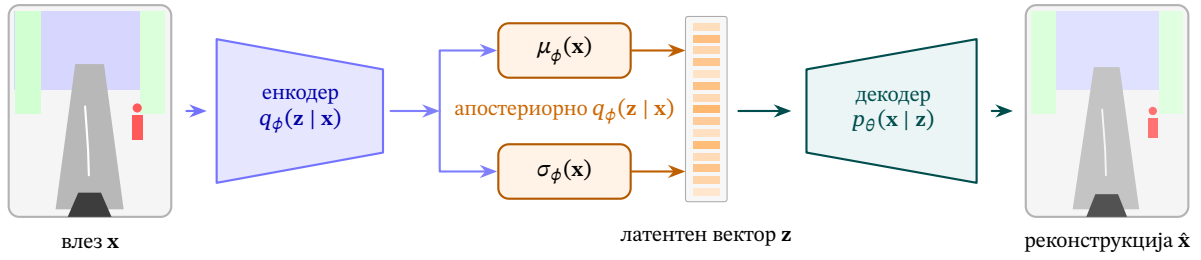
GRU служи како детерминистичко рекурентно јадро на претставувачкиот модел. Тој ја одржува временската меморија на сцената, комбинирајќи историја на набљудувања и акции во скриена состојба од која подоцна се изведуваат предвидувањата. Затоа квалитетот на оваа меморија директно влијае врз тоа колку добро агентот ги следи и предвидува пешаците низ време.

3.1.5 Варијациски автоенкодери

Варијациски автоенкодер (VAE, англ. Variational Autoencoder) е веројатносен генеративен модел што учи компактна и структурирана латентна претстава на една распределба на податоци [40]. Тој се состои од две компоненти: *енкодер* $q_\phi(\mathbf{z}|\mathbf{x})$, со параметри ϕ , што пресликува набљудување $\mathbf{x}$ во распределба над латентни кодови $\mathbf{z}$, и *декодер* $p_\theta(\mathbf{x}|\mathbf{z})$, со параметри θ , што го пресликува латентниот код назад во просторот на набљудување. Латентниот код не е точката проценка, туку случајна променлива, што ја регуларизира претставата и му овозможува на моделот да создава нови примероци преку семплирање од латентната априорна $p(\mathbf{z})$. Моделот се тренира да ја максимизира лог-веродостојноста на податоците $\log p_\theta(\mathbf{x})$. Бидејќи директната максимизација е нерешлива — бара интеграција по сите можни латентни кодови — VAE наместо тоа максимизира долна граница на лог-веродостојноста, позната како ELBO:

$$\mathcal{L}_{\text{ELBO}} = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})} [\log p_\theta(\mathbf{x}|\mathbf{z})] - D_{\text{KL}}(q_\phi(\mathbf{z}|\mathbf{x}) \parallel p(\mathbf{z})) \quad (3.8)$$

Првиот член е *зајубајба на реконструкција*: очекуваната лог-веродостојност на податоци-



Слика 3.5: Како VAE работи. VAE го компресира набљудувањето во распределба, а не во една точка. Енкодерот q_ϕ создава апостериорна средина и варијанса. Примерокот $\mathbf{z}$ се добива преку трикот за репараметризација и се предава на декодерот p_θ , кој го реконструира влезот.

те под декодерот, усреднета преку латентни кодови семплирани од енкодерот. Овој член го наградува моделот кога може да го реконструира влезот од латентниот код. Вториот член е *KL дивергенцијата* меѓу распределбата на енкодерот и априорната распределба: таа го казнува енкодерот ако создава распределба што значително се разликува од априорната распределба, со што го регуларизира латентниот простор.

KL дивергенцијата $D_{\text{KL}}(q\|p)$ го мери информациското засилување кога за опис на случајна променлива се користи q наместо p . За Гаусови распределби $q = \mathcal{N}(\mu, \sigma^2 \mathbf{I})$ и $p = \mathcal{N}(\mathbf{0}, \mathbf{I})$, таа има затворена форма:

$$D_{\text{KL}}(q\|p) = \frac{1}{2} \sum_j (\mu_j^2 + \sigma_j^2 - \log \sigma_j^2 - 1) \quad (3.9)$$

Минимизирањето на овој член го турка енкодерот кон априорната распределба: средната кон нула, а варијансата кон единица. Загубата ELBO ги урамнотежува овие две цели: квалитетот на реконструкцијата и регуларизацијата на латентниот простор. Модел што ја минимизира само загубата на реконструкција би колабирал во детерминистички енкодер со нулта варијанса. Модел што ја минимизира само KL дивергенцијата, пак, целосно би ги игнорирал податоците и би произведувал деградирано усогласување со априорната распределба.

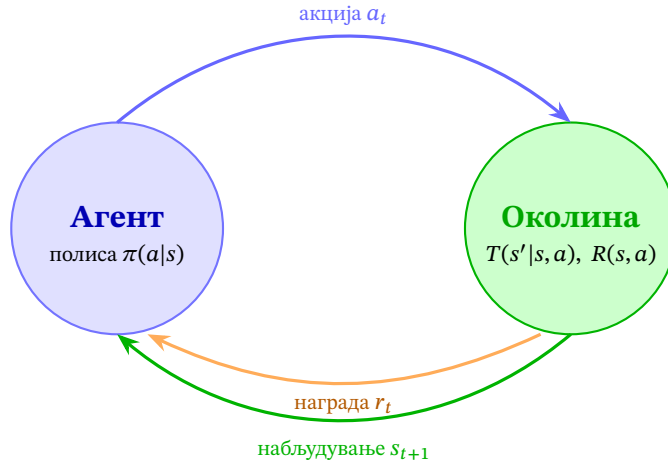
Тренирањето на VAE бара пресметување градиенти на ELBO во однос на параметрите на енкодерот ϕ . Членот на реконструкција вклучува очекување под $q_\phi(\mathbf{z}|\mathbf{x})$, кое зависи од ϕ . Стандардната Монте Карло проценка на градиентот на ова очекување има висока варијанса. *Трикој за репараметризација* го решава ова така што латентниот примерок го изразува како детерминистичка функција од параметрите шум од независна променлива:

$$\mathbf{z} = \mu_\phi(\mathbf{x}) + \sigma_\phi(\mathbf{x}) \odot \epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (3.10)$$

Тука $\odot$ означува множење по елементи. Бидејќи случајноста сега е во ϵ , која не зависи од ϕ , градиентот $\nabla_\phi \mathcal{L}_{\text{ELBO}}$ може да се пресмета со стандардно враќање на градиентот низ μ_ϕ и σ_ϕ . Со тоа тренирањето на VAE станува ефикасно и стабилно.

Континуираната Гаусова формулација погоре е класичниот VAE, но стохастичката латентна променлива може да зема и дискретни категоријални вредности. Категоријален VAE ја заменува Гаусовата распределба со категоријална распределба над K класи. Бидејќи операторот argmax на една категоријална распределба не е диференцијабилен, при тренирањето се користи *straight-through проценувач на градиент*: во напредниот премин се семплира засебно-кодирачки вектор од категоријалната распределба, но во повратниот премин градиентот се пропушта директно низ softmax, третирајќи го примерокот како да е самиот softmax-излез [5]. Со тоа дискретниот VAE може да се тренира од крај до крај, а притоа ги задржува придобивките од компактоста и раздвојливоста на структурирана дискретна претстава.

Рамката VAE, илустрирана на слика 3.5, лежи во основата на стохастичката компонента на RSSM. Енкодерот q_ϕ ја прима претставата на сцената од CNN — компресирана претста-



Слика 3.6: Циклусот на интеракција меѓу агентот и околината. На секој чекор, агентот ја применува полисата π за да избере акција, како управување и гас, која околината потоа ја извршува. Околината го враќа следното набљудување и скаларна награда. Целта на агентот е да ја пронајде полисата што го максимизира очекуваниот дисконтиран поврат $G_t = \sum_k \gamma^k r_{t+k}$ низ неговото работење. Во оваа теза, околината е симулаторот CARLA.

ва на каналите за сегментација кодирана според ризикот, инверзната длабочина и водилките — и создава апостериорна распределба над стохастичка латентна состојба $\mathbf{z}_t$. Следејќи го дизајнот на DreamerV3, ова апостериорно е множество од 32 независни категоријални распределби, секоја над 32 класи, со што се формира 1024-димензионална дискретна претстава што е истовремено компактна и изразна. Декодерот p_θ го реконструира набљудувањето од семплираната $\mathbf{z}_t$ во комбинација со скриената состојба на GRU, а истата латентна состојба се користи и за предвидување награди и сигнали за продолжување на епизодата. KL-членот ја регуларизира $\mathbf{z}_t$ кон научено априорно, што е токму распределбата што се користи за време на проекции со имагинација кога нема реално набљудување. Straight-through проценувачот на градиент го прави овој чекор на дискретно семплирање диференцијабилен, така што целата низа од сурови пиксели до актер и критичар се тренира од крај до крај преку ELBO-целта.

3.2 Поттикнувано учење

Поттикнуваното учење го формализира проблемот на секвенциско одлучување под неизвесност [81]. Агентот стапува во интеракција со околината така што ја набљудува нејзината состојба, избира акција, прима скаларна награда и преминува во нова состојба. Целта на агентот е да открие полиса — пресликување од состојби во акции — што ја максимизира кумулативната награда низ времето. Основниот циклус на интеракција е прикажан на слика 3.6. За разлика од надгледуваното учење, кое бара означени влезно-излезни парови, поттикнуваното учење учи од последиците на сопствените одлуки на агентот.

Стандардната математичка рамка за RL е Марков процесот на одлучување (MDP, англ. Markov Decision Process), дефиниран како петорка $(\mathcal{S}, \mathcal{A}, T, R, \gamma)$:

- $\mathcal{S}$: множеството состојби што може да ги заземе околината.
- $\mathcal{A}$: множеството акции што му стојат на располагање на агентот.
- $T(s' | s, a)$: транзициската функција, што ја дава веројатноста да се достигне состојба s' по преземање акција a во состојба s .
- $R(s, a)$: функцијата за награда, што го дава непосредниот скаларен сигнал за преземање акција a во состојба s .

- $\gamma \in [0, 1)$: дисконтната стапка, што ги тежи идните награди во однос на непосредните.

Марковото својство вели дека транзициската функција и функцијата за награда зависат само од тековната состојба и акција, а не од историјата на претходните состојби. Токму ова својство го прави MDP корисен: оптималната акција во дадена состојба е независна од тоа како агентот стигнал дотаму.

Агентот се стреми да го максимизира очекуваниот дисконтиран поврат од секоја состојба:

$$G_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k} \quad (3.11)$$

каде што r_{t+k} е скаларната награда примена k чекори по времето t . Дисконтната стапка γ го определува ефективниот хоризонт на планирање. Кога $\gamma = 0$, агентот е целосно кусоглед и се грижи само за непосредната награда. Како што γ се приближува кон 1, агентот ги вреднува далечните награди речиси исто колку и непосредните.

Во многу реални проблеми, агентот не ја набљудува целата состојба на околината. Делумно набљудлив Марков процес на одлучување (POMDP, англ. Partially Observable Markov Decision Process) го проширува MDP со воведување функција на набљудување $O(o | s)$: агентот прима набљудување o семплирано од распределба условена со вистинската состојба s , наместо директно да го набљудува s . Под делумна набљудливост, Марковото својство повеќе не важи за набљудувањата — оптималната акција зависи од историјата на претходни набљудувања, а не само од тековното. Во автономното возење, делумната набљудливост се појавува природно: намерата на пешакот, односно дали има намера да го премине патот, е скриена, а набљудливи се само неговата положба и движење. За да се адресира тоа, потребен е модел што одржува внатрешна состојба што ја сумира историјата на набљудувањата и му овозможува на агентот да формира верувања за ненабљудуваните аспекти на околината. Рекурентниот претставувачки модел применет во оваа теза има токму таква форма.

Две основни величини во RL се функцијата на вредност на состојба V^π и функцијата на вредност на акција Q^π :

$$V^\pi(s) = \mathbb{E}_\pi[G_t | s_t = s] \quad (3.12)$$

$$Q^\pi(s, a) = \mathbb{E}_\pi[G_t | s_t = s, a_t = a] \quad (3.13)$$

Функцијата $V^\pi(s)$ е очекуваниот поврат од состојбата s под полисата π , додека $Q^\pi(s, a)$ е очекуваниот поврат кога акцијата a се презема во состојбата s , а полисата π потоа продолжува да се следи. Овие величини му даваат на агентот начин да оцени колку е добра дадена состојба или акција, без експлицитно да ги набројува сите идни траектории.

Двете функции на вредност задоволуваат рекурзивно условување, познато како *Белманови равенки* [81], кое ја разложува вредноста на една состојба на непосредната награда плус дисконтираната вредност на следбената состојба:

$$V^\pi(s) = \mathbb{E}_\pi[r_t + \gamma V^\pi(s_{t+1}) | s_t = s] \quad (3.14)$$

$$Q^\pi(s, a) = \mathbb{E}_\pi[r_t + \gamma V^\pi(s_{t+1}) | s_t = s, a_t = a] \quad (3.15)$$

Клучниот увид во Белмановите равенки е дека тие го претвораат проблемот на планирање на бесконечен хоризонт во едночекорен проблем: ако агентот ја знае вредноста на сите следбени состојби, може да ја пресмета вредноста на тековната состојба без да гледа подалеку во иднината. Ова својство на потпирање врз сопствени процени е темел на речиси сите практични RL алгоритми — и на пристапите засновани на вредност, што директно решаваат за V или Q , и на методите со градиент на полиса, што ги користат проценките на вредност како основа за намалување на варијансата.

Наместо секоја транзиција (s_t, a_t, r_t, s_{t+1}) да се отфрли по една единствена употреба, RL агентите обично ги складираат транзициите во *бафер со искуства* — голема меморија со фик-

сна големина, од која при тренирањето се земаат мини-пакети примероци. Овој бафер служи за две цели: ја прекинува временската корелација меѓу последователните транзиции, со што ја подобрува стабилноста на учењето со градиент, и овозможува секоја транзиција да се употреби повеќепати, со што се подобрува ефикасноста по примероци. Во DreamerV3, баферот со искуства е централен за двете фази на тренирање: зачуваните секвенци се користат за тренирање на претставувачкиот модел, а апостериорните состојби пресметани од тие секвенци служат како почетни точки за замислените проекции со кои се тренираат актерот и критичарот.

Формализмот на MDP го дава математичкиот јазик за задачата на возење: состојбата ја опфаќа сцената околу возилото, акцијата е континуирана наредба за управување, а наградата го спојува напредокот со избегнувањето судири. Проширувањето кон POMDP е важно бидејќи намерата на пешакот не е директно набљудлива. Затоа во оваа теза е потребен рекурентен претставувачки модел што ја акумулира историјата на набљудувањата и формира верување за динамиката на пешаците.

3.2.1 Поттикнувано учење без модел

Поттикнуваното учење без модел учи полиса или функција на вредност директно од интеракциите со околината, без да конструира експлицитен модел на транзициската динамика. Агентот ја третира околината како црна кутија: испробува акции, набљудува награди и следни состојби и ги ажурира своите параметри за да ги подобри идните перформанси. Постојат две широки групи на такви пристапи, разграничени според тоа што точно се учи.

Методите засновани на вредност ја учат функцијата на вредност на акција $Q^\pi(s, a)$ и потоа ја изведуваат полисата со избирање на акцијата со највисока проценета вредност. Основен пример е Q-учењето [90]. Тоа го ажурира Q кон цел добиена со самопроцената, користејќи ја Белмановата равенка за оптималност:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t) \right] \quad (3.16)$$

каде што $\alpha \in (0, 1]$ е стапката на учење, а a' ги опфаќа сите достапни акции во состојбата s_{t+1} . Длабока Q-мрежа (DQN, англ. Deep Q-Network) го проширува Q-учењето на високодимензионални простори на состојби така што го претставува Q со невронска мрежа и користи бафер со искуства и периодично ажурирана целна мрежа за стабилност при тренирањето. DQN покажа дека RL може да учи директно од сурови пикселски набљудувања [55], но неговото потпирање на max-операторот го ограничува на дискретни простори на акции, што го прави несоодветен за континуираното управување при возење.

Методите со градиент на полиса директно ја параметризираат полисата π_θ и ја оптимизираат со движење по градиентот на очекуваниот поврат $J(\theta) = \mathbb{E}_\pi[G_t]$. REINFORCE е основниот алгоритам во оваа група [91]. Тој дава непристрасна проценка на овој градиент:

$$\nabla_\theta J = \mathbb{E}_\pi [\nabla_\theta \log \pi_\theta(a_t | s_t) \cdot G_t] \quad (3.17)$$

каде што G_t е Монте Карло повратот од времето t . Иако е непристрасна, оваа проценка има висока варијанса бидејќи G_t содржи шум во однос на вистинската вредност. Намалувањето на варијансата е централниот предизвик: практичните алгоритми го заменуваат G_t со понисковаријансна основа, што води до рамката актер-критичар.

Proximal Policy Optimisation (PPO) [75] ги стабилизира ажурирањата на градиентот на полиса така што го отсекува односот на веројатности меѓу новата и старата полиса, спречувајќи разрушително големи чекори. Soft Actor-Critic (SAC) [28] максимизира комбинирана цел составена од очекуван поврат и ентропија на полисата, поттикнувајќи истражување и обезбедувајќи робусни резултати на задачи со континуирано управување. И двата алгоритма се

широко употребувани основи за споредба во истражувањето на автономното возење.

Алгоритмите без модел бараат секое ажурирање на полисата да се темели на реални транзиции во околината. Во домени во кои интеракцијата со околината е скапа, како симулацијата на возење, каде што секој чекор бара физичка пресметка, рендерирање и обработка на сензори, оваа неефикасност по примероци станува практичен теснец. Методите без модел, исто така, не нудат механизам за интроспекција на тоа што агентот очекува за иднината: полисата е реактивно пресликување од состојби во акции, без достапен внатрешен модел на идната динамика. Токму оваа непрозирност ги оправдува пристапите со модел.

Поттикнуваното учење без модел е референтната класа за споредба наспроти која е поставен DREAMERV3. Неефикасноста по примероци кај методите без модел ги прави буџетите за тренирање од 10^7 чекори, што се користат во оваа теза, неприпрактични без претставувачки модел. Отсуството на внатрешен модел што може да се разгледува значи дека анализата на имажинацијата не би била можна со агент без модел.

3.2.2 Поттикнувано учење со модел

Поттикнуваното учење со модел учи експлицитен модел на динамиката на околината — функција што ги предвидува следната состојба и наградата од тековната состојба и акција — и го користи овој модел за да ја подобри полисата без дополнителни реални транзиции. Научениот модел може да се употреби на два суштински различни начина.

Управувањето со предвидување базирано на модел (MPC, англ. Model Predictive Control) го користи научениот модел на динамика во моментот на одлучување: агентот евалуира секвенци од акции така што ги симулира нивните исходи низ моделот, а потоа ја извршува првата акција од најдобрата секвенца. Овој пристап планира во реално време и не бара параметриска полиса т.е. самиот модел служи како подлога за одлучување. MPC е директен, но е пресметковно скап при извршување, бидејќи планирањето мора повторно да се спроведува на секој чекор.

Алтернативата, воведена од серијата алгоритми DREAMER, го користи моделот за да создава синтетичко искуство — *замислени траектории* — врз кое параметриска полиса се тренира со стандардни методи актер-критичар. Полисата се учи за време на тренирањето, па при изведба агентот само ја евалуира мрежата на полисата без пребарување или проекции. Со тоа цената на планирањето се амортизира во фазата на тренирање и се добива реактивна полиса што може да работи со брзината потребна за управување во реално време.

Пристапите со модел нудат две клучни предности. Првата е *ефикасност по примероци*: со учење на модел, агентот може да извлече многу градиентни чекори од секоја реална транзиција, бидејќи моделот генерализира низ баферот со искуства и создава ново замислено искуство. Втората е *интроспекција*: замислените траектории на претставувачкиот модел можат да се декодираат назад во просторот на набљудување, со што се добива прозорец кон тоа што агентот очекува за иднината. Таква интроспекција не е можна кај агенти без модел, кои не одржуваат експлицитен модел на динамиката на околината.

Парадигмата со модел не е само пресметковна погодност, таа е научната основа на оваа теза. Способноста да се декодираат замислени иднини овозможува да се постават прашања за тоа што агентот замислува за пешаците за време на планирањето и дали квалитетот на имажинацијата е во корелација со однесувачките исходи — прашања што не може да се одговорат со агенти без модел. DREAMERV3 е конкретниот алгоритам за поттикнувано учење со модел што се користи низ целата теза.

3.3 DreamerV3

DreamerV3 [33] е алгоритам за поттикнувано учење со модел изграден околу три заеднички тренирани невронски мрежи: *прејдвидувачки модел* што учи компактни претстави на сензорските влезови и ги предвидува исходите од можни акции; *критичар* што ја проценува вредноста на секој предвиден исход; и *актер* што избира акции за да стигне до највредните исходи. Сите три компоненти се тренираат истовремено од репродуцирано искуство додека агентот стапува во интеракција со околината. Клучно е тоа што ниту актерот ниту критичарот никогаш не примаат градиенти од реални транзиции — целото учење на полисата се одвива преку замислени проекции создадени од претставувачкиот модел.

DreamerV3 е трета генерација од серијата Dreamer, при што секоја нова верзија ја гради истата основна архитектура додека ги адресира ограничувањата на претходната:

- *DreamerV1* го вовеле рекурентниот просторен модел на состојби (RSSM, англ. Recurrent State-Space Model) со континуирани Гаусови латентни состојби [31]. Тој покажа дека актер-критичар, трениран целосно на замислени траектории, може да постигне и да надмине перформанси на пристапи без модел на задачи со визуелно управување.
- *DreamerV2* [32] ги замени Гаусовите латентни состојби со дискретни категоријални распределби, со што го подобри квалитетот на претставата и му овозможи на агентот да го совлада Atari бенчмаркот директно од пиксели.
- *DreamerV3* [33] вовеле симетрично-логаритамски предвидувања, нормализација со експоненцијален подвижен просек (ЕМА, англ. Exponential Moving Average) на повратите и балансирање на KL, со што доби стабилно тренирање во домени со различни размери на награда без нагонување хиперпараметри по задача.

Оваа секција ја следи структурата на алгоритмот: најнапред претставувачкиот модел, прво архитектурата па целите на тренирање, потоа учењето на однесувањето, односно актерот и критичарот тренирани на замислени проекции, и на крај тренинг-процесот што ги испреплетува двете фази.

3.3.1 Рекурентниот модел на простор на состојби

RSSM е архитектурата на претставувачкиот модел во јадрото на DreamerV3 [31]. Тој учи компактни претстави на сензорските влезови преку автоенкодирање [40] и овозможува планирање преку предвидување секвенци идни претстави врз основа на можни акции. Неговата клучна одлика е латентна состојба поделена на две компоненти со комплементарни улоги: *детерминистичка* компонента што носи временски контекст низ чекорите и *стохастичка* компонента што ги опфаќа непредвидливите аспекти на околината во секој миг. Заедно, овие компоненти обезбедуваат компактна и секвенциска претстава на светот што поддржува и точно предвидување и ефикасни замислени проекции.

На секој временски чекор t , RSSM одржува моделска состојба $s_t = (\mathbf{h}_t, \mathbf{z}_t)$, каде што:

- $\mathbf{h}_t \in \mathbb{R}^{d_h}$ е *детерминистичката рекурентна состојба*, што ја одржува GRU, при што d_h е димензијата на скриената состојба. Таа ја сумира историјата на минатите состојби и акции и се ажурира детерминистички.
- $\mathbf{z}_t \in \{0, 1\}^{32 \times 32}$ е *стохастичката латентна состојба*, односно примерок од категоријална распределба условена со $\mathbf{h}_t$ и, при обработка на набљудувања, и со тековната претстава на набљудувањето.

Детерминистичката компонента е неопходна затоа што чисто стохастички латентен модел не може да пренесе долгорочни временски зависности: стохастичкиот примерок на секој

чекор е независен, а нема механизам преку кој едно минато набљудување сигурно би влијаело на предвидување многу чекори подоцна. Стохастичката компонента е неопходна затоа што чисто детерминистички модел не може да претстави неизвесност: тој колеба во една единствена предвидена иднина и не може да ја моделира распределбата на можни исходи. Дводелниот дизајн ги обезбедува двете својства истовремено.

RSSM работи во два режима, зависно од тоа дали е достапно тековно набљудување. Кога е достапно реално набљудување $\mathbf{x}_t$ — за време на интеракција со околината и при репродукција од баферот — стохастичката состојба е изведено од рекурентната состојба и од кодираната претстава на набљудувањето:

$$\mathbf{h}_t = \text{GRU}(\mathbf{h}_{t-1}, [\mathbf{z}_{t-1}, \mathbf{a}_{t-1}]) \quad (3.18)$$

$$\mathbf{z}_t \sim q_\phi(\mathbf{z}_t | \mathbf{h}_t, \mathbf{e}_t) \quad (3.19)$$

каде што $\mathbf{e}_t = \text{enc}(\mathbf{x}_t)$ е CNN-кодираната претстава на набљудувањето, $\mathbf{a}_{t-1}$ е акцијата преземена при претходниот чекор, а ϕ ги означува параметрите на претставувачкиот модел што ги делат и апостериорната распределба q_ϕ и априорната распределба p_ϕ . Распределбата q_ϕ е *апостериорната распределба*: таа користи реални сензорски податоци за прецизно да ја определи стохастичката состојба. Овој режим се користи за време на тренирањето за да се конструираат точни состојби за гранките за реконструкција. Кога нема достапно набљудување — за време на проекции со имагинација — стохастичката состојба се предвидува само од рекурентната состојба:

$$\mathbf{h}_t = \text{GRU}(\mathbf{h}_{t-1}, [\mathbf{z}_{t-1}, \mathbf{a}_{t-1}]) \quad (3.20)$$

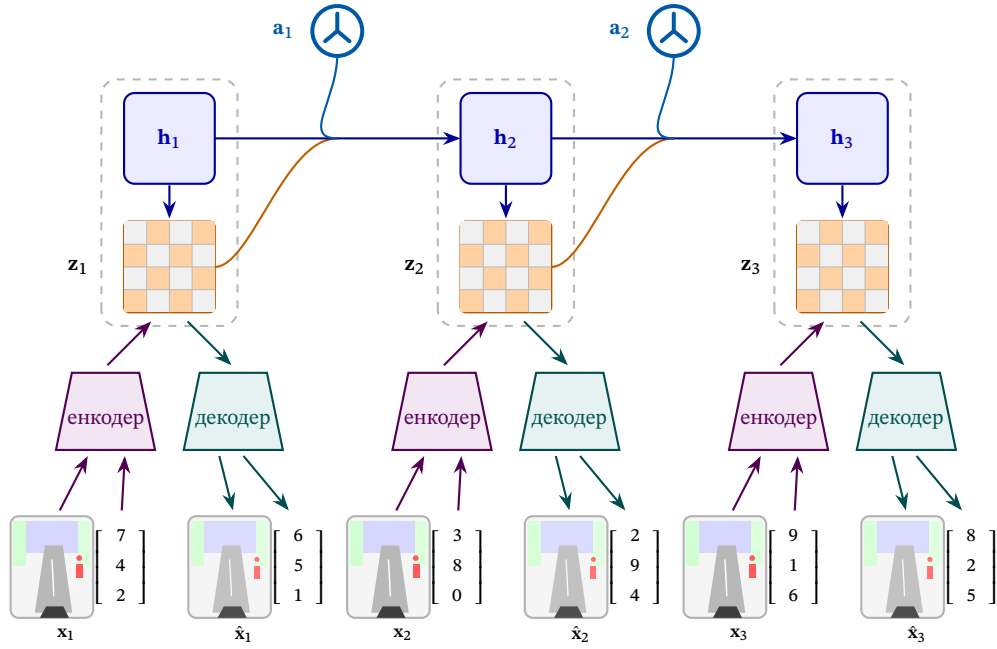
$$\mathbf{z}_t \sim p_\phi(\mathbf{z}_t | \mathbf{h}_t) \quad (3.21)$$

Распределбата p_ϕ е *априорната распределба*: таа мора да предвиди каква треба да биде стохастичката состојба без да ја набљудува околината. Овој режим е механизам на имагинацијата: агентот се придвижува низ времето така што повекепати го применува GRU и семплира од априорната распределба, создавајќи секвенца моделски состојби $\{(\mathbf{h}_t, \mathbf{z}_t)\}_{t=1}^H$, каде што H е хоризонтот на имагинација, без да презема реални акции и без да прима реални набљудувања. Квалитетот на овие замислени секвенци зависи целосно од тоа колку добро априорната распределба p_ϕ научила да ги предвидува стохастичките латентни состојби што апостериорната распределба q_ϕ би ги извела ако ги имаше вистинските набљудувања.

Разликата меѓу тоа што q_ϕ го прима набљудувањето а p_ϕ не, воведува критична асиметрија. За време на тренирањето, апостериорната распределба q_ϕ го прима тековното набљудување и може да произведе точна стохастичка состојба дури и за невообичаени или неочекувани настани. За време на имагинацијата, априорната распределба p_ϕ мора да екстраполира само од рекурентната историја. Ако пешак нагло влезе во сцената, апостериорната распределба може веднаш да го регистрира неговото присуство. Априорната распределба, за време на имагинацијата, може да го предвиди појавувањето на пешакот само ако слична динамика е доволно често среќавана при тренирањето, така што скриената состојба на GRU сигурно ја предвидува. Ретко среќавани динамики — како пешак што зачекорува на патот под кос агол — може да бидат замислени неточно, поради што замислената иднина ќе се разидува од стварната.

3.3.2 Тренирање на претставувачкиот модел

Претставувачкиот модел се тренира врз секвенци семплирани од баферот со искуства составен од реални транзиции во околината $\{(\mathbf{x}_t, \mathbf{a}_t, r_t)\}$, како што е шематски прикажано на слика 3.7. За секоја секвенца, RSSM се развива во режим на набљудување за да создаде секвенца



Слика 3.7: Учење на претставувачкиот модел во DREAMerV3, шематски прикажано низ три временски чекори. На секој чекор, тековното набљудување $\mathbf{x}_t$ — овде илустрирано како слика од гледната точка на возилото заедно со мал вектор — се обработува од енкодер и се пресликува во моделската состојба, претставена како детерминистичка рекурентна компонента $\mathbf{h}_t$ и стохастичка латентна компонента $\mathbf{z}_t$. Рекурентната состојба се пренесува напред низ времето по горната сина врска, додека латентната состојба и акцијата $\mathbf{a}_t$ влегуваат во транзицијата кон следната рекурентна состојба. Декодер реконструира набљудување $\hat{\mathbf{x}}_t$ од тековната моделска состојба, обезбедувајќи го реконструкцискиот сигнал што го тренира претставувачкиот модел да задржи информација релевантна за набљудувањето. Сликата е адаптирана од [33].

апостериорни моделски состојби $\{(\mathbf{h}_t, \mathbf{z}_t)\}$. На овие состојби се прикачени повеќе предвидувачки гранки, при што секоја е тренирана да предвидува различен аспект на околината.

Прикачените предвидувачки гранки се:

- *Декодер*: конволуциски декодер го реконструира набљудувањето $\hat{\mathbf{x}}_t$ од моделската состојба. Реконструкцијата го принудува моделот да задржи доволно информација за да го репродуцира влезот — без овој сигнал, моделот би можел да научи компактна, но неинформативна претстава. За влезните канали што се користат во оваа теза, се применува загуба од вкрстена ентропија на категоријалниот канал за сегментација и загуба од средна квадратна грешка на континуираниот канал на инверзна длабочина.
- *Гранка за награда*: Повеќеслојна мрежа (MLP, англ. Multilayer Perceptron) ја предвидува непосредната награда $\hat{r}_t$ од моделската состојба. Точното предвидување на наградата е суштинско за учењето на однесувањето: замислените награди произведени при тренирањето на актерот и критичарот се сигналот врз кој се учи полисата, па грешките во предвидувањето на наградата директно се претвораат во подоптимално однесување.
- *Гранка за продолжување*: MLP ја предвидува *веројатноста на продолжување* $\hat{\gamma}_t \in (0, 1)$, што ја претставува веројатноста епизодата да продолжи на следниот чекор. За време на проекциите со имагинација, $\hat{\gamma}_t$ ја модулира ефективната дисконтна стапка и ги скратува замислените траектории кога е веројатен прекин.

Целосната загуба на претставувачкиот модел ги спојува сите цели на предвидување со членот за KL регуларизација:

$$\mathcal{L}_{\text{WM}} = \mathcal{L}_{\text{recon}} + \mathcal{L}_{\text{KL}} + \mathcal{L}_{\text{reward}} + \mathcal{L}_{\text{cont}} \quad (3.22)$$

KL-членот ја регуларизира апостериорната распределба кон априорната распределба, така што истовремено ги тренира и мрежата за претстава и предвидувачот на динамика. При тоа DreamerV3 го дели овој член на две асиметрични компоненти со коефициент на рамнотежа $\alpha = 0.8$:

$$\mathcal{L}_{KL} = \alpha \cdot \text{sg}(q) \cdot \log \frac{\text{sg}(q)}{p} + (1 - \alpha) \cdot q \cdot \log \frac{q}{\text{sg}(p)} \quad (3.23)$$

каде што $\text{sg}(\cdot)$ го означува операторот за запирање на градиентот, а q и p се кратенки за апостериорната распределба $q_\phi(\mathbf{z}_t | \mathbf{h}_t, \mathbf{e}_t)$ и априорната распределба $p_\phi(\mathbf{z}_t | \mathbf{h}_t)$, соодветно. Првиот член, со тежина $\alpha = 0.8$, ја ажурира само априорната распределба p кон апостериорната распределба q , тренирајќи го предвидувачот на динамика да ги предвиди идните претстави. Вториот член, со тежина $1 - \alpha = 0.2$, ја ажурира само апостериорната распределба q кон априорната распределба p , регуларизирајќи ја претставата. Оваа асиметрија е намерна: точните априорни распределби се суштински за квалитетот на имагинацијата, додека преголема регуларизација на апостериорната распределба може да ја спречи претставата верно да го кодира тековното набљудување.

Во DreamerV3, за да се спречи пререгуларизација во димензии што носат малку информација, KL-придонесот на секоја категоријална единица се отсекува одоздола на 1 nat. Единиците кај кои априорната и апостериорната распределба веќе се блиски не примаат градиент, со што на претставата ѝ се дозволува слободно да ги користи тие единици за кодирање информација специфична за набљудувањето наместо да го троши капацитетот на усогласување со априорната распределба што и онака е блиска.

Стохастичката состојба $\mathbf{z}_t$ се состои од 32 независни категоријални распределби, секоја над 32 класи, што создаваат 1024-димензионална бинарна претстава кога се конкатенираат како засебно-кодирачки вектори. Во споредба со Гаусова латентна состојба со иста димензионалност, категоријалната претстава има две практични предности: KL дивергенцијата меѓу две категоријални распределби има едноставна затворена форма, без потреба од апроксимација, а дискретната структура создава теснец што ја поттикнува претставата да кодира квалитативни разлики меѓу сцените, наместо континуирани интерполации.

