Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет інформаційно-аналітичних технологій та менеджменту
(повна назва)

Кафедра прикладної математики
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Виявлення магазинних крадіжок на відео за допомогою нейронних мереж
(тема)

Виконав:
студент _2_ курсу, групи _ПМм-22-1
Сидоренко Б.Ю.
(прізвище, ініціали)
Спеціальність
113 Прикладна математика
(код і повна назва спеціальності)
Тип програми _освітньо-професійна
(освітньо-професійна або освітньо-наукова)
Освітня програма
Прикладна математика
(повна назва освітньої програми)
Керівник проф. Кіріченко Л.О.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри ПМ _____ Сидоров М.В.
(підпис) (прізвище, ініціали)

2024 р.

Харківський національний університет радіоелектроніки

Факультет _інформаційно-аналітичних технологій та менеджменту
Кафедра _прикладної математики
Рівень вищої освіти _другий (магістерський)
Спеціальність _113 Прикладна математика
(код і повна назва)
Тип програми _освітньо-професійна
(освітньо-професійна або освітньо-наукова)
Освітня програма _Прикладна математика
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри ПМ _____
(підпис)
“_06_” _листопада_ 2023 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові _Сидоренку Богдану Юрійовичу
(прізвище, ім’я, по батькові)

1. Тема роботи _Виявлення магазинних крадіжок на відео за допомогою нейронних мереж
затверджена наказом по університету від 2 листопада 2023 р. № 1276 Ст

2. Термін подання студентом роботи до екзаменаційної комісії 10 січня 2024 р.

3. Вихідні дані до роботи _відеоряд з камери спостереження в магазині роздрібної торгівлі

4. Перелік питань, що потрібно опрацювати в роботі
1. Аналіз предметної області
2. Вибір і обґрунтування методу розв’язання
3. Програмна реалізація
4. Результати обчислювального експерименту
5. Аналіз можливих застосувань

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій

1. Актуальність теми роботи
2. Постановка задачі
3. Аналіз предметної області
4. Метод чисельного аналізу
5. Результати обчислювального експерименту

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Підбір та вивчення технічної літератури за темою роботи	6 – 12 листопада 2023 р.	виконано
2	Вибір та обґрунтування методу	13 – 26 листопада 2023 р.	виконано
3	Розробка алгоритму і програми	27 листопада – 10 грудня 2023 р.	виконано
4	Проведення аналітичних досліджень та розрахунків	11 грудня – 24 грудня 2023 р.	виконано
5	Робота над текстом пояснювальної записки	25 грудня 2023 р. – 9 січня 2024 р.	виконано
6	Представлення роботи на рецензію в ЕК	10 січня 2024 р.	виконано

Дата видачі завдання _6_ листопада 2023 р.

Студент _____
(підпис)

Керівник роботи _____ проф. Кіріченко Л.О.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 84 с., 6 табл., 5 рис., 1 дод., 53 джерела.

ВІДЕО КЛАСИФІКАЦІЯ, ВІДЕОСПОСТЕРЕЖЕННЯ, ЗГОРТКОВІ НЕЙРОННІ МЕРЕЖІ, КОМП'ЮТЕРНИЙ ЗІР, МАШИННЕ НАВЧАННЯ, РЕКУРЕНТНІ НЕЙРОННІ МЕРЕЖІ, РОЗПІЗНАВАННЯ ВІДЕО, ШТУЧНИЙ ІНТЕЛЕКТ.

Об'єкт дослідження – задача розпізнавання шопліфтингу в магазині роздрібної торгівлі.

Мета роботи – розробка класифікатору відео для виявлення шопліфтингу за даними відеоспостереження на основі нейронної мережі.

Методи дослідження – методи попередньої обробки даних, знаходження людей в кадрі (YOLO-NAS), відстеження людей на послідовності кадрів (DeepSort), класифікація відео за допомогою нейронної мережі SlowFast з Res-Net основою.

У даній кваліфікаційній роботі в ході дослідження використовувалися глибокі нейронні мережі, що дозволяють автоматично вивчати та розпізнавати складні закономірності у поведінці осіб на відеозаписах. Застосування таких алгоритмів дозволяє моделі адаптуватися до різноманітних умов і отримувати високу точність виявлення крадіжок.

Сфера застосування даної роботи розширюється від торгових мереж і супермаркетів до різноманітних галузей, де важливо забезпечити високий рівень безпеки та захисту майна. Такий підхід може бути корисним для бізнесу, охоронних служб, транспортних підприємств тощо.

Під час виконання кваліфікаційної роботи було досліджено та проаналізовано низку різноманітних існуючих алгоритмів для розпізнавання відео, також було обрано алгоритм для відео класифікації та на власному наборі даних модель нейронної мережі було навчено та оцінено.

ABSTRACT

Introductory note: 84 pages, 6 tables, 5 figures, 1 appendix, 53 sources.

ARTIFICIAL INTELLIGENCE, COMPUTER VISION, CONVOLUTIONAL NEURAL NETWORKS, MACHINE LEARNING, RECURRENT NEURAL NETWORKS, VIDEO CLASSIFICATION, VIDEO RECOGNITION, VIDEO SURVEILLANCE.

Object of research – the task of recognizing shoplifting in a retail store.

Purpose of work – to develop a video classifier for detecting shoplifting using video surveillance data based on a neural network.

Methods of research – data preprocessing methods, person detection (YOLO-NAS), person tracking (DeepSort), video classification using the SlowFast neural network with ResNet backbone.

In this qualification work, deep neural networks were used in the study, which allow to automatically learn and recognize complex patterns in the behavior of people in video recordings. The use of such algorithms allows the model to adapt to various conditions and obtain high accuracy of theft detection.

The scope of this work extends from retail chains and supermarkets to various industries where it is important to ensure a high level of security and property protection. This approach can be useful for businesses, security services, transportation companies, etc.

During the qualification work, several different existing algorithms for video recognition were researched and analyzed, an algorithm for video classification was selected, and a neural network model was trained and evaluated on its own data set.

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАК, ОДИНИЦЬ І ТЕРМІНІВ

2D-CNN – двовимірні згорткові нейронні мережі;

3D-CNN – тривимірні згорткові нейронні мережі;

CMOT – структура оптимального транспортування;

CNN – згорткові нейронні мережі;

COSNet – нейронна мережа, чутлива до вартості;

CRAVED – абревіатура, що перекладається як Прихована, Знімна, доступна, Цінна, Приємна та Одноразова (англ. Concealable, Removable, Available, Valuable, Enjoyable, and Disposable);

CV – комп’ютерний зір;

DeepSORT – глибоке просте онлайн відстеження в режимі реального часу (англ. Deep Simple Online Realtime Tracking);

OrViT – модель відеотрансформерів об'єктно-області;

RCNN – регіональна згортка нейронної мережі;

ResNet – залишкова нейронна мережа;

RNN – рекурентні нейронні мережі;

StagNet – мережа просторово-часової уваги та семантичних графів;

YOLO-NAS – ти дивишся лише раз - пошук нейронної архітектури (англ. You Only Look Once - Neural Architecture Search).

ВСТУП

Актуальність теми. Розпізнавання шопліфтингу в магазинах залишається надзвичайно актуальною проблемою, оскільки вона впливає на різні аспекти роздрібної торгівлі та суспільства в цілому. Економічні втрати, спричинені крадіжками, мають негативний вплив на прибутковість підприємств, що може призводити до підвищення цін на товари та зменшення ефективності бізнесу. Це, в свою чергу, може вплинути на споживачів.

