ХЕРСОНСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ТА ДИЗАЙНУ
КАФЕДРА ПРОГРАМНИХ ЗАСОБІВ І ТЕХНОЛОГІЙ

Пояснювальна записка

до кваліфікаційної роботи магістра

на тему:

«ДОСЛІДЖЕННЯ ГЛИБОКОГО НАВЧАННЯ З ПІДКРІПЛЕННЯМ У СЕРЕДОВИЩІ VIZDOOM»

Виконав: здобувач 6 курсу, групи 6ПР1
спеціальності 121 «Інженерія
програмного забезпечення»
Вдовиченко В.С.
(прізвище та ініціали)

Керівник: Доровський В.О.
(прізвище та ініціали)

Рецензент_____
(прізвище та ініціали)

Хмельницький – 2025 р.

Факультет **Інформаційних технологій та дизайну**
Кафедра **Програмних засобів і технологій**
Освітньо-кваліфікаційний рівень **бакалавр**
Галузі знань **12 «Інформаційні технології»**
Спеціальність **121 «Інженерія програмного забезпечення»**
Освітньо-професійної програми **«Програмна інженерія»**

ЗАТВЕРДЖУЮ
Завідувач кафедри ПЗіТ,
професор
_____О.Є.Огнєва
«____» _____ 20__ року

З А В Д А Н Н Я
НА ДИПЛОМНУ РОБОТУ ЗДОБУВАЧУ
_____Вдовиченко Валерія Сергіївна_____
(прізвище, ім'я, по батькові)

1. Тема роботи «Дослідження глибокого навчання з підкріпленням у середовищі VizDoom»_____
_____
_____
керівник роботи д.т.н., професор Доровський В.О_____,
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)
затверджені наказом ХНТУ від «___» _____ 20__року №___.
2. Строк подання здобувачем роботи _____
3. Вихідні дані до роботи:
– Середовище VizDoom та його стандартні сценарії.
– Базова архітектура DQN.
– Підходи до покращення DQN.
– Засоби збору статистики та побудови кривих навчання.
4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)
– Обґрунтування мети й завдань дослідження.
– Огляд сучасних методів RL та архітектури DQN.
– Аналіз покращень DQN і їх очікуваного впливу.
– Опис середовища VizDoom та вибір сценаріїв.
– Методика реалізації базового та покращених агентів.
– Опис експериментального плану й метрик оцінювання.
– Представлення кривих навчання та порівняльний аналіз результатів.
– Висновки та пропозиції щодо подальшого розвитку.
5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)

– Схема архітектури базового DQN-агента.
– Блок-схема процесу навчання агента в середовищі VizDoom. (обов’язково)
– Криві навчання (середня винагорода, швидкість збіжності) для всіх агентів. (обов’язково)

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1.	д.т.н., професор Доровський В.О	10.10.25	20.10.25
2.	д.т.н., професор Доровський В.О	21.10.25	29.11.25
3.	д.т.н., професор Доровський В.О	3.12.25	7.12.25
4.	д.т.н., професор Доровський В.О	7.12.25	14.12.25

7. Дата видачі завдання_____28.09.25_____

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного роботи	Строк виконання етапів роботи	Примітка
1.	Аналіз предметної області та джерел літератури.	10.10.25	виконано
2.	Вибір і налаштування середовища VizDoom.	26.10.25	виконано
3.	Реалізація базового DQN-агента.	16.11.25	виконано
4.	Реалізація покращених варіантів DQN.	19.11.25	виконано
5.	Планування експериментів та вибір метрик.	1.12.25	виконано
6.	Проведення тренувань і збір статистики.	5.12.25	виконано
7.	Аналіз результатів і побудова графічних матеріалів.	10.12.25	виконано
8.	Формування висновків і написання записки.	14.12.25	виконано

здобувач _____ _Вдовичено В.С._____
(підпис) (прізвище та ініціали)

Керівник роботи _____ ___Доровський В.О.____
(підпис) (прізвище та ініціали)

АНОТАЦІЯ

У магістерській роботі досліджено методи глибокого навчання з під-кріпленням у контексті середовища VizDoom, що є тривимірною платфор-мою для моделювання агентів у шутері від першої особи. Основну увагузосереджено на аналізі, розробці та експериментальному оцінюванні аген-тів, побудованих на базі алгоритму Deep Q-Network (DQN) та його по-кращених модифікацій. Розглянуто ключові компоненти DQN: нейроннуапроксимацію Q-функції, використання цільової мережі, буфера досвідута механізмів стабілізації навчання.