3.3.3 Учење на однесувањето

DreamerV3 ја тренира својата полиса со архитектура актер-критичар — општа класа методи што може да постои и независно од кој било претставувачки модел, но која DreamerV3 ја применува целосно внатре во имагинацијата. Архитектурата одржува две компоненти: *актер* (полисата π_θ , со параметри θ) што избира акции и *критичар* (мрежата на вредност V_ψ , со параметри ψ) што го проценува очекуваниот поврат од секоја состојба. Проценките на вредност на критичарот служат како нисковаријансна основа за градиентот на полисата на актерот, адресирајќи ја високата варијанса на чистите методи со градиент на полиса, како REINFORCE [91]. Двете мрежи работат врз моделски состојби $s_t = (\mathbf{h}_t, \mathbf{z}_t)$ и никогаш не се ажурираат од реални транзиции во околината.

Теоремата за градиент на полиса [81] утврдува дека градиентот на очекуваниот поврат во однос на параметрите на полисата е:

$$\nabla_\theta J(\pi_\theta) = \mathbb{E}_\pi [\nabla_\theta \log \pi_\theta(a_t | s_t) \cdot A^\pi(s_t, a_t)] \quad (3.24)$$

каде што $A^\pi(s, a) = Q^\pi(s, a) - V^\pi(s)$ е *функцијата на предност*: $Q^\pi(s, a)$ е очекуваниот поврат при преземање акција a во состојба s , по што се следи полисата π , а $V^\pi(s) = \mathbb{E}_\pi[Q^\pi(s, a)]$ е очекуваниот поврат усреднет преку сите акции. Предноста мери колку е акцијата a подобра од просечната акција во состојбата s . Акции со позитивна предност се засилуваат, а акции со негативна предност се потиснуваат. Ваквата формулација го центрира градиентот околу нула, намалувајќи ја варијансата во споредба со употребата на сурови поврати G_t .

Критичарот се тренира да го предвидува λ -повратот, односно тежинска мешавина од повеќечекорни цели добиени со самопроцената:

$$R_t^\lambda = r_t + \gamma \left[(1 - \lambda) V_\psi(s_{t+1}) + \lambda R_{t+1}^\lambda \right] \quad (3.25)$$

Тука r_t е непосредната награда, $\gamma \in (0, 1)$ е дисконтната стапка, а $V_\psi(s_{t+1})$ е проценката на вредност на критичарот за следната замислена состојба. Параметарот $\lambda \in [0, 1]$ интерполира меѓу едночечно учење со временска разлика, при $\lambda = 0$, кое има ниска варијанса, но висока пристрасност од апроксимациската грешка на критичарот, и целосни Монте Карло поврати, при $\lambda = 1$, кои имаат висока варијанса, но нулта пристрасност. Критичарот ја минимизира средната квадратна грешка кон λ -повратни цели со запирање на градиентот:

$$\mathcal{L}_{\text{critic}} = \mathbb{E} \left[\left(V_\psi(s_t) - \text{sg}(R_t^\lambda) \right)^2 \right] \quad (3.26)$$

Во практика, $\lambda \approx 0.95$ дава добар компромис меѓу пристрасност и варијанса.

За задачи со континуирано управување, какво што е возењето, актерот параметризира распределба над акциите и го оптимизира очекуваниот λ -поврат преку градиент добиен со репараметризација, аналогно на енкодерот кај VAE. Со тоа се овозможува оптимизација на полисата со градиент без потреба од дискретно множество акции.

Актерот и критичарот не се тренираат врз реални транзиции во околината, туку врз замислени траектории создадени од претставувачкиот модел, како што е шематски прикажано на слика 3.8. На секој чекор на тренирање:

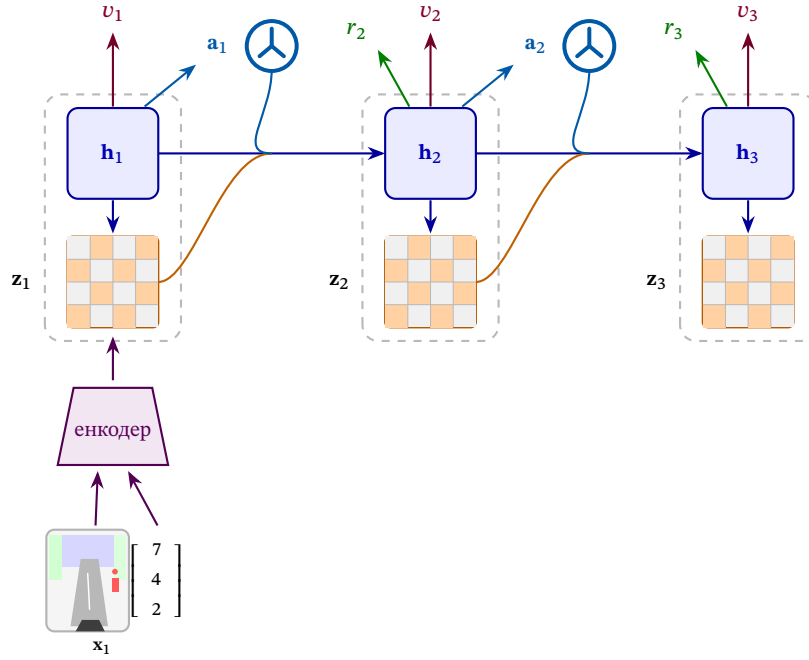
1. Множество почетни моделски состојби $\{s_\tau\}$ се семплира од баферот со реално искуство.
2. Од секоја почетна состојба, претставувачкиот модел се развива во априорен режим во должина од H чекори, при што акциите ги избира тековниот актер: $\mathbf{a}_t \sim \pi_\theta(\cdot | s_t)$. Така се добиваат замислени траектории $\{s_t, \mathbf{a}_t, \hat{r}_t, \hat{\gamma}_t\}_{t=\tau}^{\tau+H}$.
3. Критичарот $V_\psi(s_t)$ се тренира со минимизирање на средната квадратна грешка кон целите R_t^λ . Тие се λ -поврати пресметани од замислените награди и веројатностите на продолжување.
4. Актерот се тренира да го максимизира очекуваниот λ -поврат преку градиент добиен со репараметризација.

Во оваа теза, хоризонтот на имагинација е $H = 32$ чекори, што одговара на 1.6 секунди симулирано возење на 20 Hz — прозорецот во кој актерот мора да ги предвиди и да одговори на опасностите поврзани со пешаци преку замислени идни состојби.

Стандардните невронски мрежи се чувствителни на размерот на своите цели. Кога наградата значително варира по големина — како во функцијата за награда на оваа теза, која се движи од прирасти на напредок низ порти блиски до нула до терминални казни од -300 — наивната регресија може да стане нестабилна поради цели со голема апсолутна вредност. DreamerV3 ја применува симетричната логаритамска трансформација на сите скаларни предвидувања:

$$\text{symlog}(x) = \text{sign}(x) \cdot \log(|x| + 1) \quad (3.27)$$

Декодерот, гранката за награда и критичарот сите предвидуваат цели трансформирани со symlog , а инверзната трансформација се применува при отчитувањето. Symlog ги компресира големите апсолутни вредности, а притоа ги зачувува знакот и поредокот. Таа е приближно линеарна близу до нула и логаритамска за големи вредности, што ѝ овозможува на мрежата да учи предвидувања низ многу редови на големина со една иста архитектура и без посебна нормализација на наградата.



Слика 3.8: Учење на однесувањето во DreamerV3, прикажано низ кратка замислена проекција. Едно реално набљудување $\mathbf{x}_1$ се кодира во почетната моделска состојба $(\mathbf{h}_1, \mathbf{z}_1)$. Од таа почетна точка, претставувачкиот модел се развива кон замислени состојби $(\mathbf{h}_2, \mathbf{z}_2)$ и $(\mathbf{h}_3, \mathbf{z}_3)$, при што акциите $\mathbf{a}_1$ и $\mathbf{a}_2$ ги условуваат транзициите. Критичарот ги предвидува вредностите v_1, v_2 и v_3 од моделските состојби, додека гранката за награда ги предвидува замислените награди r_2 и r_3 за идните чекори. Така, актерот и критичарот се тренираат на траектории во латентен простор создадени од претставувачкиот модел, а не на директни транзиции во околината. Сликата е адаптирана од [33].

Покрај symlog, DreamerV3 врши и динамичка нормализација пред пресметката на загубата на критичарот. Нормализацијата се применува врз вредностите на λ -повратот. Пар експоненцијални подвижни средини (ЕМА) ги следи 5-от и 95-от перцентил на распределбата на повратите. Целите се нормализираат со опсегот меѓу овие перцентили, со што се добива преразмерениот поврат $\tilde{R}_t^\lambda$:

$$\tilde{R}_t^\lambda = \frac{R_t^\lambda}{\max(1, \hat{S}_{95} - \hat{S}_5)} \quad (3.28)$$

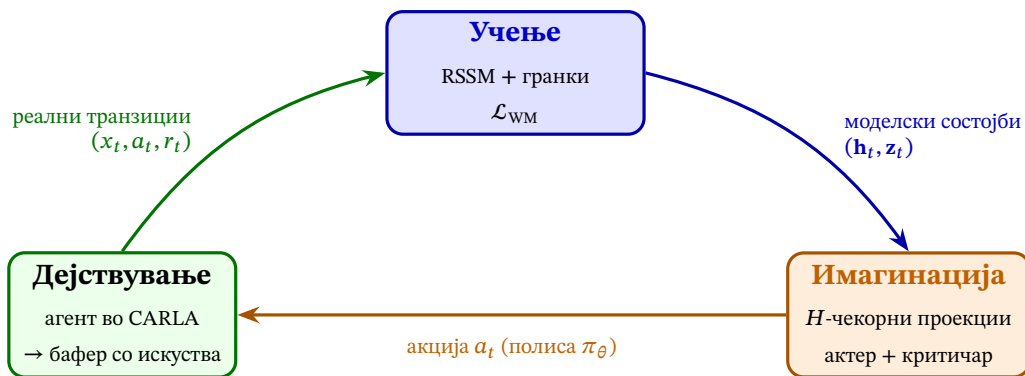
каде што $\hat{S}_{95}$ и $\hat{S}_5$ се тековните ЕМА-проценки на 95-от и 5-от перцентил, соодветно. Изразот $\max(1, \cdot)$ спречува делење со речиси нулт опсег во раните фази на тренирањето. Оваа нормализација осигурува дека размерот на градиентот на критичарот е инваријантен на апсолутната големина на повратите — важно својство кога функцијата за награда вклучува ретки терминални казни со голема вредност заедно со густии награди за напредок со мала вредност.

Актерот, исто така, се тренира преку нормализирани предности. Предноста $A_t = R_t^\lambda - V_\psi(s_t)$ се нормализира со истиот ЕМА-опсег, осигурувајќи дека градиентот на актерот не е доминиран од епизоди со невообичаено големи, позитивни или негативни, поврати.

3.3.4 Тренинг-процес

Целосниот тренинг-циклус на DreamerV3, прикажан на слика 3.9, ја испреплетува интеракцијата со околината со учењето на претставувачкиот модел и на однесувањето. На секоја итерација:

1. Агентот избира акција користејќи го актерот и тековната моделска состојба, што се одржува од претходниот чекор во режим на набљудување на RSSM.



Слика 3.9: Тренирањето на DreamerV3 како циклус од три фази. Во фазата **Дејствување** (зелено), полисата стапува во интеракција со CARLA и ги складира транзициите во баферот со искуства. Во фазата **Учење** (сино), секвенци од баферот го тренираат претставувачкиот модел, односно RSSM и гранките за предвидување, преку $\mathcal{L}_{\text{WM}}$. Во фазата **Имагинација** (портокалово), претставувачкиот модел создава H -чекорни проекции во априорен режим, врз кои се тренираат актерот и критичарот преку замислени награди. Сите три фази се испреплетуваат во секоја итерација на тренирање.

2. Акцијата се извршува во околината, а добиеното набљудување, наградата и ознаката за прекин се складираат во баферот со искуства.
3. Од баферот со искуства се семплира пакет секвенци и се минимизира загубата на претставувачкиот модел, односно равенка 3.22.
4. Користејќи моделски состојби од истиот пакет, се создаваат замислени траектории и се минимизираат загубите на актерот и критичарот.

Чекорите 3 и 4 се изведуваат со тренинг-однос од 512 репродуцирани рамки по секој чекор во околината: на секои два чекора во околината се семплира пакет од 16 секвенци по 64 чекори (1024 рамки) и се изведува по едно градиентно ажурирање на претставувачкиот модел и на мрежите за однесување. Така секоја реална транзиција се употребува многупати низ тренингот, па претставувачкиот модел и полисата може да се тренираат многу потемелно отколку што би дозволил стандарден алгоритам без модел. Овој однос е изводлив затоа што градиентните чекори врз репродуцирани и замислени податоци се пресметковно поевтини од реална интеракција со околината и затоа што RSSM може да генерализира низ баферот со искуства наместо да бара свежо искуство за секое ажурирање.

Моделската состојба на RSSM, $s_t = (\mathbf{h}_t, \mathbf{z}_t)$, е единствената претстава на околината со која располага агентот во секој миг. Врз неа работат сите надолни компоненти — гранките за предвидување, актерот и критичарот. Затоа способноста на RSSM да одржува корисни претстави на пешаците и да ги пренесува низ проекциите со имагинација е клучна за квалитетот на планирањето и однесувањето на агентот.

3.4 Симулаторот CARLA

CARLA (Car Learning to Act) е урбан симулатор за возење со отворен код, развиен во Центарот за машински вид во Барселона [21]. Изграден врз Unreal Engine 4, тој спојува 3D рендерирачки систем со целосен стек за сообраќајна симулација, сеопфатна рамка за сензори со точни референтни податоци и скриптибилен апликациски програмски интерфејс (API, англ. Application Programming Interface) во Python, со што се наметнува како референтна платформа за истражување во автономното возење и како стандардна бенчмарк-околина за агенти за поттикнувано учење засновани на вид. Пример за сцена рендерирана во CARLA е прикажан на слика 3.10.

Физичката симулација на CARLA се потпира на PhysX решавачот за крути тела во Unreal Engine 4. Секое возило е моделирано со физички модел на четири тркала што ги зема предвид распределбата на масата, коефициентите на триење меѓу гума и пат, одот и придушувањето на потпирањето, аеродинамичниот отпор и кривите на вртежен момент на моторот за секое возило. Резултатот е симулатор во кој растојанието на сопирање, подуправувањето и одзивот на возилото на промени на подлогата се физички уверливи, наместо кинематички зададени. Пешаците и велосипедистите се управувани од систем за физика на ликови со целосна геометрија на судир, пасивна артикулирана динамика при удар и двоножно движење управувано од состојбени машини на анимации засновани на снимање на движење.

CARLA ги рендерира сцените преку системот за физички засновано рендерирање (PBR, англ. Physically Based Rendering) на Unreal Engine 4, кој ги моделира одразните својства на материјалите, глобалното осветлување и динамичкото фрлање сенки со стапки погодни за реално време. Системот за временски услови е целосно параметарски: облачност, интензитет на врнежи, акумулација на локви, одразност на влажен пат, густина на магла, брзина на ветер, висина на сонцето и азимут на сонцето можат независно да се поставуваат или континуирано да се менуваат низ опсег, создавајќи услови од јасно пладне до густа магла и силен ноќен дожд. Обезбедени се повеќе од десет именувани временски поставки, како ClearNoon, HardRainSunset и WetCloudyNight, а сите параметри можат динамички да се интерполираат за да се симулираат променливи услови во текот на една епизода. Резолуцијата на рендерирање, квалитетот на текстурите и нивото на антиалијасинг се прилагодливи, што овозможува компромис меѓу визуелната веродостојност и пропусноста на симулацијата при тренирање во голем обем.

CARLA доаѓа со растечка библиотека градски мапи со различен размер и сложеност. Поедноставните мапи, како Town01 и Town02, содржат дволентни патишта со благи кривини и минимална сигнализација. Сложените урбани мапи, како Town03, Town05 и Town10, вклучуваат повеќелентни коловози, Т-раскрсници, четирикраки крстосници, кружни текови, издигнати автопатни делници, тунели и мостовски сегменти. Секоја мапа вклучува целосна патна инфраструктура: линиски ознаки на коловозот, како полни, непрекинати, средишни и рабни линии, семафори со прилагодлив тајминг, знаци за стоп, знаци за давање првенство, знаци за ограничување на брзината, пешачки премини, автобуски постојки и паркинг-простори. Тротоарите, фасадите на зградите, уличната опрема и вегетацијата се присутни со ниво на детал што поддржува 28 различни семантички класни ознаки во излезот на сегментацијата со точни референтни податоци.

CARLA поддржува широк спектар типови учесници во сообраќајот. Флотата возила опфаќа патнички автомобили, спортско-корисни возила, комбиња, камиони, автобуси, мотоцикли и велосипеди, секое со различни физички параметри и високо детални визуелни мрежи. Пешаците се реализирани како артикулирани ликови извлечени од множество модели што се разликуваат по пол, возраст и облека, анимирани со циклични анимации на одење засновани на снимање на движење при различни брзини. Сите агенти се полноправни физички објекти: возилата и пешаците се судираат меѓусебно, со статичната геометрија на сцената и со возилото, создавајќи физички уверлива динамика по ударот и настани на судир што може да се детектираат преку сензорската рамка.

Вградениот Traffic Manager на CARLA обезбедува автопилот за возила што се справува со следење лента, почитување ограничувања на брзината, минување низ раскрсници, почитување светлосна сигнализација, престигнување со менување лента и избегнување судири со други возила. Пешаците имаат поедноставен вграден навигациски режим што ги води по тротоари и преку означени премини, користејќи го внатрешниот граф на водилки на CARLA — доволно за однесување на позадинска толпа, но без модел на намера, двоумење или конфликтни одлуки за преминување. Затоа конфликтните пешачки сценарија мора експлицитно да ги конструира корисникот преку скриптибилното API, што е јазот што ја оправдува Библиотеката на конфликтни пешаци.



Слика 3.10: Сцена рендерирана од симулаторот CARLA, врз Unreal Engine 4. Околината вклучува целосна патна мрежа со линиски ознаки, светлосна сигнализација, пешаци и возила, сето тоа рендерирано со PBR материјали и динамички временски услови.

CARLA користи модел сервер-клиент преку Transmission Control Protocol (TCP). Серверскиот процес ја хостира симулацијата на Unreal Engine: ја придвижува физиката, управува со сообраќајните агенти, ги пресметува сензорските излези и ги испраќа резултатите. Клиентот, најчесто Python процес, се поврзува со серверот, создава актери, издава команди за управување и ги чита сензорските набљудувања. Два режима одредуваат како се синхронизираат овие два процеса. Во *асинхрон режим*, што е стандардниот режим, серверот работи со највисоката стапка на кадри што може да ја постигне, без оглед на брзината на обработка на клиентот. Клиентот испраќа команди и чита податоци со стапката што може да ја постигне, без гаранција дека последователните клиентски набљудувања одговараат на последователни симулациски кадри. Овој режим е погоден за рачно телеуправување и визуелизација. Во *синхрон режим*, серверот застанува на крајот на секој симулациски кадар и чека клиентот експлицитно да го побара следниот чекор пред да продолжи. Токму оваа гаранција дека клиентот го обработува секој симулациски кадар точно еднаш и по редослед е суштинска за поттикнуваното учење, бидејќи прескокнатите кадри би го расипале баферот со искуства и би ја нарушиле временската доследност што ја претпоставува алгоритмот за учење.

CARLA обезбедува модуларен сензорски состав што може да се прикачи на кој било актер во сцената, при што секој сензор испраќа пакет податоци до клиентот на секој чекор на симулацијата. Достапни се камери за RGB, семантичка и инстанциска сегментација, длабочина и оптички тек, LiDAR и семантички LiDAR, радар, GNSS и IMU, како и настански сензори за судир, пречекорување на лента и блиски пречки.

Во оваа теза CARLA служи како експериментална околина за целото тренирање и евалуација. Неговиот синхрон режим на извршување обезбедува секој симулациски кадар да биде зафатен во баферот со искуства, зачувувајќи ја временската доследност што ја бара RL алгоритмот. Неговиот сензорски состав со точни референтни податоци ги испорачува каналите на набљудување — сегментација (дел 3.1.2), инверзна длабочина (дел 3.1.3) и водилки — што директно го хранат претставувачкиот модел. API-то за скриптирање на пешаци, иако е ограничено во вградената способност за конфликтно однесување, обезбедува доволно нисконивоиска контрола за да се конструираат систематски сценарија на преминување.

Глава 4

Експериментална платформа и конфигурација на агентот

Оваа глава ја документира системско-инженерската подлога на тезата: сè што мора да биде поставено пред да започне тренирањето. Нејзиното средиште е *Библиотека на конфликтни пешаци* — збирка сценарија од три архетипа на конфликтни пешаци што ѝ дава емпириска потпора на остатокот од работата — и главата започнува со неа за читателот од самиот почеток да знае со какво однесување на пешаци се бара агентот да се справи. Потоа следи подлогата што ја поддржува библиотеката: претворањето на симулаторот во Gymnasium-компатибилна околина за поттикнувано учење, дизајнот на набљудувањето и акцискиот простор, минималната функција за награда и конфигурацијата на DreamerV3. Главата завршува со тренинг-постапката и инженерската инфраструктура. Алгоритамските основи и општиот опис на CARLA се опфатени во глава 3. Резултатите добиени со оваа платформа се наоѓаат во глава 5.

4.1 Библиотека на конфликтни пешаци

Библиотеката на конфликтни пешаци е компонентата за дизајн на сценарија во експерименталната платформа и еден од оригиналните придонеси на оваа теза. Таа е библиотека од *шесетови на интеракција*: скриптирани однесувања на пешаци што го поставуваат возилото во прецизно дефинирани конфликтни ситуации и откриваат каков репертоар за безбедност на пешаци стекнала научената полиса. Библиотеката ја претвора CARLA од симулатор со општа намена во инструмент за оптоварување на проактивната безбедност на пешаци под поттикнувано учење со модел. Библиотеката содржи три конфликтни архетипа — *Најрајник*, *Нерешителен преминувач* и *Прошеситан*. Секој архетип е репродуцибилен и намерно едноставен на ниво на локомоција. Скелетната анимација на пешаците во CARLA се извршува секогаш кога на интерфејсот WalkerControl му се задава брзина — преку API-то ги управуваме брзината и насоката, додека симулаторот го изведува одењето, така што одот не е нешто што библиотеката го управува или го изоставува. Сложеноста е во логиката на интеракција.

4.1.1 Мотивација и принципи на дизајн

Проактивната безбедност се тестира преку средби во кои пешакот создава временски релевантен конфликт со планираната траекторија на возилото, така што агентот мора навреме да го препознае ризикот и да избере безбеден одговор. Сценариото треба да го надмине простото присуство на пешак во сликата и да создаде интеракција во која времето, траекторијата и одговорот на возилото се меѓусебно зависни. Затоа релевантното прашање за оваа теза е по-

Табела 4.1: Дијагностичка улога на типот архетипа од Библиотекајта на конфликтни пешаци. Библиотекајта ја варира структурата на интеракцијата, а не изледот или одењето на пешакој.

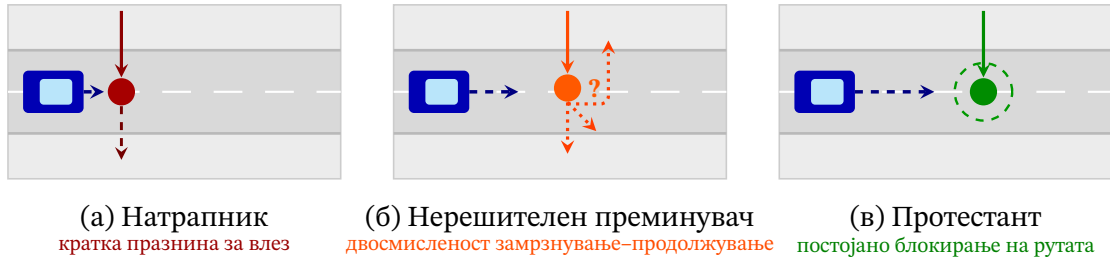
Архетип	Структура на интеракцијата	Примарна тестирана способност
Натрапник	Доцно преминување темпирано според пристигнувањето на возилото	Брзо препознавање и избегнување под тесно ТТС
Нерешителен	Рано преминување проследено со замрзнување–продолжување активирани од ТТС	Адаптација на промена во намерата на пешакој и на грешка во предвидувањето
Протестант	Постојано заземање на центарот на лентата	Кочење и давање првенство или пак странично избегнување

остро: дали агент со претставувачки модел може да предвиди и разреши конфликт со пешак пред судирот да стане неизбежен?

Вградената интелигенција за пешаци во CARLA не може сигурно да произведе доцно влегување пред возилото, двоумење среде преминување или постојано блокирање на лента на зададено растојание по рутата. Причината е структурна: вградениот управувач за пешаци е врзан за дискретна навигациска мрежа од водилки за пешаци по тротоари и означени пешачки премини, а неговиот планер е пристрасен кон рути што се усогласени со картата и го почитуваат тој граф. Графот го прави пронаоѓањето пат ефикасно за обична навигација, но не остава чист начин пешак да се постави на коловоз ниту пак да се активира влегување во лента условено од проценетото време на пристигнување на возилото. Овие однесувања бараат директно управување со движењето на пешакој во секој чекор. Библиотекајта затоа не се обидува да ја репродуцира целата варијабилност на човечкото одење: параметризацијата на одењето, позата на телото, распределбите на реакциско доцнење и социјалната сигнализација намерно се надвор од опфат, а контролираната променлива е *структурата на средбата*: кога се појавува ризикот, како се развива и каков одговор на возилото би го разрешил. Пешакојот може да се постави во однос на конфликтната точка на произволен број метри по рутата, а пред преминувањето може да се движи по тротоарот. Динамичките архетипови потоа го следат напредокот на возилото и времето на пристигнување до конфликтната точка, така што нивното однесување се активира од состојбата на возилото, а не од фиксен временски печат во симулацијата. Така се спречуваат тривијални заобиколувања во кои побрзо или побавно возило едноставно го промашува наменетиот конфликтен прозорец. Секој архетип е дизајниран да оптовари различна површина на неуспех: Натрапникот ја тестира реакцијата под кратко време, Нерешителниот преминувач го тестира предвидувањето под неизвесност во однесувањето, а Протестантот тестира дали агентот може да даде првенство на постојано заземање на лентата наместо да се потпира на странично избегнување. Одржувањето на архетиповите одделени еден од друг го прави секое забележано однесување интерпретабилно: судир во случајот Протестант значи нешто различно од судир во случајот Натрапник. Табела 4.1 ја сумира дијагностичката улога на секој архетип, а слика 4.1 ја прикажува просторната конфигурација на секој од нив.

4.1.2 Конструкција на сценарија

Сите три архетипа ја делат истата просторна конструкција. На почетокот од епизодата се гради рутата и се избира конфликтна точка A на фиксна должина $s_A = 45\text{ m}$ од почетокот на рутата, $\mathbf{p}_A = \mathbf{r}(s_A)$, каде што $\mathbf{r}(s)$ е централната линија на рутата параметризирана според поминатото растојание по рутата. Пешакојот се создава на едната или на другата страна од патот релативно на локалната тангента на рутата во A , при што се пробуваат неколку странични поместувања и се проектираат врз блиски тротоарски водилки каде што се достапни. Создадениот пешак се управува директно преку интерфејсот `WalkerControl` на CARLA без вграден



Слика 4.1: Просторна конфигурација на трите архетипа од Библиотеката на конфликтни пешаци. (а) Натрапникот чека покрај патот и влегува во лентата кога возилото е блиску, создавајќи средба со кратко време за реакција. (б) Нерешителниот преминувач почнува да преминува порано, но може да замрзне во лентата додека возилото се приближува, така што од една рамка не е јасно дали лентата ќе се расчисти или ќе остане блокирана. (в) Протестантот оди до фиксна точка во лентата и ја држи, претставувајќи постојана блокада што бара кочење, чекање или широк страничен маневар. Поглед одозгора: синиот правоаголник е возилото, исполнетиот круг е пешакот. Стрелките го означуваат наменетото движење на пешакот.

навигациски управувач, а детекцијата на судир е овозможена така што контакт со возилото создава терминален настан на судир со пешак.

Динамичките архетипови дополнително зависат од проценети времиња на пристигнување во A :

$$\tau_{\text{ped}}(A) = \frac{\|\mathbf{p}_A - \mathbf{p}_{\text{ped}}\|}{\max(v_{\text{ped}}, \epsilon)}, \quad \tau_{\text{veh}}(A) = \frac{\|\mathbf{p}_A - \mathbf{p}_{\text{front}}\|}{\max(v_{\text{close}}, \epsilon)}, \quad (4.1)$$

каде што v_{ped} е зададената брзина на одење, v_{close} е брзината на возилото проектирана врз насоката од предниот дел на возилото до A , а ϵ е мала позитивна константа што спречува делење со нула. Со други зборови, $\tau_{\text{ped}}(A)$ е проценетото време за пешакот да стигне до A , а $\tau_{\text{veh}}(A)$ е проценетото време за предниот дел на возилото да стигне до истата точка. Овие величини се внатрешни за генераторот на сценарија и не му се изложуваат на агентот. Од перспектива на агентот, внатрешната состојба на пешакот е скриена: истиот видлив пешак може да чека, да биде пред обврзување на преминување, пред замрзнување или пред продолжување.

4.1.3 Натрапник: доцно влегување на рутата

Натрапникот е сценариото со пристап со кратко време во библиотеката. Тој тестира дали агентот може да одговори кога пешак влегува во планираната патека доволно доцна за достапниот прозорец за реакција да биде мал.

Натрапникот започнува на тротоар близу конфликтната точка A додека возилото се приближува. Во оваа фаза пешакот е на тротоарот, уште не на коловозот. Може да стои или да изведува мали мирувачки движења по должина на тротоарот, но не влегува во лентата. Пешакот се обврзува да премине кога возилото е доволно блиску за пристигнувањето на пешакот во A да се преклопи со пристигнувањето на возилото:

$$\tau_{\text{veh}}(A) \leq \tau_{\text{ped}}(A) + t_{\text{lead}}. \quad (4.2)$$

Отстапното време t_{lead} е нагодено така што, во отсуство на какво било кочење од агентот, пешакот и предниот дел на возилото ја достигнуваат конфликтната точка A во истиот миг, без да го закачи од страна. Вредноста се одржува мала за средбата да остане со кратко време за реакција, но доволно голема за да ја апсорбира вградената рампа на забрзување на пешакот во CARLA од мирување, така што пешакот ја достигнува зададената брзина на преминување до моментот кога влегува во лентата. По обврзувањето, Натрапникот не преиспитува: оди кон A и продолжува неколку метри понатаму во истата насока на преминување.

Натрапникот има три концептуални состојби. Во состојбата *Чекање*, пешакот е на тротоа-

рот и го следи приближувањето на возилото. Не мора да биде неподвижен и може слободно да се движи по должина на тротоарот, но сè уште не влегол во лентата. Во состојбата *Обврзано преминување*, равенката (4.2) се активирала и пешакот оди кон А. Во состојбата *Расчистување*, пешакот продолжува покрај А за да ја напушти лентата на возилото. Отсуството на состојба за прекин е намерно. Натрапникот не е социјален пешак што преговара со возилото — тој е контролиран генератор на конфликт. Штом се обврзе, средбата е задача што агентот треба да ја разреши.

Успешен одговор бара агентот да го детектира пешакот, да заклучи дека неговата траекторија ја пресекува рутата и да започне маневар во краток ТТС прозорец. Натрапникот ја испитува границата на реактивен одговор при кратко време за реакција. Успешно избегнување преку странично управување со мало намалување на брзината укажува на брз рефлекс за странично избегнување. Успешно избегнување со значајно забавување укажува на порано регистрирање на опасноста. Двата исхода се видливо различни во сигналите за брзина и странично поместување анализирани во подоцнежните глави.

4.1.4 Нерешителен преминувач: однесувачка неизвесност

Нерешителниот преминувач е сценариото за промена на намера во библиотеката. Тој тестира дали агентот може да се справи со пешак чија видлива траекторија станува несигурна во текот на средбата.

Нерешителниот преминувач започнува како Натрапникот, но тргнува порано. Неговиот услов за тргнување е

$$\tau_{\text{veh}}(A) \leq \tau_{\text{ped}}(A) + t_{\text{depart}}, \quad (4.3)$$

со $t_{\text{depart}} \in [1.2, 2.8]$ s поголемо од отстапното време на Натрапникот, па пешакот влегува на патот додека возилото сè уште е подалеку. На почеток пешакот изгледа како пешак што преминува и чија идна положба може да се екстраполира од тековното движење. Клучната разлика е механизмот замрзнување–продолжување. Додека преминува, пешакот го следи $\tau_{\text{veh}}(A)$. Ако тоа падне под праг за замрзнување,

$$\tau_{\text{veh}}(A) \leq \tau_{\text{freeze}}, \quad (4.4)$$

пешакот застанува. Продолжува само откако ќе помине минимално времетраење на замрзнувањето и проценетото време на пристигнување (ETA, англ. Estimated Time of Arrival) на возилото ќе се врати над поголем праг за продолжување,

$$\tau_{\text{veh}}(A) \geq \tau_{\text{resume}}, \quad \tau_{\text{resume}} > \tau_{\text{freeze}}. \quad (4.5)$$

Јазот меѓу τ_{freeze} и τ_{resume} е хистереза. Без него, пешакот би можел брзо да се префрла меѓу движење и застанување низ последователни чекори. Со хистереза, замрзнувањето е однесувачки настан на кој агентот мора да се приспособи.

Нерешителниот преминувач има пет состојби: *Чекање*, *Преминување*, *Замрзнај*, *Продолжено преминување* и *Расчистување*. Како и кај Натрапникот, *Чекање* значи дека пешакот е на тротоарот и сè уште не зачекорил во лентата. Во оваа фаза може да се движи по должина на тротоарот наместо да биде држен совршено мирен. Визуелното набљудување на почетокот од средбата е компатибилно со неколку иднини: пешакот може да ја премине лентата, да замрзне во лентата или да продолжи откако возилото веќе започнало маневар. На агентот не му се кажува која можност е активна.

Имплементацијата се менува меѓу два семантички профила. Во *безбедниот профил* пешакот тргнува порано, се движи побрзо (1.8–2.4 m/s) и замрзнува само накратко (0.3–0.6 s) — под внимателен пристап пешакот ја расчистува лентата. Во *замрзнувачкиот профил* пешакот тргнува подоцна, се движи побавно (1.1–1.8 m/s) и замрзнува подолго (0.7–1.3 s), што ја зголемува веројатноста пешакот да остане во лентата за време на конфликтниот прозорец.

Менувањето низ епизоди спречува полисата секој Нерешителен преминувач да го третира како безопасен или како секогаш блокирачки.