Застосування штучного інтелекту дозволяє аналізувати величезний обсяг даних та ефективно розпізнавати аномалії в поведінці покупців, що може свідчити про потенційний шопліфтинг.

Важливим викликом для індустрії є постійна адаптація до нових стратегій шопліфтерів, які шукають інноваційні способи вчинення крадіжок. Таким чином, набуттям нових технологій, таких як системи відеоспостереження та розпізнавання крадіжок, відкривається можливість для вдосконалення систем безпеки та попередження випадків шопліфтингу. Використання цих інструментів може забезпечити більш ефективний контроль і виявлення потенційних крадіжок, що сприятиме зменшенню збитків для бізнесу.

Мета і завдання кваліфікаційної роботи. Метою кваліфікаційної роботи є розробка розробка класифікатору відео для виявлення шопліфтингу за даними відеоспостереження на основі нейронної мережі.

Для досягнення поставленої мети необхідно виконати наступні завдання:

– провести огляд і аналіз сучасного стану задач «розпізнавання шопліфтингу»;

– розглянути методи розпізнавання відео;

– вибрати та дослідити найбільш підходящий під виконання задачі алгоритм;

– ознайомитися з SlowFast попередньо навченою моделлю для класифікації відео;

– на основі обраної моделі побудувати нову з використанням існуючих ваг;

– розробити програмну реалізацію для навчання, підбору гіперпараметрів та оцінки якості роботи моделі;

– провести аналіз роботи класифікатора;

– на основі отриманих даних зробити висновок про проведену роботу.

Об'єктом дослідження є задача розпізнавання шопліфтингу в магазині роздрібної торгівлі.

Предметом дослідження є програмна реалізація архітектури моделі SlowFast та її навчання для класифікації відео.

Методи дослідження. У кваліфікаційній роботі використовуються методи попередньої обробки даних, знаходження людей в кадрі (YOLO-NAS), відстеження людей на послідовності кадрів (DeepSort), класифікація відео за допомогою нейронної мережі SlowFast з ResNet основою.

Публікації. Результати, отримані у кваліфікаційній роботі, були на друковані в журналі «Progress in Polish Artificial Intelligence Research 4» [1], представлені на 27-му Міжнародному молодіжному форумі «Радіоелектроніка та молодь у XXI столітті» (м. Харків, 10-12 квітня 2023 р.) [2], та подані на конкурс студентських наукових робіт зі штучного інтелекту.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧ ДОСЛІДЖЕННЯ

1.1 Означення об’єкта дослідження та аналіз проблеми

Крадіжка з магазинів – це розкрадання товарів із закладів роздрібної торгівлі. Якщо говорити точніше, це вчиняється не працівниками в робочий час установи. Це також широко відоме як «крадіжка в магазині» у Великобританії, «зменшення» в галузі роздрібної торгівлі та «розвиток» вуличних злочинців. На відміну від інших форм крадіжок, крадіжки в магазинах зазвичай залишаються непоміченими протягом деякого часу або можуть не бути виявлені тижнями або ніколи – особливо у великих роздрібних мережевих магазинах. Більшість крадіжок у магазинах є аматорами, які ховають товари в кишенях, сумках або іншим чином при собі. Однак є й інші люди та групи, які заробляють на життя крадіжками в магазинах, які набагато більш вправні у крадіжках [3].

У разі крадіжки товару рітейлери зазнають як прямих, так і непрямих збитків. Очевидно, що вартість об'єкта та прибуток, який він міг би принести, зникли. Але те саме стосується і капіталу, вкладеного в купівлю, обробку, маркетинг і демонстрацію цього предмета, а також до альтернативної вартості капіталу. Примітно, що коли злодії знищують популярний товар, покупці стають незадоволеними і іноді звертаються до конкуруючих магазинів та/або товарів [4]. Крім того, коли роздрібні продавці неодноразово стають об'єктами крадіжок у магазинах та інших злочинів у сфері роздрібної торгівлі, законні покупці можуть почуватися небезпечно та демонструвати уникну поведінку, таку як зниження купівельної активності, обмеження покупок у нічний час, скорочення відвідувань магазинів та перехід до конкурентів через страх перед злочинністю [5]. Тоді є юридичні міркування. Притягнення до відповідальності магазинних злодіїв – дороге і трудомістке завдання [6, 7]. Деякі роздрібні торговці очікують і навіть приймають певну суму втрат від крадіжки (зазвичай звану «усадкою») як прикрі, але неминучі витрати на ведення бізнесу. Але

втрати від крадіжок можуть бути настільки більшими, що життєздатність роздрібного торговця опиняється під загрозою. Джонсон і Кейм [8] стверджують, що чистий прибуток роздрібних фірм, ймовірно, погіршиться під тиском високого рівня конкуренції та змін у поведінці споживачів, що зробить такі тяжкі втрати нестерпними та потребуватиме цілеспрямованих рішень [9].

Існує типова послідовність подій, за якою слідкують крадії. Їхні рішення ґрунтуються на середовищі та ситуації, з якою вони стикаються. Гілл [10] розділив процес прийняття рішень крадієм на шість концептуальних етапів:

а) вибір магазину. Крамні злодії ґрунтують свій вибір на таких факторах, як цільові продукти, близькість магазину та ймовірний ризик бути впізнаним персоналом або зіткнутися з видимими заходами безпеки;

б) магазинні злодії прагнуть залишитися непоміченими, підкреслюючи важливість відчуття себе непоміченими. Такі фактори, як взаємодія персоналу, зайнятість магазину та загальна атмосфера, впливають на їхнє рішення продовжити крадіжку;

в) визначення розташування продукту. Залежно від конкретної людини злодії можуть знати місцезнаходження потрібного продукту чи їм потрібно шукати їх у магазині;

г) приховування товару. Цей крок відрізняє професійних злодіїв від аматорів. Професіонали віддають перевагу швидкості та секретності, швидко приховуючи предмети. Аматори можуть ходити з продуктом, виглядати менш гладкими та потенційно виявляти ознаки нервозності;

ґ) догляд з магазину. Фахівці прагнуть піти швидко, змішавшись з іншими покупцями. Аматори можуть проявляти нервову поведінку, наприклад відступати на виході, щоб виділитися;

д) утилізація товарів. Професіонали часто продають чи «огорожують» вкрадені речі, покладаючись на те, що інші сплатять за їхні послуги. Аматори зазвичай зберігають продукти для особистого використання або діляться ними з друзями та сім'єю [10].

Незважаючи на те, що крадіжка в магазині зазвичай сприймається як незначне правопорушення, вона незмінно вважається злочином проти власності, яка потребує великих витрат. Лише у 2011 році крадіжки у магазинах призвели до збитків роздрібної торгівлі на суму близько 51 мільярда доларів [10], що робить їх серйозною економічною проблемою. Більше того, це поширений злочин: приблизно кожна одинадцята людина регулярно чинить крадіжку у роздрібних торговців [12]. Крадіжка в магазині відносно легше зробити непоміченою: приблизно один із 150 інцидентів призводить до арешту та втручання поліції [13].

Наслідки крадіжок у магазинах поширюються на багато аспектів. По-перше, це накладає фінансовий тягар на магазини, виробників та споживачів. По-друге, система кримінального правосуддя має виділяти ресурси для переслідування магазинних злодіїв, включаючи арешт, судовий розгляд та виправлення. По-третє, крадіжки в магазинах спричиняють різні матеріальні і нематеріальні витрати для суспільства, виступаючи як потенційний провісник більш серйозних злочинів і пов'язаних зі зловживанням наркотиками. Цей злочин визнається "перехідним злочином", при цьому неповнолітні часто переходять до більш серйозних правопорушень, а для активних грабіжників – «резервним злочином», коли крадіжка зі зломом недоцільна [14, 15]. Існує також чітко встановлений зв'язок між крадіжками в магазинах та зловживанням наркотиками, оскільки люди вдаються до крадіжки, щоб отримати товари для продажу чи обміну, щоб підтримати свою пристрасть до наркотиків.