У роботі детально вивчено вплив стекування кадрів, n-крокових по-вернень та пріоритетного повторного відтворення досвіду на швидкістьзбіжності та кінцеву ефективність агентів. Проведено систематичний ана-ліз простору станів і дій, параметрів середовища VizDoom, а також чинни-ків, що визначають складність навчання в умовах часткової спостережу-ваності та високої динаміки. Запропоновано експериментальну методику,що включає порівняння базового DQN з його модифікаціями за метрика-ми середньої та мінімальної винагороди, стабільності та варіативності.

Отримані результати демонструють суттєве покращення ефективно-сті агентів при використанні комбінації вдосконалень, зокрема n-кроковихповернень і пріоритетного досвіду, що забезпечують швидшу адаптаціюдо складних сценаріїв VizDoom. Робота робить внесок у практичне засто-сування глибокого навчання з підкріпленням у 3D-середовищах та можебути використана як основа для подальших досліджень алгоритмів на-вчання з підкріпленням у високодинамічних задачах.

ABSTRACT

This master’s thesis investigates deep reinforcement learning methods within the VizDoom environment, a three-dimensional first-person shooter platform designed for training and evaluating autonomous agents. The study focuses on the analysis, development, and empirical evaluation of agents based on the Deep Q-Network (DQN) algorithm and its improved variants. The core components of DQN—neural approximation of the Q-function, target network stabilization, experience replay, and learning regularization techniques—are examined in depth.

The research thoroughly explores the effects of frame stacking, n-step returns, and prioritized experience replay on both the convergence speed and asymptotic performance of agents. Special attention is given to analyzing the state and action spaces of VizDoom, the reward structure, and the challenges posed by partial observability and high visual dynamics. An experimental framework is proposed to compare the baseline DQN with its enhanced versions using performance metrics such as mean and minimal rewards, variance and learning stability.

The results show significant improvements when advanced techniques, particularly n-step returns and prioritized replay, are incorporated, leading to faster and more robust learning in challenging VizDoom scenarios. The thesis contributes to the practical application of deep reinforcement learning in complex 3D environments and provides a foundation for further research on reinforcement learning algorithms in visually rich and dynamic tasks.

ЗМІСТ

ВСТУП
10

Розділ 1. **АНАЛІТИЧНИЙ ОГЛЯД МЕТОДІВ ГЛИБОКОГО НАВЧАННЯ З ПІДКРІПЛЕННЯМ У СЕРЕДОВИЩАХ КОМП'ЮТЕРНИХ ІГОР15**

1.1. Класичні підходи до навчання з підкріпленням 15

1.1.1. Формальна постановка задачі НП та ціннісні функції 15

1.1.2. Політика, функція виграшу та дилема дослідження і використання 16

1.1.3. Часткова спостережуваність та POMDP 17

1.2. Глибоке навчання з підкріпленням 17

1.2.1. Поєднання нейронних мереж та НП 17

1.2.2. Проблеми корельованих даних та нестабільності Беллмана 17

1.2.3. Типові архітектури глибоких мереж для НП 18

1.3. Архітектура Deep Q-Network (DQN) 19

1.3.1. Ідея DQN та апроксимація Q-функції 19

1.3.2. Цільова мережа та відтворення досвіду 19

1.4. Огляд середовищ для досліджень глибокого навчання з підкріпленням 20

1.4.1. Atari та інші 2D-середовища 20
1.4.2. 3D-середовища та шутери від першої особи 20
1.5. Аналіз підходів до використання DQN та модифікацій у VizDoom 21
1.5.1. DQN-агенти у VizDoom 21
1.5.2. Обмеження та відкриті проблеми 21
1.5.3. Порівняльний аналіз підходів DQN та його модифікацій у VizDoom 22
1.5.4. Висновки 23