Овој архетип е особено информативен за теза за претставувачки модели. Актерот се тренира во замислени иднини. Ако претставувачкиот модел екстраполира пешак што преминува како да продолжува во еден потег, може да замисли дека лентата се расчистува дури и кога реалниот пешак замрзнува. Обратно, ако научи дека пешаците често застануваат, може да замисли постојана блокада дури и во случаи со безбеден профил. Затоа Нерешителниот преминувач ја испитува врската меѓу набљудуваното движење, латентното предвидување и изборот на акција.

4.1.5 Протестант: постојано заземање на лента

Протестантот е сценариото за давање првенство и справување со постојана блокада во библиотеката. Тој проверува дали агентот може да даде првенство и да се справи со постојана блокада на лентата. Во ова сценарио проблемот е намерно јасен: еден или повеќе пешаци ја заземаат лентата кај конфликтната точка и ја држат положбата фиксно времетраење.

На почетокот од епизодата Протестантот оди однапред доделена точка за блокирање во или близу *A* со константна брзина на пристап од 2.0 m/s. Штом ќе се најде на 0.8 m од целта, пешакот застанува и ја држи положбата 10 s. По истекот на блокадата, пешакот излегува странично со 1.0 m/s. Сценариото поддржува групна конфигурација во која повеќе пешаци се создаваат истовремено, секој со доделена точка за блокирање странично поместена. Група од двајца или тројца преку ширината на лентата ја претвора централната блокада во целосно заземање на лентата. Конфигурацијата со еден пешак се користи за експериментите во оваа теза, каде што заобиколување околу пешакот останува геометриски можно по цена на големо странично отстапување.

Протестантот е реализиран во два профила, двата центрирани на истата точка за блокирање. Во профилот *сџајично-ослободување* пешакот ја држи точката за блокирање додека не истече времетраењето и потоа заминува. Страничната положба е фиксна низ целата блокада, па доволно широко заобиколување ја разрешува средбата геометриски. Во профилот *сџранично-поместување* пешакот дополнително се преместува лево и десно во мал домен околу точката за блокирање, го следи страничното поместување на возилото и претвора делумно заобиколување во повторно блокирана средба. Пешакот сè уште не ја модулира средбата врз основа на надолжната брзина или растојание на возилото. Она што се менува меѓу профилите е колку строго се брани заземањето на лентата странично, а не дали блокадата се држи.

Самата точка за блокирање е фиксна и не се преговара со возилото. Оваа ограничена реактивност е намерна: сценариото изолира едно прашање — дали научената полиса може да забави или да застане пред видлива пречка што ја брани својата положба во лентата?

Протестантот има четири состојби: *Присџајување*, *Блокирање*, *Заминување* и *Расчистење*. Состојбата *Блокирање* е дијагностичкото средиште на архетипот. Движечки пешак често може да се избегне преку предвидување на неговата идна патека. Неподвижен пешак во лентата не може да се разреши само со привремено кочење — единствените безбедни одговори се да се кочи и чека, или да се преземе широк страничен заобиколен маневар што прифаќа голема казна за напуштање на рутата.

Протестантот е најјасната проверка дали агентот научил кочење како безбедносно однесување. Ако агентот се судри со Натрапник, настанот може да се припише на недоволен прозорец за реакција. Ако се судри со Нерешителен преминувач, промената на намерата на пешакот нуди делумно објаснување. Ако се судри со Протестант, ниту едно оправдување не важи: пешакот е видлив од моментот кога влегува во лентата, не се движи неочекувано и останува неподвижен доволно долго за да успее секој разумен одговор со кочење.

Табела 4.2: Компактна параметризација на типите архетипа од Библиотека на конфликтни пешаци.

Архетип	Параметар	Вредност
Натрапник	Брзина на преминување	[2.0, 3.2] m/s
	Водечко време t_{lead}	0.6 s
	Продолжување по A	[5.0, 8.0] m
Нерешителен	Брзина на преминување	[1.1, 2.4] m/s
	Водење при тргнување t_{depart}	[1.2, 2.8] s
	Праг за замрзнување τ_{freeze}	[0.7, 1.8] s
	Праг за продолжување τ_{resume}	[1.2, 2.6] s
	Времетраење на замрзнување	[0.3, 1.3] s
Протестант	Брзина на пристап	2.0 m/s
	Времетраење на блокирање	10.0 s
	Радиус на блокирање	0.8 m

4.1.6 Параметризација

Секој архетип има компактно множество параметри што ја контролира сериозноста на средбата без да го промени нејзиното значење. Табела 4.2 ги наведува вредностите користени во експериментите. Нумеричките опсези не се избрани произволно. Тие се темелат на емпириската литература за однесување на пешаци, земајќи примероци од конфликтниот крај на набљудуваните распределби наместо да ја репродуцираат централната тенденција на која било поединечна студија. Брзините на преминување за Натрапникот и Нерешителниот преминувач одговараат на горниот крај од распределбите на брзина на преминување на пешаци пријавени во теренски студии на урбани раскрсници [2, 41], каде што типичните брзини на одење се околу 1.2–1.5 m/s, а брзиот крај надминува 2 m/s за конфликтни пешаци кои преминуваат во многу кратко време. Отстапното време на Натрапникот од 0.6 s и времето на на тргнување на Нерешителниот преминувач од 1.2–2.8 s се поставени во режимот на преминување со кратко време документиран во студиите за периоди на преминување кај пешаци, каде што минималните прифатени периоди се групираат околу 1–3 s за наметливи преминувачи [38]. Праговите за замрзнување и продолжување и времетраењата на замрзнување на Нерешителниот преминувач ги следат кинематичките и замрзнување–продолжување профили документиран за нерешителни пешаци на несигнализирани премини [47, 59]. Постојаното заземање на лента од Протестантот ги одразува набљудувањата дека пешаците стекнуваат стратешка доминација над автономни возила обврзани да дадат првенство кога избираат да останат на коловозот [53], и дека ненадејното искористување празнина е најчестото сценарио непосредно пред судир со пешак [56], што ја мотивира геометријата на доцно влегување на Натрапникот. Затоа параметарските опсези на библиотеката го опфаќаат опасниот крај на документираниот простор на однесување, а не инстанцираат конкретна референтна распределба.

4.2 Од симулатор до околина

CARLA, воведена во глава 3, обезбедува физика, рендерирање и управување со пешаци. Сама по себе не дефинира што претставува епизода, како се структурирани набљудувањата или кога агентот се наградува. Оваа секција документира како симулаторот се конфигурира за тренирање и како се обвива во околината Gymnasium [85] со која интерагира DreamerV3. Низ целата работа се користи CARLA издание 0.9.16. Библиотеката на конфликтни пешаци воведена во секција 4.1 се приклучува во оваа околина како сценариски слој за време на из-

вршување: при `reset()`, пешаците се инстанцираат и им се прикачуваат нивните управувачи на однесување. При секој чекор, управувачите пресметуваат наредби и ги испраќаат преку нискоинивовското API за пешаци на CARLA.

4.2.1 Конфигурација на CARLA

Серверот на CARLA работи во *неграфички режим*, со рендерирање пренасочено кон бафер надвор од екран. Тренирањата опфаќаат милиони чекори низ многу часови. Прозорец за приказ би трогел GPU меморија и би барал поврзан монитор. Python клиентот се поврзува со серверот преку TCP на `localhost:2000`, при што двата процеса се на истата машина. Серверот е врзан за GPU, додека клиентот, кој ја обработува логиката на околината, сензорската обработка и комуникацијата со тренинг-циклусот на DreamerV3, работи во одделен Python процес на истата машина.

Сите експерименти го користат *синхрониот режим* на CARLA: симулацијата е замрзната додека клиентот не повика `tick()`. Асинхрониот режим дозволува рамки тивко да се прескокнат, што го корумпира баферот со искуства и ја нарушува временската надежност што ја бара RSSM. Синхрониот режим гарантира подредени сензорски податоци без празнини и репродукцибилност под фиксно случајно семе. Физичкиот чекор е фиксиран на $\Delta t = 0.05 \text{ s}$ (20 Hz) — максимумот на кој PhysX backend-от на CARLA одржува стабилна динамика на возилото. Агентот издава нова акција на секој чекор, со фреквенција на управување од 20 Hz, стапка значително побрза од доброволните човечки реакциски времиња и усогласена со временските размери на конфликтите со пешаци што се проучуваат тука.

Тренирањето се одвива во Town02. Одлучувачкиот фактор при изборот е меморискиот простор на GPU: Town02 е најмалиот од градовите испорачани со CARLA, со компактен буџет на вчитани средства, па една GPU може истовремено да ги хостира рендерерот на симулаторот и PyTorch тренинг-процесот на DreamerV3 без да ја исцрпи меморијата. Покрај тоа што се вклопува на еден уред, картата е добро приспособена на прашањата што се истражуваат тука: обезбедува Т-раскрсници, урбани улици со две ленти, континуирани тротоари и разновидна густина на згради по должина на патот — сите геометрии на средби што ѝ се потребни на библиотеката се присутни. Нејзиниот компактен, повторлив распоред има секундарна придобивка: претставувачкиот модел гледа конзистентна структура на патот низ епизоди, што ѝ помага на латентната претстава да конвергира при тренирањето.

Town01 се користи исклучиво како околина за тестирање и генерализација и никогаш не се појавува при тренирање. Тој е поголем град со поинаква структура на раскрсници и поинаков распоред на згради од Town02. Евалуацијата во Town01 проверува дали полисата и претставувачкиот модел се пренесуваат на невидена патна мрежа, наместо да ги меморизираат конкретните рути и визуелни признаци на тренинг-картата. Двете карти се прикажани на слика 4.2.

CARLA се користи тука затоа што ја поддржува експерименталната поделба што ѝ е потребна на оваа теза: перцепцијата може да се направи доверлива, додека одлучувањето останува тешко. Сензорите за семантичка сегментација и длабочина даваат визуелни канали со точни референтни податоци, па неуспех за безбедност на пешак во подоцнежните глави може да се толкува како ограничување на научениот претставувачки модел или на полисата, а не како пропуштен излез од детектор. Во исто време, симулаторот ги изложува патниот граф, динамиката на возилото и состојбата на пешаците преку Python API, што овозможува да се генерираат разновидни рути за возење од 200 m, а експериментот сепак прецизно да се контролира. Нискоинивовскиот API за пешаци дава управување со положбата и брзината на пешакот во секој чекор, што е суштинско за логиката на обврзување и замрзнување-продолжување во Библиотеката на конфликтни пешаци. Синхрониот режим комбиниран со генератори на случајни броеви со семе овозможува точна репродукција на епизоди, користена обемно во анализите на имагинацијата во подоцнежните глави. Отворениот код на



Town02: тренирање



Town01: тестирање и генерализација

Слика 4.2: Погледи од птичја перспектива на двете CARLA карти користени во оваа теза. **Лево:** Town02, компактна станбена област со решеткаст распоред, користена за целото тренирање. Малиот буџет на вчитани средства овозможува симулаторот и тренерот на DreameerV3 да делат една GPU. **Десно:** Town01, поголем град со поинаква структура на раскрсници и распоред на згради, користен исклучиво за евалуација на генерализација и никогаш не вииден при тренирање.

CARLA исто така е важен за платформа за докторска теза: однесувањето на симулаторот и претпоставките врзани за конкретно издание може да се проверат, наместо да се третираат како непроѕирни детали на добавувач. Конечно, широката употреба на CARLA во литературата за поттикнувано учење за автономно возење ја олеснува споредбата со сродна работа и обезбедува дека дизајнерските избори специфични за симулаторот се добро документирани. Затоа изборот е практичен, а не тврдење дека CARLA е најреалистичниот симулатор за возење во секој поглед: тој обезбедува доволно урбана структура и управливост врз актерите за Библиотека на конфликтни пешаци, додека одржува репродукцибилност за анализа со поттикнувано учење.

Вградената интелигенција за пешаци во CARLA ги движи пешаците по тротоари и низ означени премини со умерена брзина, без да создава систематски предизвик за возилото. Агент што избегнува такви пешаци можеби не научил ништо повеќе од забавување пред неподвижни пречки. Безбедносните прашања што се истражуваат тука бараат доцно обврзување, двоумење среде преминување и одржано заземање на патот — ништо од ова вградениот систем не може да го произведе. Затоа сите пешачки актери во оваа теза се имплементирани преку нискоинивовското API за управување со пешаци, кое дозволува положбата и брзината да се поставуваат програмски во секој чекор.

4.2.2 Животен циклус на епизода

Секоја епизода започнува со фиксна иницијализациска секвенца во `reset()`: актерите од претходната епизода се уништуваат, возилото се создава на случајна патна водилка со мала почетна брзина од 5–8 km/h, се генерира случајна рута, се создаваат пешаци и им се прикачуваат нивните управувачи за конфликтни пешаци, а сензорските бафери се тестираат во неколку чекори пред да се издаде првото набљудување. Почетната брзина напред обезбедува агентот веднаш да ја искуси корелацијата гас–движење–награда. Студен старт од мирување би го оставил сигналот за награда рамен неколку чекори. Откако `reset()` ќе врати резултат, чекорниот циклус работи до терминација: ограничување и декодирање на акцијата, управување со возилото, напредување на симулаторот, ажурирање на пешаците, пресметка на награда, проверка на терминација и составување на набљудувањето. Агентот издава нова акција



(а) Целосна рута од 200 m.



(б) Набљудувачки водилки под должина на рутата.

Слика 4.3: Поглед од птичја перспектива на создадената тренинг-рута од 200 m во Town02. Магента точките ја следат рутата во двата панела. Во горниот панел, зелената точка го означува почетокот на рутата, а црвената точка го означува нејзиниот крај. Во долниот панел, заокружените магента точки ги означуваат погрубите набљудувачки водилки семплирани од рутата и користени како претстава на патеката што ја гледа агентот. Прикажаната рута е една од многуте: планерот избира случајно на секоја раскрсница, па различни епизоди поминуваат низ различни патеки во истата карта.

на секој чекор на симулаторот преку обвивката на DreamerV3, со фреквенција на управување од 20 Hz.

4.2.3 Создавање рути

Секоја епизода е организирана околу рута создадена од патниот граф на CARLA. Почнувајќи од точката на создавање на возилото, планерот го следи графот напред, избирајќи случајно на секоја раскрсница, додека не се акумулираат приближно 200 m должина. Рутата се чува во две резолуции. *Фина руша* со растојание 0.5 m се користи внатрешно за пресметка на страничната и аголната грешка e_{lat} и e_{ψ} . Погруба секвенца од *набљудувачки водилки* со растојание 3.0 m ги обезбедува $K = 10$ претстојни позиции што влегуваат во каналот на картата на водилки. Слика 4.3 прикажува пример на целосна рута и набљудувачките водилки семплирани од неа.

4.2.4 Услови за терминација

Епизодите завршуваат под пет услови, сумирани во табела 4.3. Условите се делат во две групи според природата на завршувањето. Завршување на целта и двата режима на судир се *дефинитивни*: нешто навистина се случило во светот што ја завршува епизодата и нема

Табела 4.3: Услови за терминација на епизода. Терминалните направи се наведени во табела 4.6.

Услов	Тип на завршување	Активирање
Достигната цел	Дефинитивно	Агентот го достигнува крајот на рутата
Судир со пешак	Дефинитивно	Контакт со кој било пешак
Судир со возило	Дефинитивно	Контакт со кое било возило
Надвор од рута	Прекин	Надминати странични/аголни прагови
Истек на време	Прекин	Поминати 1500 чекори (≈ 75 s)

продолжение што треба да се замислува по тоа. Напуштањето на рутата и истекот на време се *прекини*: светот би продолжил да работи ако не сме го запреле вештачки, но извршувањето е скратено — затоа што возилото залутало предалеку од планираната рута, или затоа што епизодата го потрошила временскиот буџет. Разликата е важна за претставувачкиот модел, не само за евиденција — таа контролира дали агентот се учи дека светот завршува во оваа состојба или само дека снимањето запира.

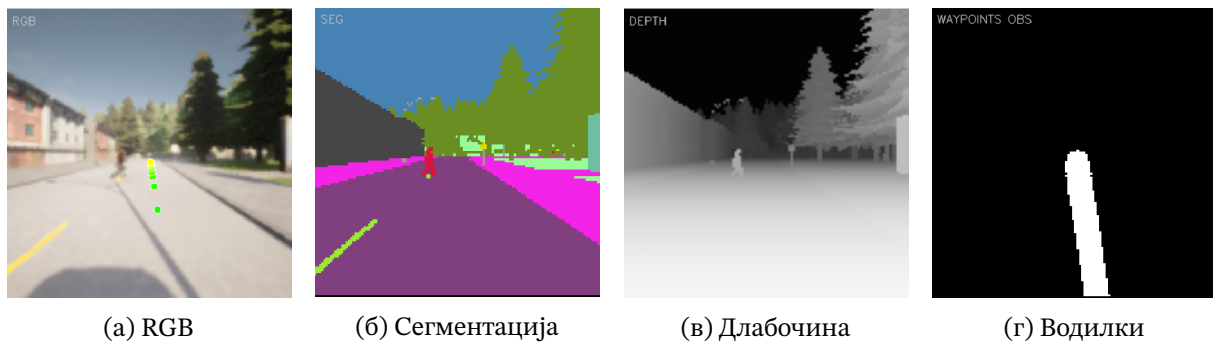
Епизода се прекинува поради напуштање на рутата кога возилото ја напушта планираната рута на начин од кој е малку веројатно да се опорави. Се проверуваат три услови: чист страничен праг $|e_{\text{lat}}| > 3.0$ m, чист аголен праг $|e_{\psi}| > 90^\circ$ и комбиниран услов $|e_{\text{lat}}| > 2.5$ m и $|e_{\psi}| > 75^\circ$. Се користат три услови наместо еден затоа што ниту еден поединечен праг не ја опфаќа добро геометријата на напуштање на рутата. Единечен страничен праг доволно тесен за да означи вистински напуштања се активира при легитимни широки свртувања на раскрсници. Единечен аголен праг доволно тесен за да означи возење во погрешна насока се активира кога возилото е кратко косо при менување лента. Комбинираниот услов ја фаќа геометријата што ги пробива двата поединечни прага: умерено странично поместување спарено со значајна аголна неусогласеност, што се смета за избегнувачки маневар од кој не е направено враќање.

4.3 Дизајн на набљудувања и акции

Агентот не го набљудува светот преку сурови пиксели. RGB сликата намерно не се користи како влез, за проблемот да не се меша со учење перцепција од текстура, боја и осветлување. Секој сензорски излез од CARLA се трансформира пред да се проследи до невронската мрежа, кодирајќи доменско знаење за тоа што е важно за безбедно урбано возење во самата структура на набљудувањето. Активната конфигурација на набљудување произведува наслагана сликовна матрица $128 \times 128 \times 3$ (каналите за сегментација, инверзна длабочина и водилки), заедно со 3-димензионален вектор на состојба на возилото и 2-димензионален вектор на претходна акција. Големината на сликата од 128×128 е компромис диктиран од хардверот, а не идеална перцепциска резолуција: 256×256 Би ги претставила далечните пешаки полесно, но го надминувае практичниот VRAM буџет кога DreamerV3 и CARLA ја делеа истата работна станица. Предноста на пониската резолуција е што ги одржува визуелниот енкодер, декодерот и пакетите од replay-баферот доволно мали за стабилно долго тренирање со избраната големина на пакет и тренинг-однос. Слика 4.4 ја прикажува суровата RGB рамка за контекст и трите сликовни канали што ги користи агентот.

4.3.1 Сензори, во симулација и во реален свет

Два од трите сликовни канали на слика 4.4 се произведуваат од сензорите со точни референтни податоци на CARLA. Картата на семантичка сегментација се чита директно од класната доделба по пиксел на рендерерот, а картата на длабочина се чита од растеризираниот z-бафер. Затоа агентот добива совршена класна идентификација и точно метричко растојание



Слика 4.4: Слика на набљудување за репрезентативна рамка од епизода во Town02. Панел (а) го прикажува суровиот RGB рендер за просторен контекст. Тој не е влез за агентот. Панелите (б)–(г) се трите сликовни канали што се наслагуваат во визуелното набљудување на агентот: семантичка сегментација подредена по ризик, инверзна длабочина и проектирана карта водилки.

за секој пиксел во своето видно поле, без грешки во детекција, без класификациски грешки и без испаѓање зависно од растојание. Третиот сликовен канал, карта на водилки, се пресметува од планираната рута и модел на камера и е идентичен во симулација и во извршување. Само каналите за сегментација и длабочина се специфични за симулаторот.

Одлуката да се користи перцепција со точни референтни податоци наместо да се тренира перцепциски систем заедно со полисата е намерна и методолошки централна. Перцепцијата во реален свет е неуредна и склона кон грешки на карактеристични начини. Мрежите за семантичка сегментација погрешно класифицираат делумно прекриени пешаци, ја заматуваат границата меѓу пешак и соседен структурен елемент на растојание и не успеваат на карактеристични начини на работ од својата тренинг-распределба. Дури и на умерено растојание, пешак може да се подели на два фрагмента на ознака, да биде проголтан во ознака за зграда или да добие класа со ниска доверба што испаѓа при праг. Стерео и LiDAR-проектираната длабочина деградираат на големо растојание и на површини со ниска рефлективност, а научените монокуларни мрежи за длабочина произведуваат систематски грешки на нова геометрија. Сите овие режими на неуспех се добро проучени истражувачки проблеми сами по себе, но тие не се истражувачките прашања на оваа теза. Тезата истражува дали агент со поттикнувано учење и претставувачки модел може да развие проактивно-безбедносно однесување од ретка награда. Ако шумот од перцепција беше дел од системот, секој судир би имал двосмислена причина — неуспех во планирање, неуспех во перцепција или некоја интеракција меѓу двете. Фиксирањето на перцепцијата на точни референтни податоци дава чисто припишување: секој неуспех да се избегне пешак може да биде само неуспех на полисата или претставувачкиот модел, никогаш неуспех на сензор. Оваа изолација го прави прашањето за проактивна безбедност одговорливо како самостоен проблем пред врз него да се наслага дополнителниот слој на робусност на перцепција.

Во извршување на реално возило, истото набљудување би се реконструирало од неидеални сензори, а влезниот формат што го гледа полисата би останал непроменет. Метричката длабочина би доаѓала од едно или фузија од: калибриран пар стереокамери (каде дисперзијата меѓу два синхронизирани RGB текови се претвора во длабочина по пиксел), LiDAR скенирање проектирано во рамката на камерата за да пополни ретки мерења на длабочина, или монокуларна мрежа за длабочина. Системите за автономно возење обично ги фузираат сите три за да ги балансираат режимите на неуспех на секој сензор еден против друг. Семантичката сегментација би доаѓала од мрежа на возилото во реално време што работи на RGB стримот — научен модел трениран надвор од возилото да емитува класни веројатности по пиксел, при што најверојатната класа се зема на секој пиксел. Чекорот на кодирање на ризик во равенка (4.6) и нормализацијата на инверзна длабочина во равенките (4.7)–(4.8) потоа би се примениле врз излезите на мрежата точно како што се применуваат врз излезите од симу-

Табела 4.4: Кодирање на семантичка сегментација поредена по ризик. Секоја класа на CARLA се мапира во скаларна безбедносна вредност што ја одразува сериозноста на повреда при судир. Големите јаз меѓу вредностите во околните на возила (230–255) и вредностите на структурни опасности (55–100) спречува грешките во реконструкција на имажинацијата појасно да класифицираат структурен елемент како возило или пешак.

Ниво	Класи	Вредности	Образложение
Висок ризик	Pedestrian Rider, Motorcycle, Bus, Car, Truck	255 230–250	Повреда во сите случаи
Структурна опасност	Building, Wall, Bridge, Train GuardRail, Fence, TrafficLight, TrafficSign	80–100 55–75	Повреда при удар
Мала опасност	Static objects, Dynamic objects Bicycle, Vegetation, RailTrack	35–50 27–40	Мала или никаква повреда
Без опасност	SideWalk, Terrain RoadLine, Ground, Road, Sky	10–25 0–2	Нема ризик од судир

латорот тука. Енкодерот по перцепцискиот блок не знае и не го интересира дали неговиот влез е z-бафер или резултат од стерео дисперзија. Картата на водилки воопшто не бара перцепција и е непроменета: се пресметува од планираната рута и калибриран модел на камера, и двете достапни на реално возило без модификација.

Реалната перцепција е шумна, ограничена по опсег и латентност и кршлива надвор од тренинг-распределбата. Затворањето на тој јаз е посебна истражувачка програма и е надвор од опфатот на оваа теза. Редоследот е намерен: *како однесување може да научи агенти со претставувачки модел и рејка на прага* мора да се одговори пред *колку е робусно што однесување по шум од перцепција* да може разумно да се постави.

4.3.2 Канал за семантичка сегментација подредена по ризик

Камерата за семантичка сегментација на CARLA му доделува на секој пиксел класна ознака од фиксен речник од 29 категории на урбани сцени. Директното проследување на класни ID до конволуциска мрежа би било семантички бесмислено — растојанието меѓу ID 12 (пешак) и ID 14 (автомобил) не носи информација за опасност. Наместо тоа, секое класно ID се мапира во *вредност на ризик* во [0, 255] што ја кодира сериозноста на повреда при судир со таа класа објекти:

$$I_{\text{seg}}(x, y) = r(\text{class_id}(x, y)), \quad (4.6)$$

каде $r : \mathcal{C} \rightarrow [0, 255]$ е функцијата за ризик дефинирана во Табела 4.4. Добиената едноканална слика ги зачувува просторната локација и безбедносниот приоритет, а ги отфрла текстурата и бојата. Табела 4.4 го прикажува целото мапирање групирани по ниво на ризик. Две својства се намерни. Пешаците ја добиваат максималната вредност 255, со што недвосмислено се поставуваат на врвот на ризичната скала и максимално се одвојуваат од сите други класи. Јазот меѓу нивото на возила (230–255) и нивото на структурни опасности (55–100) е намерно голем, што спречува грешки во реконструкција на имажинацијата да префрлат структурен елемент во опсегот на детекција на пешаци кога декодерот на претставувачкиот модел го реконструира каналот за сегментација.

4.3.3 Канал за инверзна длабочина

Камерата за длабочина на CARLA ја кодира длабочината на сцената со 24-битна прецизност, обезбедувајќи точно референтно метричко растојание за секој пиксел. Суровата длабочина во метри не е соодветен директен влез: линеарна скала на длабочина им дава еднаква

тежина на објекти на 2 m и објекти на 40 m, додека избегнувањето судир бара многу поголема резолуција во блиското поле. Затоа длабочината се трансформира во нормализирана вредност на инверзна длабочина, калибрирана така што работниот опсег за пешаци паѓа на средината од каналот. Две растојанија го параметризират кодирањето: блиска рамнина $d_{\text{near}} = 1$ m, на која каналот се заситува со максимална осветленост, и растојание на половина интензитет $d_{\text{half}} = 20$ m, на кое каналот ја зема вредноста 128/255. Суровата длабочина $d[x, y]$ прво се ограничува оддолу до блиската рамнина,

$$d_c = \max(d[x, y], d_{\text{near}}), \quad (4.7)$$

а потоа се мапира преку

$$I_{\text{depth}}[x, y] = \frac{1}{1 + \frac{d_c - d_{\text{near}}}{d_{\text{half}} - d_{\text{near}}}}. \quad (4.8)$$

Резултатот се скалира во $[0, 255]$ и се зачувува како uint8 канал. Објекти на 1 m или поблиску се мапираат во 255, објекти на 20 m во 128, а одговорот континуирано опаѓа кон нула за поголеми длабочини. Изборот $d_{\text{half}} = 20$ m доделува половина од опсегот на каналот на длабочини во прозорецот на интеракција со пешаци, а преостанатата половина на подолги растојанија, што задржува доволен градиент во далечното поле за претставувачкиот модел да го следи далечниот контекст, а сепак го нагласува блискиот опсег каде што се одлучува физиката на судир.

4.3.4 Канал на карта на водилки

Агентот мора да знае каде оди рутата. Наместо да се обезбедуваат информации за рутата преку векторот на состојба на возилото — што би барало MLP мрежата да научи просторна навигација од вектор на скалари — водилките се кодираат како просторен сликовен канал што CNN енкодерот го обработува заедно со каналите за сегментација и длабочина. Важно е прецизно да се каже што се овие водилки. Тие ја кодираат патеката што возилото би ја следело во отсуство на кој било пешак: планот на централната линија пресметан од патниот граф на почетокот од епизодата, пред кој било пешак да биде поставен во сцената. Тие не се поместуваат како одговор на детектирани опасности, не означуваат каде се пешаците и не му кажуваат на агентот како да ги заобиколи. Каналот на водилки кажува само *ова е ѝаѝекаѝа шѝо би ја следел ако никој не ѝи сѝои на ѝаѝоѝ*. Да се одлучи што да се направи кога некој е на патот е задача на полисата. $K = 10$ Најблиски непосетени водилки се изразуваат во референтната рамка на возилото и се проектираат на рамнината на сликата со модел на камера што ја одразува физичката поставка на камерата. Континуирана линија со ширина 9 пиксели ги поврзува последователните проектирани водилки, со исполнет круг со радиус 4 пиксели на секоја локација на водилка. Пикселите на водилки се поставуваат на 255, а сите други пиксели се 0. Овој канал е и единствениот извор на агентот за геометрија релативна на рутата. Страничното поместување на возилото од рутата и аголната неусогласеност меѓу возилото и рутата не се обезбедуваат како скаларни влезови никаде на друго место во набљудувањето — и двете се содржани во положбата и ориентацијата на проектираната линија на водилки во сликата, а од CNN енкодерот се очекува да ја извлече оваа геометрија визуелно. Со тоа што информацијата за рутата е дадена во еден сликовен канал, а не поделена меѓу слика и однапред пресметани скалари, претставувачкиот модел мора во истиот простор да ги предвидува и пешаците и геометријата на лентата. Ова просторно кодирање е важно и при имагинација: декодерот го реконструира каналот на водилки заедно со сегментацијата и длабочината. Ако реконструираниите водилки отстапат од вистинската рута низ хоризонтот на планирање, актерот добива погрешна претстава за тоа каде треба да управува.

Табела 4.5: Компоненти на векторот на состојба на возилото. Сите величини се изразени во соодветната референзна рамка на возилото. На агентиот не му се изложуваат глобални координати.

Компонента	Симбол	Единица
Брзина	v_t	m/s
Надолжно забрзување	a_t^{lon}	m/s ²
Стапка на свртување	ω_t	rad/s

4.3.5 Вектор на состојба на возилото

Векторот на состојба на возилото се означува со $e_t = (v_t, a_t^{\text{lon}}, \omega_t)$ и ги содржи надолжната брзина, надолжното забрзување и стапката на свртување околу вертикалната оска. Овие три величини обезбедуваат самосвесност за кинематичката состојба на возилото и не се изводливи од една рамка од камера. Табела 4.5 ги наведува нивните единици и улоги. Векторот не содржи глобална положба, координати од карта или величини релативни на рутата. Информацијата за држење лента — каде се наоѓа возилото релативно на планираната патека и како неговата насока се усогласува со тангентата на рутата — не се обезбедува како скалар. Наместо тоа, таа е содржана во картата на водилки: проектираните позиции на претстојните водилки во рамката на камерата на возилото веќе ги кодираат страничното поместување и аголната неусогласеност во просторот на сликата, а од CNN енкодерот се очекува да ја извлече оваа геометрија визуелно наместо да ја прими како однапред пресметан влез. Така просторната страна на свесноста на агентот живее целосно во доменот на сликата — истиот домен преку кој тој ги перципира пешаците — и се избегнува изложување на полисата на изведен локализациски производ што не би бил рамномерно достапен во реално извршување.

Векторот на претходна акција $u_{t-1} = (a_{\text{steer}}, a_{\text{long}})$, четвртата компонента на набљудувањето, му дава на агентот меморија за неговата последна наредба. Ова е особено корисно за RSSM: знаењето на претходната акција ѝ помага на детерминистичката GRU состојба h_t да го моделира доцнењето меѓу издавањето наредба и одговорот на возилото, и му дозволува на актерот да избегнува нагли пресврти (на пример, целосно кочење веднаш по целосен гас).

Компонентите во табела 4.5 се и најчистиот дел од набљудувањето за пренос од симулација во реалност. Брзината доаѓа од енкодери на брзина на тркалата, надолжното забрзување од IMU акцелерометарот, а стапката на свртување од IMU жироскопот — сите три се сензори што веќе постојат на секое модерно автономно возило и се изложени на CAN магистралата од модулот за стабилносно управување, со грешка под еден процент и стапки на ажурирање далеку над управувачкиот циклус на агентот од 20 Hz. За разлика од каналите за сегментација и длабочина, овој дел од набљудувањето не зависи од научена перцепциска мрежа, локализациски систем или карта со висока дефиниција.

4.3.6 Простор на акции

Просторот на акции е дводимензионален и континуиран:

$$\mathcal{A} = [-1, 1]^2, \quad a = (a_{\text{steer}}, a_{\text{long}}), \quad (4.9)$$

каде што $a_{\text{steer}} \in [-1, 1]$ ја управува насоката, а $a_{\text{long}} \in [-1, 1]$ кодира обединета надолжна наредба. Овој избор е намерен, иако многу студии за поттикнувано учење во автономно возење користат дискретни множества на акции бидејќи го поедноставуваат истражувањето и проценката на вредност. Дискретизирањето на управувачот би барало однапред да се избере мал каталог од примитиви за управување, гас и кочење, што би вовело артефакти од квантизација и би направило дел од однесувањата да бидат последица од дизајнот на акциите наместо од научената полиса. Сценаријата со пешаци што се проучуваат тука зависат од континуално

управување: лесно кочење, целосно итно кочење, мали корекции за центрирање во лента и пошироко избегнувачко управување се различни однесувачки исходи. Континуиран простор на акции е потежок за оптимизација, но ја зачувува оваа разлика и се совпаѓа со основниот интерфејс на CARLA за управување со возилото.

Разложувањето од скаларната надолжна наредба во одделните сигнали за гас и кочница на CARLA е:

$$(\text{гас}, \text{кочница}) = \begin{cases} (a_{\text{long}}, 0) & \text{ако } a_{\text{long}} \geq 0 \\ (0, -a_{\text{long}}) & \text{ако } a_{\text{long}} < 0. \end{cases} \quad (4.10)$$

Оваа дефиниција јасно го мапира континуираниот простор: $a_{\text{long}} = +1$ е целосен гас, $a_{\text{long}} = -1$ е целосна кочница, а $a_{\text{long}} = 0$ е движење без гас. Излезите на актерот на DreamerV3 се во $[-1, 1]$ и мора да поминат низ равенка (4.10) пред да влијаат врз динамиката на возилото. Секое заситување во излезната распределба на актерот кон позитивни вредности подразбира структурна пристрасност против кочење.