Наслідки крадіжок у магазинах виходять за межі економічних втрат для споживачів та роздрібних продавців. Це впливає на систему кримінального правосуддя, яка витрачає ресурси на поширені, але ненасильницькі злочини. Видимі наслідки для рітейлерів включають зниження прибутків, потенційне закриття магазинів, а також несприятливий вплив на можливості працевлаштування та благоустрій району [16]. Більше того, витрати, пов'язані з крадіжками в магазинах, не обмежуються фінансовими втратами, але включають зниження морального духу персоналу, фізичну та психологічну шкоду, що веде до

втрати роботи і навіть до людських жертв [17]. Багатогранні наслідки крадіжок у магазинах наголошують на її глибоких і широко поширених наслідках для окремих людей, спільнот та ширшої соціальної структури [3].

Також можна провести порівняння між частотою шопліфтингу та звичайними крадіжками (для порівняння візьмемо відкриті данні США за період січня 2018 – червень 2023). Злочини, пов’язані з магазинними крадіжками, становили приблизно 20% усіх крадіжок протягом періоду дослідження. Хоча деяке зниження крадіжок можна пояснити зниженням крадіжок у магазинах, крадіжки, не пов’язані з крадіжками, незалежно від того зменшилися на початку пандемії. Крадіжки, не пов’язані з крадіжками в магазинах, і крадіжки в магазинах, ймовірно, значно знизилися через обмеження, пов’язані з COVID-19, закриття магазинів і зменшення відвідуваності торгових та інших закладів, які залишалися відкритими [18].

Таким чином, крадіжки в магазинах створюють серйозні проблеми для роздрібних продавців, що призводить до суттєвих втрат, недоступності продуктів, збільшення витрат та насильства в магазинах, яке зачіпає всю спільноту. Крадіжка в магазинах має довгу історію, в якій беруть участь люди різного походження. Хоча демографічні чинники, такі як вік, раса та економічний стан, були ретельно вивчені, вік особливо корисний для роздрібних продавців. Дослідники розробили типології злочинців, щоб зрозуміти психологічні мотиви та класифікувати злодіїв на основі мотивів, методів та цільових товарів, що допомагає сфокусувати профілактичні програми. Окрім типологій, кримінологічні теорії та теорії споживчої поведінки розвиваються, щоб пояснити окремі правопорушення та спосіб дій ймовірних правопорушників, що сприяє розробці більш цілеспрямованих захисних заходів. Такі теорії, як рутинна діяльність, ситуативне запобігання злочинам, запланована поведінка та трикутник крадіжок обіцяють допомогти роздрібним торговцям збільшити продажі та скоротити втрати [9].

В даний час роздрібні торговці використовують поєднання персоналу, програм та систем для запобігання крадіжкам у магазинах. Навчання персоналу

магазинів, менеджерів та фахівців із захисту активів тому, як приділяти увагу покупцям, впроваджувати процедури запобігання втратам та повідомляти про підозрілі дії, є звичайною практикою. Планування магазинів розроблено для покращення спостереження і включає ширші проходи, нижні полиці, яскравіше освітлення та видимі робочі зони співробітників. В даний час вивчаються методи захисної упаковки, включаючи стійкі до злому наклейки, обтискання, друк та використання міцних матеріалів. Такі технології, як відеоспостереження, електронне виявлення, маркування відмови у пільгах та сигналізація про вилучення товарів, підвищують ризик виявлення магазинних злодіїв. Контрольований доступ до предметів із високими втратами та методи зниження мобільності забезпечують захист у середовищах високого ризику. Крадіжка у магазинах не тільки знижує доступність товарів та життєздатність роздрібних продавців, а й призводить до насильства у магазинах. Незважаючи на багатство існуючих досліджень і теорій магазинних крадіжок, подальше вивчення динаміки індивідуальних та організованих крадіжок, а також ефективності захисних програм та технологій у різних умовах залишається важливою областю, яка потребує більшого вивчення.

1.2 Огляд попередніх досліджень

Шопліфтинг, тобто крадіжка товарів із торгових точок, є постійною проблемою як для бізнесу, так і для правоохоронних органів. Як складне та багатогранне явище, мотиви, методи та наслідки крадіжок у магазинах зацікавили дослідників різних дисциплін. Метою цього розділу є надання всебічного огляду попередніх досліджень крадіжок у магазинах, вивчення ключових тем, методологій та результатів, які сформували наше розуміння цієї незаконної поведінки.

Історично вивчення крадіжок у магазинах перетворилося на спрощене вивчення злочинної поведінки на більш детальне дослідження соціальних,

психологічних та економічних факторів, що сприяють цьому явищу. Ранні дослідження часто були зосереджені на кримінальному профілюванні та стратегіях правоохоронних органів з метою виявити загальні характеристики магазинів та розробити ефективні заходи стримування. Однак сучасні дослідження розширили сферу своєї діяльності та включили ширший спектр впливів, починаючи від індивідуальних психологічних факторів та закінчуючи соціальними та економічними умовами, що сприяють поширенню магазинних крадіжок.

Аби всебічно проаналізувати попередні дослідження шопліфтингу, важливо вивчити різноманітні методологічні підходи, які використовуються вченими для дослідження цього складного явища. Ранні дослідження переважно спиралися на кількісні методи, використовуючи опитування та статистичний аналіз для виявлення закономірностей та кореляцій, пов'язаних із випадками крадіжок у магазинах. Ці зусилля були спрямовані на кількісну оцінку поширеності крадіжок у магазинах, визначення злочинців та оцінку ефективності профілактичних заходів. В останні роки дослідники все частіше використовують якісні методології для вивчення суб'єктивного досвіду та мотивацій, що лежать в основі крадіжки у магазинах. Глибинні інтерв'ю, тематичні дослідження та етнографічні підходи дозволили отримати цінну інформацію про соціальні та психологічні аспекти крадіжок у магазинах, розкриваючи нюанси взаємодії особистих обставин, емоційних тригерів та соціальних впливів. Інтеграція як кількісних, і якісних методів стала помітною тенденцією, що дозволяє отримати цілісне розуміння магазинних крадіжок, що виходить за рамки простого статистичного уявлення. У міру розвитку літератури з магазинних крадіжок стає очевидним, що інтеграція різних методологічних підходів збагатила глибину та широту знань у цій галузі.

Магазинні злодії неоднорідні: вони різного віку, раси, статі та походження. Їх мотиви різноманітні – багато хто краде, очевидно, з жадібності, іноді з потреби, а часом їх дії здаються незрозумілими [19, 20, 21, 22]. Дехто вважає, що в основі проблеми крадіжок у магазинах лежить феномен споживчого бажання [23]. Сьогоднішні ритейлери та постачальники їхньої

продукції стикаються з парадоксальною проблемою: як підвищити попит на продукцію у законних споживачів та одночасно стримувати потенційних магазинних злодіїв. Крадіжка в магазинах залишається серйозною соціальною та економічною проблемою, яка потребує більш цілеспрямованих досліджень та рішень. Навіть великомасштабні дослідження змогли продемонструвати тісний зв'язок між крадіжками в магазинах і демографічними характеристиками [24]. Незважаючи на цю труднощі, експерти постійно прагнуть розділити магазинних злодіїв на окремі групи або типології, щоб краще зрозуміти проблему. Більшість із них класифікують крадіжку в магазинах як аномальні злочинні діяння, хоча деякі схильні розглядати її як поведінку, що випливає з крайнього споживання та ощадливості: абсолютний покупець за вигідною ціною. Єдині загальноприйняті узагальнення полягають у тому, що більшість магазинних злодіїв варіюються від молодого до середнього віку, схильні до інших форм поведінки, що відхиляються, і знаходяться з іншими людьми в момент скоєння злочину [23].