Розділ 2. **СИСТЕМНИЙ АНАЛІЗ ОБ’ЄКТА ДОСЛІДЖЕННЯ ТА ОБҐРУНТУВАННЯ ПРОБЛЕМИ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ DQN-АГЕНТІВ У СЕРЕДОВИЩІ VIZDOOM** **26**
2.1. Середовище VizDoom як об’єкт дослідження 26
Загальна характеристика середовища 28
2.1.1. Можливості конфігурурування сценаріїв 29
2.2. Формалізація задачі навчання агента у VizDoom як задачі НП 33
2.2.1. Простір станів (кадри, стек кадрів, додаткові ознаки) 34
2.2.2. Простір дій та структура винагороди 36
2.3. Аналіз предметної області та вимог до програмної системи . 38
2.3.1. Вимоги до швидкості та стабільності навчання агентів 40
2.3.2. Вимоги до збору, зберігання та обробки експериментальних даних 42
2.4. Аналіз факторів, що впливають на ефективність DQN-агентів у VizDoom 44
2.4.1. Вибір архітектури мережі та гіперпараметрів 46
2.4.2. Вплив розміру досвіду, частоти оновлення та інших налаштувань 48
2.5. Постановка та обґрунтування проблеми 50
2.5.1. Формулювання проблеми підвищення швидкості навчання 51
2.5.2. Необхідність порівняння базового DQN з покращеними варіантами 53
2.6. Критерії ефективності та показники якості роботи агентів . 55
2.6.1.

Вимоги до інтерпретації та візуалізації результатів . 57 2.7. Постановка проблеми і обґрунтування розроблення програмного продукту 58

Розділ 3. **МЕТОДИ ТА ЗАСОБИ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ НАВЧАННЯ DQN-АГЕНТІВ У СЕРЕДОВИЩІ VIZDOOM** **65** 3.1. Вибір та обґрунтування методів вирішення проблеми 65

3.1.1. Специфіка та особливості інформації у програмному продукті 65
3.1.2. Методи і засоби обробки даних: препроцесинг візу- . . . . . .
66альних кадрів 3.1.3. Вимоги до подання фактів і “правил” у даних: кор-
тежі досвіду . . . . . .
66 3.1.4. Вибір форми подання знань: Q-функція як параме- тричне подання політики . . . . . 67 3.1.5. Порівняння і обґрунтування переваг вибраних мето-
дів над альтернативами . . . . . .
67
68 3.2. Покращення DQN у програмному продукті . . . . . .
3.2.1. Стекування кадрів для обробки візуальної інформації 68 3.2.2. n-
крокові повернення для прискорення поширення
винагороди . . . . . .
68
Пріоритетне повторне відтворення досвіду (PER) . . 69 3.2.3.
3.2.4. Double DQN як метод зменшення переоцінки . . . . .
70 3.2.5.
Dueling-архітектура для стабільнішого оцінювання
70станів . . . . . .
3.2.6. Цільова мережа та частота її оновлення . . . . . .
70 3.3. Опис механізмів логічного накопичення даних . . . . . .
71
71 3.3.1.
Специфіка накопичення даних у задачі VizDoom . . .
3.3.2. Вимоги до порядку та способів пошуку даних у буфері 71 3.3.3.
Обґрунтування вибору інформації для програмного
продукту . . . . . .
72

Способи та засоби створення даних . . . . . .
72 3.3.4.

3.4. Вибір та обґрунтування засобів вирішення проблеми . .
. . 72 3.4.1. Вимоги до платформи
реалізації 72
3.4.2. Обґрунтування вибору середовища реалізації та
бі- бліотек 72
3.4.3. Характеристика обраного середовища з погляду

73функціонування продукту . . . . . .