Гасот и кочницата во принцип би можеле да се изложат како две независни димензии на акција. Тие намерно се составени во една. Причината е специфична за тоа како DreamerV3 истражува. Актерот на DreamerV3 е континуирана стохастичка полиса, и секоја акција што ја емитува носи шум за истражување на секоја излезна димензија во секој чекор. Со одделни гранки за гас и кочница, шумот ги одржува двете гранки ненулта практично во секој чекор, па агентот постојано издава мала позитивна наредба за гас заедно со мала позитивна наредба за кочница. Управувачот на возилото во CARLA ја третира секоја ненулта кочница како надвладување на гасот, па гасот тивко се отфрла секогаш кога примерокот на шум во кочницата е позитивен — агентот не може конзистентно да набљудува што навистина прават неговите наредби за гас, а сигналот за истражување во насоката на гас станува несигурен. Спојувањето на гасот и кочницата во еден скалар ги прави меѓусебно исклучиви по конструкција: само едно од нив е ненулта во даден момент, шумот за истражување паѓа на една надолжна оска, а градиентот што ги поврзува акциите со последователната динамика на возилото се зачувува. Обединувањето се совпаѓа и со физичката реалност дека притискање на педалата за гас и кочница во исто време не е смислена возачка наредба.

Слика 4.5 го прикажува целосниот систем набљудување-до-акција.

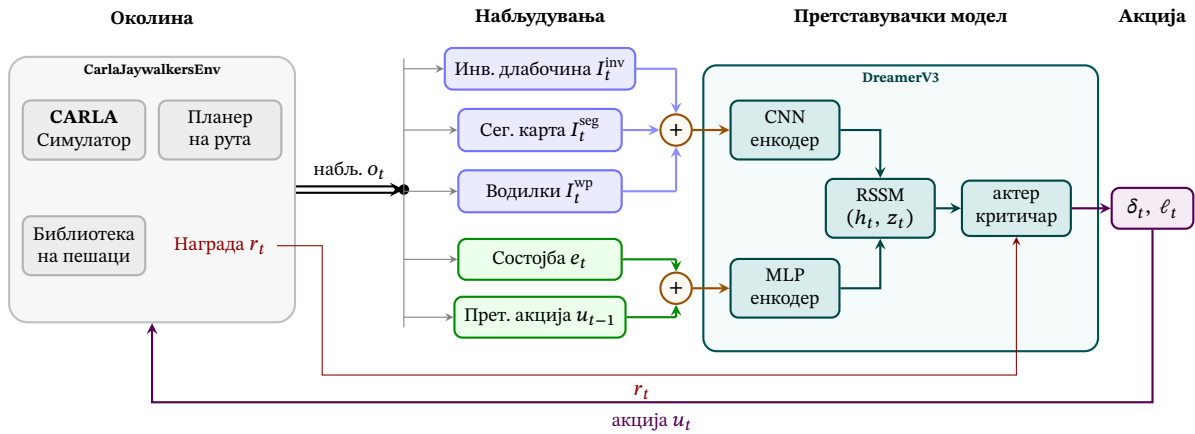
4.4 Архитектура на награда

Функцијата за награда е дизајнирана да кодира возачка компетентност — завршување на рута со безбедни брзини — и ништо повеќе. Ниту еден член не наградува давање првенство, не казнува близина до пешаци и не се повикува на време до судир. Ова е намерен истражувачки избор: истражувањето прашува какво однесување кон пешаци произлегува под минимален безбедносен сигнал, а инженерско вградување безбедност на пешаци во наградата би одговорило на друго прашање. Затоа наградата е чисто раздвојување меѓу возачката задача и безбедносниот феномен што се проучува.

Целосната спецификација е дадена во табела 4.6. Наградата r_t во секој временски чекор t е збир од активните компоненти:

$$r_t = -k_{\text{speed}} \max(0, v - v_{\text{max}})^2 + r_{\text{terminal}}, \quad (4.11)$$

каде што v е моменталната брзина на возилото во km/h, $v_{\text{max}} = 30$ km/h е прагот над кој се активира казната за брзина, $k_{\text{speed}} = 0.01$ е нејзиниот коефициент, а $r_{\text{terminal}} = 0$ во нетерминални чекори. Терминалната награда, применета еднаш на крајот од епизодата, има два дела: бонус или казна специфична за настанот, определена од тоа како завршила епизодата, и редок член за напредок еднаков на вкупното растојание што возилото го покрило по должина



Слика 4.5: Дизајниран систем набљудување-до-акција. Обвивката CarlaJaywalkersEnv (лево) го опфаќа симулаторот CARLA, Библиотеката на конфликтни пешаци и планерот на рути. На секој чекор, околината емитира мултисигнално набљудување O_t и скаларна награда r_t . Дизајнот на набљудување-то е придонес на оваа работа: три сликовни канали (инверзна длабочина, сегментација подредена по ризик и канал за проекција на рутата) кодираат геометрија, семантички ризик и намера за навигација во компактна претстава која ги истакнува структурите релевантни за судир. Тие се проследуваат заедно со нумеричката состојба на возилото и претходната акција до DreamerV3, кој применува споделен CNN и MLP енкодер, ја одржува латентната состојба на RSSM (h_t, z_t) преку имагинација на претставувачкиот модел и управува со пар актер-критичар кој ги произведува акциите за управување и надолжно движење (δ_t, ℓ_t) , вратени назад до симулаторот.

Табела 4.6: Активни компоненти на наградата. Терминалните награди се меѓусебно исклучиви и се применуваат само на крајот од епизодата.

Компонента	Израз	Белешки
Казна за брзина	$-0.01 (v - 30)^2$	По чекор; само кога $v > 30$ km/h
Изминато растојание	$+ s_{\text{end}}$	Терминално; по должина на на рутата, во метри
Достигната цел	$+100$	Терминален настан
Истек на време	-50	Терминален настан
Надвор од рута	-100	Терминален настан
Судир со возило	-100	Терминален настан
Судир со пешак	$-300 \left(1 + \left(\frac{v}{30}\right)^2\right)$	Терминален настан; зависен од брзина

на рутата во текот на епизодата,

$$r_{\text{terminal}} = r_{\text{event}} + s_{\text{end}}, \quad (4.12)$$

каде што s_{end} е конечната положба на возилото по должина на рутата, во метри, а r_{event} зема една од вредностите во табела 4.6. Членот s_{end} е единствениот сигнал што агентот го добива дека движење понатаму по рутата е добро. Нема награда за напредок по чекор: секој чекор пред терминација не придонесува ништо кон повратот освен казната за брзина, а наградата за движење нанапред пристигнува во една грутка на крајот. Ова е намерно најминималниот сигнал за напредок што сè уште е усогласен со целта на следење рута и е еден од експерименталните услови анализирани во подоцнежните глави.

Изборот да се вклучи s_{end} како позитивен терминален сигнал, наместо казните за судир и истек на време да ја носат целата награда, е намерен одговор на добро познат режим на неуспех кај дизајни на награда доминирани од казни. Со големи негативни казни за судир, без позитивна награда за движење и со слобода да не се прави ништо, функцијата на вредност учи

дека стоење во место има нула очекуван поврат, додека секоја акција што го движи возилото има некаков ненулт ризик од судир и затоа негативен очекуван поврат. Белман-оптималната полиса под таква награда е да замрзне и никогаш да не се движи — единствената акција без лоша страна. Ова е оптимумот на замрзнување, понекогаш наречен *мрзлив ајенџ*, и е реален емпириски феномен во поттикнуваното учење за возење, не теоретска љубопитност. Наивниот лек е казна за мала брзина или минимална брзина што го казнува агентот затоа што е неподвижен, но тоа ја внесува спротивната патологија: агентот сега е обврзан да продолжи да се движи и се казнува за застанување, дури и кога застанувањето е безбедниот одговор на пешак во лентата, па она што изгледа како непромисленост е агентот што почитува награда која вели дека стоење во место е лошо. Тезата се наоѓа меѓу овие две крајности. Нема казна за мала брзина (казната за брзина во равенка (4.11) се активира само над 30 km/h, па целосно застанување е бесплатно), но терминалниот s_{end} дава позитивен поттик да се напредува по должина на рутата. Затоа замрзнувањето не е Белман-оптимално — го губи целиот придонес s_{end} — но кочење и чекање пешак исто така не се казнува, затоа што агентот сè уште може да заврши со висок s_{end} ако на крај продолжи. Обликот на наградата е калибриран така што внимателно возење што понекогаш застанува доминира и над замрзнувањето и над непромисленото пробивање.

4.4.1 Регулација на брзина

Квадратна казна се применува над 30 km/h:

$$r_{\text{speed}} = -k_{\text{speed}} \max(0, v - 30)^2, \quad k_{\text{speed}} = 0.01. \quad (4.13)$$

Коефициентот е намерно мал: на 50 km/h казната е -4.0 по чекор, што е значајно релативно на сигналот за брзина, но мало релативно на терминалната награда за изминато растојание s_{end} акумулирана низ типична рута од 200 m. Казната не треба да наметне строго почитување на брзина — таа спречува опасно високи брзини, оставајќи го агентот слободен да ја избере својата оптимална брзина.

4.4.2 Терминални награди

Пет меѓусебно исклучиви настани ја терминираат епизодата (табела 4.6). Достигнувањето на целта носи бонус од +100, додека истекот на време носи казна од -50 , со што се раздвојуваат успешно завршена рута и епизода во која агентот не успеал да стигне до крајот навреме. Напуштањето на рутата (-100) обесхрабрува големи странични патешествија што завршуваат надвор од возната површина. Судир со возило носи фиксна казна од -100 . Судир со пешак се скалира со брзината на удар:

$$r_{\text{ped}}(v) = -300 \left(1 + \left(\frac{v}{30} \right)^2 \right). \quad (4.14)$$

При мирување казната е -300 , на 30 km/h се удвојува на -600 , а на 50 km/h достигнува $-1,133$. Зависноста од брзина ја одразува биомеханичката реалност дека сериозноста на повредите расте стрмно со брзината на удар. Како обликување на наградата, оваа зависност ја прави казната сè построга како што расте брзината при приближување, наместо судирите со пешак да имаат иста фиксна цена без оглед на брзината.

Поделбата меѓу дефинитивни завршетоци и прекини директно одредува како моделот учи да го толкува крајот на епизодата. DreamerV3 одржува гранка на продолжување чија задача е, на секој замислен чекор, да процени дали светот продолжува да работи. Кога епизода завршува дефинитивно — се случува судир, целта е достигната — гранката на продолжување се тренира да го препознае тој чекор како вистински терминален: нема иднина, дисконтниот

фактор на самопроценките се поставува на нула, а пропагацијата на вредност по таа точка се отсекува. Кога епизода е прекината — истек на време, скратување поради напуштање на рутата — на предвидуваачот на продолжување му се кажува спротивното: светот би продолжил да работи, отсекувањето е артефакт на снимањето, а дисконтниот фактор се држи на еден за да може идната вредност и понатаму да се проценува. Оваа разлика е важна конкретно за напуштање на рутата и истек на време. Ако двете класи на завршување се третираат како дефинитивни, претставувачкиот модел би научил дека странично залутување или едноставно трошење на времето предизвикува светот да заврши на ист начин како судир, што е и физички погрешно и стратешки заведувачко. Нивното третирање како прекини наместо тоа ги зачувува судирот и завршувањето на целта како единствени настани што внатрешниот модел на агентот ги смета за навистина терминални.

4.4.3 Зошто наградата завршува тука: отфрлени алтернативи

Наградата намерно не содржи член што реагира на присуство на пешак пред да се случи судир. Нема ТТС казна, нема казна за близина, нема позитивна награда за давање првенство или забавување во близина на пешак, ниту пак казна за мала брзина што би го оттурнувала агентот од „ползење“. На прв поглед, сите овие идеи изгледаат разумно. Дел од нив дури и се вообичаени во RL литературата за автономно возење. Сепак, секоја беше свесно отфрлена, бидејќи би го променила самото истражувачко прашање. Целта овде не е да се проектира безбедно однесување преку наградата, туку да се набљудува што може да произлезе од минимален сигнал за судир.

Казната за странично или аголно отстапување од рутата беше една од првите разгледувани опции. Таа е привлечна бидејќи држењето лента е тежок потпроблем, а таквата грешка лесно се мери. Но токму тука се појавува проблемот: избегнување пешак често бара краткотрајно напуштање на идеалната траекторија. Ако секое странично отстапување се казнува, тогаш полисата се оттурнува од маневри за избегнување уште пред тие да имаат шанса да се развијат. Наместо да се охрабри адаптивно однесување, наградата би го фаворизирала крулото следење на рутата. Поради тоа, доволна се сметаше терминалната заштита — епизодата и онака завршува ако возилото ја напушти дозволената област.

Слична дилема се појави и кај густата награда за напредок. Во практична смисла, ваков дизајн работи добро. Виртуелни порти поставени долж рутата му даваат на агентот постојан сигнал дека се движи во правилна насока, па тренирањето станува значително постабилно и побрзо. Вакви верзии беа користени и во претходниот развој на платформата. Но тука густината на сигналот создава несакан ефект: агентот е постојано наградуван за продолжување напред, дури и кога треба да му даде првенство на пешак. Во таков случај, забавувањето или запирањето стануваат директен опортунитетен трошок. Полисата може да научи да „протурка“ низ конфликтната ситуација бидејќи секој дополнителен метар носи награда сè до моментот на судир. Тоа подоцна би ја направило безбедносната анализа двосмислена: дали агентот не дава првенство затоа што не разбира ситуација или затоа што наградата активно го турка напред? Затоа, информацијата за изминато растојание останува само во терминалниот член, без густ поттик по чекор.

Разгледувана беше и експлицитна награда за давање првенство. Од човечка перспектива тоа изгледа природно — добар возач успорува или застанува пред пешак. Но со тоа одговорот однапред би бил вграден во наградата. Ако агентот добива позитивен сигнал секогаш кога кочи пред пешак, тогаш повеќе не се испитува дали такво однесување може да произлезе само од ретка казна за судир. Наместо да се мери појава на безбедно однесување, наградата директно би го диктирала.

Казната за мала брзина беше отфрлена од слична причина. Во развојните експерименти таа навистина помогна агентот побрзо да излезе од режим на стоење и порано да почне да ја поминува рутата. Но истовремено создава конфликт со кочењето како безбедносно од-

несување. Во сценарија како Протестант, каде што целта е да се види дали агентот ќе запре пред пречка, казнувањето на ниска брзина би значело дека безбедното однесување носи негативен сигнал. Токму затоа казната во равенка (4.13) се активира само при поголеми брзини над 30 km/h, каде што ризикот е поврзан со прекумерно брзо движење, а не со внимателно сопирање.

Најважната отфрлена алтернатива се ТТС и казните за близина до пешак. Овие сигнали се многу чести во литературата бидејќи директно кодираат опасност: мал ТТС или мало растојание речиси секогаш значат ризична ситуација. Сепак, нивното користење овде би довело привилегирана информација што агентот реално не ја добива. Полисата гледа само сегментација и длабочина и не добива однапред пресметан ТТС или геометриска анализа на интеракцијата. Ако такви величини се внесат во наградата, тогаш расудувањето што треба да го научат претставувачкиот модел и полисата веќе е делумно решено од инженерот. Дополнително, ТТС подоцна се користи како дијагностичка метрика во евалуацијата. Ако истиот сигнал се оптимизира при тренирање, тогаш ТТС повеќе не може чисто да се интерпретира како доказ за произлезена безбедност.

Заедничката нишка зад сите овие одлуки е иста. Секоја од отфрлените алтернативи или директно би проектирала посакувано однесување, или би внесла информација што агентот сам не ја перципира, или би ја избришала границата меѓу тренинг-сигнал и евалуациска метрика. Минималната награда во равенка (4.11) е конструирана токму за да ги избегне тие три проблеми.

4.5 Конфигурација на DreamerV3

Алгоритмот за учење е DreamerV3 [33]. Глава 3 дава целосен опис. Оваа секција ги набројува примарните конфигурациски ставки што управуваат со инстанца на DreamerV3, и оние оставени на референтните почетни вредности и оние нагодени за урбаното возачко окружување. Пасусите што следуваат го даваат образложението за изборите специфични за извршување и белешка во една реченица за секој алгоритамски специфичен трик што DreamerV3 го наследува од својата линија.

Физиката на CARLA работи на 20 Hz, а агентот избира нова акција на секој чекор. Оваа фреквенција е доволно брза за да одговори на пешаци што преминуваат со до 3.2 m/s — максималната брзина на одење во Библиотеката на конфликтни пешаци — додека и дозволува на рекурентната состојба на претставувачкиот модел да акумулира значаен контекст меѓу одлуките.

Актерот се тренира со замислување $H = 32$ идни чекори од тековната латентна состојба. На 20 Hz, ова се 1.6 s замислена траекторија. Хоризонтот беше избран за да ја опфати целата динамика на средба со пешак: пристапување, прозорец на преминување и разрешување. Хоризонт од половина секунда би му дал на актерот недоволно гледање напред за да расудува за пешак што сè уште се приближува до патот. Хоризонт од неколку секунди би ги исполнил повеќето замислени чекори со сцени без пешаци, разредувајќи го безбедносно релевантниот сигнал.

Дисконтирањето $\gamma = 0.997$ подразбира ефективен хоризонт на планирање од $1/(1 - \gamma) \approx 333$ чекори, приближно 16.7 s на 20 Hz. Тренинг-рутите опфаќаат 200 m, а при целна брзина од 30 km/h една рута трае околу 24 s. Дисконтирањето е калибрирано така што кумулативната награда од завршување рута останува значајна релативно на казните по чекор, спречувајќи агентот да го третира истекот на време како прифатлив исход.

Сликите на набљудување се 128×128 пиксели. Ова ја балансира видливоста на пешаци со пресметковниот трошок. При видно поле од 100° , пешак на 30 m зафаќа приближно 4 пиксели во висина и само 1–2 пиксели во ширина — блиску до границата на она што CNN енкодерот може сигурно да го детектира. Поблиски пешаци (10–15 m) опфаќаат 8–12 пиксели во висина, обезбедувајќи му на енкодерот доволна просторна структура за да ги разликува пикселите

од класата пешак од позадината. Влезот 128×128 поминува низ 5 конволуциски степени (големина на јадро 4, прескок 2 секој), создавајќи најдлабока карта од признаци 4×4 : $128 \rightarrow 64 \rightarrow 32 \rightarrow 16 \rightarrow 8 \rightarrow 4$.

Тренинг-односот е $\text{train_ratio} = 512$: за секој собран чекор во околината се репродуцираат 512 рамки од баферот со искуства низ градиентните ажурирања. Со пакети од 16 секвенци по 64 чекори (1024 рамки по ажурирање), тоа одговара на едно градиентно ажурирање на претставувачкиот модел и на мрежите за однесување на секои два чекора во околината. Ова е високо според општите RL стандарди, но е соодветно за CARLA, каде што секој чекор во околината бара цел чекор на симулаторот, вклучувајќи рендерирање и физика. Секоја собрана рамка така се употребува многупати низ тренингот наместо да се отфрли по едно ажурирање. Поставувањето висок тренинг-однос ја максимизира информацијата извлечена од секоја скапа интеракција со околината.

Сликовните компоненти — семантичка сегментација, инверзна длабочина и карта на водилки — се наслагуваат во триканален тензор 128×128 и се обработуваат од конволуцискиот енкодер. Нумеричките компоненти — состојба на возилото и претходна акција — се конкатенираат во вектор и се обработуваат од MLP енкодерот. Проследувањето ја следи структурата на самите набљудувања: CNN за просторно организирани сликовни податоци и MLP за неподредени скаларни мерења.

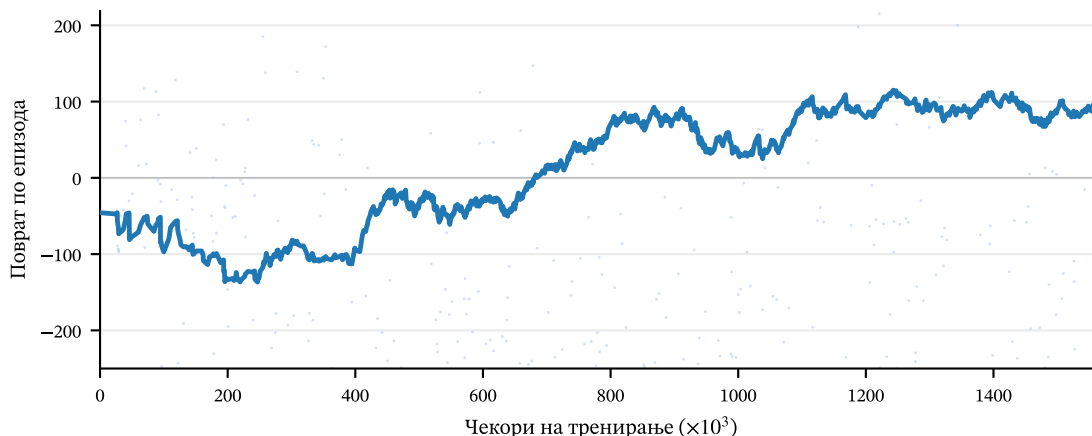
4.6 Поставки за тренирање

Тренирањето се одвива исклучиво во Town02 под конфигурацијата на околината од секција 4.2, интерфејсот за набљудување/акција од секција 4.3, наградата од секција 4.4 и конфигурацијата на DreameerV3 од секција 4.5. Генераторот на рути генерира рути рамномерно низ типовите патишта во Town02 на почетокот од секоја епизода. Распределбата на рути е доволно разновидна што две тренинг-епизоди ретко поминуваат низ идентични секвенци на водилки.

Тренирањето трае до 2×10^7 чекори во околината на едно семе, на една работна станица. Хардверот е ASUS ROG Strix RTX 3090 (24 GB VRAM), AMD Ryzen 9 3900X (12 јадра, 24 нишки) и 32 GB системска RAM. И серверот на CARLA и тренинг-процесот на DreameerV3 ја делат единствената GPU — ограничување што поврзува неколку претходни избори на платформа: Town02 (најмалата карта на CARLA според буџет на вчитани средства, секција 4.2.1), резолуцијата на слика 128×128 (секција 4.5) и чекорната стапка во синхрон режим (секција 4.2.1) се калибрирани така што симулаторот и тренерот се вклопуваат во 24 GB. Извршувањето од 2×10^7 чекори изнесува приближно 5 денови континуирано реално време. Слика 4.6 го прикажува повратот по епизода низ текот на тренирањето. Почетната фаза е доминирана од чести судири и терминации поради напуштање на рутата што создаваат големи негативни поврати. Како што претставувачкиот модел акумулира искуство и актерот се подобрува, повратите растат и се стабилизираат. Високата варијанса низ целото тренирање е карактеристична за DreameerV3 во стохастичка околина: повратот зависи од конкретно извлечената рута, времето на преминувањето на пешаци и моменталниот квалитет на полисата на агентот.

Баферот со искуства се иницијализира со мал број епизоди генерирани од рамномерна случајна полиса пред да започнат какви било ажурирања на актерот. Ова самоотпочнување е релативно кратко во споредба со вкупното тренирање. Целта е да му се даде на претставувачкиот модел почетна популација од траектории што го опфаќаат целиот простор на акции, наместо тесната распределба што свежо иницијализиран актер би ја произвел.

Ниту еден конфликтен архетип од Библиотеката на конфликтни пешаци не се создава при тренирање. Тренинг-распределбата користи намерно неконфликтна популација на пешаци што сепак ги одржува пешаците како постојан признак на сцената. Популацијата има две компоненти. Првата е множество од *едноставни пешаци што преминуваат* поставени



Слика 4.6: Поврат по епизода низ чекорите на тренирање. Бледите точки ги прикажуваат суровите поврати по епизода. Задебелената линија прикажува експоненцијален подвижен просек со $\alpha = 0.99$. Трендот го одразува растечкото владеење на следење рута како што претставувачкиот модел и актерот се развиваат заедно.

на фиксни растојанија по должина на секоја тренинг-рута (30, 60, 90, 120, 150 и 180 m од почетокот). Секој отпочнува од оваа точка на рутата, но не е врзан за неа: едноставен пешак што преминува се менува меѓу странично минување преку лентата меѓу тротоарот и центарот на патот со брзина по пешак извлечена од $[0.3, 2.0]$ m/s, кратко чекање на тротоарската страна и одење по должина на самиот тротоар пред да го започне следното преминување. Нивното движење е независно од возилото, па агент што се движи брзо обично ќе расчисти премин без блиска интеракција. Нивната улога е да обезбедат пешаците да бидат постојан признак на тренинг-распределбата, а не да наметнат систематски предизвик. Втората компонента е стандардната популација на пешаци на CARLA создадена низ целата карта, која се движи меѓу случајни дестинации по должина на тротоари и преку означени премини користејќи ја вградената навигација за пешаци на симулаторот. За бројот на живи пешаци на картата да се одржува близу целна густина — избрана да апроксимира реалистичен урбан волумен на пешаци — околината изложува механизам *зайри-и-йродолжи* за пешаците: кога бројот на пешаци се искачува над целта, создавањето се пазира. Кога паѓа под целта (затоа што пешаците ги достигнуваат дестинациите и се отстрануваат), создавањето продолжува.

4.6.1 Инфраструктура за тренирање

Оваа подсекција ја документира инженерската подлога што го направи тренирањето во оваа теза изводливо.

Агентот е имплементиран преку библиотеката PyTorch, врз референтна имплементација на DreamerV3 што ги изложува претставувачкиот модел, актерот и критичарот како одделни модули со сопствени оптимизатори. CARLA 0.9.16 се користи за симулаторот. Python клиентот, обвивката на околината, Библиотеката на конфликтни пешаци и тренинг-циклусот на DreamerV3 се сите во една кодна база. Околината му се изложува на DreamerV3 преку Gymnasium-компатибилен интерфејс, така што заменување со друг агент или друга библиотека не бара допирање на тренинг-кодот.

Серверот на CARLA и тренинг-процесот на DreamerV3 секој работи во Podman контејнер. Двата контејнери, за CARLA и за DreamerV3, ја делат GPU на работната станица преку NVIDIA container toolkit и комуницираат преку приватна контејнерска мрежа: серверот на CARLA го изложува својот TCP порт внатрешно, а тренинг-контејнерот се поврзува со него како со мрежно прикачен симулатор. Контејнерските слики ги фиксираат зависностите за CARLA, CUDA и Python така што околината е репродуцибилна низ рестартирања и низ ма-

шини.

Скаларните тренинг-метрики — компоненти на загубата на претставувачкиот модел, загуби на актерот и критичарот, KL членови, криви на поврат — се логираат во TensorBoard, кој е монтиран како одделен сервис достапен преку веб-прелистувач. Прилагодена Streamlit контролна табла работи заедно со TensorBoard и прикажува повисоконивовски тренинг-резимеа: кумулативни поврати по епизода, подвижна распределба на причините за терминација, најновите епизоди со судир со исечоци од траекторија што може да се репродуцираат и вредности во живо од тековниот тренинг-процес.

Глава 5

Анализа на безбедноста: однесување и имагинација

Оваа глава го евалуира тренираниот основен агент како референтна полиса за безбедноста анализа. Целта не е само да се пријави стапката на судири, туку да се објасни механизмот преку кој компетентна полиса за следење рута станува небезбедна околу конфликтни пешаци. Затоа главата напредува како сè подлабока истрага. Прво го опишува надворешното однесување на возилото, потоа прашува што останува скриено во тој надворешен запис, и на крај ги прегледува кратките замислени иднини што ги произведува претставувачкиот модел на Dreamer во секој чекор на одлучување.

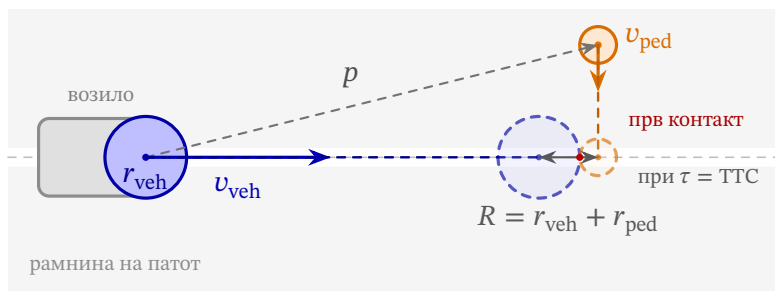
Истражувањето е организирано околу едно водечко прашање: кога полисата не успева да избегне пешак, од каде потекнува неуспехот — од перцепцијата, од управувањето на самото возило или од внатрешниот план на агентот? Истата терминологија се користи низ целата глава: *надворешно однесување* се однесува на реализираното движење на возилото, наредбите, ТТС и исходите од судирите. *Имагинација* се однесува на предвидените состојби на претставувачкиот модел кои интерно ги генерира полисата за време на евалуацијата.

5.1 Протокол за евалуација

За да се одговори на водечкото прашање, главата ги разгледува имагинациите на претставувачкиот модел зачувани при секој реален чекор на одлучување и анализира што агентот „замислувал“ пред и за време на секоја интеракција. Протоколот за евалуација ги поставува условите под кои тоа се прави.

Сите евалуации користат фиксирани тежини од мрежите на моделот. Учењето е исклучено: за време на евалуацијата нема ажурирање на претставувачкиот модел, актерот, критичарот, баферот со искуства, ниту пак онлајн адаптација. Секој неуспех пријавен во оваа глава затоа е својство на евалуираната полиса и нејзиниот научен претставувачки модел, а не минлив ефект од продолжен тренинг. Ова е важно за подоцнежната анализа на имагинацијата, бидејќи проекциите што се прегледуваат таму се проекциите на замрзнатиот основен агент додека тој навистина се однесува во тестирање.

Евалуацијата е поделена во две фази. Фаза 0 ја мери нормалната способност за возење. Таа содржи епизоди без пешаци и епизоди со четири пасивни случајни минувачи во зоната на рутата, распоредени по тротоарите. Тие можат да одат или да стојат, но не ја преминуваат планираната рута. Фаза 1 воведува конфликтни пешаци и ја мери безбедноста на пешаците. Оваа разделба спречува честа двосмисленост: ако полисата не може да ја вози рутата во обични услови, судир со пешак не би бил специфично информативен. Фаза 0 мора прво да утврди дека основниот агент е функционален возач.



Слика 5.1: Геометрија на TTC со модел на дискови што одговара на равенката (5.1). Возилото е претставено со диск, а пешакот со пешачки диск. При константни брзини, TTC е првото позитивно време во кое релативното поместување на центрите го достигнува збирниот радиус $R = r_{veh} + r_{ped}$.

Фаза 0 користи 1,000 епизоди за секој од четири услови: без пешак во Town02, без пешак во Town01, со четири случајни минувачи во Town02 и со четири случајни минувачи во Town01. Фаза 1 користи 1,000 епизоди по архетип на конфликтен пешак за надворешната анализа на однесувањето. Анализата на имагинацијата користи посебен запис што содржи 300 епизоди од Фаза 1, со 100 епизоди за секој архетип. Поголемиот примерок од Фаза 1 ги поддржува статистиките за исходи и однесување по рута. Помалиот примерок за имагинација ги поддржува внатрешните дијагностики на проекциите. Овој помал број епизоди е практичен избор, бидејќи логирањето со вклучена имагинација е приближно десет пати пресметковно поскапо од обичното логирање на надворешното однесување.

Времето до судир (TTC) се пресметува со истиот модел на диск што се користи за време на евалуацијата. Возилото е претставено со диск на предната половина на возилото, пешакот со диск со радиус 0.35 m, а TTC е првото позитивно решение на равенката за пресек на дискови со константна брзина

$$\|p + \tau u\|^2 = R^2, \quad \tau > 0, \quad (5.1)$$

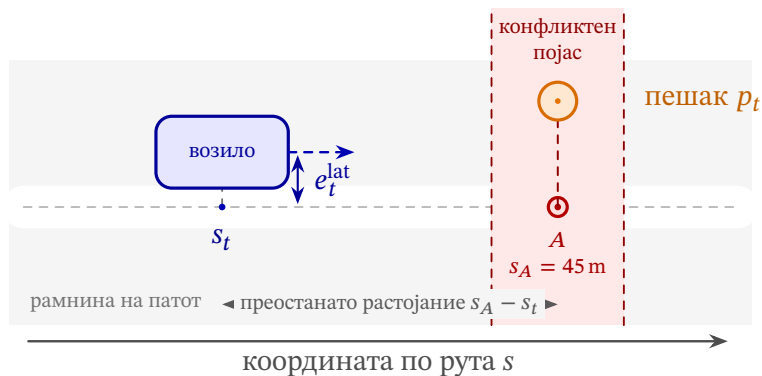
каде што p е релативната позиција на пешакот од предниот диск на возилото, u е релативната брзина на пешакот, а R е збирот од радиусите на дисковите на возилото и пешакот. Ако нема позитивно решение, TTC е бесконечно. Конфликтниот прозорец е дефиниран со конечно TTC под 3 s. TTC под 1.5 s се третира како блиско промашување или критичен регион. Чекорот на најблиското приближување е чекорот со конечно TTC во кој TTC е минимално во епизодата. Геометријата е прикажана на слика 5.1.

Епизодите од Фаза 1 користат планирана координата по рутата s во метри. Судирната точка на конфликтниот пешак е фиксирана на

$$s_A = 45 \text{ m}. \quad (5.2)$$

Оваа заедничка судирна точка овозможува надворешните мерења по рута и внатрешните проекции на имагинацијата да се порамнат со истата физичка конфликтна локација. Сликите по рута го интерполираат сигналот од секоја епизода по чекори на заедничка оска за напредок по рута и ги агрегираат медијаната, интерквartilниот опсег и опсегот од 10-ти до 90-ти перцентил. Ова одговара на просторното прашање: што прави возилото додека се приближува до пешачката точка? Геометриската поставка е прикажана на слика 5.2.

Реализираното избегнување се разложува на надолжни и странични компоненти. Надолжното избегнување вклучува кочење, давање првенство и намалување на брзината. Страничното избегнување вклучува отстапување на насоката и поместување релативно на рутата што е конзистентно со поминување околу пешакот. На ниво на реализирано однесување, падот на брзина и страничното поместување се користат за да се опишат анатомијата на судирите и геометријата на избегнувањето. На ниво на имагинација, истата идеја е формализирана



Слика 5.2: Геометрија на напредокот по рута што се користи за порамнување во Фаза 1. Состојбата на возилото се проектира врз централната линија на рутата за да се добие s_t , а страничното отстапување се мери како e_t^{lat} . Судирната точка на пешакот A е фиксирана на $s_A = 45$ m, па преостанатото растојание по рута до конфликтната локација е $s_A - s_t = 45 - s_t$. Реализираната позиција на пешакот p_t може да се движи релативно на судирната точка во текот на епизодата.

со ознаки на проекциите: директно продолжување, давање првенство, странично избегнување и кочење-плус-странично. Подоцнежната поделба ги прави овие ознаки експлицитни.

5.2 Основна способност за возење (Фаза 0)

Фаза 0 ја утврдува обичната способност за возење на основниот агент. Агентот се евалуира без конфликтни учесници во сообраќајот на видениот град од тренингот и на невидениот град за евалуација. Условите со случајни минувачи додаваат четири пасивни пешаци во зоната на рутата, распоредени по тротоарите. Тие можат да одат или да стојат, но не ја преминуваат планираната рута. Овој дизајн поставува две прашања пред да се разгледа која било конфликтна средба: дали агентот може да ја заврши рутата, и дали неконфликтните пешаци го нарушуваат неговото нормално возење?