Перелічені типології виникли з урахуванням спостережень і теорій різної суворості. Більшість класифікаційних систем були розроблені або для лікування правопорушників, або для їх стримування або для того й іншого. Кемерон та Клемке [19, 23] не повністю засновані на конкретних дослідженнях, а є результатом систематичних спостережень. Усі перелічені типології однаково корисні як дослідників роздрібної торгівлі, так практиків захисту активів. Системи психологічної класифікації можуть особливо допомогти фахівцям у галузі лікування та соціальних працівників.

Подібно до кримінологічних досліджень, дослідження крадіжок у магазинах, як правило, вивчають або індивідуальну злочинність, або походження та динаміку кримінальних подій. З цією метою корисно розглянути деякі відповідні теорії девіантності: дослідження Роберта Мертона [25] з вивчення девіантної поведінки як результату розриву розриву між прагненням до економічного успіху і відсутністю засобів для його досягнення. Дослідники також застосували теорії соціального контролю для пояснення крадіжок у

магазинах. Основна передумова полягала в тому, що слабкі соціальні зв'язки – такі як соціальні порушення чи невдачі, особливо сімейні – призводять до правопорушень чи девіантної поведінки. Таким чином, щоб бути ефективними, стримуючі фактори, які використовують роздрібні продавці, мають бути спрямовані на те, щоб вплинути на сприйняття магазинними злодіями серйозності крадіжки. Останнім часом дедалі більше досліджень намагаються пояснити, як злочинність пов'язана із ситуаційними чинниками чи чинниками довкілля. Один із таких прикладів – теорія раціонального вибору – стверджує, що злочинці, такі як магазинні злодії, насправді найчастіше є нормальними, розумними людьми, які зважують відносні ризики та вигоди, пов'язані зі злочином, перш ніж ухвалити рішення про його вчинення [9].

Концепція ситуаційного запобігання злочинності [20, 26, 27] поєднує в собі як рутинну теорію, так і теорії раціонального вибору в розвивається, заснованому на фактичних даних наборі методів запобігання злочинності. Методи ситуаційного запобігання злочинності підкреслюють скорочення можливостей, вплив на мотивацію та підвищення особистого ризику для потенційних злочинців [28]. Працюючи в рамках ситуаційної профілактики злочинів, трикутник крадіжок розпізнає та поєднує «фонові фактори» потенційного злочинця з «факторами переднього плану» у конкретній ситуації. Перш ніж спробувати вчинити крадіжку, правопорушник може взяти до уваги деякі першочергові фактори [24]: мотив чи намір вкрасти, передбачуваний рівень особистого ризику та рівень можливостей.

Найбільш популярними типами магазинів для злодіїв є ті, в яких продаються дуже бажані товари, легко доступні і здаються вразливими [24, 29]. До цих типів місць належать торгові центри, універмаги, дисконтні магазини, комп'ютерні магазини, господарські магазини, взуттєві магазини, магазини одягу, магазини музики та відео, продуктові магазини, аптеки та книгарні [8, 30]. Сучасні магазини роздрібної торгівлі пропонують багатий асортимент товарів, але більшість крадіжок посідає відносно невелику групу товарів («гарячі товари»). «Спекотність» предмета частково пояснюється абревіатурою

CRAVED (англ Concealable, Removable, Available, Valuable, Enjoyable, and Disposable); ці предмети найчастіше легше вкрасти і обміняти на готівку, ніж звичайні товари [29]. Приклади нинішніх CRAVED предметів включають леза для гоління, чорнильні картриджі для принтерів, ароматизатори, дитяче харчування і ліки, що відпускаються без рецепта [7].

Огляд попередніх досліджень крадіжок у магазинах підкреслює еволюцію наукового дослідження від спрощених характеристик до більш детального та міждисциплінарного розуміння цієї поширеної поведінки. Дослідники перейшли від переважно кількісних оцінок злочинних моделей до комплексного дослідження багатогранних факторів, які сприяють крадіжкам у магазинах. Інтеграція як якісних, так і кількісних методологій уможливила більш цілісне дослідження, пропонуючи більш багатий гобелен уявлення про мотивацію, досвід і суспільну динаміку, що оточує крадіжки в магазинах.

1.3 Формальна та змістовна постановка задачі

Задача виявлення шопліфтингу у поведінці покупця в магазині роздрібної торгівлі зводиться до задачі класифікації відео.

В формальному виді постановка задачі класифікації має такий вигляд. Нехай $x_i \in X$, $i = \overline{1,n}$, – множина об'єктів ознак, входів моделі, $y_i \in Y$, $i = \overline{1,n}$, – множина об'єктів відповідей, виходів моделі. Пара $(x_i, y_i) \in X \times Y$ називається розмічений об'єкт, або прецедент. Кінцева множина $\{x_i\}$, $i = \overline{1,n}$, представляє собою матрицю $\{x_{i,j}\}$, $i = \overline{1,n}$ $j = \overline{1,m}$, розміром $n \times m$, де рядок матриці – це масив ознак одного об'єкта, $\{y_i\}$, $i = \overline{1,n}$, – вектор відповідей, елемент якого є значення номеру класу. Комбінація $\{x_i\}$, $i = \overline{1,n}$, та $\{y_i\}$, $i = \overline{1,n}$, називається навчальною вибіркою. Задача класифікації полягає у визначенні функції залежності $f : X \to Y$ котра пророкує по $x \in X$ відповіді $y \in Y$.

Крадіжка є особливою формою поведінки для покупця, відмінною від звичайної, тому так як ми маємо дві різні поведінки, то наша задача полягає у класифікації відео.

У відео класифікації ми ставимо перед собою завдання визначення класу, до якого відноситься кожен відеокадр. У нашому випадку це буде відзначення відео як «Шопліфтинг» чи «Не Шопліфтинг». Для досягнення цієї мети ми обираємо глибоку нейронну мережу із спеціалізованою архітектурою SlowFast, яка відзначається високою ефективністю у вирішенні задач відео класифікації.

SlowFast враховує інформацію як просторову, так і часову, але з важливим акцентом на високу швидкість обробки відеоданих. Замість згорткової нейронної мережі та рекурентної нейронної мережі ми використовуємо Slow шари для просторової обробки та Fast шари для часової обробки.

Наша модель SlowFast стане основою для класифікації відео, де вихідним значенням буде один із двох класів: «Шопліфтинг» чи «Не Шопліфтинг». Такий підхід дозволяє нам ефективно розрізняти між цими двома сценаріями за допомогою спеціалізованої архітектури SlowFast.

1.4 Постановка задач дослідження

Метою дослідження є розробка класифікатору відео для виявлення шопліфтингу за даними відеоспостереження на основі нейронної мережі.