3.4.4. Вимоги та обґрунтування застосування сервісних
і допоміжних засобів . . . . . .
73 3.5. Висновки до розділу . .
. . . . . .
74

Розділ 4. **ПРАКТИЧНАРЕАЛІЗАЦІЯ СИСТЕМИ НАВЧАННЯ DQN-АГЕНТІВ У СЕРЕДОВИЩІ VIZDOOM 75**

4.1. Описання реалізації завдання 75

4.1.1. Структура та склад бази даних 75

4.1.2. Структура програмного забезпечення 76

4.1.3. Реалізація механізмів прийняття рішень 77

4.1.4. Технологічний процес обробки інформації 77

4.2. Аналіз отриманих результатів 79

4.2.1. Верифікація розробленої системи 79

4.2.2. Валідація та оцінювання ефективності 80

4.2.3. Налаштування експериментів 80

4.2.4. Висновки до розділу 85

Висновки

86

Список використаних джерел

89

ВСТУП

Глибоке навчання з підкріпленням (ГНП) упродовж останнього де-сятиліття стало однією з ключових технологій у сфері штучного інтеле-кту. Його відмінність полягає у здатності агентів навчатися оптимальнійповедінці не через аналіз статичних даних, а через послідовну взаємодіюз динамічним середовищем, що є предметом навчання з підкріпленням(НП). Це робить ГНП надзвичайно привабливим для задач, де необхі-дно приймати рішення в реальному часі, адаптуючись до змінних умові неповної інформації. Одним із найвпливовіших алгоритмів у цій галузіє Deep Q-Network (DQN), який уперше дав змогу агентам опановуватискладні візуальні задачі без використання ручних ознак, працюючи без-посередньо з необробленими піксельними зображеннями.

Успіхи DQN у середовищах ігор Atari спричинили стрімкий розви-ток підходів НП на основі глибоких нейронних мереж. Проте перехід віддвовимірних середовищ до тривимірних значно ускладнює задачу, оскіль-ки збільшується як розмірність станів, так і складність динаміки середо-вища. Саме тому середовище VizDoom, побудоване на рушії класичноїгри Doom, стало однією з ключових платформ для дослідження ГНП ускладних 3D просторах. VizDoom надає агентам завдання, що вимагаютьнавігації, ухилення, виявлення об’єктів, керування ресурсами та швидко-го реагування на візуальні стимули.

На відміну від 2D-середовищ, тут агент стикається з частковою спо-стережуваністю, коли інформація про стан доступна лише через обмеже-ний кут огляду, а важливі об’єкти можуть бути тимчасово прихованими.Крім того, динамічність середовища, наявність перешкод і висока швид-кість змін станів створюють умови, за яких класичні підходи НП можутьвиявитися недостатньо ефективними. Це зумовлює необхідність розробкимодифікацій базового DQN, які здатні покращити якість представленнястану, стабільність навчання та здатність агента робити прогноз на основікороткочасної динаміки середовища.

Таким чином, дослідження DQN у VizDoom є важливим як для тео-

ретичного розвитку ГНП, так і для практичного застосування алгоритміву реальних системах, що працюють з візуальною інформацією в умовахневизначеності.

Актуальність теми. Сучасні тенденції у сфері штучного інтеле-кту свідчать про все більшу інтеграцію автономних інтелектуальних си-стем у реальні застосування: від мобільної робототехніки до промисловихавтоматизованих комплексів, від систем автоматичного керування транс-портом до розумних систем відеоаналітики. У всіх цих випадках критичноважливим є здатність системи працювати з візуальною інформацією таприймати дії в непередбачуваних умовах.

Саме середовище VizDoom дає змогу моделювати такі задачі у зру-чній експериментальній формі. Агент, що навчається у VizDoom, змуше-ний працювати з шумними та складними візуальними спостереженнями,швидко реагувати на загрози, прогнозувати майбутні стани й адаптува-тися до непередбачуваних подій. Таким чином, VizDoom є ближчим дореальних задач, ніж класичні навчальні середовища, і тому може слугу-вати тестовим майданчиком для перевірки життєздатності сучасних ал-горитмів ГНП.

Попри свій потенціал, базовий DQN має істотні обмеження:

• низька стабільність навчання у середовищах із високою варіа- тивністю;

• залежність від короткострокових спостережень, що є критичним недоліком у випадку часткової спостережуваності;

• повільна збіжність у задачах із багатим простором можливих станів;

• нерівномірність корисності переходів у буфері досвіду.