Табела 5.1 ги прикажува исходите и метриките за квалитет на возењето. Завршувањето на рутата е над 98.7% во секој услов. Неуспесите поради излегување од рута и истек на време се ретки, а судирите со пешаци се нула. Ова ја утврдува првата неопходна основа: полисата може да ја вози рутата без конфликтна интеракција. Полисата исто така генерализира кон Town01, кој не бил користен за тренинг, без какво и да е влошување на стапката на завршување.

Метриките за квалитет на возењето покажуваат средно брзо, но стабилно следење на рутата. Медијаната на брзината по епизода е околу 22–24 km/h, додека медијаната на максималната брзина е приближно 34 km/h. Средната странична грешка е само околу 0.16–0.20 m, а средната грешка во насоката е околу 2.4°–3.4°. Агентот затоа не е слаб следач на лента. Тој ја следи средишната линија на рутата блиску, додека вози со релативно наметлива брзина.

Агрегираните вредности во табела сепак би можеле да скријат распределбено поместување предизвикано од пасивните пешаци. Слика 5.3 го проверува ова директно: распределбите без пешак и со случајни минувачи речиси целосно се преклопуваат за брзина, странична грешка и грешка во насоката. Нема доказ дека пасивните пешаци предизвикуваат двоумење, деградација на следењето на рутата или различен режим на брзина. Фаза 0 затоа ја поддржува интерпретацијата што е потребна за остатокот од главата: подоцнежните неуспеси се неуспеси на интеракцијата со конфликтен пешак, а не неуспеси на обичното следење на рутата или реакции на објекти со форма на пешак во набљудувањето.

Основниот агент треба да се третира како функционален следач на рута. Тој ја завршува рутата, генерализира низ евалуациската карта и ги игнорира пасивните случајни минувачи. Ова ја прави Фаза 1 безбедносен стрес-тест, а не тест за способност за возење. Следното прашање е што се случува кога пешакот повеќе не е пасивен и влегува во патеката на возилото.

Табела 5.1: Сјособносии за возење во Фаза 0 низ четири услови (по 1,000 епизоди). Условиите со случајни минувачи вклучуваат пасивни пешаци распоредени по тротоариите, кои можат да одат или да стојат, но не ја преминуваат планираната рута. Условиите без пешак немаат пешаци. Судирот со пешак е нула во сите услови.

Метрика	Без пешак Town02	Без пешак Town01	Со пешаци Town02	Со пешаци Town01
Завршување (%)	99.3	99.6	98.7	99.6
Излегување од рута (%)	0.4	0.1	0.6	0.1
Истек на време (%)	0.3	0.3	0.6	0.3
Друг судир (%)	0.0	0.0	0.1	0.0
Судир со пешак (%)	0.0	0.0	0.0	0.0
Медијанско време до цел (s)	8.85	8.25	8.90	8.43
Медијанска брзина (km/h)	22.5	24.0	22.2	23.5
Медијанска максимална брзина (km/h)	34.4	34.6	34.1	34.0
Максимална брзина (km/h)	43.4	42.1	44.5	40.8
Средна странична грешка (m)	0.192	0.161	0.196	0.169
Средна грешка во насока (°)	3.39	2.44	3.36	2.56

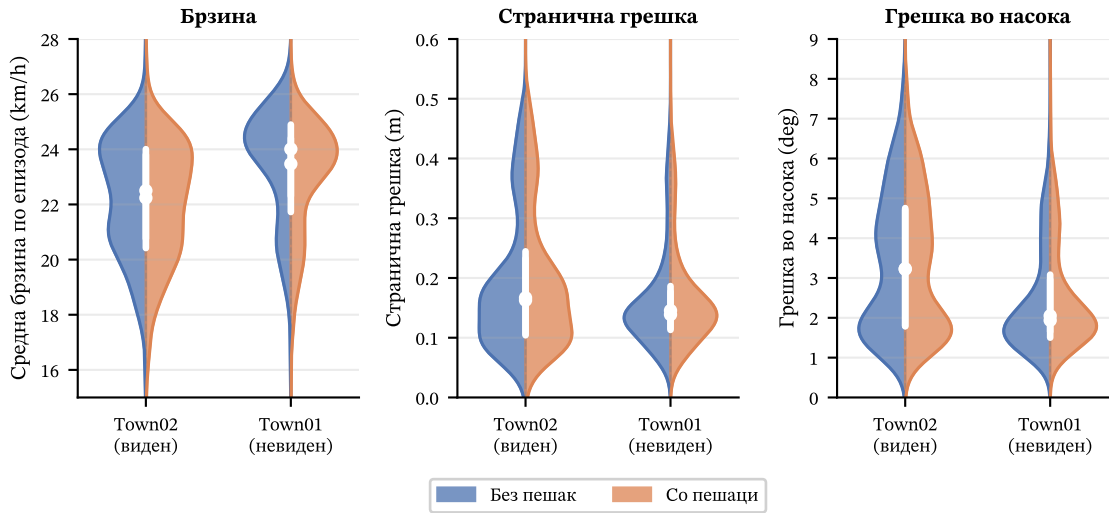
5.3 Агрегирани безбедносни исходи во Фаза 1

Фаза 1 ги заменува пасивните случајни минувачи со еден од трите архетипа на конфликтни пешаци. *Найрајникот* влегува во лентата во тесен влезен период темпиран според возилото. *Нерешителниот преминувач* започнува да преминува порано, но може да замрзне кога возилото се приближува. *Протестантот* оди до судирната точка и се задржува во мал локален регион во лентата. Кај дел од епизодите може бавно да се придвижува во тој регион за да го блокира возилото кога возилото пристапува со мала брзина. Овие архетипови оптоваруваат различни безбедносни способности: кратко време за реакција, одговор на променлива намера на пешак и давање првенство на локално ограничена пречка.

Табела 5.2 ги прикажува статистиките за исходите на највисоко ниво за 1,000 епизоди по архетип. Контрастот со Фаза 0 е очигледен. Стапките на судир со пешак растат до 35.7% за Натрапникот, 14.5% за Нерешителниот преминувач и 39.7% за Протестантот. Резултатот за Протестантот е најдијагностички бидејќи пешакот не ја расчистува рутата, туку останува во мал регион околу судирната точка и може бавно да се поместува во него. Агентот не само што не успева против брз или двосмислен преминувач — тој исто така удира лице што намерно ја задржува блокадата во лентата.

Стапката на средба помага да се објасни редоследот на архетиповите. Протестантот создава конфликт во 96.7% од епизодите бидејќи пешакот останува во локалниот регион околу судирната точка наместо да ја расчисти рутата. Натрапникот создава конфликт во 90.0% од епизодите бидејќи тесниот јаз обично го поставува во патеката на возилото. Нерешителниот преминувач создава конфликт во 75.1% од епизодите. Во преостанатите епизоди, временското поставување и однесувањето замрзни-продолжи можат да го разрешат преминувањето пред да се отвори ТТС прозорецот. Броевите за исходите затоа покажуваат фреквенција на неуспех, но сè уште не ја покажуваат тежината или темпирањето на тие неуспеси.

Слика 5.4 ја покажува распределбата на минималното ТТС низ епизодите со средба. Медијаната на минимално ТТС е 0.41 s за Натрапникот, 1.06 s за Нерешителниот преминувач и 0.17 s за Протестантот. Овие вредности не се маргинално небезбедни — тие ги поставуваат типичните епизоди со средба во режим блиску до судир. Распределбата за Протестантот е особено концентрирана блиску до нула, што потврдува дека локална блокада на лента не се обработува со удобна безбедносна маргина. Оваа концентрација на ниско ТТС не треба да се чита само како ризик од удар со висока брзина: таа може да одразува и бавно приближување или повторливи блиски средби додека Протестантот останува на патот и се движи само



Слика 5.3: Распределби на квалитетот на возење по епизода во Фаза 0, прикажани како поделени виолини. За секоја метрика, левата група е Town02 (виден), а десната група е Town01 (невиден). Во секоја група левата половина-виолина е условот без пешак, а десната половина е условот со случајни минувачи. Речиси идентичната форма на секој поделен пар потврдува дека пасивните пешаци немаат мерлив ефект врз однесувањето на агентот при возење на ниту една карта.

во својата мала зона на блокирање. Резултатот за Натрапникот е различен. Бидејќи влезот на пешакот е темпиран конфликтно, ситуацијата идеално треба да се препознае и обработи порано, пред тесниот јаз да се распадне.

Табелата исто така покажува дека агентот може силно да кочи откако опасноста станува непосредна. Медијанскиот почеток на забавување откако ТТС ќе падне под 1.5 s е 0.30 s за Натрапникот, 0.15 s за Нерешителниот преминувач и 0.25 s за Протестантот. Средното максимално забавување е приближно 7–8 m/s². Безбедносниот проблем затоа не е едноставно недостиг на реакција со кочење. проблемот е што реакцијата се појавува откако епизодата веќе влегла во критичен ТТС режим.

Медијански ред за почеток на свртување намерно не е пријавен во Табела 5.2. Во суровиот канал за акција, прагот за свртување често е веќе надминат при активирањето на критично ТТС бидејќи нормалното следење рута содржи леви-десни корекции. Ова прави медијанското доцнење на почетокот да колабира на 0 s за сите три архетипа, што е математички точно, но не е информативно. Страничното однесување затоа се анализира преку слики за насока-по-рута и геометрија на избегнување, а не како единечен медијански број за почеток на свртување.

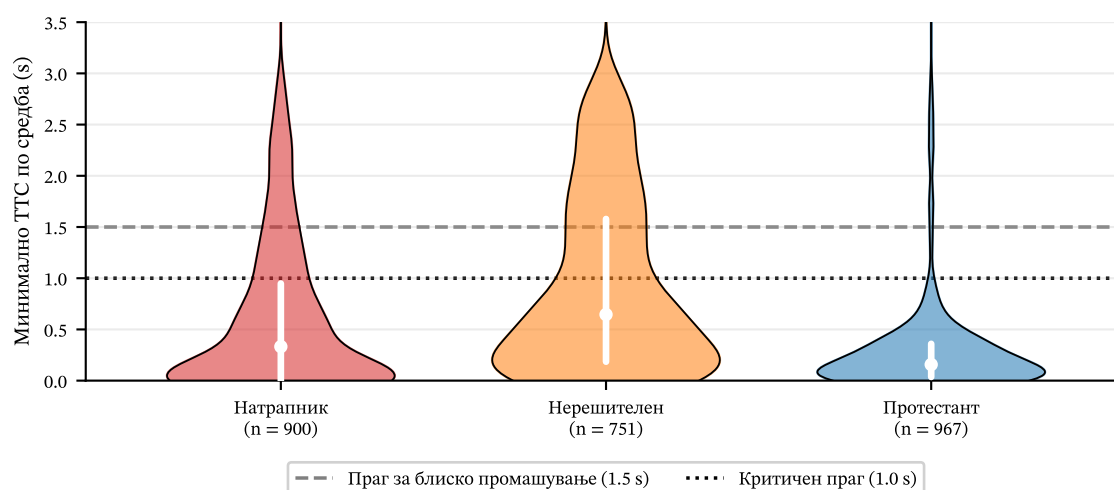
Агентот е компетентен во Фаза 0, но небезбеден во Фаза 1. Неуспесите се доволно сериозни за да произведат ниско ТТС и високи стапки на судир. Резултатот за Протестантот спречува лесно објаснување засновано само на неизвесност во темпирањето на преминувачки пешак: дури и кога пешакот останува локално ограничен околу судирната точка, полисата често не создава доволна маргина. Она што останува скриено е каде се развива небезбедната траекторија. Следната секција затоа преминува од исходи на ниво на епизода кон однесување по рута по должина на пристапот кон $s_A = 45$ m.

5.4 Однесување по рута

Анализата по рута ги порамнува сите епизоди според физичкиот напредок по должина на планираната рута. Ова открива дали полисата почнува да се подготвува за пешакот на време или го задржува нормалното возење до последните неколку метри. Основниот панел во сликите по рута ги здружува условите од Фаза 0 и служи како референца без конфликт.

Табела 5.2: Исоходи по архетипот во Фаза 1 (по 1,000 епизоди). Стапка на средба: дел од епизодите во кои пешакот влегол во конфликтниот прозорец ($TTC < 3\text{ s}$). Конфликтна брзина: средна брзина на возилото за време на конфликтниот прозорец. Почеток на забавување: медијанско доцнење од активирањето t^* ($TTC < 1.5\text{ s}$) до првиот чекор на забавување. Максимално забавување: средно максимално забавување измерено во прозорецот од 2 s по t^* .

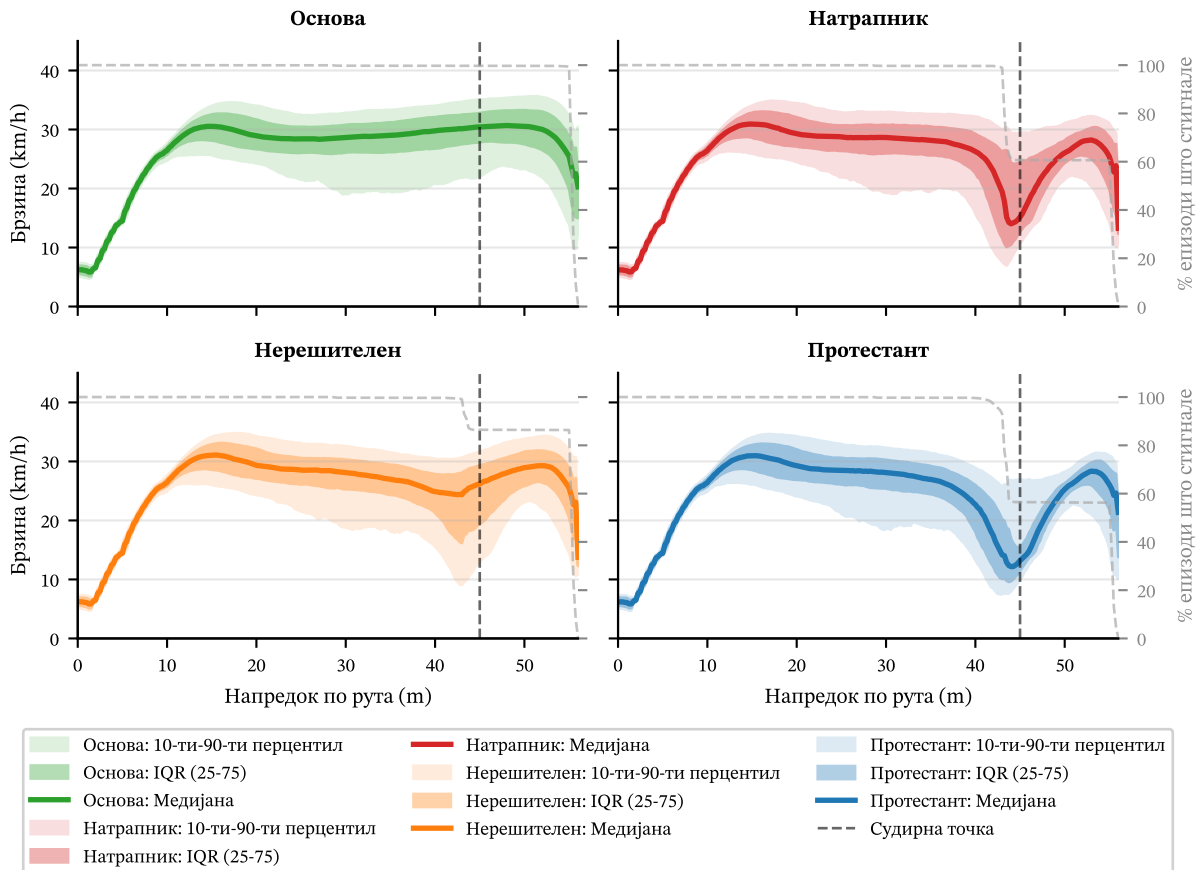
Метрика	Натрапник	Нерешителен	Протестант
Завршување на рута (%)	63.3	84.5	52.8
Излегување од рута (%)	0.5	0.6	0.7
Истек на време (%)	0.5	0.3	6.8
Судир со пешак (%)	35.7	14.5	39.7
Стапка на средба (%)	90.0	75.1	96.7
Медијанско минимално TTC (s)	0.41	1.06	0.17
Средна конфликтна брзина (km/h)	27.4	25.2	22.2
Средна максимална брзина (km/h)	33.5	33.7	33.0
Медијански почеток на забавување (s)	0.30	0.15	0.25
Средно врвно забавување (m/s^2)	8.34	7.15	8.02



Слика 5.4: Распределба на минимално TTC по епизода со средба, по една виолина за секој архетип. Испрекинатите линии ги означуваат праговете за блиско промашување (1.5 s) и критично TTC (1.0 s). Вклучени се само епизоди во кои TTC паднало под 3 s барем еднаш.

Слика 5.5 ја покажува брзината на возилото како функција од напредокот по рута. Основниот профил забрзува до опсег на стабилна брзина околу 22–24 km/h и останува таму во поголемиот дел од рутата. Во споредба со оваа референца, сите три профили со конфликтен пешак покажуваат суптилно намалување на брзината низ средниот дел од рутата. Ова навес-тува одредена навремена чувствителност на сцената со пешак. Сепак, главната фаза на коче-ње започнува само околу 10 m пред судирната точка на пешакот, особено за Натрапникот и Протестантот. Одговорот затоа изгледа најмногу слабо проактивен: видлива е одредена рана моду-лација, но одлучувачкото намалување на брзината е компресирано во крајниот пристап.

Распрснувањето околу s_A е исто така информативно. Некои епизоди значајно забавуваат, додека други остануваат блиску до слободната брзина. Кривата на опаѓање за Протестантот паѓа остро близу судирната точка бидејќи многу епизоди судираат или застануваат пред завршување на рутата. Дополнителниот пад на брзина по приближно 50 m треба внимателно да се интерпретира: номиналната рута е 60 m, но целта се регистрира во радиус од 4 m, околу 56 m. Во овој терминален дел од рутата, каналот со водилки практично завршува, а полисата научила да го намали гасот или да кочи блиску до завршувањето на рутата дури и без пешак.

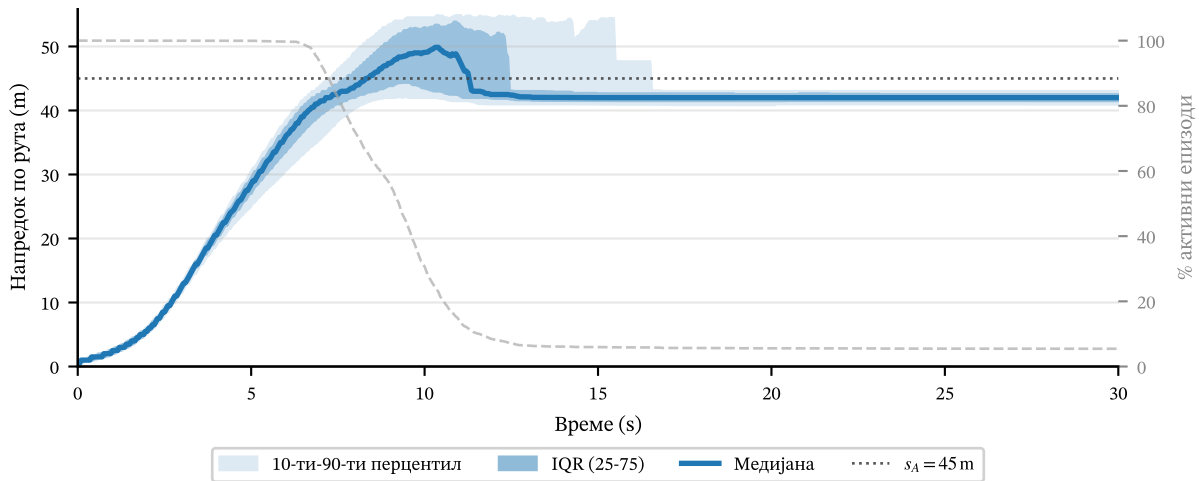


Слика 5.5: Распределба на брзината на возилото по напредок по рута за секој архетип (медијана, интерквартилен опсег (IQR, англ. Interquartile Range) и опсег 10-ти–90-ти перцентил). Испрекинатата вертикална линија го означува појавувањето на пешакот на $s_A = 45$ m. Десната оска (сива испрекинатата) го покажува процентот на епизоди што стигнале до секоја позиција по рутата.

Безбедносно релевантниот дел од сликата затоа е пристапот кон s_A , а не крајното забавување кај завршетокот на рутата. Брзината покажува делумна рана модулација, но доцно одлучувачко кочење, па следниот чекор е да се прегледа наредбата што го произведува падот на брзина.

Слика 5.6 ја изолира трагата на напредок по рута за Протестантот и ја проширува временската оска на 30 s. Хоризонталната линија на $s_A = 45$ m ја означува судирната точка и го разделува поминувањето на пешачкиот регион од останувањето пред блокираната точка на рутата. Проширениот поглед покажува дека нетривијален дел од епизодите со Протестант останува активен и под судирната точка долго по првиот пристап. Ова е надворешниот пандан на резултатот за блокада во чекање: полисата понекогаш може веднаш да стане безбедна со застанување или бавно движење, но потоа не го завршува вториот стадиум на однесувањето, односно разрешување на локално ограничениот блокада и продолжување.

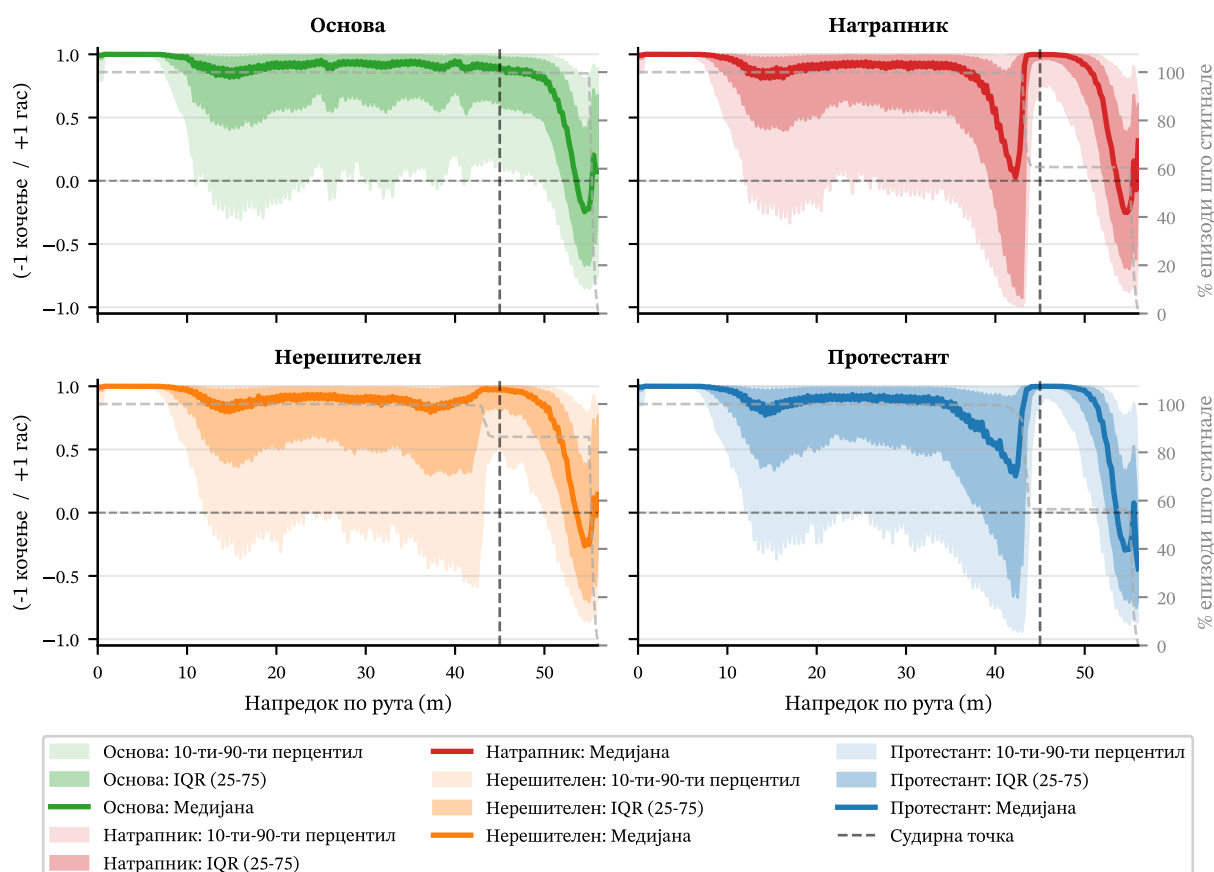
Слика 5.7 ја прикажува надолжната наредба. Основната наредба е блиску до целосен гас низ поголемиот дел од рутата, а конфликтните панели остануваат блиску до тој образец до истиот терминален регион каде што брзината почнува да паѓа. За Натрапникот и Протестантот, кочењето е нагло и доцно. Случајот Протестант е особено откривачки: пешакот ја зазема лентата од рано во епизодата, но надолжната наредба останува висок гас до крајниот пристап. Нерешителниот преминувач нејасно се опишува со овој образец. Многу епизоди изгледаат како да се саморазрешуваат преку временското замрзнување–продолжи однесување на пешакот, па агрегираниот надолжен сигнал не го покажува истото остро итно кочење. Профилот на наредбата затоа поддржува интерпретација водена од близина за Натрапникот и Протестантот, додека Нерешителниот случај мора да се чита заедно со темпирањето на пешакот и ТТС.



Слика 5.6: Проширен напредок по рута низ време за Протестантот. Обоените појаси ги покажуваат перцентилите на напредокот по рута во 30 s, хоризонталната точкеста линија ја означува судирната точка на пешакот $s_A = 45\text{ m}$, а сивата испрекината крива го покажува процентот на епизоди што сè уште се активни. Сликата го разделува ризикот од судир од дефанзивниот застој: некои епизоди избегнуваат непосреден удар, но остануваат блокирани пред пешакот.

Кочењето само по себе не одредува дали агентот се обидува да избегнува странично. Сепак, агрегирана странична анализа е суптилна. Суровото свртување не е доверлива мерка по рута бидејќи агентот постојано коригира лево и десно додека ја следи рутата. Просечно низ епизоди, позитивните и негативните корекции се поништуваат. Страничната грешка релативна на рутата го има истиот проблем на поништување, додека нејзината апсолутна вредност го губи насочниот контекст: движење подалеку од пешак од десната страна и лизгање кон пешак од левата страна можат да имаат иста апсолутна големина. Поради ова, свртувањето и страничното поместување се посоодветни за преглед на единечни епизоди отколку за агрегирани докази по рута. Слика 5.8 затоа ја користи апсолутната грешка во насоката, $|\psi_{\text{err}}|$. Конфликтните панели бавно и умерено се издигнуваат над основниот аголен шум пред судирната точка на пешакот. Агентот затоа не е чисто кочечки — понекогаш почнува да врти пред конфликтната точка. Ова покажува дека едноставното тврдење „нема страничен одговор“ не е доволно прецизно. Сепак, ефектот останува мал во агрегатниот приказ, бидејќи страничните маневри од различни епизоди не се случуваат на истата страна, во ист момент, ниту со иста геометрија. Графиконот за насока затоа треба да се третира како доказ за можна странична намера, не како доказ за валидно странично поминување. Потребен е поблизок преглед на единечни епизоди и на ниво на имагинација за да се види дали овие рани промени во насоката стануваат одржливи планови за заобиколување.

Сликите по рута ги претвораат агрегатните исходи во просторна приказна. Основниот агент не е целосно слеп за сцената со конфликтен пешак: и брзината и насоката покажуваат мали навремени промени. Но одлучувачкиот одговор сè уште е концентриран блиску до судирната точка, страничното однесување останува тешко за толкување при агрегација, а временската слика за Протестантот открива втор надворешен режим на неуспех во кој возилото останува заглавено пред пречката. Надворешно, ова е однесување на возач што почнува да го чувствува проблемот, но не создава доволна маргина доволно рано, ниту план за закрепнување по застанувањето. Она што останува скриено е дали управувањето во секој момент е врзано за непосредната опасност. Следната секција затоа го порамнува однесувањето според ТТС, а не според напредокот по рута.



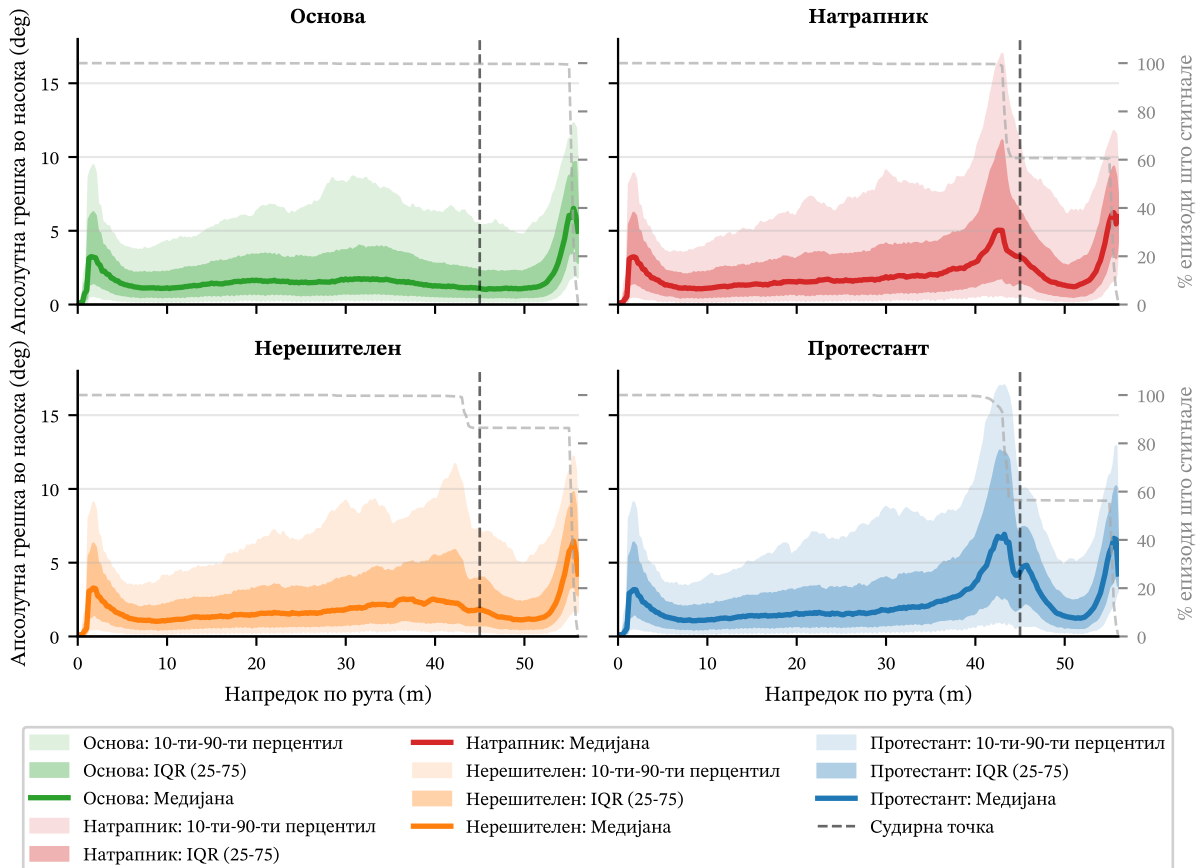
Слика 5.7: Распределба на надолжната наредба по напредок по рута (медијана, IQR, појас 10-ти–90-ти перцентил). Хоризонтална испрекината линија на нула ја означува границата помеѓу гас и кочница. Распоредот и десната оска се како во Слика 5.5.

5.5 Конфликтна динамика на ниво на чекор

Графиконите по рута покажуваат каде се случува одговорот во простор. Конфликтната динамика на ниво на чекор прашува кога наредбите се менуваат релативно на непосредниот ризик од судир. Ова е важно бидејќи позиција по рута блиску до s_d не е автоматски опасна во секоја епизода. Опасноста зависи од позицијата и брзината на пешакот, како и од состојбата на возилото.

Слика 5.9 ги прикажува реализираните вредности на акцијата наспроти моменталното ТТС низ сите чекори. Надолжните панели покажуваат висок гас при високо ТТС и пад кон кочење кога ТТС станува мало. Ова е конзистентно со резултатот по рута: полисата навистина реагира на опасност, но голем дел од одговорот е врзан за веќе непосредни состојби. Панелот за Протестантот содржи многу примероци со кочење при ниско ТТС: во конфликтниот прозорец $TTC < 3s$, 23,927 од 52,033 примероци за Протестантот имаат негативна надолжна наредба, а 21,600 имаат силно кочење под -0.3 . Ова покажува дека агентот често ја притиска кочницата пред пешак што веќе ја зазема лентата. Тоа не е бескорисно однесување, но е доцно.

Панелите за свртување ја усложнуваат приказната. Изразено свртување при ниско ТТС се појавува за Натрапникот, каде што останува малку време штом конфликтот стане видлив. За Протестантот и Нерешителниот преминувач, свртување е видно и при ТТС вредности над $3s$, што сугерира дека полисата понекогаш почнува да прави нешто пред да се отвори формалниот конфликтен прозорец. Ова е истата најава за проактивност видена во графиконот за насока по рута. Сепак, одговорот не е доволно конзистентно организиран за да гарантира расчистување, а густината на свртување сама по себе не може да покаже дали добиената



Слика 5.8: Распределба на апсолутна грешка во насоката $|\psi_{\text{err}}|$ по напредок по рута (медијана, IQR, појас 10-ти–90-ти перцентил). Основниот панел го утврдува аголниот шум на нормално следење рута. Издигнување над Основата во конфликтните панели одразува вистински одговор со вртење кон или од пешакот. Распоредот и десната оска се како во Слика 5.5.

патека безбедно поминува околу пешакот. Затоа графиконот на густина мотивира поблизок поглед кон анатомијата на судирите и геометријата на избегнување.