Для досягнення поставленої мети необхідно виконати наступні завдання:

– провести огляд і аналіз сучасного стану задач «розпізнавання шопліфтингу»;

– розглянути методи розпізнавання відео;

– вибрати та дослідити найбільш підходящий під виконання задачі алгоритм;

– ознайомитися з SlowFast попередньо навченою моделлю для класифікації відео;

– на основі обраної моделі побудувати нову з використанням існуючих ваг;

– розробити програмну реалізацію для навчання, підбору гіперпараметрів та оцінки якості роботи моделі;

– провести аналіз роботи класифікатора;

– на основі отриманих даних зробити висновок про проведену роботу.

2 ВИБІР ТА ОБҐРУНТУВАННЯ МЕТОДУ РОЗВ’ЯЗАННЯ

2.1 Основні відомості з використання нейронних мереж в комп’ютерному зорі

Слово «нейронний» символізує існування нейронних мереж як штучної спроби скопіювати нейрони людського мозку. Термінологія, яка використовується для частин цієї мережі, така сама, як і в нейронній мережі людини. Тисячі нейронів працюють разом у нейронній мережі, щоб генерувати результати та стежити за процесом навчання та покарання, щоб зробити модель більш ефективною та точною. Нейронні мережі зазвичай використовуються для вирішення завдань, пов'язаних із розпізнаванням або порівнянням нової точки даних з величезним набором даних існуючих точок даних. В нейронних мережах на дію прихованих шарів впливають приховані раніше шари, накопичуючи інформацію, зібрану нейронами.

Значною перевагою використання нейронних мереж для завдань глибокого навчання є те, що глибоке навчання зазвичай має справу з величезними наборами даних, що містять мільйони точок даних. Стандартні алгоритми машинного навчання вимагають значного часу для їх вивчення та навчання. Однак через великі взаємозв'язки між нейронами нейронні мережі усувають необхідність у таких ефективних проміжках часу і можуть ефективно корелювати між різними точками даних. Це стає дуже вигідним, оскільки навчання моделі та точність збільшуються при наданні великого набору даних, що дозволяє моделі відповідати більшості точок даних та точно прогнозувати результати.

Модель нейронної мережі є зображенням людського мозку, оскільки в людському мозку присутні мільярди нейронів, пов'язаних між собою, нейронна мережа, а також набір прихованих шарів, де кожен прихований шар складається з певної кількості нейронів, які комп'ютерний зір або інженер з машинного навчання можуть визначити самі.

Вхідними для нейронної мережі є характеристики точок даних. Важливим аспектом нейронних мереж є функція активації, яка включає модель нелінійність і дозволяє їй вирішувати, чи повинен бути активований нейрон чи ні. Це означає визначення того, чи досить значущі знання нейрона, щоб їх можна було передати на наступний прихований рівень. Інший введений термін називається зміщенням, який використовується для запобігання перенавчанню та недостатньому підбору і дозволяє функції активації зміщувати та усувати такі порушення. Вага – важлива частина алгоритмів машинного навчання і навіть нейронних мереж, яка допомагає визначити необхідність функції, яка є у вхідних даних. Вхідні дані множаться на ваги, а потім додається член зміщення та передається через функцію активації перед подачею на наступний рівень. Перший рівень нейронної мережі називається вхідним шаром, куди подаються дані, які передбачається використовуватиме навчання моделі. Наступні шари відомі як приховані шари та містять нейрони. Кількість прихованих шарів та нейронів у кожному прихованому шарі визначає складність нейронної мережі. Традиційно зменшення кількості нейронів відбувається за рахунок прогресуючих прихованих шарів. Вихідний шар є останнім шаром нейронної мережі, що дає результати нашої моделі. Кількість нейронів у вихідному шарі дорівнює кількості класів у нашій постановці задачі. Вихід розраховується із застосуванням моделей лінійної регресії [31].

Комп’ютерний зір (англ. Computer vision) – наукова галузь, яка визначає, як машини інтерпретують значення зображень і відео. Його основна фундаментальна концепція – це нейронна мережа. Алгоритми комп’ютерного зору аналізують певні критерії в зображеннях і відео, а потім застосовують інтерпретації до завдань прогнозування або прийняття рішень. Він є важливою галуззю штучного інтелекту та машинного навчання, яка має справу з цифровими носіями, такими як зображення та відео. Незважаючи на те, що дослідження щодо подальшої оптимізації штучних нейронних мереж все ще тривають, ефективність сучасних методів у поєднанні з нейронними мережами є взірцевою.

Комп'ютерний зір отримав свою назву через те, що використовується, щоб навчити комп'ютери «бачити» та використовувати візуальну інформацію для виконання візуальних завдань, які можуть виконувати люди. Моделі комп'ютерного зору призначені для перекладу візуальних даних на основі особливостей та контекстної інформації, виявлених під час навчання. Це дозволяє моделям інтерпретувати зображення та відео та застосовувати ці інтерпретації до завдань прогнозування чи прийняття рішень. Хоча обидва вони пов'язані з візуальними даними, обробка зображень – це не те саме, що комп'ютерний зір. Обробка зображень включає зміну або покращення зображень для отримання нового результату. Це може включати оптимізацію яскравості або контрастності, збільшення роздільної здатності, розмиття конфіденційної інформації або обрізання. Різниця між обробкою зображень та комп'ютерним зором полягає в тому, що перше не обов'язково потребує ідентифікації контенту [32].

У комп’ютерному зорі нейронна мережа працює за допомогою шарів взаємопов’язаних вузлів або нейронів для обробки візуальних даних. Мережа отримує вхідне зображення, яке потім пропускається через кілька шарів нейронів. Кожен нейрон застосовує математичну операцію до вхідних даних і передає результат на наступний рівень. У міру того, як дані переміщуються мережею, вони стають все більш абстрактними і зрештою можуть бути використані для прогнозування або класифікації вхідного зображення. Цей процес часто навчається з використанням великих наборів даних, щоб навчитися точно інтерпретувати та розуміти візуальну інформацію. Нейронні мережі можна використовувати для таких завдань, як розпізнавання об’єктів, класифікація зображень і створення зображень у програмах комп’ютерного зору.

Сучасні алгоритми комп’ютерного зору базуються на згорткових нейронних мережах (англ: Convolutional neural networks), які забезпечують різке покращення продуктивності порівняно з традиційними алгоритмами обробки зображень. CNN – це нейронні мережі з багаторівневою архітектурою, яка використовується для поступового скорочення даних і обчислень до найбільш

відповідного набору. Потім цей набір порівнюється з відомими даними, щоб ідентифікувати або класифікувати вхідні дані. CNN зазвичай використовуються для завдань комп’ютерного зору, хоча аналіз тексту та аудіо також можна виконувати. Однією з перших архітектур CNN була AlexNet (описана нижче), яка виграла конкурс візуального розпізнавання ImageNet у 2012 році [32].

Розглянемо наступні види завдань комп’ютерного зору. Одним із найпоширеніших додатків є класифікація зображень. Це передбачає надання комп'ютера можливості ідентифікувати основний об'єкт на зображенні та надання мітки для класифікації зображення. Також можна дозволити комп'ютеру визначити розташування об'єкта на зображенні. Це досягається шляхом укладання об'єкта в «обмежувальну рамку», яку можна ідентифікувати за числовими параметрами, пов'язаними з краями зображення. Класифікація об'єктів обмежена одним об'єктом зображення. Виявлення об'єктів є більш складним і вимагає, щоб комп'ютер виявив та визначив місцезнаходження всіх різних об'єктів на зображенні. Семантична сегментація включає маркування кожного пікселя зображення класом, відповідним тому, що він представляє. Це також відомо як «щільне передбачення», оскільки необхідно передбачити кожен піксель. На відміну від інших завдань комп'ютерного зору, семантична сегментація не просто створює мітки та рамки, що обмежують. Він генерує зображення високої роздільної здатності, у якому класифікується кожен піксель. Сегментація екземплярів йде ще далі, класифікуючи кожен екземпляр одного й того самого класу окремо. Наприклад, якщо на зображенні зображено трьох собак, кожен собака є екземпляром класу «Собака». Кожен із них класифікуватиметься окремо, наприклад, з використанням різних кольорів. Завдяки цим різним завданням комп'ютер «розуміє» зміст зображень з дедалі точнішим рівнем деталізації. У цьому випуску ми зосередимося на задачі семантичної сегментації [33].