Це зумовлює необхідність дослідження таких модифікацій, як сте-кування кадрів, n-крокові повернення та пріоритетне повторне відтворен-ня досвіду, які спрямовані на покращення роботи агентів НП у складнихтривимірних середовищах.

Мета і завдання дослідження. Метою магістерської роботи єрозробка та експериментальне дослідження удосконалених підходів до на-вчання агентів у VizDoom на основі алгоритму DQN, а також визначеннявпливу різних модифікацій на ефективність навчання в рамках НП.

Для досягнення поставленої мети сформульовано такі завдання:• провести детальний огляд класичних методів НП та сучасних підходів ГНП;
• описати особливості середовища VizDoom та формалізувати йо- го як задачу НП;
• реалізувати базову версію DQN для роботи зі вхідними візуаль- ними даними;
• впровадити стекування кадрів, n-крокові повернення та відтво- рення досвіду за пріоритетом;
• виконати експериментальне порівняння різних конфігурацій агентів;
• визначити закономірності впливу параметрів на швидкість збі- жності та якість політики;
• сформувати рекомендації щодо практичного застосування мо- дифікованих DQN-агентів у 3D середовищах.

Об’єкт дослідження. Об’єктом дослідження є процес навчанняагентів з підкріпленням у динамічних середовищах із частковою спосте-режуваністю та високорозмірними візуальними станами.

Предмет дослідження. Предметом дослідження є DQN алго-ритм, методи його вдосконалення та вплив цих методів на якість навчанняагента у тривимірному середовищі VizDoom.

Наукова новизна одержаних результатів. У магістерськійкваліфікаційній роботі одержано результати, що становлять наукову но-визну та формують теоретичне підґрунтя для розв’язання поставленогонаукового завдання.

1. **Вперше одержано** емпіричне обґрунтування того, що дляоцінювання стабільності збіжності DQN-агентів у середовищіVizDoom інформативним є не лише середнє значення резуль-татів тестових епізодів, а і найнижчий отриманий результат се-ред усіх таких епізодів. Встановлено, що для відносно простихсередовищ, середня винагорода є малочутливою до випадковихдеградацій політики, тоді як найменший зафіксований резуль-тат відображає реальну надійність агента при різних початкових

умовах, заданих генератором випадкових чисел.

2. **Удосконалено** методику експериментального аналізу модифі-кацій алгоритму DQN завдяки впровадженню багатокритеріаль-ного підходу до оцінювання агентів. На відміну від традицій-них методів, які орієнтуються переважно на середню продуктив-ність, запропоновано оцінювати також стійкість поведінки аген-та через аналіз найгірших тестових результатів. Такий підхіддозволяє точніше порівнювати конфігурації у частково спосте-режуваних тривимірних середовищах.

3. **Подальшого розвитку** набула методика системного вивченнявпливу модифікацій DQN у візуальних середовищах. У роботівперше здійснено повний перебір комбінацій трьох компонентів(стекування кадрів, використання багатокрокових повернень тазастосування пріоритетного повторного відтворення досвіду) якокремо, так і у взаємодії між собою. Це дало змогу виявитинизку нелінійних закономірностей та встановити, що поведінкаагентів залежить не лише від наявності окремої модифікації, ай від поєднання кількох параметрів одночасно.

4. **Вперше одержано** практично орієнтовані рекомендації що-до вибору конфігурацій DQN у середовищах класу VizDoom наоснові поєднання оцінки середньої ефективності та стійкості донайгірших випадків. Показано, що надійність політики визнача-ється не лише її середньою продуктивністю, а й тим, наскількисильно змінюються результати між окремими тестовими епізо-дами та між різними початковими умовами. Отримані рекомен-дації заповнюють методичну прогалину між теоретичними опи-сами модифікацій DQN та їх реальною практичною поведінкоюу тривимірних середовищах із частковою спостережуваністю.

Практичне значення одержаних результатів Практичне значення роботи полягає у можливості використання отриманих резуль-татів для:

• побудови надійних агентів НП для середовищ із візуальною ди- намікою;

• підвищення ефективності прототипів автономних систем, що

працюють у складних умовах;

• оптимізації алгоритмів на основі DQN у робототехніці, відеоана- літиці та системах безпілотного керування.