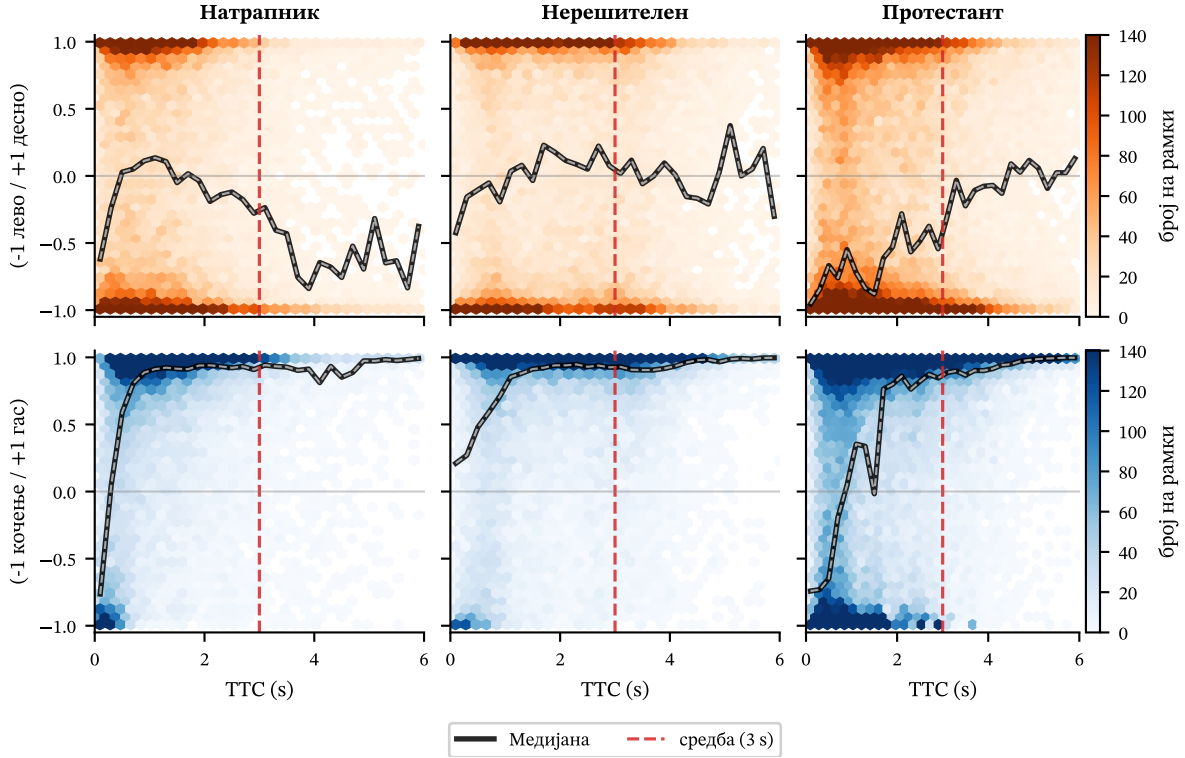
Ова сè уште остава една двосмисленост. Доцен одговор може да биде физички силен, но едноставно да нема време да проработи, или може да биде слаб и отсутен. Анатомијата на судирите ги раздвојува овие можности.

5.6 Анатомија на судири

Анатомијата на судири се фокусира само на епизоди што достигнуваат физички контакт со пешак. Таа прашува што правело возилото при удар и која модалност на избегнување, ако воопшто постоела, била присутна.

Слика 5.10 ја споредува брзината при најблиско приближување со брзината при удар. Распределбите на брзина при судир зависат од архетипот. Судирите со Натрапник се случуваат при највисоки преостанати брзини, што е конзистентно со конфликтен влезен јаз што остава малку време за кочење штом конфликтот стане видлив. Судирите со Протестант се случуваат при многу пониски брзини, што потврдува дека агентот често кочи пред удар. Но контакт со ниска брзина со пешак што ја задржува блокадата во мал регион околу судирната точка сè уште е неуспех: полисата ја намалила брзината без да создаде доволно растојание за запирање или странична маргина. Одговорот што постои е реален, но не создава доволна маргина пред контакт.

Слика 5.11 ги класифицира епизодите со судир во груби реализирани режими на неуспех.



Слика 5.9: Реализирани акции како функција од моменталното време до судир. Секој хексагон собира чекори со слично TTC и слична вредност на акцијата. Прикажани се само чекори со конечно TTC до 6 s. Горниот ред ја покажува акцијата за свртување, а долниот ред надолжната наредба, каде позитивни вредности значат гас, а негативни вредности кочење. Црната линија ја следи медијанската вредност на акцијата низ TTC, а вертикалната испрекината линија го означува влезот во конфликтниот прозорец при $TTC = 3$ s.

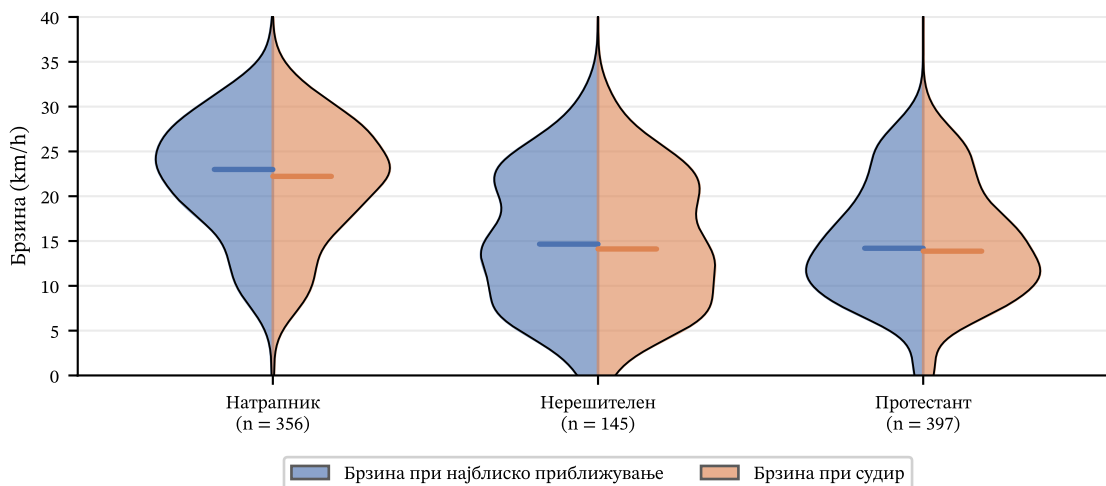
Класификацијата се применува само на епизоди со судир што влегле во конфликтниот прозорец. За секоја таква епизода, нека $\mathcal{W}$ е множеството временски чекори со конечно TTC под 3 s. Брзината пред средбата v_{pre} е медијанската брзина на возилото во 1 s пред првиот чекор во $\mathcal{W}$. Реализираните странични и надолжни големини на одговорот потоа се

$$L = \max_{i \in \mathcal{W}} |e_i^{lat}|, \quad \Delta v = \max\left(0, v_{pre} - \min_{i \in \mathcal{W}} v_i\right). \quad (5.3)$$

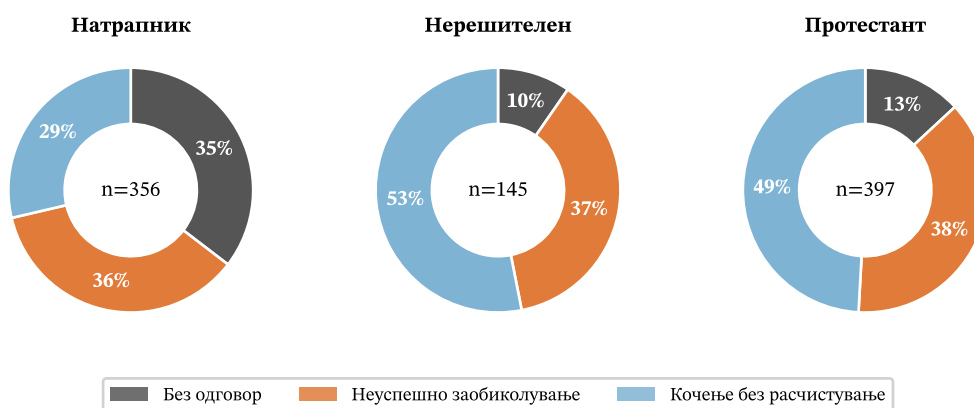
Страничен одговор се детектира кога $L > 0.5$ m, а надолжен одговор се детектира кога $\Delta v > 1.0$ m/s. Епизодите без ниту еден одговор се означуваат како *без одговор*. Епизодите со страничен или комбиниран одговор, но сепак со судир, се означуваат како *неуспешно заобиколување*. Епизодите само со надолжен одговор се означуваат како *кочење без расчистување*. Така, *без одговор* значи дека нема детектибилен реализиран одговор над овие прагови, а не нужно дека нема ненулта сурова наредба за управување.

Доминантниот неуспех е кочење без расчистување. Ова значи дека агентот често иницира надолжен одговор, но останува на курс кон судир. Неуспешното заобиколување е поретко, па типичниот судир не е чисто испланирано свртување што за влакно промашува. Почесто е доцен настан на кочење без доволна странична или временска маргина. Ова се вклопува со претходниот резултат за насока: вртење се појавува, но не како робусна стратегија за навремено заобиколување.

Слика 5.12 ја покажува континуираната размена помеѓу кочење и странично поместување. Повеќето епизоди со средба се групираат при умерен до висок пад на брзина, но ниско странично поместување. Маркерите за судир се вградени во овој кластер, што покажува дека кочењето само по себе често не е доволно. Епизодите со поголема странична маргина се



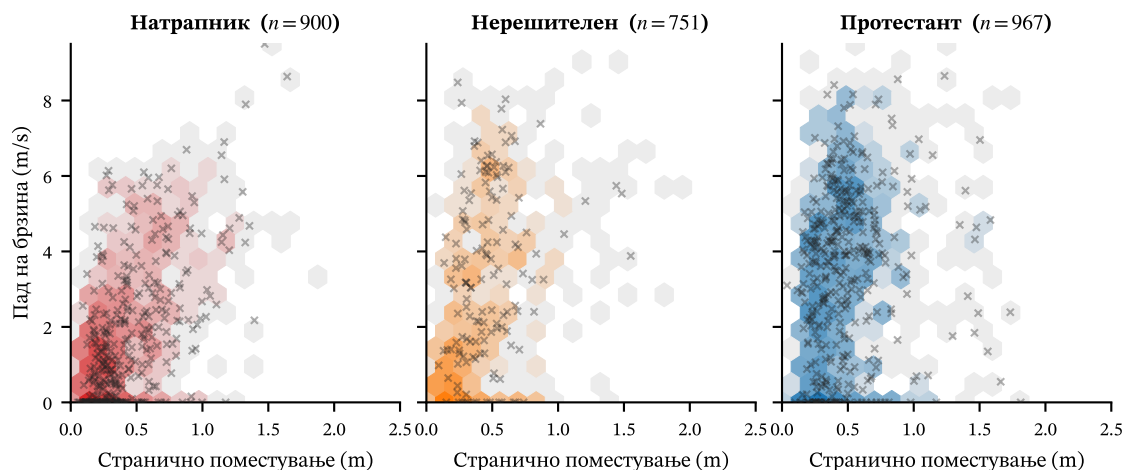
Слика 5.10: Брзина при најблиско приближување (лева половина од секоја поделена виолина) наспроти брзина при удар (десна половина), само за епизоди со судир. По една поделена виолина за секој архетип. Близината на двете половини покажува колку намалување на брзината се случува помеѓу моментот на најблиско приближување и физичкиот контакт.



Слика 5.11: Разложување на режими на неуспех за епизоди со судир. Прстенестите дијаграми го покажуваат делот од епизодите со судир во секој архетип класифицирани во три меѓусебно исклучиви режими: без детектибилен избегнувачки одговор, обид за странично избегнување без доволна маргина (неуспешно заобиколување) и надолжно кочење без странично расчистување.

поретки и имаат тенденција да бидат побезбедни. За Протестантот, геометријата е особено непростлива бидејќи пешакот останува локално ограничен околу судирната точка на средишната линија, наместо да ја расчисти рутата: агентот мора или да застане рано или да создаде доволна странична маргина за да помине околу блокадата. Основната полиса обично не прави ниту едно од двете доволно рано.

Надворешните докази сега се внатрешно конзистентни. Агентот има авторитет за итно управување: може силно да кочи и понекогаш врти. Но темпирањето и геометријата се погрешни. Тој задржува брзина блиску до основната и висок гас до близу судирната точка на пешакот, почнува да кочи и да врти откако конфликтниот прозорец веќе се отворил, и судира главно преку кочење без расчистување. Ова го поставува клучното каузално прашање: дали безбеден маневар бил замислен порано, но не бил извршен, или самите замислени иднини останале небезбедни додека не било предоцна? Следната секција ги отвора проекциите на Dreamer за да одговори на тоа прашање.



Слика 5.12: Динамика на избегнување за епизоди со средба. Оските се максималното странично поместување (m) и максималниот пад на брзина (m/s) постигнати за време на конфликтниот прозорец. Маркерите со x ги преклопуваат епизодите со судир. Заедничката распределба открива кои комбинации на кочење и странично поместување биле доволни за да се избегне удар.

5.7 Технички вовед во проекциите и реконструкциите на Dreamer

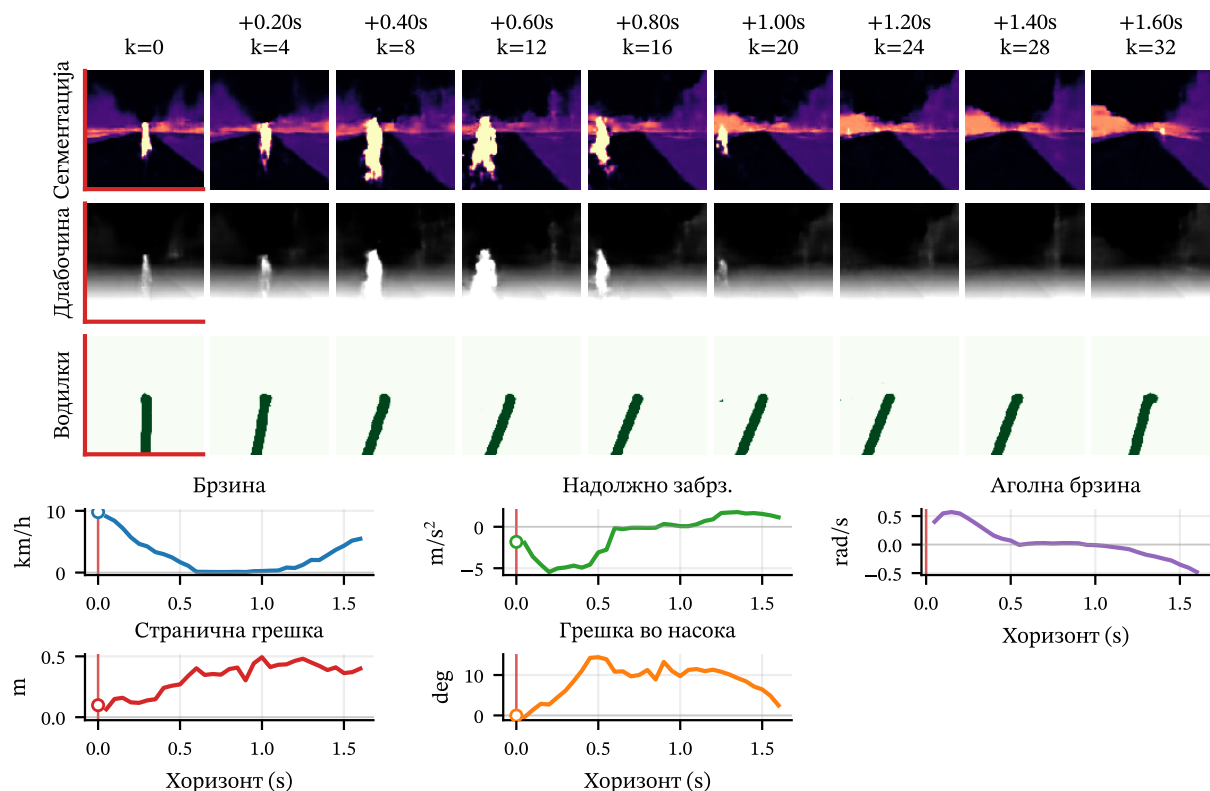
Dreamer е агент со модел. Во секој реален чекор на управување, историјата на набљудувања се енкодира во рекурентно латентно апостериорно верување што ја претставува тековната претстава на агентот за сцената. Актерот и критичарот се тренираат преку замислени идни генерирани од такви латентни состојби.

Во реалниот чекор t , агентот има латентно апостериорно верување условено од реализираната секвенца на набљудувања до тој чекор. Почнувајќи од тоа апостериорно верување, научениот рекурентен динамички модел се одмотува нанапред. Ова значи дека проекцијата не е дополнително рачно зададено геометриско предвидување. Таа е замислена иднина условена од полисата, произведена од истиот научен претставувачки модел што го користи агентот.

Секоја проекција има $H = 32$ замислени чекори при 20 Hz, што одговара на приближно 1.6 s. Овој хоризонт е избран намерно: иако динамичкиот модел технички може да се одмотува подолго, агентот е трениран со проекции од 32 чекори, па по тој хоризонт се очекува значително влошување на предвидувањата. Затоа евалуацијата го користи истиот хоризонт на кој полисата научила да расудува. Во секој замислен чекор, актерот ја обезбедува замислената акција, динамичкиот модел ја предвидува следната латентна состојба, а декодерските гранки реконструираат величини поврзани со замислената иднина. Конечниот хоризонт е важен: конфликт со пешак што лежи повеќе од околу 1.6 s напред можеби сè уште не е достиген во проекцијата. Затоа поделбата подолу прво прашува дали замислената рута воопшто ја достигнува конфликтната точка.

Квантитативната анализа користи реконструирани состојби на возилото и акциски величини: замислена брзина, замислена странична грешка, замислена грешка во насоката и замислена надолжна наредба. Овие сигнали се достапни во секој чекор на хоризонтот и можат да се сведат на детерминистички класни ознаки. Декодираните рамки за сегментација/длабочина се достапни за квалитативен преглед, но не се користат како метрики.

Декодираните рамки за сегментација/длабочина се вредни бидејќи даваат визуелен впечаток за тоа што претставувачкиот модел очекува да види. Тие можат и да сугерираат зошто агентот очекува сцената да се расчисти. Сепак, нивна употреба како автоматски детектор на верување за пешак би барала посебна студија за валидација. Малите објекти можат да се за-



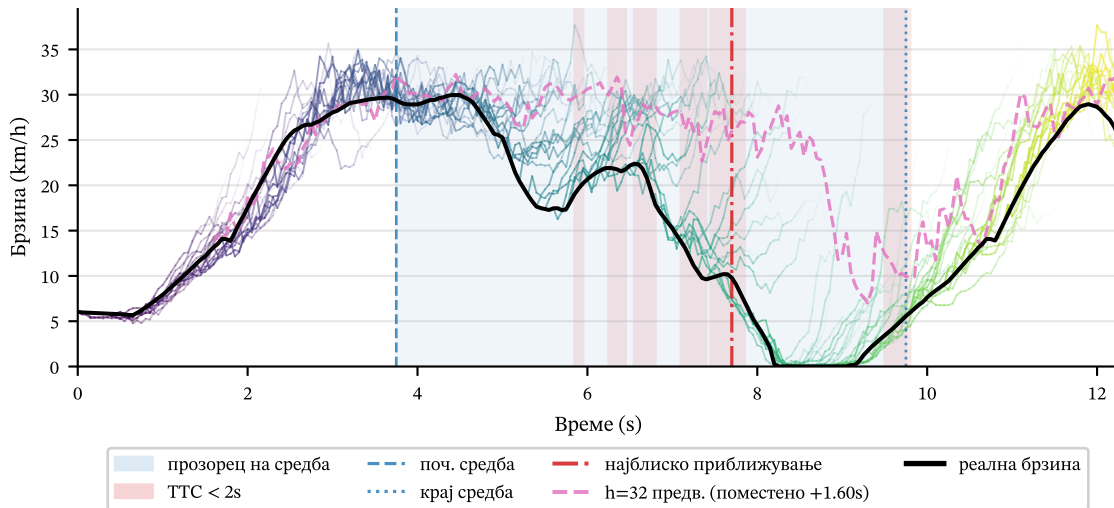
Слика 5.13: Декодирана реконструкција на проекција за епизода 24 на *Нерешителен ѓреминувач* во чекор 154 (најблиско приближување, ТТС ≈ 0.52 s, исход: цел). Колоните ја покажуваат апостериорната реконструкција во тековниот реален чекор ($k = 0$, црвена рамка) и замислените идни реконструкции при $k = 4, 8, \dots, 32$ во хоризонтот од 1.6 s. Редовите со слики ги покажуваат каналите за сегментациски ризик, инверзна длабочина и водилки. Долниот ред ги покажува декодираните компоненти на векторот на возилото (брзина, надолжно забрзување, аголна брзина на вртење, странична грешка, грешка во насоката) над истиот замислен хоризонт, со кружен маркер што ја означува реализираната вредност во чекор 154. Таксономската ознака во овој чекор е *давање ѓрвенсџиво*. Сликата е квалитативна: илустрира што претставувачкиот модел реконструира и предвидува во моментот на најблиско приближување, додека квантитативните тврдења во оваа глава користат метрики од проекциите на состојбата на возилото и акции.

матат, семантичките канали можат да бидат несовршени, а реконструирана форма слична на пешак не е исто што и калибрирана проценка на состојба на пешак. Поради тоа, главата ги задржува декодираните слики како квалитативен материјал за студија на случај, а сите агрегирани тврдења ги темели на проекции на состојбата на возилото и акцијата.

5.8 Истражувачки траги на имагинација во единечна епизода

Пред секоја замислена иднина да се сведе на класна ознака, корисно е да се покаже суровата форма на имагинацијата во една епизода. Овие слики се истражувачки и компаративни, а не главен популациски доказ. Епизодата прикажана низ оваа секција е епизода 24 на *Нерешителен ѓреминувач* (исход: цел, најблиско приближување во чекор 154, ТТС ≈ 0.52 s). Слика 5.13 прво го покажува декодираниот поглед на претставувачкиот модел во самиот чекор на најблиско приближување. Следните две слики го прошируваат тој поглед низ целата временска линија на епизодата и низ целата просторна патека.

Слика 5.14 го поставува најдиректното внатрешно прашање: дали агентот замислува заба-



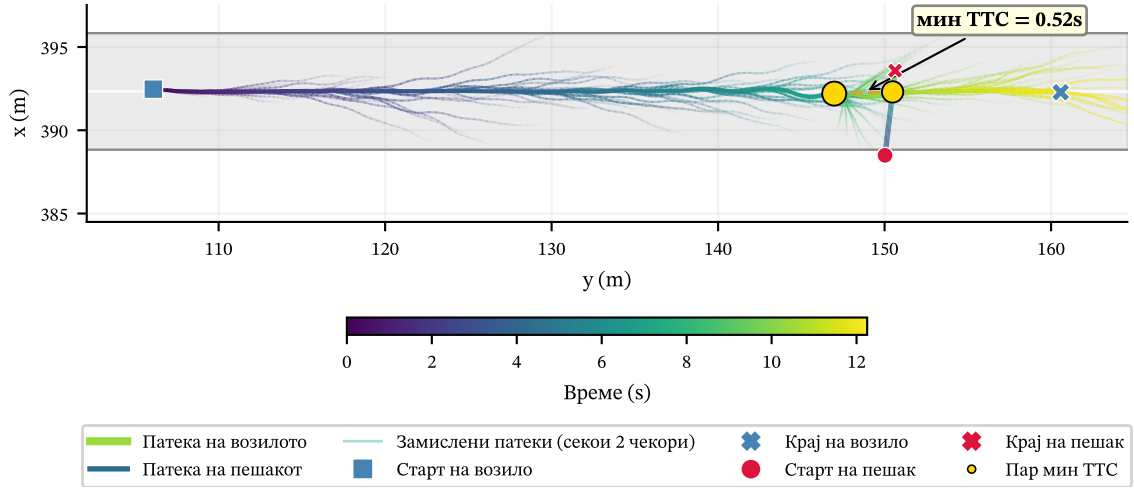
Слика 5.14: Замислена брзина низ време на епизодата за *Нерешителен ѝреминувач*, епизода 24. Секоја избледувачка нишка е замислен хоризонт на брзина од 1.6 s лансиран од еден реален чекор на одлучување (нацртан е секој втор чекор). Бојата го кодира времето на епизодата од рано (темно) до доцно (светло). Црната полна линија е реализираната брзина. Синото засенчување го означува прозорецот на средба, а црвеното засенчување ги означува чекорите каде $TTC < 2$ s. Розовата испрекинатата линија е предвидената брзина на далечниот хоризонт ($k = 32$) поместена нанапред за 1.6 s. Замислената брзина останува слична на брзина на слободно возење во поголемиот дел од пристапот и паѓа само кога конфликтот е веќе непосреден, совпаѓајќи се со доцниот надворешен профил на кочење.

вување пред средбата со пешакот да стане итна? Нишките во *viridis* бои го прават одговорот видлив: низ раниот и средниот пристап, хоризонтите лансирани од секој реален чекор остануваат на или над реализираната брзина на слободно возење. Замислената брзина не се свиткува надолу сè додека регионот $TTC < 2$ s (црвено засенчување) веќе не е активен. Ова значи дека внатрешниот план на агентот се совпаѓа со доцниот надворешен профил на кочење виден во сликите по рута: и реализираната акција и замислената иднина остануваат на брзина на слободно возење додека конфликтот не е блиску. Розовата трага на далечниот хоризонт, поместена нанапред за 1.6 s, го прави темпирањето експлицитно — предвиденото забавување ја следи, наместо да ја антиципира, вистинската промена на брзината.

Брзината сама по себе не може да покаже дали агентот планира да помине околу пешакот. Слика 5.15 ја додава просторната димензија. Замислените нишки ја покриваат целата епизода: раните нишки (темни) покажуваат право движење по лентата. Нишките блиску до средбата остануваат центрирани на рутата без странично свиткување околу s_A . Само последните нишки блиску до најблиското приближување покажуваат одредено странично ширење. Нема доказ за одржлив страничен план што се формира навремено пред конфликтот — замислените патеки на агентот не прикажуваат свртување. Ова е конзистентно со декодираната реконструкција во чекор 154 (Слика 5.13), каде што таксономската ознака беше *давање ѝрвенсѝво*, а не странично избегнување. Заедно, двете истражувачки слики сугерираат дека внатрешните иднини на агентот се надолжно реактивни и странично пасивни, што мотивира таксономија што може да ги измери двете димензии низ секој чекор од секоја епизода.

5.9 Таксономија на проекции

Таксономијата ја претвора секоја замислена иднина во една ознака на однесување. Ова овозможува споредба на имагинацијата низ чекори, епизоди, исходи и архетипови на конфликтни пешаци. Класификацијата користи само реконструирани величини од состојбата на возилото и замислени акции.



Слика 5.15: Замислени траектории преклопени врз реализираната патека за *Нерешителен преминувач*, епизода 24. Дебелата линија во viridis бои е реализираната патека на возилото. Потенката viridis линија е реализираната патека на пешакот (круг: почеток, x: крај). Од секој втор реален чекор, краткорхоризонтна замислена патека се одредува и се црта како избледувачка нишка обоена според времето на потекло. Златните маркери поврзани со испрекинатата линија го означуваат парот на најблиско приближување. Замислените патеки остануваат тесно центрирани на лентата во поголемиот дел од пристапот, со странично отстапување што се појавува само блиску до чекорот на најблиско приближување.

Нека $H = 32$ е хоризонтот на имагинацијата, $\Delta t = 0.05 \text{ s}$ е чекорот на симулаторот, s_t е реализираниот напредок по рута во чекор на одлучување t , а $s_A = 45 \text{ m}$ е конфликтната судирна точка на пешакот. За чекор на хоризонтот $h \in \{1, \dots, H\}$, замислената брзина ја означуваме со $\hat{v}_{t,h}$, страничната грешка релативна на рутата со $\hat{e}_{t,h}^{\text{lat}}$, грешката во насоката со $\hat{e}_{t,h}^{\psi}$, а замислената надолжна наредба со $\hat{a}_{t,h}^{\ell}$.

Замислениот напредок по рута се апроксимира со интегрирање на напредната компонента на замислената брзина:

$$\hat{s}_{t,h} = s_t + \Delta t \sum_{j=1}^h \max\left(0, \hat{v}_{t,j} \cos\left(\text{clip}(\hat{e}_{t,j}^{\psi}, -1.2, 1.2)\right)\right). \quad (5.4)$$

Кратењето на $\pm 1.2 \text{ rad}$ ($\approx \pm 69^\circ$) е нумеричка заштита: ако деградиран чекор од проекцијата произведе екстремна грешка во насоката, нескратениот косинус би можел да стане силно негативен и да одземе од акумулираниот напредок по рута, создавајќи недоверлива проценка на напредок. Кратењето ја одржува проекцијата добро условена без да влијае на нормалните чекори на проекцијата, каде што грешките во насоката се мали.

Примерокот од хоризонтот најблизок до судирната точка е

$$k_t = \arg \min_{h \in \{1, \dots, H\}} |\hat{s}_{t,h} - s_A|. \quad (5.5)$$

Проекцијата достигнува конфликт кога замислената рута влегува во конфликтниот појас $\pm 2.5 \text{ m}$ или завршува доволно блиску од него:

$$C_t = \mathbf{1} \left[\min_h |\hat{s}_{t,h} - s_A| \leq \delta_s \vee \hat{s}_{t,H} \geq s_A - \delta_s \right], \quad \delta_s = 2.5 \text{ m}. \quad (5.6)$$

Множеството примероци од хоризонтот во конфликтниот појас е

$$\mathcal{K}_t = \{h : |\hat{s}_{t,h} - s_A| \leq \delta_s\}. \quad (5.7)$$

Овој прв тест е суштински. Од конечна проекција од 1.6 s не може да се очекува да класифицира локација на пешак до која сè уште не стигнала.

Страничната намера не може да се заснова само на крајното поместување. Замислената странична грешка е величина со предзнак, а една проекција може да се движи кон двете страни на средишната линија на рутата во истиот хоризонт. Таксономијата затоа користи апсолутен страничен признак:

$$e_{t,A}^{\text{lat}} = |\hat{e}_{t,k_t}^{\text{lat}}|, \quad (5.8)$$

$$Q_t^{\text{lat}} = \text{P90}_h(|\hat{e}_{t,h}^{\text{lat}}|), \quad (5.9)$$

$$F_t^{\text{lat}} = \frac{1}{H} \sum_{h=1}^H \mathbf{1}[|\hat{e}_{t,h}^{\text{lat}}| \geq 0.6], \quad (5.10)$$

$$\Psi_t = \max_h |\hat{e}_{t,h}^{\psi}| \cdot \frac{180}{\pi}. \quad (5.11)$$

Индикаторот за странична намера е

$$L_t = \mathbf{1}[e_{t,A}^{\text{lat}} \geq 0.8 \vee F_t^{\text{lat}} \geq 0.25 \vee (\Psi_t \geq 10^\circ \wedge Q_t^{\text{lat}} \geq 0.6)]. \quad (5.12)$$

Првиот услов се активира кога апсолутната странична грешка во најблискиот хоризонтски примерок k_t надминува 0.8 m, што укажува дека замислената рута поминува покрај судирната точка. Вториот се активира кога странично поместување над 0.6 m зазема најмалку една четвртина од хоризонтот, опфаќајќи проекција која содржи привремено отстапување од рутата. Третиот бара и максимално отстапување на насоката од најмалку 10° и 90-ти перцентил на апсолутно странично поместување од најмалку 0.6 m, така што проекција со вртење се означува само кога промената на насока е придружена со вистинско странично поместување, а не со отстапување на насоката што се враќа на средишната линија.

Намерата за давање првенство се детектира преку замислено намалување на брзина, одржливо кочење или брзина блиску до запирање во конфликтниот појас:

$$\Delta v_t = \max\left(0, \hat{v}_{t,1} - \min_h \hat{v}_{t,h}\right), \quad (5.13)$$

$$B_t = \frac{1}{H} \sum_{h=1}^H \mathbf{1}[\hat{a}_{t,h}^{\ell} < -0.2], \quad (5.14)$$

$$v_{t,A}^{\text{band}} = \min_{h \in \mathcal{K}_t} \hat{v}_{t,h}. \quad (5.15)$$

Индикаторот за намера за давање првенство е

$$Y_t = \mathbf{1}[\Delta v_t \geq 1.0 \vee B_t \geq 0.25 \vee v_{t,A}^{\text{band}} \leq 1.0]. \quad (5.16)$$

Првиот услов се активира кога замислената брзина паѓа за најмалку 1 m/s од почетокот на хоризонтот до неговиот минимум, опфаќајќи каква било значајна намера за забавување. Вториот се активира кога акциите за кочење под -0.2 заземаат најмалку една четвртина од хоризонтот, така што кратките импулси на кочење се разликуваат од одржливо давање првенство. Третиот се активира кога минималната замислена брзина во конфликтниот појас $\mathcal{K}_t$ паѓа на или под 1 m/s. Оценувањето низ целиот појас, а не само во k_t , гарантира дека проекција блиску до запирање не се пропушта кога примероците од хоризонтот паѓаат малку пред или по судирната точка.

Доделувањето класа е детерминистичко:

C_t	L_t	Y_t	Класа на проекција
0	*	0	Конфликт на двор од хоризонт
0	*	1	Давање првенство пред конфликт
1	0	0	Право низ конфликтот
1	0	1	Давање првенство
1	1	0	Странично избегнување
1	1	1	Кочење + странично

Тука * значи дека страничната намера се игнорира додека конфликтната точка е надвор од хоризонтот. Со зборови, *Конфликт на двор од хоризонт* значи дека замислената иднина завршува пред пешачкиот регион да стане релевантен. *Давање првенство пред конфликт* значи дека агентот замислува забавување пред хоризонтот да ја достигне судирната точка. *Право низ конфликтот* значи дека конфликтот е достиген, но проекцијата не содржи ниту намера за давање првенство ниту странична намера. *Давање првенство* е надолжно избегнување што го достигнува конфликтот. *Странично избегнување* е странично избегнување што го достигнува конфликтот без значајно забавување. *Кочење + странично* ги комбинира двата режима.

Оваа таксономија ја раздвојува свесноста за конфликт од намерата за избегнување. Ниско реално ТТС не значи нужно дека замислената рута од 1.6 s го достигнала s_A . Обратно, штом $C_t = 1$, ознаката право низ конфликтот е директен доказ дека внатрешната иднина на агентот не содржи избегнување кај судирната точка на пешакот.

Слика 5.16 ја применува таксономијата на студиската епизода користена низ оваа глава — *Нерешителен преминувач*, епизода 24, исход: цел. Хоризонталната оска е времето на епизодата. Секој обоен сегмент е континуиран период од истата класа на проекција. Сликата ја прави таксономијата конкретна пред да се воведат агрегатните метрики: раната епизода е доминирана од *Конфликт на двор од хоризонт* (судирната точка на пешакот сè уште не е достижна во 1.6 s), *Давање првенство пред конфликт* и *Право низ конфликтот* се појавуваат кога конфликтот влегува во хоризонтот, а епизодата завршува со *Давање првенство* близу најблиското приближување — што се совпаѓа со таксономската ознака пријавена во декодираната реконструкција (Слика 5.13). Брзото префрлање меѓу класите видливо на сликата е изразот по епизода на нестабилноста на намерата што агрегатните метрики ќе ја квантифицираат.

5.10 Агрегатни метрики за проактивност

Таксономијата му дава ознака на секој реален чекор на одлучување. Метриците за проактивност ги сумираат тие секвенци на ознаки на ниво на епизода. Нека t_i^{ca} е чекорот на најблиско приближување на епизодата i , а $\ell_{i,t}$ е класата на проекцијата доделена во чекор t .