Продуктивність і ефективність CNN визначається її архітектурою. Це включає структуру шарів, те, як елементи розроблені та які елементи присутні в кожному шарі. Було створено багато CNN, але нижче наведено деякі з найефективніших проектів.

Розвиток технологій глибокого навчання дозволило створювати точніші і складніші моделі комп'ютерного зору. З розвитком цих технологій використання додатків комп'ютерного зору стає дедалі кориснішим.

Далі розглянемо кілька способів використання глибокого навчання поліпшення комп'ютерного зору.

а) виявлення об'єктів: існує два поширені типи виявлення об'єктів, що виконуються за допомогою методів комп'ютерного зору:

1) двоетапне виявлення об'єктів – для першого кроку потрібна мережа пропозицій регіонів , що надає ряд регіонів-кандидатів, які можуть містити важливі об'єкти. Другим кроком є передача пропозицій регіонів до архітектури нейронної класифікації, зазвичай це алгоритм ієрархічного угруповання на основі RCNN або об'єднання областей інтересу у Fast RCNN. Ці підходи є досить точними, але можуть бути дуже повільними;

2) одноетапне виявлення об'єктів – у зв'язку із необхідністю виявлення об'єктів у реальному часі з'явилися архітектури одноетапного виявлення об'єктів, такі як YOLO, SSD та RetinaNet. Вони поєднують етап виявлення та класифікації шляхом регресії прогнозів рамки, що обмежує. Кожна рамка, що обмежує, представлена всього кількома координатами, що спрощує об'єднання етапів виявлення і класифікації і прискорює обробку;

3) локалізація та виявлення об'єктів – локалізація зображення використовується визначення розташування об'єктів на зображенні. Після ідентифікації об'єкти відзначаються рамкою, що обмежує. Виявлення об'єктів розширюється та класифікує ідентифіковані об'єкти. Локалізація та виявлення об'єктів можуть використовуватись для ідентифікації кількох об'єктів у складних сценах. Потім це можна застосувати до таких функцій, як інтерпретація діагностичних зображень у медицині;

4) семантична сегментація. Семантична сегментація, також відома як сегментація об'єктів, аналогічна до виявлення об'єктів, за винятком того, що вона заснована на певних пікселях, пов'язаних з об'єктом. Це дозволяє більш ретельно визначати об'єкти зображення і не потребує рамок, що обмежують.

Семантична сегментація часто виконується з використанням повністю згорткових мереж або U-мереж. Одним із популярних застосувань семантичної сегментації є навчання автономних транспортних засобів. За допомогою цього методу дослідники можуть використовувати зображення вулиць або проїздів із чітко визначеними межами об'єктів;

5) оцінка пози. Оцінка пози – це метод, який використовується для визначення того, де знаходяться суглоби на зображенні людини або об'єкта та на що вказує розташування цих суглобів. Його можна використовувати як із 2D, так і з 3D зображеннями. Основною архітектурою, яка використовується для оцінки пози, є PoseNet, заснована на CNN. Оцінка пози використовується визначення того, які частини тіла можуть відображатися на зображенні, і може використовуватися створення реалістичних поз чи рухів людських фігур. Часто цей функціонал використовується для доповненої реальності, дзеркального відображення рухів за допомогою робототехніки чи аналізу ходи [32].

Дослідження основ використання нейронних мереж у комп'ютерному зір виявляє важливі аспекти в галузі візуального аналізу. Ці технології не лише надають ефективні методи обробки зображень, а й відчиняють двері для глибшого розуміння сприйняття візуальної інформації. Розуміння того, як нейронні мережі взаємодіють із зображеннями на базовому рівні, аналізуючи їх структуру та зміст, формує фундамент для подальших досліджень та інновацій у галузі комп'ютерного зору. У цьому контексті оцінка застосування нейронних мереж у комп'ютерному зорі являє собою значний внесок у розвиток методів обробки візуальних даних, що залишає перспективні перспективи для майбутніх досліджень і технологічного розвитку.

2.2 Основні відомості з теорії розпізнавання відео

Розпізнавання відео є важливим компонентом комп’ютерного зору, і в останні роки воно користується значним сплеском популярності. З широкою

доступністю цифрових камер, смартфонів і систем відеоспостереження відбувся вибух відеоданих у різних сферах, від розваг і спорту до безпеки та охорони здоров'я. Як наслідок, розпізнавання відео стало важливим інструментом для аналізу та отримання інформації з великих наборів відеоданих. Візуальний досвід за допомогою цифрових носіїв, таких як відео, стає все більш звичним явищем у сучасному технологічно орієнтованому світі. Використання відео стрімко зростає в ногу з технологічним прогресом. Кількість відеозаписів, які мають люди, надзвичайно велика. Але це також тягне за собою додаткові проблеми з моніторингом і аналізом відео.

Розпізнавання відео – це процес аналізу та розуміння вмісту відеопотоку, який зазвичай передбачає виявлення, відстеження та розпізнавання об'єктів, сцен і дій. Це важливий компонент комп'ютерного зору, який пов'язаний з автоматичною інтерпретацією візуальних даних з навколишнього світу. Основна мета розпізнавання відео – отримати значущу інформацію з необроблених відеоданих, перетворивши її на структуроване представлення, яке можна використовувати для аналізу та прийняття рішень [34].

Хоча за останні роки було досягнуто значного прогресу, все ще існує кілька проблем, які необхідно вирішити, щоб побудувати точні та надійні системи розпізнавання відео. Деякі з основних проблем розпізнавання відео:

– висока розмірність даних. Відеодані зазвичай мають велику розмірність, кожен кадр містить мільйони пікселів. Це ускладнює ефективну обробку та аналіз даних. Наприклад, для аналізу 1-хвилинного відеоролика зі швидкістю 30 кадрів на секунду знадобиться обробка понад 100 мільйонів пікселів;

– висока варіативність. Іншою значною проблемою в розпізнаванні відео є мінливість зовнішнього вигляду. Один і той самий об'єкт може виглядати по-різному залежно від умов освітлення або ракурсу камери. Подібним чином різні об'єкти можуть виглядати схожими або навіть ідентичними, що ускладнює їх розрізнення;

– складність взаємодії та діяльності об'єктів. Розпізнавання відео передбачає ідентифікацію не лише окремих об'єктів, а й їх взаємодії та

діяльності. Наприклад, розпізнавання людини, яка йде, передбачає ідентифікацію людини в кадрі, а також визначення її руху в часі та напрямку. Подібним чином розпізнавання групи людей, які грають у футбол, передбачає визначення гравців, м'яча, їхніх рухів і взаємодії. Ці складні взаємодії та дії можуть ускладнити створення точних і надійних систем розпізнавання відео;

– обмежена доступність позначених даних. Іншою проблемою при розпізнаванні відео є обмежена доступність даних, які були анотовані мітками або тегами, що описують об'єкти або дії, присутні у відео. Дані з мітками необхідні для навчання алгоритмів машинного навчання розпізнаванню об'єктів і дій у відео. Однак маркування відеоданих є трудомістким і дорогим процесом, що ускладнює отримання великих обсягів мічених даних. Щоб вирішити цю проблему, дослідники розробили такі методи, як напівконтрольоване навчання та активне навчання, які можуть допомогти зменшити кількість позначених даних, необхідних для навчання;

– продуктивність у реальному часі. Розпізнавання відео зачасту потрібно виконувати в режимі реального часу, наприклад, у системах спостереження або автономних транспортних засобах. Продуктивність у реальному часі вимагає, щоб система розпізнавання відео могла обробляти та аналізувати відеодані, як правило, зі швидкістю 30 кадрів на секунду або вище в реальному часі. Досягти такої продуктивності може бути складно, особливо для методів глибокого навчання, які можуть потребувати інтенсивних обчислень.

Розпізнавання відео з глибоким навчанням передбачає навчання нейронних мереж автоматично вивчати відповідні функції з відеоданих і використовувати їх для розпізнавання об'єктів і дій. Тривимірні згорткові нейронні мережі і рекурентні нейронні мережі – це два популярні типи нейронних мереж, які використовуються для розпізнавання відео.

Глибоке навчання зробило революцію у сфері розпізнавання відео, надаючи потужні інструменти для автоматичного виявлення та класифікації об'єктів і дій у відеоданих. Існує кілька поширених підходів глибокого навчання, які використовуються для різноманітних завдань розпізнавання відео, наприклад розпізнавання дій, пошук відео та субтитри до відео.

Розглянемо найбільш популярні з них:

а) тривимірні згорткові нейронні мережі (англ. 3D Convolutional Neural Networks).

3D-CNN є розширенням двовимірних згорткових нейронних мереж (2D-CNN) і можуть обробляти просторово-часові дані у відео. 3D-CNN можуть навчитися витягувати характеристики з кількох відеокадрів одночасно, на відміну від 2D-CNN, які обробляють кадри по одному, дозволяючи їм захоплювати часову динаміку відео.

Скомпрометована метрика через оптимальне транспортування (англ Compromised Metric via Optimal Transport) – це цікавий метод, який використовує 3D-CNN у структурі навчання з кількома кадрами для розпізнавання дій. CMOT одночасно порівнює відмінності у змісті та порядку двох відео, щоб дати компромісний вимір у рамках структури оптимального транспортування , таким чином балансуючи семантичну та тимчасову інформацію у відео.

CMOT вибирає кілька сегментів відео та отримує їх впровадження за допомогою 3D-CNN для формування послідовності уявлень контенту. Матриця семантичної вартості обчислюється між їхніми уявленнями контенту. Наприклад, сегменти того самого відео матимуть низьку вартість, а два сегменти, взяті з різних відео, матимуть високу вартість. Щоб зберегти властиву інформацію про часовий порядок (який сегмент з'являється після чого), CMOT додатково вносить поправки в матрицю семантичної вартості, караючи її позиційною відстанню між парою сегментів. Наприклад, останній кадр одного сегмента буде дуже близький до першого кадру наступного сегмента, якщо вони послідовно вибираються з одного й того самого відео. Потім створюється метричний класифікатор для оптимізації 3D-CNN шляхом розрахунку відстані між відео як вартості транспортування. Тут «відстань» можна розуміти як вираз подібності. Дуже схожі сегменти відео матимуть низькі значення відстані, і навпаки. Це оптимальна транспортна структура, прийнята у CMOTi;

б) рекурентні нейронні мережі: RNN – це тип моделі глибокого навчання, яка зазвичай використовується для обробки послідовних даних (див. рис.2.1). Його можна застосувати до завдань розпізнавання відео, розглядаючи кожен кадр відео як часовий крок у послідовності та обробляючи кадри послідовно.

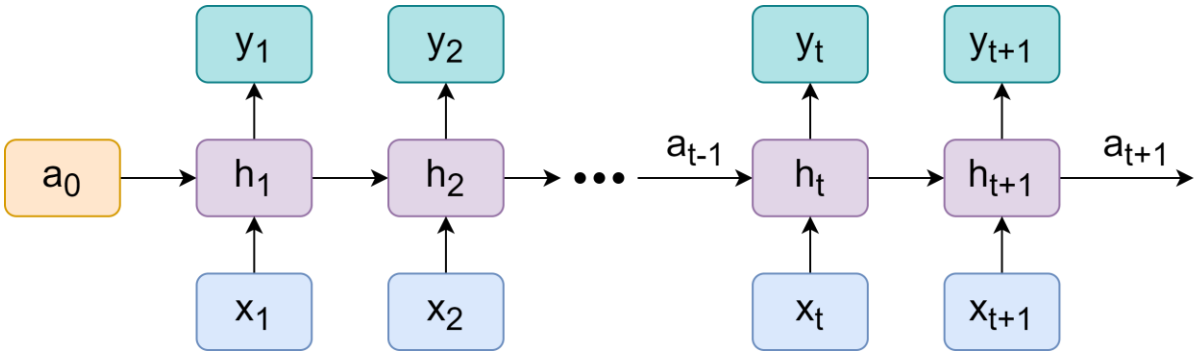

Рисунок 2.1 – Базова структура RNN

Вихідні дані RNN на кожному тимчасовому кроці зазвичай є набором функцій, які фіксують просторову та тимчасову інформацію в поточному кадрі. Ці функції можна комбінувати по кадрах за допомогою операції об'єднання або окремого класифікатора для прогнозування відео, наприклад розпізнавання конкретних дій або дій.

Мережа просторово-часової уваги та семантичних графів (англ. Spatio-temporal attention and semantic graph network) – це архітектура RNN, призначена для вирішення проблеми розпізнавання групової активності. Розпізнавання групової діяльності фокусується на визнанні дій, які виконуються кількома людьми у групі. Мета завдання – автоматично виявляти та розпізнавати дії кожної людини на відео, а також групову динаміку та взаємодії, що відбуваються під час активності. StagNet виводить окремі дії та їх просторові відносини та представляє їх у вигляді явного семантичного графа. Тимчасові взаємодії інтегруються за допомогою структурної моделі RNN. Поверх нього вбудований просторово-часовий механізм уваги, що дозволяє надавати різні рівні важливості різним людям/кадрам у відеопослідовності. Тобто дії деяких об'єктів на відео більш помітні, наприклад, зміни відбуваються швидше у просторі чи часі.

StagNet складається з дворівневої RNN, яка об'єднує два типи модулів RNN (тобто nodeRNN та EdgeRNN) у свою структуру, яка навчається наскрізно. Зокрема, перша частина полягає у побудові семантичного графа із вхідних кадрів, а потім додається тимчасовий фактор з використанням структурної RNN. Висновок досягається за допомогою механізмів передачі повідомлень і спільного використання факторів. У графічних моделях вузли графа є випадковими величинами, а ребра становлять залежності між ними. Основна ідея передачі повідомлень полягає у поширенні інформації між вузлами графа на основі даних, що спостерігаються, і припущень моделі. Передача повідомлень включає передачу розподілу ймовірностей між сусідніми вузлами графа на основі умовних ймовірностей, визначених моделлю. Це дозволяє вузлам оновлювати свої уявлення про змінні на основі даних їхніх сусідів.