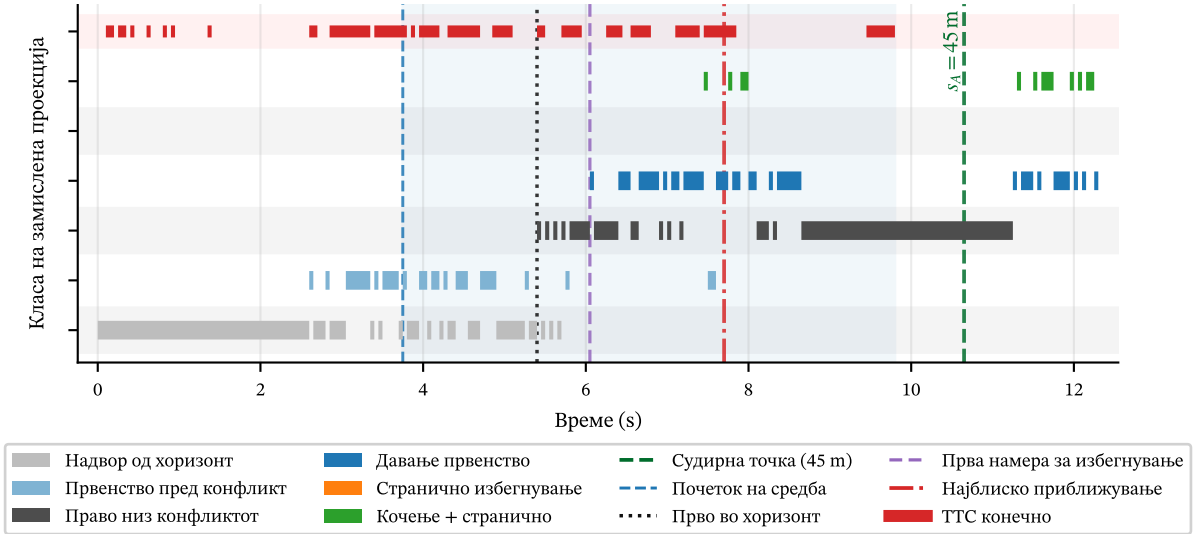
Множеството пред-најблиско достигнување конфликт $\mathcal{R}_i$ ги собира сите чекори на одлучување пред најблиско приближување во кои замислената иднина ја достигнува судирната точка на пешакот, односно чекорите каде агентот има каква било внатрешна информација за конфликтот:

$$\mathcal{R}_i = \{t \leq t_i^{ca} : C_{i,t} = 1\}. \quad (5.17)$$

Активното избегнувачко подмножество $\mathcal{A}_i$ е подмножеството од тие чекори каде замислената иднина содржи експлицитна намера за избегнување, надолжна, странична или комбинирана:

$$\mathcal{A}_i = \{t \in \mathcal{R}_i : \ell_{i,t} \in \{\text{давање првенство, странично избегнување, кочење + странично}\}\}. \quad (5.18)$$

Време на претходно на конфликтната свесност τ_i^C мери колку порано од најблиското приближување замислената иднина првпат го достигнува пешачкиот регион, односно колку рано



Слика 5.16: Временска линија на класите на проекции за *Нерешителен преминувач*, епизода 24 (исход: цел). Секој хоризонтален сегмент е континуиран период од истата класа на замислена проекција. Црвената линија го покажува реализираното TTC. Вертикалните испрекинати линии го означуваат влезот во прозорецот на средба и најблиското приближување. Раната епизода е доминирана од *Конфликт надвор од хоризонт*. *Право низ конфликтот* и *Давање првенство* се менуваат штом судирната точка влегува во хоризонтот. Епизодата се разрешува со *Давање првенство* при најблиско приближување.

агентот станува внатрешно свесен за конфликтот:

$$\tau_i^C = (t_i^{ca} - \min \mathcal{R}_i) \Delta t. \quad (5.19)$$

Време на претходно намера за избегнување τ_i^A мери колку порано од најблиското приближување замислената иднина првпат содржи активен план за избегнување, односно колку рано свесноста се претвора во намера:

$$\tau_i^A = (t_i^{ca} - \min \mathcal{A}_i) \Delta t. \quad (5.20)$$

Истите величини можат да се изразат и во простор по рута, како растојанија наместо времетраења, што е корисно кога епизодите се разликуваат по брзина на пристап. Растојанието на конфликтна свесност d_i^C е преостанатиот напредок по рута до судирната точка во првиот чекор што го достигнува конфликтот,

$$d_i^C = s_A - s_{i,\min \mathcal{R}_i}, \quad (5.21)$$

а растојанието на намера за избегнување d_i^A е соодветното преостанато растојание во првиот чекор каде се појавува намера за избегнување:

$$d_i^A = s_A - s_{i,\min \mathcal{A}_i}. \quad (5.22)$$

Упорност на небезбедна имација U_i е делот од чекорите што го достигнуваат конфликтот каде замислената иднина сè уште вози право низ пешакот, мерејќи колку често свесноста не станува избегнување:

$$U_i = \frac{1}{|\mathcal{R}_i|} \sum_{t \in \mathcal{R}_i} \mathbf{1}[\ell_{i,t} = \text{право низ конфликтот}]. \quad (5.23)$$

Дел на активно избегнување A_i е комплементарниот дел од чекорите што го достигнуваат

Табела 5.3: Метрики за проактивност засновани на имитација, пресметани од таксономијата на проекции во Фаза 1. Времињата на претходење се мерат до најблиското приближување. Небезбедната упорност е средниот дел од замислениите проекции што го достигнуваат конфликтот и се означени како право низ конфликтот. Активношто избегнување ги комбинира давање првенство, странично избегнување и кочење + странично. Блокадата во чекање мери речиси нулти напредок по рушта блиску до пешакош додека замислената проекција останува во класа на давање првенство.

Метрика	Натрапник	Нерешителен	Протестант
Епизоди	100	100	100
Претходење на свесност, медијана (s)	1.70	1.50	2.15
Претходење на намера за избегнување, медијана (s)	1.48	1.52	2.00
Растојание на свесност, медијана (m)	16.0	16.0	16.0
Растојание на намера, медијана (m)	15.0	15.0	15.0
Небезбедна упорност (%)	44.3	50.3	45.6
Дел со активно избегнување (%)	55.7	49.7	54.4
Епизоди со блокада во чекање (%)	0.0	0.0	9.0
Време поминато во чекање (s/еп.)	0.0	0.1	6.2
Стапка на промена (s ⁻¹)	5.96	6.48	6.46
Средно време на задржување (s)	0.17	0.13	0.14
Доминантен режим што стигнува до конфликт	право низ конфликтот	право низ конфликтот	право низ конфликтот

конфликтот каде е присутна каква било намера за избегнување:

$$A_i = \frac{1}{|\mathcal{R}_i|} \sum_{t \in \mathcal{R}_i} \mathbf{1}[t \in \mathcal{A}_i]. \quad (5.24)$$

Распределба по режими го разложуваат A_i на трите избегнувачки класи — давање првенство, странично избегнување и кочење-плус-странично — откривајќи дали внатрешното избегнување на агентот е претежно надолжно, странично или комбинирано. *Стабилноста на намерата* се мери со стапката на промена на ознака во секунда и средното време на престој во континуирани периоди со иста ознака во $\mathcal{R}_i$. Кратки престои укажуваат дека замислениот план трепери меѓу класи наместо да се посвети на одржлив маневар.

Протестантот исто така бара метрика за безбедно, но нерешено чекање, бидејќи епизода со локално ограничена пречка може да заврши со истек на време без судир и без безбедна замислена проекција право низ конфликтот. Нека $q_{i,t} = s_A - s_{i,t}$ е преостанатото растојание по рута до судирната точка, а $\Delta s_{i,t}^{1s}$ е реализираниот напредок по рута во центриран прозорец од 1 s. *Чекор на дефанзивна блокада во чекање* $D_{i,t}$ означува момент кога возилото е блиску до пешак што ја држи блокадата околу судирната точка, речиси не се движи и замислува план за давање првенство — заедно образец на агент што држи позиција без да ја разреши блокадата:

$$D_{i,t} = \mathbf{1}[0 \leq q_{i,t} \leq 10 \wedge \Delta s_{i,t}^{1s} \leq 0.25 \\ \wedge \ell_{i,t} \in \{\text{давање првенство пред конфликт,} \\ \text{давање првенство, кочење + странично}\}]. \quad (5.25)$$

Времето доминало во чекање T_i^D е вкупното време поминато во чекори на блокада во чекање, мерејќи колку долго епизодата останува во нерешено дефанзивно однесување:

$$T_i^D = \Delta t \sum_t D_{i,t}. \quad (5.26)$$

Епизода се брои како блокада во чекање кога најдолгиот континуиран период со $D_{i,t} = 1$ трае најмалку 5 s, разликувајќи одржлив застој од кратки паузи.

Табела 5.3 покажува дека агентот не е едноставно слеп. Медијанското време на претходење на конфликтната свесност е 1.70 s за Натрапникот, 1.50 s за Нерешителниот преминувач и 2.15 s за Протестантот. Времето на претходење на намерата за избегнување е 1.48 s, 1.52 s и 2.00 s. Во однос на преостанатото растојание по рута до судирната точка, свесноста првпат се појавува при медијана од околу 16 m пред судирната точка, а намерата за избегнување при околу 15 m.

Ова е само слабо проактивно. При набљудуваните конфликтни брзини, 15 m одговараат на приближно 2.0–2.5 s патување, само малку подолго од хоризонтот на имагинација од 1.6 s. Избегнување се појавува, но се појавува блиску до физичките граници на краткохоризонтниот планер.

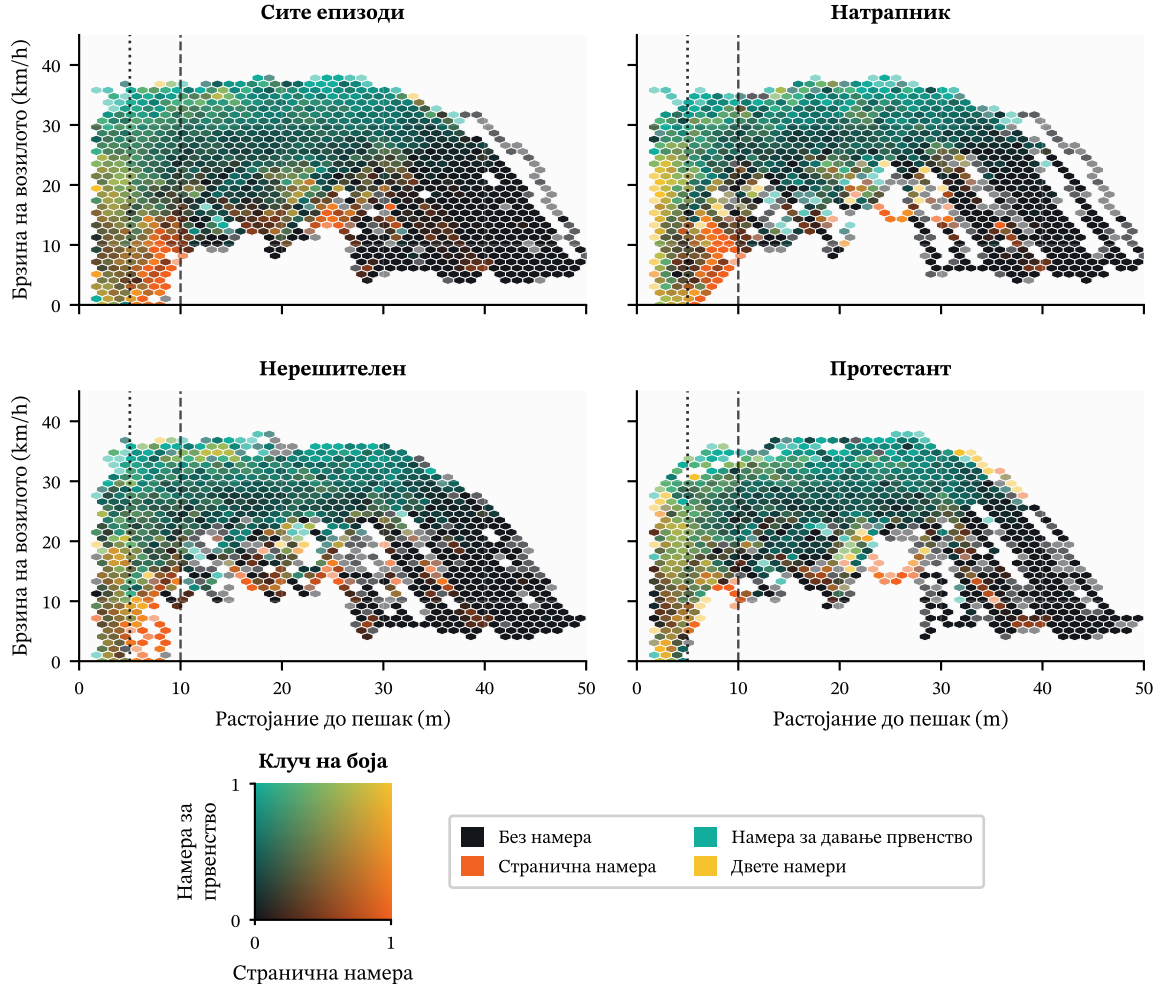
Поважниот резултат е што се случува откако конфликтот станува достиген. Небезбедната упорност на директно продолжување останува висока: 44.3% за Натрапникот, 50.3% за Нерешителниот преминувач и 45.6% за Протестантот. Во приближно половина од чекорите на одлучување каде замислената иднина го достигнува пешачкиот регион, замислената иднина сè уште вози низ него. Квалитативниот преглед на декодираните рамки на проекции сугерира веројатен механизам: претставувачкиот модел често го реконструира пешакот како веќе да ја расчистил патеката до моментот кога замислената рута пристигнува на s_A , па не се активира намера за избегнување иако реалниот пешак останува присутен. Ова не е валидирано квантитативно откритие — декодираните слики не се користат како детектор на пешак во оваа глава — но образецот е доволно конзистентен низ прегледаните епизоди за да вреди да се забележи како можно објаснување за високата стапка на директно продолжување. Активното избегнување го зазема преостанатиот дел, главно преку надолжно давање првенство. Чистото странично избегнување е мало, додека кочење-плус-странично се појавува како секундарен комбиниран режим.

Ова откритие го менува начинот на кој треба да се интерпретираат успешните епизоди. Во поделбата според исход, епизодите што стигнуваат до целта сè уште содржат значајна небезбедна упорност: 44.2% за цели кај Натрапник, 52.0% за цели кај Нерешителен преминувач и 50.3% за цели кај Протестант. Успехот затоа не значи неопходно дека агентот држел чист проактивен план. Некои успеси се случуваат и покрај нестабилни внатрешни иднини, додека некои судири се случуваат откако избегнување се појавило, но предоцна или премногу прекинато.

Стабилноста на намерата е слаба. Средните стапки на промена се 5.96, 6.48 и 6.46 промени на класа во секунда за Натрапник, Нерешителен преминувач и Протестант. Средните времиња на престој во иста класа се само 0.17 s, 0.13 s и 0.14 s. Овие времиња на престој се кратки во однос на физичкото време потребно за извршување значаен маневар за избегнување пешак. Внатрешната иднина на агентот трепери меѓу некомпатибилни алтернативи наместо да се смири во одржлив план за избегнување.

Метриката за блокада во чекање го издвојува Протестантот од архетиповите што ја преминуваат рутата. Натрапникот и Нерешителниот преминувач немаат епизоди со блокада во чекање подолга од 5 s. Протестантот има стапка на блокада во чекање од 9.0% и средно време поминато во чекање од 6.2 s по епизода. Меѓу епизодите со Протестант што завршуваат со истек на време, конкретно, 70.0% содржат блокада во чекање, а средното време поминато во чекање е 59.0 s. Ова е безбедно, но нецелосно однесување: агентот престанува да биде непосредно опасен, но не произведува план што ја разрешува локално ограничениот блокада и ја завршува рутата.

Претходните метрики се сумирања на епизоди. Комплементарен поглед е да се праша како замислената намера се распределува над реализираната состојба на средбата. Слика 5.17 го поставува секој реален чекор на одлучување од преминот за имагинација во хексагон според растојание до пешак и брзина на возилото. За секој зафатен хексагон, бојата ги кодира средниот индикатор за замислена странична намера и средниот индикатор за замислена намера за давање првенство на проекциите лансирани од тие реални состојби. Сликата е корисна



Слика 5.17: Замислена намера условена од состојба над реализирано растојание до пешак и брзина на возилото. Секој хексагон содржи реални чекори на одлучување. Хоризонталната оска е реализираното растојание од возилото до пешакот, а вертикалната оска е реализираната брзина на возилото. Бојата ја кодира средната замислена намера на проекциите лансирани од тие состојби: темно означува ниту една намера, портокалово странична намера, тиркизно намера за давање првенство, а златно комбинирана намера за давање првенство и странично избегнување. Хексагоните со помалку од три примероци се благо избледени, но не се отстранети. Испрекинатите и точкестите вертикални линии означуваат растојание до пешак од 10 m и 5 m. Панелот *Сите епизоди* е здружен по чекори, па ја одразува и зафатеноста на состојбите и фреквенцијата на архетиповите.

бидејќи ја претвора таксономијата на проекции во внатрешно поле на одговор. Десната страна на секој панел одговара на големи растојанија до пешак. Левата страна одговара на мали растојанија до пешак. Проактивна полиса треба да покажува зголемено активно избегнување како што растојанието се намалува, а брзината останува ненулта. Полиса што останува темна во блиски, подвижни состојби е внатрешно тивка токму во состојбите каде безбеден план е најпотребен.

Континуираната боја во ова поле не е нов континуиран излез од моделот. Таа е агрегат од бинарните индикатори дефинирани во таксономијата. Ако $\mathcal{S}_b$ е множеството реални чекори што паѓаат во хексагон b , тогаш средната странична намера и средната намера за давање првенство се

$$L_b = \frac{1}{|\mathcal{S}_b|} \sum_{(i,t) \in \mathcal{S}_b} L_{i,t}, \quad Y_b = \frac{1}{|\mathcal{S}_b|} \sum_{(i,t) \in \mathcal{S}_b} Y_{i,t}. \quad (5.27)$$

Полето може да се сведе на скаларни бенчмарк метрики што одговараат на четири пра-

Табела 5.4: Бенчмарк метрики за имажинација изведени од йоле, йресметани од йолето на намера условено од состојба. DWISI: ојасносно-тежински дел од замислени иднини со каква било намера за избеѓнување (йовисоко е йодобро). DSR: ојасносно-тежински тивок дел во блиски йодвижни состојби $d \leq 10 \text{ m}$, $v \geq 5 \text{ km/h}$ (йониско е йодобро). D_{50} : најолемо растојание до йешак йри кое избеѓнувањето е мнозинско замислено однесување (йовисоко е йодобро). AOA: дел од хексагоните со средно избеѓнување над 0.5 (йовисоко е йодобро). $\bar{\theta}$: ојасносно-тежински агол на режим на избеѓнување (90° =само давање йрвенство, 0° =само стйранично, 45° =балансирано). YSCI: индекс на стйрејнајост давање йрвенство-сврйување (йовисоко = йовеќе комбинирани йланови).

Метрика	Насока	Натрапник	Нерешителен	Протестант
DWISI	↑	0.588	0.549	0.566
DSR	↓	0.238	0.332	0.288
D_{50} (m)	↑	30.7	31.2	31.2
AOA	↑	0.490	0.473	0.473
$\bar{\theta}$ (deg)	—	74.1	75.8	70.2
YSCI	↑	0.206	0.136	0.216

шања: колку избегнување се замислува во опасни состојби, колку рано се појавува во простор на растојание, каков облик има замислената стратегија, и дали небезбедната имажинација е упорна или само трепери. Табела 5.4 ги пријавува сите вредности. Пасусите подолу ја дефинираат секоја метрика и го интерпретираат резултатот.

Нека $b \in \mathcal{B}^{(k)}$ индексира зафатени хексагони за архетип k , со централно растојание d_b , брзина на возилото v_b , број на примероци n_b , средна намера за давање првенство Y_b и средна странична намера L_b . Активното избегнување и тишината по хексагон се

$$M_b = 1 - (1 - L_b)(1 - Y_b), \quad (5.28)$$

$$S_b = 1 - M_b. \quad (5.29)$$

Тежината на опасност

$$\omega_b = n_b \frac{v_b}{d_b + 1}, \quad (5.30)$$

монотонно ги нагласува блиските, брзи состојби без да биде физички TTC модел. *Тежинскиот индекс на ојасноост йри имажинација* (DWISI, англ. *Danger-Weighted Imagination Safety Index*)

$$\text{DWISI}^{(k)} = \frac{\sum_b \omega_b M_b}{\sum_b \omega_b} \quad (5.31)$$

го мери опасносно-тежинскиот дел од замислени иднини што содржат каква било намера за избегнување. Вредностите се 0.588 (Натрапник), 0.549 (Нерешителен преминувач) и 0.566 (Протестант): агентот замислува избегнување во мало мнозинство од опасносно-тежинските состојби, но маргината е тесна.

Стйайкајта на ојасна тйишина (DSR, англ. *Dangerous Silence Rate*) го ограничува вниманието на блискиот опасен регион

$$\mathcal{D}^{(k)} = \{b \in \mathcal{B}^{(k)} : d_b \leq 10 \text{ m}, v_b \geq 5 \text{ km/h}\} \quad (5.32)$$

и го мери опасносно-тежинскиот дел од тие состојби каде агентот не замислува ниту давање првенство ниту свртување:

$$\text{DSR}^{(k)} = \frac{\sum_{b \in \mathcal{D}^{(k)}} \omega_b S_b}{\sum_{b \in \mathcal{D}^{(k)}} \omega_b}. \quad (5.33)$$

Вредностите се 0.238 (Натрапник), 0.332 (Нерешителен преминувач) и 0.288 (Протестант). Приближно една четвртина до една третина од блиските, подвижни состојби се внатрешно

тивки: нема план за кочење и нема план за свртување, иако пешакот е во рамки на 10 m.

Колабирањето на полето по должина на оската на брзина дава профил по растојание $M^{(k)}(d)$ на средно избегнување како функција од растојанието до пешак. *Растојанието на активирање намера* (англ. *Intent Onset Distance*, D_{50})

$$D_{50}^{(k)} = \max\{d : M^{(k)}(d) \geq 0.5\} \quad (5.34)$$

е најголемото растојание до пешак при кое агентот сè уште замислува избегнување почесто отколку не. Тоа е 30.7 m за Натрапникот и 31.2 m за Нерешителниот преминувач и Протестантот. Ова значи дека над приближно 31 m од пешакот доминантната замислена иднина е право низ конфликтот. Избегнувањето станува мнозинство само кога возилото веќе е во рамки на околу 31 m. *Зафатената површина на избеѓнување* (АОА, англ. *Avoidance Occurance Area*) $AOA_{0.5}^{(k)}$ — делот од хексагоните каде средното избегнување надминува 0.5 — е 0.490, 0.473 и 0.473 за Натрапник, Нерешителен преминувач и Протестант, соодветно: избегнувањето е мнозинско замислено однесување во помалку од половина од зафатените хексагони на просторот на состојби.

Аголот на режим на избеѓнување (анг. *Avoidance Mode Angle*)

$$\bar{\theta}^{(k)} = \frac{\sum_b \omega_b \operatorname{atan2}(Y_b, L_b)}{\sum_b \omega_b} \quad (5.35)$$

изразува дали избегнувањето е доминирано од давање првенство ($\bar{\theta}$ блиску до 90°), доминирано од странично избегнување (блиску до 0°), или балансирано (блиску до 45°). Вредностите се 74.1° , 75.8° и 70.2° за Натрапник, Нерешителен преминувач и Протестант: замислената стратегија е силно доминирана од давање првенство низ сите архетипови, нумерички потврдувајќи ја квалитативната слика од надворешната рамнина на избегнување. *Индексот на спрејнајност давање првенство-свртување* (YSCI, англ. *Yield-Steer Coupling Index*)

$$YSCI^{(k)} = \frac{\sum_b \omega_b L_b Y_b}{\sum_b \omega_b (L_b + Y_b - L_b Y_b)} \quad (5.36)$$

мери колку често агентот ги замислува двете намери истовремено. Вредностите се 0.206, 0.136 и 0.216: комбинираниите планови заземаат мал дел од замислените иднини, што значи дека намерата за давање првенство и страничната намера најчесто се замислуваат одвоено, а не како координиран маневар за заобиколување.

5.11 Дискусија на резултатите

Површно читање на статистиките за судири повикува директна интерпретација: агентот кочи предоцна и недоволно силно. Профилите на брзина по рута и сликите за анатомија на судири ја поддржуваат оваа интерпретација на прв поглед. Но доказите од имагинацијата ја прават неодржлива. Агентот може да произведе максимални забавувања од $7-8 \text{ m/s}^2$ и тоа го прави конзистентно штом ТТС падне под 1.5 s. Физичкиот авторитет за кочење е присутен. Она што недостасува е одлуката да се употреби доволно рано — а таа одлука се донесува, или не се донесува, во замислените иднини на претставувачкиот модел, не на ниво на моторна наредба. Доцниот надворешен одговор е последица од доцен внатрешен план. Третирањето на тоа како проблем на управување, а не како проблем на планирање, би водело кон погрешна поправка.

Доказите од имагинацијата откриваат посуптилен неуспех од слепост. Проекциите на агентот навистина ја достигнуваат судирната точка на пешакот пред најблиско приближување — времето на претходење на конфликтната свесност е 1.50–2.15 s зависно од архетипот —

па претставувачкиот модел не потфрла едноставно во претставувањето на конфликтот. Проблемот е што свесноста не се претвора во стабилен план. Јазот помеѓу првата проекција што го достигнува конфликтот и првата проекција со намера за избегнување е приближно еден метар преостанато растојание по рута. До моментот кога замислената иднина на агентот првпат содржи безбеден маневар, возилото веќе е доволно блиску што хоризонтот од 1.6 s едвај го покрива времето потребно за да го изврши. Свесноста и намерата речиси се совпаѓаат, што значи дека агентот нема внатрешен бафер помеѓу препознавањето на конфликтот и потребата да го разреши. Кај проактивен агент, тој бафер треба да се мери во секунди и десетици метри, а не во делови од секунда и единечни метри.

Резултатот за време на престој го заострува ова уште повеќе. Средните времиња на престој во иста класа од 0.13–0.17 s значат дека замислениот план не е само доцен — тој е и непосветен. Претставувачкиот модел изгледа како да осцилира помеѓу замислена иднина во која пешакот го расчистува патот и иднина во која не го расчистува, а актерот ја наследува оваа осцилација како треперлива намера. Квалитативно читање на декодираните рамки на проекциите сугерира зошто: претставувачкиот модел често го реконструира пешакот како да ја напуштил патеката до моментот кога замислената рута пристигнува до судирната точка, конзистентно со априорно верување научено за време на тренинг врз неконфликтни пешачки сценарија. Ова би објаснило зошто имагинацијата право низ конфликтот опстојува дури и кога реалното набљудување јасно покажува пешак сè уште во лентата. Латентната динамика предвидува иднина без пешак не затоа што набљудувањето го вели тоа, туку затоа што наученото преодно априорно верување на претставувачкиот модел вели дека тоа е веројатно. Ако оваа интерпретација е точна, коренот на причината не е во актерот, туку во самиот претставувачки модел, конкретно во распределбеното несовпаѓање помеѓу тренингот и конфликтната евалуација.

Архетипот Протестант е аналитички важен бидејќи ја намалува главната збунувачка променлива во евалуацијата на безбедноста на пешаците: неизвесноста околу тоа дали пешакот ќе ја расчисти рутата. Пешакот не е неподвижна точка. Како што е дефинирано во Глава 4, дел од протестантите може бавно да се придвижуваат во мал локален регион за да го блокираат возилото кога возилото пристапува со мала брзина. Но неговото движење е ограничено околу судирната точка. Тоа не е преминувачка траекторија што нормално треба да ја напушти лентата. Неуспех на полисата против таква локално ограничена блокада е силен доказ дека проблемот е внатрешен за агентот, а не само производ на надворешна неизвесност во темпирањето на пешачко преминување. Стапката на судири од 39.7% ја покажува оваа поента директно. Агентот судира со локална блокада со висока стапка, а полето на имагинација не покажува квалитативно различен и посигурен избегнувачки одговор за таа блокада отколку за архетиповите што ја преминуваат рутата. Ова е токму она што би се очекувало ако неуспехот е лоциран во априорното верување на претставувачкиот модел за иднините на пешаците, а не во одговорот на полисата на дадена иднина: априорното верување не ја кодира доволно силно информацијата дека овој пешак останува во зоната на конфликт, бидејќи латентната динамика на состојбата генерализира преку обрасци на движење на пешаци, наместо да се услови на тековната набљудувана улога на пешакот како локална блокада.

Главата раздвојува два надворешно различни режими на неуспех што имаат ист внатрешен корен. Судир преку небезбедна упорност се случува кога замислената иднина не одржува избегнување доволно долго за реализираната траекторија да излезе од геометријата на судир. Дефанзивен застој се случува кога замислената иднина навистина одржува план за давање првенство, но тој план е ограничен на хоризонтот од 1.6 s и не може да го претстави подалечниот проблем дали чекањето на крај ќе ја расчисти пречката. Двата неуспеси се изрази на хоризонт на планирање што е премногу краток и нестабилен за да поддржи проактивно, посветено избегнување какво што бара безбедноста на пешаците. Режимот на судир го одразува тоа што хоризонтот е недоволен за да изгради безбедна траекторија. Режимот на застој го одразува тоа што хоризонтот е недоволен за да процени дали тековната безбедна состојба

може да се претвори во завршување на рутата. Полиса што замислува само 1.6 s напред не може да праша „дали овој пешак на крај ќе ја расчисти патеката?“ — може само да праша „дали мојата замислена рута во моментот е блокирана?“ — а кога одговорот е неизменично да и не, резултатот е осцилирачката намера видлива во временските слики.

Пошироката импликација е дека за агенти со поттикнувано учење со модел во безбедносно критични поставки, квалитетот на имагинацијата на претставувачкиот модел не е само прашање на перформанси, туку прашање на безбедност. Евалуација заснована на исходи може да забележи дека агентот не успева. Таа не може да каже дали неуспехот потекнува од изборот на акција на полисата или од предвидувањето на претставувачкиот модел за тоа што следува. Дијагностичкиот протокол воведен во оваа глава, таксономија на проекции, метрики за проактивност и поле на имагинација условено од состојба, дава начин да се направи таа разлика. Тој покажува дека за овој основен агент, самата имагинација е слабата алка. Подобрување на актерот без подобрување на способноста на претставувачкиот модел да одржува точни, свесни за пешак проекции би го оставило коренот на причината недопрен. Обратно, претставувачки модел што сигурно замислува пешак како останува во лентата, дури и кога неговото научено априорно верување вели поинаку — би го променил влезот на актерот и веројатно би можел да го промени неговиот излез без никаква модификација на самиот актер. Ова го преобликува проблемот на безбедносно подобрување од обликување на награда или регуларизација на акција кон верност на претставувачкиот модел под распределбено поместување.

Глава 6

Заклучок

Оваа теза истражуваше дали агент со поттикнувано учење со модел, трениран со минимална награда за возачка компетентност и изложен за време на тренинг само на неконфликтни пешаци, развива проактивно однесување кон пешаци како произлезено својство на планирање преку претставувачки модел. Истражувањето беше спроведено врз агент DreamerV3 трениран во CARLA и евалуиран врз *Библиотекаџија на конфликтни пешаци* — бенчмарк со конфликтни пешаци воведен како дел од оваа работа, составен од три архетипови (Натрапник, Нерешителен преминувач, Протестант) што испитуваат квалитативно различни безбедносни неуспеси. Евалуацијата отиде подалеку од статистиките за исходи и навлезе во сопствените замислени иднини на агентот, третирајќи ги проекциите од претставувачкиот модел како главна дијагностичка површина. Ова завршно поглавје сумира што постигна работата, се враќа на истражувачките хипотези наведени во Глава 1, и ги наведува ограничувањата на тековното истражување и насоките во кои треба да се прошири.

6.1 Сумирање на истражувањето

Тезата следеше со намерна структура од три слоја. Првиот слој го карактеризираше надворешното однесување на агентот низ обични и конфликтни услови. Вториот слој ги претвори проекциите од претставувачкиот модел во интерпретабилни ознаки на однесување преку детерминистичка таксономија и ги сведе секвенците од ознаки на метрики за проактивност: времиња на претходење на конфликтната свесност и на намерата за избегнување, стапка на небезбедна упорност, времиња на престој за стабилност на намерата и индикатор за дефанзивна блокада во чекање. Третиот слој ја проектираше замислената намера во реализираниот простор растојание-до-пешак/брзина-на-возилото и го сведе добиеното поле на скаларни бенчмарк метрики што мерат колку избегнување замислува агентот во опасни состојби, колку рано се појавува и каков облик има замислената стратегија.

Фаза 0 утврди дека тренираниот агент е компетентен обичен возач и на видената и на невидената карта, со високи стапки на завршување рута и без мерливо нарушување од пасивни пешаци на тротоарот. Ова го исклучува објаснувањето дека последователните неуспеси се општи возачки неуспеси маскирани како пешачки конфликти. Фаза 1 го изложи агентот на трите конфликтни архетипови и најде стапки на судир што значајно се разликуваат по архетип, при што Протестантот создава исходи споредливи со Натрапникот што брзо преминува. Клучно е тоа што агентот ја задржа физичката способност за управување низ целата евалуација: способноста за кочење беше присутна, но се применуваше предоцна и без странична маргина.

Доказите од имагинацијата го дадоа објаснувањето што надворешниот запис не може да го даде. Агентот го перципира конфликтот, замислените проекции го достигнуваат пешачкиот регион пред најблиското приближување, но свесноста не се претвора сигурно во посвете-

ност. Значаен дел од проекциите што го достигнуваат конфликтот остануваат право низ конфликтот, времињата на престој во иста класа се кратки, а значаен дел од блиските подвижни состојби создаваат замислени иднини без давање првенство и без свртување. Избегнувањето што се појавува е силно доминирано од давање првенство. Комбинираниите планови со кочење и странично избегнување се најреткиот замислен исход. Сличноста во стапките на судир меѓу локално ограничена пречка и брз подвижен пешак ја исклучува неизвесноста на пешачкото движење како доминантна причина. Неуспехот е внатрешен за агентот: недоволна *проактивна замислена намера*.

Посебен режим на неуспех се појави кај Протестантот: дефанзивен застој. Дел од епизодите со Протестант завршуваат во одржливи блокади во чекање, а мнозинството епизоди со истек на време содржат таква блокада. Агентот може да стане непосредно безбеден со запирање, но неговиот краток хоризонт на имагинација не го поддржува прашањето дали чекањето на крај ќе ја расчисти пречката. Судирот и застојот се надворешно различни исходи со ист внатрешен корен: хоризонт на планирање што е прекраток и премногу нестабилен за да поддржи проактивно, посветено избегнување.

6.2 Наоди во однос на истражувачките хипотези

Четири хипотези наведени во Глава 1 повторно се разгледуваат тука според собраните докази.

Резултатите покажуваат дека архитектурата навистина поседува способност за проактивно планирање преку претставувачки модел. Замислените иднини често го достигнуваат регионот на пешакот уште пред најблиското приближување, што укажува дека системот не реагира исклучиво рефлексно на непосреден судир. Сепак, ова однесување останува ограничено и нестабилно. Проактивното избегнување најчесто се сведува на кратки корекции или доцно давање првенство, а судирите со локално ограничени пречки покажуваат дека минималната награда и тренинг-распределбата без конфликтна изложеност не се доволни за сигурно произлегување на безбедно однесување. Наодите затоа укажуваат дека проактивната безбедност е архитектонски можна, но не се појавува автоматски без дополнителен тренинг-притисок или изложеност на конфликтни ситуации.

Доказите исто така не укажуваат дека главниот теснец се наоѓа во перцепцијата. Во повеќе сценарија пешакот е јасно присутен во тековното набљудување, но исчезнува од замислените иднини само една секунда подоцна. Таквиот образец е поконзистентен со предиктивен отколку со перцепциски неуспех. Ова сугерира дека ограничувањето не лежи првенствено во камерскиот влез, туку во способноста на претставувачкиот модел стабилно да ги задржи и развива критичните елементи од сцената низ времето.

Систематската конфликтна евалуација се покажа неопходна за интерпретација на однесувањето на агентот. Трите пешачки архетипови не произведуваат само различни стапки на судир, туку и различни механизми на неуспех. Протестантот доминантно создава дефанзивен застој, Натрапникот резултира со највисоки преостанати брзини при судир, додека Нерешителниот преминувач најчесто предизвикува опасна тишина во планирањето. Ако сите сценарија се сведат на единечна агрегирана вредност, овие структурно различни обрасци стануваат невидливи. Во таа смисла, библиотеката на конфликтни пешаци се потврдува како дијагностички инструмент бидејќи овозможува интерпретабилна врска помеѓу однесувањето и конкретниот тип на внатрешен неуспех.

Резултатите дополнително покажуваат дека имагинацијата може да се користи како мерлив прозорец кон процесот на планирање. Таксономијата на проекции овозможува секој реален чекор на одлучување да се означи со интерпретабилна категорија, а метриците за проактивност ги сведуваат тие секвенци на скалари што ја раздвојуваат свесноста од посветеноста и судирот од застојот. Полето условено од состојба дополнително открива во кои физички

состојби агентот располага со стабилни и безбедни внатрешни планови, а во кои состојби планирањето се распаѓа или останува недоволно определено.

6.3 Придонеси

Придонесите на тезата опфаќаат инженерски, методолошки и емпириски аспекти поврзани со евалуација на MBRL агенти во конфликтни пешачки ситуации.

Истражувањето воведува библиотека на конфликтни пешаци имплементирана како сценаријски слој во време на извршување во CARLA. Трите архетипови се дефинирани преку набљудливи параметри на движење врзани за фиксна точка на рутата, што овозможува нивното однесување да остане семантички конзистентно и кога се менува тежината на сценариото. Библиотеката е целосно параметризирана и репродуцибилна, а имплементацијата е изведена преку нисконивовото API за пешаци на CARLA наместо преку вградената AI логика. Со тоа, темпирањето на конфликтот и врзувањето на пешакот за рутата можат директно да се управуваат во секој чекор од симулацијата.

Тезата дополнително предлага методологија за интроспекција на претставувачки модел кај Dreamer архитектурите. При секој реален чекор на одлучување се зачувува хоризонт на имажинација од $H = 32$, односно 1.6 s, по што состојбата на возилото и акциските големини се реконструираат и се сведуваат на детерминистичка таксономија од шест класи. Од добиените секвенци на ознаки се изведуваат метрики за проактивност на ниво на епизода, а индикаторите за намера потоа се проектираат во реализираниот простор растојание-до-пешак и брзина-на-возилото. Процесот е документиран од почеток до крај и не е врзан за конкретна агентска архитектура, туку за самиот интерфејс на претставувачкиот модел.

Глава 5 обезбедува детална повеќеслојна карактеризација на однесувањето на тренираниот агент низ сите конфликтни архетипови. Анализата ги опфаќа статистиките за исходи, профилите на движење по рута, густините акција-ТТС на ниво на чекор, анатомијата на судирите, геометријата на избегнување и анализите од страната на имажинацијата. Оваа карактеризација претставува централната емпириска основа на тезата и референтна точка за споредба со идни подобрувања на архитектурата, наградата или тренинг-распределбата.

Конечно, за DreamerV3 синтезата на резултатите дава конкретна дијагноза под минимална награда за возачка компетентност и без конфликтна изложеност на пешаци за време на тренинг. Агентот развива стабилен следач на рута во обични услови, но безбедноста кон пешаци останува ограничена првенствено во слојот на имажинација на претставувачкиот модел. Доказите не укажуваат дека главниот проблем се наоѓа во перцепцијата, управувачкиот авторитет или самиот дизајн на наградата. Наместо тоа, доминира нестабилна и доцно пристигната намера за избегнување во замислените иднини, често сведена на пасивно давање првенство и веројатно засилена од научено априорно верување дека пешаците ќе ја расчистат рутата. Дополнителен механизам на неуспех е дефанзивниот застој што се појавува кога хоризонтот на имажинација е пократок од времето потребно за разрешување на локално ограничена пречка.

6.4 Ограничувања

Тврдењата и заклучоците кои ги изложивме во оваа теза се поставени под неколку ограничувања кои се наведени во продолжение.

Анализите се засноваат на едно тренирање на DreamerV3 со едно случајно семе и еден сет хиперпараметри. Иако дијагностичкиот протокол е проектиран да биде независен од агентот, конкретниот образец на теснец — нестабилна имажинација доминирана од давање првенство со научено априорно верување дека „пешакот расчистува“ — е наод за една конкретна

тренирана полиса и нејзиниот претставувачки модел. Репликација низ семиња, хиперпараметарски режими и варијанти на Dreamer е неопходна пред конкретните квантитативни метрики да се третираат како својства на MBRL воопшто, а не на оваа реализација.

Сите експерименти се спроведени во CARLA, со тренирање на Town02 и тестирање на генерализација на Town01. Прикажувањето на пешаци, физиката и перцепциските канали во CARLA се доволни за релативните споредби пријавени тука, но тие не се реални податоци. Валидноста на симулацијата во однос на реалниот свет на кој било конкретен квантитативен наод, на пример растојанието на активирање намера од 30.7–31.2 m, е непозната и не треба да се претпоставува.

Хипотезата дека претставувачкиот модел го замислува пешакот како ја расчистува патеката пред проекцијата да стигне до судирната точка е поддржана со квалитативен преглед на декодирани рамки од сегментација и длабочина, а не со калибриран декодер за пешачко верување. Квантитативна валидација на оваа хипотеза би барала посебна студија што ќе изгради детекциски процес за пешаци врз декодирани проекции и ќе ја спореди замислената пешачка упорност со вистинската состојба низ евалуацискиот сет.

Секое сценарио од Библиотеката на конфликтни пешаци содржи еден конфликтен пешак. Групни конфигурации се поддржани од библиотеката, но не беа користени. Конфликти со повеќе пешаци, каде два или повеќе конфликтни актери истовремено интерагираат со возилото, се надвор од тековната евалуација.

Минималниот дизајн на наградата е намерен истражувачки избор што го прави проактивното однесување прашање на произлегување, а не прашање на обликување. Затоа тековната работа не кажува како безбедносно обликувана награда би го променила слојот на имагинација, што е ограничување на заклучоците: можно е експлицитни наградни сигнали да ја променат не само полисата на актерот туку и претставата на претставувачкиот модел за иднините на пешаците.

Најзначајното ограничување е отсуството на студија што директно би го менувала идентификуваниот теснец. Тезата го идентификува теснецот и предлага механизам. Сè уште не го тестира тој механизам со манипулирање на системот и набљудување дали предвиденото подобрување се појавува.

6.5 Идна работа

Заклучоците од оваа теза природно наложуваат неколку можни правци на идна работа кои би донеле системски подобрувања. Дијагнозата од Глава 5 сугерира четири насоки, секоја насочена кон конкретен слој од синџирот перцепција–имагинација–акција.

Ако доминантниот неуспех е научено априорно верување дека пешаците ја расчистуваат рутата, природната насока е целна изложеност на претставувачкиот модел на пешачки конфигурации каде ова априорно верување е погрешно. Библиотеката на конфликтни пешаци може да се премести од само-евалуациска алатка во компонента на курикулум за време на тренинг на претставувачкиот модел, со задржана разделба меѓу тренинг-распределбата на претставувачкиот модел и наградниот сигнал на актерот. Дијагностичкиот протокол воведен во оваа теза потоа може повторно да се примени за да се тестира дали вредностите DSR паѓаат, времињата на претходење растат и времињата на престој се продолжуваат во замислените иднини на повторно тренираниот агент.

Квалитативното набљудување дека претставувачкиот модел замислува пешак што ја расчистува рутата заслужува квантитативен тест. Декодер што ги пресликува латентните состојби на проекцијата во предвидувања за присуство на пешак, валидиран според вистинските позиции на CARLA актерите, би го претворил сегашното квалитативно тврдење во калибрирана метрика. Ова исто така би овозможило преглед на хипотезата за замислено расчистување по проекција и по епизода.

Иако оваа теза намерно ја задржа наградата минимална, контролирана споредба со награда проширена со ТТС или со близина е најчистиот начин да се локализира придонесот на претставувачкиот модел во однос на наградата. Ако агент со обликувана награда создаде значајно подобра надворешна безбедност, но исти метрики на имагинација, теснецот е во одговорот на актерот на имагинацијата. Ако се подобрат и метриците на имагинација, тогаш наградата ја обликува и тренинг-распределбата на претставувачкиот модел, што само по себе би било наод.

Долгорочна цел на оваа истражувачка линија е пренос од симулација во реален свет, но поизводлив посреден чекор е примената на дијагностичкиот протокол врз снимени податоци од возење со анотирани пешачки конфликти, реконструирајќи замислени иднини од претставувачки модел трениран на реални податоци. Ова би овозможило метриците за проактивност да се пресметаат врз реални конфликти и би тестирало дали обрасците забележани во CARLA опстојуваат во побучната распределба на реалниот свет.

Групните конфигурации веќе се поддржани од Библиотеката на конфликтни пешаци и претставуваат природен следен степен на тежина. Тие се особено релевантни за случајот Протестант, каде групна блокада целосно ја отстранува опцијата за странично заобиколување и ја присилува полисата на одлука за давање првенство.

6.6 Завршни забелешки

Главната порака на оваа теза не е само дека агентот може или не може безбедно да се справува со пешаци, туку дека самото набљудување на исходите не е доволно за да се разбере од каде потекнува неуспехот. Судирот е краен резултат — тој не открива дали проблемот започнал во перцепцијата, во планирањето или во ограничувањата на самата имагинација на претставувачкиот модел. Кај агенти со поттикнувано учење со модел, токму таа внатрешна имагинација го обликува однесувањето. Поради тоа, разликата меѓу безбедно и небезбедно однесување често станува видлива во замислените иднини уште пред да се појави во реализираната траекторија.

Дијагностичкиот протокол предложен во оваа теза претставува обид таа внатрешност да стане интерпретабилна. Таксономијата на проекции ги претвора латентните претстави во ознаки што можат систематски да се анализираат, метриците за проактивност овозможуваат да се разликува свесност од вистинска посветеност на избегнување, а полето условено од состојба покажува во кои физички состојби агентот располага со стабилни и безбедни внатрешни планови, а во кои планирањето се распаѓа или останува неодлучно.

Кога протоколот се применува врз основниот агент DreamerV3 во Библиотеката на конфликтни пешаци, се појавува релативно јасна слика. Агентот е компетентен во обично следење на рута, но неговата безбедност кон пешаци останува ограничена од самиот слој на имагинација на претставувачкиот модел. Замислените иднини често доцнат со избегнувањето, се потпираат на претпоставка дека пешакот ќе ја расчисти рутата или немаат доволно долг хоризонт за да разрешат упорна локална пречка. Токму затоа идните подобрувања најверојатно треба да се насочат кон квалитетот и стабилноста на внатрешното планирање, а не само кон перцепцијата или контролата.

Насоките предложени во Секција 6.5 природно произлегуваат од оваа дијагноза. Придонесот на тековната работа е пред сè во воспоставувањето на инфраструктура што овозможува вакви прашања да се поставуваат и анализираат систематски. Следниот чекор е таа инфраструктура да се искористи за да се испита дали ограничувањата на имагинацијата можат да се надминат и дали проактивната безбедност може да стане стабилно својство на агентот, наместо повремено и нестабилно однесување.

Користена литература

- [1] Ahmed Abouelazm, Jonas Michel, and Johann Marius Zöllner. A review of reward functions for reinforcement learning in the context of autonomous driving. *2024 IEEE Intelligent Vehicles Symposium (IV)*, pages 156–163, 2024.
- [2] Rolla Almodfer, Shengwu Xiong, Xiangzhen Kong, and Pengfei Duan. Pedestrian crossing speed patterns and running frequency analysis at a non-signalized marked crosswalk: quantitative and qualitative approaches. *Sustainable cities and society*, 34:183–192, 2017.
- [3] Tirthankar Bandyopadhyay, Kok Sung Won, Emilio Frazzoli, David Hsu, Wee Sun Lee, and Daniela Rus. Intention-aware motion planning. In *Workshop on the Algorithmic Foundations of Robotics*, 2013.
- [4] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2):157–166, 1994. doi: 10.1109/72.279181.
- [5] Yoshua Bengio, Nicholas Léonard, and Aaron Courville. Estimating or propagating gradients through stochastic neurons for conditional computation. *arXiv preprint arXiv:1308.3432*, 2013.
- [6] Raunak P. Bhattacharyya, Derek J. Phillips, Changliu Liu, Jessica B. Hamrick, Mykel J. Kochenderfer, and Edward Schmerling. Modeling human driving behavior through generative adversarial imitation learning. *IEEE Transactions on Intelligent Transportation Systems*, 24(3):2874–2887, 2023.
- [7] Maxime Bouton, Alireza Nakhaei, Kikuo Fujimura, and Mykel J. Kochenderfer. Safe reinforcement learning with scene decomposition for navigating complex urban environments. *2019 IEEE Intelligent Vehicles Symposium (IV)*, pages 1469–1476, 2019.
- [8] Axel Brunnbauer, Luigi Berducci, Matus Kneissl, Radu Grosu, and Ichiro Hasuo. Scenario-based curriculum generation for multi-agent autonomous driving. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2025.
- [9] Fanta Camara, Nicola Bellotto, Serhan Coşar, Florian Weber, Dimitris Nathanael, Matthias Althoff, Jingyuan Wu, Johannes Ruenz, André Dietrich, Gustav Markkula, Anna Schieben, Fabio Tango, Natasha Merat, and Charles W. Fox. Pedestrian models for autonomous driving part ii: High-level models of human behavior. *IEEE Transactions on Intelligent Transportation Systems*, 22:5453–5472, 2020.
- [10] Dian Chen, Brady Zhou, Vladlen Koltun, and Philipp Krähenbühl. Learning by cheating. In *Proceedings of the Conference on Robot Learning*, volume 100 of *Proceedings of Machine Learning Research*, pages 66–75. PMLR, 2020.
- [11] Li Chen, Peng Wu, Kashyap Chitta, Bernhard Jaeger, Andreas Geiger, and Hongyang Li. End-to-end autonomous driving: Challenges and frontiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46:10164–10183, 2023.
- [12] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin P. Murphy, and Alan Loddon Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40:834–848, 2016.
- [13] Ziyu Chen, Xiufeng Chen, Ruicong Wang, and Mengyuan Gao. Characterization of pedestrian crossing spatial violations and safety impact analysis in advance right-turn lane. *International Journal of Environmental Research and Public Health*, 19, 2022.

- [14] Pranav Singh Chib and Pravendra Singh. Recent advancements in end-to-end autonomous driving using deep learning: A survey. *IEEE Transactions on Intelligent Vehicles*, 9:103–118, 2023.
- [15] Kashyap Chitta, Aditya Prakash, Bernhard Jaeger, Zehao Yu, Katrin Renz, and Andreas Geiger. TransFuser: Imitation with transformer-based sensor fusion for autonomous driving. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(11):12878–12895, 2023.
- [16] Kyunghyun Cho, Bart van Merriënboer, Çağlar Gülçehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, 2014. doi: 10.3115/v1/D14-1179.
- [17] Amit Chougule, Vinay Chamola, Aishwarya Sam, Fei Richard Yu, and Biplab Sikdar. A comprehensive review on limitations of autonomous driving and its impact on accidents and collisions. *IEEE Open Journal of Vehicular Technology*, 5:142–161, 2024.
- [18] Elahe Delavari, Feeza Khan Khanzada, and Jaerock Kwon. A comprehensive review of reinforcement learning for autonomous driving in the carla simulator. *ArXiv*, abs/2509.08221, 2025.
- [19] Debargha Dey and Jacques M. B. Terken. Pedestrian interaction with vehicles: Roles of explicit and implicit communication. *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2017.
- [20] Christopher Diehl, Timo Sievernich, Martin Krüger, Frank Hoffmann, and Torsten Bertram. Uncertainty-aware model-based offline reinforcement learning for automated driving. *IEEE Robotics and Automation Letters*, 8:1167–1174, 2023.
- [21] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In *Proceedings of the 1st Annual Conference on Robot Learning*, pages 1–16, 2017.
- [22] Badr Ben Elallid, Nabil Benamar, Nabil Mrani, and Tajje eddine Rachidi. Dqn-based reinforcement learning for vehicle control of autonomous vehicles interacting with pedestrians. *2022 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, pages 489–493, 2022.
- [23] Tuo Feng, Wenguan Wang, and Yi Yang. A survey of world models for autonomous driving. *ArXiv*, abs/2501.11260, 2025.
- [24] Luiz G. Galvão and M. Nazmul Huda. Pedestrian and vehicle behaviour prediction in autonomous vehicle system - a review. *Expert Syst. Appl.*, 238:121983, 2023.
- [25] Luiz G. Galvao, Maysam F. Abbod, Tatiana Kalganova, Vasile Palade, and Md. Nazmul Huda. Pedestrian and vehicle detection in autonomous vehicle perception systems—a review. *Sensors (Basel, Switzerland)*, 21, 2021.
- [26] Dechen Gao, Shuangyu Cai, Hanchu Zhou, Hang Wang, Iman Soltani, and Junshan Zhang. Cardreamer: Open-source learning platform for world-model-based autonomous driving. *IEEE Internet of Things Journal*, 12:2866–2875, 2024.
- [27] Mingzhe Guo, Zhipeng Zhang, Yuan He, Ke Wang, and Liping Jing. End-to-end autonomous driving without costly modularization and 3d manual annotation. *ArXiv*, abs/2406.17680, 2024.
- [28] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, pages 1861–1870, 2018.
- [29] Azra Habibovic, Victor Malmsten Lundgren, Jonas Andersson, Maria Klingegård, Tobias Lagström, Anna Sirkka, Johan Fagerlönn, Claes Edgren, Rikard Fredriksson, Stas Simon Krupenia, Dennis Saluäär, and Pontus Larsson. Communicating intent of automated vehicles to pedestrians. *Frontiers in Psychology*, 9, 2018.

- [30] Farrukh Hafeez, Usman Ullah Sheikh, Abdullahi Abubakar Mas'ud, Saud Al-Shammari, Muhammad Hamid, and Ameer Azhar. Application of the theory of planned behavior in autonomous vehicle-pedestrian interaction. *Applied Sciences*, 2022.
- [31] Danijar Hafner, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. Dream to control: Learning behaviors by latent imagination. *arXiv preprint arXiv:1912.01603*, 2019.
- [32] Danijar Hafner, Timothy Lillicrap, Mohammad Norouzi, and Jimmy Ba. Mastering Atari with discrete world models. In *International Conference on Learning Representations (ICLR)*, 2021.
- [33] Danijar Hafner, J. Pašukonis, Jimmy Ba, and Timothy P. Lillicrap. Mastering diverse domains through world models. *ArXiv*, abs/2301.04104, 2023.
- [34] Ya-Chuan Hsu, Swaminathan Gopalswamy, Srikanth Saripalli, and Dylan A. Shell. A pomdp treatment of vehicle-pedestrian interaction: Implicit coordination via uncertainty-aware planning. *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1984–1991, 2020.
- [35] Anthony Hu, Gianluca Corrado, Nicolas Griffiths, Zak Murez, Corina Gurau, Hudson Yeo, Alex Kendall, Roberto Cipolla, and Jamie Shotton. Model-based imitation learning for urban driving. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.
- [36] J. Huang and E. A. Sowe. Semantic segmentation in self-driving cars using u-net architecture on cityscape dataset. *International Journal of Engineering Applied Sciences and Technology*, 9:24–34, 2025. doi: 10.33564/ijeast.2025.v09i09.004.
- [37] Joel Janai, Fatma Güney, Aseem Behl, and Andreas Geiger. Computer vision for autonomous vehicles: Problems, datasets and state-of-the-art. *Found. Trends Comput. Graph. Vis.*, 12:1–308, 2017.
- [38] B Raghuram Kadali and Perumal Vedagiri. Proactive pedestrian safety evaluation at unprotected mid-block crosswalk locations under mixed traffic conditions. *Safety science*, 89:94–105, 2016.
- [39] Nan Rosemary Ke, Amanpreet Singh, Ahmed Touati, Anirudh Goyal, Yoshua Bengio, Devi Parikh, and Dhruv Batra. Learning dynamics model in reinforcement learning by incorporating the long term future. *ArXiv*, abs/1903.01599, 2019.
- [40] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [41] Richard L Knoblauch, Martin T Pietrucha, and Marsha Nitzburg. Field studies of pedestrian walking speed and start-up time. *Transportation research record*, 1538(1):27–38, 1996.
- [42] Mark Koren, Saud Alsaif, Ritchie Lee, and Mykel J. Kochenderfer. Adaptive stress testing for autonomous vehicles. *2018 IEEE Intelligent Vehicles Symposium (IV)*, pages 1–7, 2018.
- [43] Amar Kulkarni, Shangdong Zhang, and Madhur Behl. Crash: Challenging reinforcement-learning based adversarial scenarios for safety hardening. *ArXiv*, abs/2411.16996, 2024.
- [44] Somnath Lahiri, Jing Ren, and Xianke Lin. Deep learning-based stereopsis and monocular depth estimation techniques: A review. *Vehicles*, 2024.
- [45] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi: 10.1109/5.726791.
- [46] Qifeng Li, Xiaosong Jia, Shaobo Wang, and Junchi Yan. Think2drive: Efficient reinforcement learning by thinking in latent world model for quasi-realistic autonomous driving (in carla-v2). *ArXiv*, abs/2402.16720, 2024.
- [47] Quan Li, Shi Shang, Xizhe Pei, Qingfan Wang, Qing Zhou, and Bingbing Nie. Kinetic and kinematic features of pedestrian avoidance behavior in motor vehicle conflicts. *Frontiers in bioengineering and biotechnology*, 9:783003, 2021.

- [48] Yueyuan Li, Wei Yuan, Songan Zhang, Weihao Yan, Qiyuan Shen, Chunxiang Wang, and Ming Yang. Choose your simulator wisely: A review on open-source simulators for autonomous driving. *IEEE Transactions on Intelligent Vehicles*, 9:4861–4876, 2023.
- [49] Zhuoren Li, Guizhe Jin, Ran Yu, Zhiwen Chen, Nan Li, Wei Han, Lu Xiong, Bo Leng, Jia Hu, Ilya V. Kolmanovsky, and Dimitar Filev. A survey of reinforcement learning-based motion planning for autonomous driving: Lessons learned from a driving task perspective. *ArXiv*, abs/2503.23650, 2025.
- [50] Jingda Liu, Ruoxuan Xiong, Xinbo Wu, Ilhan Bae, and Ching-Yao Chan. Attention-based distributional reinforcement learning for safe and efficient autonomous driving. *IEEE Robotics and Automation Letters*, 9, 2024.
- [51] Fan Luo, Tian Xu, Hang Lai, Xiong-Hui Chen, Weinan Zhang, and Yang Yu. A survey on model-based reinforcement learning. *Science China Information Sciences*, 67, 2022.
- [52] Nadya Abdel Madjid, Abdulrahman Ahmad, Murad Mebrahtu, Yousef Babaa, Abdelmoamen Nasser, Sumbal Malik, Bilal Hassan, Naoufel Werghi, Jorge Dias, and Majid Khonji. Trajectory prediction for autonomous driving: Progress, limitations, and future directions. *ArXiv*, abs/2503.03262, 2025.
- [53] Adam Millard-Ball. Pedestrians, autonomous vehicles, and cities. *Journal of planning education and research*, 38(1):6–12, 2018.
- [54] Alireza Mirzabagheri, Majid Ahmadi, Ning Zhang, Reza Alirezaee, Saeed Mozaffari, and Shahpour Alirezaee. Navigating uncertainty: Advanced techniques in pedestrian intention prediction for autonomous vehicles—a comprehensive review. *Vehicles*, 2025.
- [55] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharmashan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015. doi: 10.1038/nature14236.
- [56] Amy L O’Dell, Andrew Morris, Ashleigh J. Filtness, and Jo Barnes. The impact of pedestrian distraction on safety behaviours at controlled and uncontrolled crossings. *SSRN Electronic Journal*, 2023.
- [57] Matthew O’Kelly, Aman Sinha, Hongseok Namkoong, John C. Duchi, and Russ Tedrake. Scalable end-to-end autonomous vehicle testing via rare-event simulation. *ArXiv*, abs/1811.00145, 2018.
- [58] David Osorio-García, Gregorio Hernández-Pulgarín, and Diego Alexander Escobar. Profiles of pedestrian risk behavior while crossing the street. *SSRN Electronic Journal*, 2022.
- [59] Jennifer A Oxley, Elfriede Ihlen, Brian N Fildes, Judith L Charlton, and Ross H Day. Crossing roads safely: an experimental study of age differences in gap selection by pedestrians. *Accident Analysis & Prevention*, 37(5):962–971, 2005.
- [60] Youssef Al Ozaibi, Manolo Dulva Hina, and Amar Ramdane-Cherif. End-to-end autonomous driving in carla: A survey. *IEEE Access*, 12:146866–146900, 2024.
- [61] Athanasios S. Polydoros and Lazaros Nalpantidis. Survey of model-based reinforcement learning: Applications on robotics. *Journal of Intelligent & Robotic Systems*, 86:153–173, 2017.
- [62] Maria Priisalu, Aleksis Pirinen, C. Paduraru, and Cristian Sminchisescu. Generating scenarios with diverse pedestrian behaviors for autonomous vehicle testing. In *Conference on Robot Learning*, 2021.
- [63] Shreyas Ramakrishna, Baiting Luo, Christopher B. Kuhn, Gabor Karsai, and Abhishek Dubey. Anti-carla: An adversarial testing framework for autonomous vehicles in carla. *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, pages 2620–2627, 2022.
- [64] Mohan Ramesh, Mark Azer, and Fabian B. Flohr. Habit: Human action benchmark for interactive traffic in carla, 2025.

- [65] Amir Rasouli and John K. Tsotsos. Autonomous vehicles that interact with pedestrians: A survey of theory and practice. *IEEE Transactions on Intelligent Transportation Systems*, 21:900–918, 2018.
- [66] Amir Rasouli and John K. Tsotsos. Joint attention in driver-pedestrian interaction: from theory to practice. *ArXiv*, abs/1802.02522, 2018.
- [67] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39:1137–1149, 2015.
- [68] Saki Rezwana and Nicholas Lownes. Interactions and behaviors of pedestrians with autonomous vehicles: A synthesis. *Future Transportation*, 2024.
- [69] Stefan Riedmaier, Thomas Ponn, Dieter Ludwig, Bernhard Schick, and Frank Diermeyer. Survey on scenario-based safety assessment of automated vehicles. *IEEE Access*, 8:87456–87477, 2020.
- [70] Giulia Rizzoli, Francesco Barbato, and Pietro Zanuttigh. Multimodal semantic segmentation in autonomous driving: A review of current approaches and future perspectives. *Technologies*, 2022.
- [71] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241. Springer, 2015. doi: 10.1007/978-3-319-24574-4_28.
- [72] Luis Alberto Rosero, Iago Pachêco Gomes, Júnior Anderson Rodrigues da Silva, Carlos Andre Braile Przewodowski Filho, Denis Fernando Wolf, and Fernando Santos Osório. Integrating modular pipelines with end-to-end learning: A hybrid approach for robust and reliable autonomous driving systems. *Sensors (Basel, Switzerland)*, 24, 2024.
- [73] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986. doi: 10.1038/323533a0.
- [74] Oleh Rybkin, Chuning Zhu, Anusha Nagabandi, Kostas Daniilidis, Igor Mordatch, and Sergey Levine. Model-based reinforcement learning via latent-space collocation. *ArXiv*, abs/2106.13229, 2021.
- [75] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.
- [76] Sergio Martín Serrano, David Fernández Llorca, Iván García Daza, and Miguel Ángel Sotelo. Realistic pedestrian behaviour in the carla simulator using vr and mocap. In *International Conference on Computer-Human Interaction Research and Applications*, 2023.
- [77] Neha Sharma, Chhavi Dhiman, and S. Indu. Pedestrian intention prediction for autonomous vehicles: A comprehensive survey. *Neurocomputing*, 508:120–152, 2022.
- [78] Zilin Sheng, Zihao Liu, Yansong Sheng, Junwei Hong, Changhao Dong, Yue Cai, and Sikai Wang. VLM-RL: A unified vision language models and reinforcement learning framework for safe autonomous driving. *Transportation Research Part C: Emerging Technologies*, 2025.
- [79] Tianyu Shi, Minghui Gu, Kehui Bi, and Pinlong Cai. Offline reinforcement learning for autonomous driving with safety and exploration enhancement. *arXiv preprint arXiv:2110.07067*, 2021.
- [80] Ze Sui, Yue Zhou, Xu Zhao, Ao Chen, and Yi Ling Ni. Joint intention and trajectory prediction based on transformer. *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7082–7088, 2021.
- [81] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2 edition, 2018.
- [82] Akshaj Tammewar, Nikita Chaudhari, Bunny Saini, Divya Venkatesh, Ganpathiraju Dharahas, Deepali Rahul Vora, Shrutika A. Patil, Ketan V. Kotecha, and Sultan Alfarhood. Improving the performance of autonomous driving through deep reinforcement learning. *Sustainability*, 2023.

- [83] Juan R. Terven, Diana-Margarita Córdova-Esparza, and Julio-Alejandro Romero-González. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Mach. Learn. Knowl. Extr.*, 5:1680–1716, 2023.
- [84] Marin Toromanoff, Emilie Wirbel, and Fabien Moutarde. End-to-end model-free reinforcement learning for urban driving using implicit affordances. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7153–7162, June 2020.
- [85] Mark Towers, Jordan K. Terry, Ariel Kwiatkowski, John U. Balis, Gianluca de Cola, Tristan Deleu, Manuel Goul ao, Andreas Kallinteris, Arjun KG, Markus Krimmel, Rodrigo Perez-Vicente, Andrea Pierré, Sander Schulhoff, Jun Jet Tai, Andrew Tan Jin Shen, and Omar G. Younis. Gymnasium, 2023.
- [86] Bhargava Tejus Uppuluri, Tomislav Mihalj, and Alois Knoll. CuRLA: Curriculum learning based deep reinforcement learning for autonomous driving. *arXiv preprint arXiv:2501.04982*, 2025.
- [87] Timothé Verstraete and Naveed Muhammad. Pedestrian collision avoidance in autonomous vehicles: A review. *Comput.*, 13:78, 2024.
- [88] Guan Wang, Haoyi Niu, Desheng Zhu, Jianming Hu, Xianyuan Zhan, and Guyue Zhou. Model: A modularized end-to-end reinforcement learning framework for autonomous driving. *ArXiv*, abs/2110.11573, 2021.
- [89] Hong Wang, Wenbo Shao, Chen Sun, Kai Yang, Dongpu Cao, and Jun Li. A survey on an emerging safety challenge for autonomous vehicles: Safety of the intended functionality. *Engineering*, 2024.
- [90] Christopher J. C. H. Watkins and Peter Dayan. Q-learning. *Machine Learning*, 8(3–4):279–292, 1992. doi: 10.1007/BF00992698.
- [91] Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8(3–4):229–256, 1992. doi: 10.1007/BF00992696.
- [92] Penghao Wu, Xiaosong Jia, Li Chen, Junchi Yan, Hongyang Li, and Yu Qiao. Trajectory-guided control prediction for end-to-end autonomous driving: A simple yet strong baseline. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.
- [93] Taokai Xia and Hui Chen. A survey of autonomous vehicle behaviors: Trajectory planning algorithms, sensed collision risks, and user expectations. *Sensors (Basel, Switzerland)*, 24, 2024.
- [94] Jie Yang, Nirajan Gauli, Nirajan Shiwakoti, Richard Tay, Hepu Deng, Jian Saturn Chen, Bharatpur Nepal, and Jimmy Li. Examining the factors influencing pedestrian behaviour and safety: A review with a focus on culturally and linguistically diverse communities. *Sustainability*, 2025.
- [95] Zhenjie Yang, Xiaosong Jia, Qifeng Li, Xue Yang, Maoqing Yao, and Junchi Yan. Raw2drive: Reinforcement learning with aligned world models for end-to-end autonomous driving (in carla v2). *ArXiv*, 2025.
- [96] Lisha Yuan. End-to-end system architectures in autonomous driving: Comparative analysis against modular design and technological exploration. *Applied and Computational Engineering*, 2024.
- [97] Dmytro Zabolotnii, Yar Muhammad, and Naveed Muhammad. Pedestrian motion prediction evaluation for urban autonomous driving. *ArXiv*, abs/2410.16864, 2024.
- [98] Zhejun Zhang, Alexander Liniger, Dengxin Dai, Fisher Yu, and Luc Van Gool. End-to-end urban driving by imitating a reinforcement learning coach. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.
- [99] Zhengming Zhang, Md Fazle Elahi, Joshua E. Domeyer, and Renran Tian. Driver temporal segmentation of pedestrian crossing intentions during negotiations. *Transportation Research Part F: Traffic Psychology and Behaviour*, 2025.

- [100] Ziyuan Zhong, Yun Tang, Yuan Zhou, Vânia de Oliveira Neves, Yang Liu, and Baishakhi Ray. A survey on scenario-based testing for automated driving systems in high-fidelity simulation. *ArXiv*, abs/2112.00964, 2021.
- [101] Zeyu Zhu and Huijing Zhao. A survey of deep rl and il for autonomous driving policy learning. *IEEE Transactions on Intelligent Transportation Systems*, 23:14043–14065, 2021.
- [102] Xinrui Zou. A review of object detection techniques. *2019 International Conference on Smart Grid and Electrical Automation (ICSGEA)*, pages 251–254, 2019.