Спільне використання факторів – це споріднена концепція, яка передбачає використання одного і того ж розподілу ймовірностей для представлення кількох змінних моделей. Завдяки спільному використанню факторів у змінних модель може ефективніше фіксувати кореляції між змінними, скорочуючи кількість параметрів, які необхідно оцінити. Також StagNet використовує просторово-часовий механізм уваги для виявлення ключових людей та кадрів для подальшого підвищення продуктивності. Тобто механізм уваги ідентифікує кадри, які більш важливі для відеопослідовності (кадри, в яких відбувається фактична дія в довгому відео), а також рамки, що обмежують, що представляють об'єкти, які роблять ці кадри такими важливими (реальні люди виконують дії);

Сіамські мережі (англ. Siamese Networks) зазвичай складаються з двох ідентичних глибоких нейронних мереж, які мають спільні ваги та навчені виводити подібні характеристики для пар схожих відео та різні функції для пар різнорідних відео.

У разі розпізнавання відео сіамські мережі часто використовуються для завдань, пов'язаних з виявленням подібностей або відмінностей між парами відео, таких як пошук відео, пошук подібності відео та перевірка відео. У цих

задачах Сіамська мережа приймає як вхідні дані пари відеокліпів і виводить оцінку подібності, яка показує, наскільки схожі два кліпи.

Прикладом такої мережі є COSNet або сіамська мережа CO-attention, розроблена для неконтрольованої сегментації відеооб'єктів. На етапі навчання COSNet приймає пару кадрів з одного і того ж відео як вхідні дані і вчиться фіксувати їх багаті кореляції. Це досягається за рахунок диференційованого, закритого механізму спільної уваги (методу виявлення взаємозалежностей між двома кадрами), який дозволяє мережі звертатися до подібних компонентів, визначати і диференціювати функції, що не збігаються. Під час тестування COSNet визначає основну мету у глобальному масштабі. Іншими словами, він використовує інформацію спільної уваги (взаємозалежні відносини, змодельовані у спільному просторі впровадження, що фіксує контекстну інформацію) між тестовим кадром та кількома опорними кадрами. COSNet пропонує уніфіковану, комплексну структуру, що навчається, яка ефективно отримує багату контекстну інформацію у відеопослідовності.

Оскільки відеодані широко доступні, їхня покадрова обробка займає винятково багато часу, в результаті чого RNN досягають своєї межі. Трансформери – це клас глибоких мереж, які не обробляють вхідні дані послідовно, а натомість обробляють усі кадри паралельно, що робить їх швидше та ефективніше. Це досягається за рахунок використання механізмів самообслуговування, які дозволяють перетворювачу вибірково фокусуватися на різних частинах відеопослідовності та звертати увагу на найважливішу інформацію.

Відеотрансформери останнім часом стали потужними інструментами для розуміння відео. Прикладом такої мережі є модель відеотрансформерів об'єктно-області (англ Object-Region Video Transformers) для розуміння відео за допомогою відстеження об'єктів. Основна мета OrViT – явно поєднати об'єктно-орієнтовані уявлення з просторово-часовими уявленнями архітектур відеотрансформерів і зробити це на всіх рівнях моделі, починаючи з більш ранніх шарів.

Об'єктно-орієнтовані уявлення фокусуються на ідентифікації та локалізації окремих об'єктів у зображенні або відео (низькорівневе уявлення), тоді як просторово-часові уявлення є функціями високого рівня, призначені для таких завдань, як розпізнавання дій. Таким чином, об'єднання об'єктно-орієнтованих уявлень робить просторово-часові характеристики багатшими. OrViT досягає цього, адаптуючи блок самообслуговування для включення інформації про об'єкт. У блоці самообслуговування кожна вхідна функція використовується для обчислення оцінок уваги (що представляють важливість функцій для конкретної задачі) стосовно кожної іншої вхідної функції, що дозволяє моделі фіксувати довгострокові залежності та взаємодії між функціями.

OrViT приймає як вхідні дані обмежувальні рамки та токени виправлень (також звані просторово-часовими уявленнями) і виводить уточнені токени виправлень на основі інформації про об'єкт. Токени виправлень можна як спосіб розбити зображення більш дрібні, більш керовані частини, що дозволяє моделі обробляти зображення ефективніше і результативно. Представляючи кожен патч як вивчений вектор ознак модель здатна фіксувати візуальні особливості зображення як низького, так і високого рівня, а також розмірковувати про взаємозв'язки між різними патчами. Усередині блоку інформація обробляється двома окремими потоками об'єктного рівня: потоком «Увага до області об'єкта», що моделює зовнішній вигляд, та потоком «Модуль об'єктної динаміки», що моделює траєкторії. Потік появи спочатку витягує дескриптори кожного об'єкта з урахуванням координат об'єкта і токенів виправлень. Потім дескриптори об'єктів додаються до токенів виправлень, і до всіх цих токенів спільно застосовується самовладання, таким чином включаючи інформацію про об'єкт у токени виправлень.

Потік траєкторії використовує лише координати об'єкта для моделювання геометрії руху та обробляє їх самостійно. Нарешті, обидва потоки реінтегруються в набір удосконалених токенів виправлень, які мають ту ж

розмірність, що й вхідні дані блоку ORViT, що дозволяє викликати багаторазово блок. Огляд архітектури ORViT показано вище. ORViT досяг найсучасніших результатів у вирішенні завдань розпізнавання композиційних та малокадрових дій, а також задач виявлення просторово-часових дій. Вивчений модуль уваги до області об'єкта візуально представлений нижче [34].

Технологія розпізнавання відео широко поширена в сучасну епоху. До найефективніших застосувань цієї технології можна віднести:

– охорона та спостереження;

– автономне водіння;

– аналітика поведінки клієнтів в роздрібній торгівлі;

– моніторинг руху.

Підсумовуючи, розпізнавання відео – це галузь, що швидко розвивається, і має численні застосування в різних галузях. Завдяки швидкому прогресу в техніці глибокого навчання та наявності великомасштабних наборів даних розпізнавання відео досягло значних успіхів за останні роки, досягнувши продуктивності людського або навіть надлюдського рівня в певних завданнях. Під час автономного водіння розпізнавання відео відіграє вирішальну роль у забезпеченні безпечної та ефективної навігації транспортних засобів і взаємодії з навколишнім середовищем. У моніторингу дорожнього руху розпізнавання відео забезпечує аналіз у реальному часі та розуміння транспортного потоку, покращуючи безпеку та ефективність дорожнього руху. Хоча все ще є труднощі, які потрібно подолати, наприклад мінливість і непередбачуваність даних, методи глибокого навчання продемонстрували великий потенціал у вирішенні цих проблем і покращенні точності та ефективності систем розпізнавання відео. Оскільки розпізнавання відео продовжує розвиватися, воно має потенціал для революції в багатьох галузях, від транспорту до безпеки та розваг, покращуючи наше повсякденне життя незліченною кількістю способів.

2.3 Тонке настроювання як підхід передавального навчання

У галузі штучного інтелекту та машинного навчання, що швидко розвивається, передавальне навчання та тонке настроювання стали потужними методами для прискорення розробки моделей і досягнення чудової продуктивності.

Глибоке навчання показало надзвичайний успіх у багатьох задачах комп'ютерного зору, але сучасні методи часто покладаються на великі обсяги позначених навчальних даних [35, 36, 37]. Передавальне навчання, де метою є передача знань із пов'язаного вихідного завдання, зазвичай використовується для компенсації відсутності достатніх навчальних даних у цільовому завданні [38, 39]. Тонке настроювання, мабуть, є найбільш широко використовуваним підходом для передавального навчання при роботі з моделями глибокого навчання. Він починається з попередньо навченої моделі на вихідному завданні та навчає її далі на цільовому завданні. Для завдань комп'ютерного зору звичайною практикою є робота з попередньо підготовленими моделями ImageNet для точного налаштування [40]. Порівняно з навчанням з нуля, точне налаштування попередньо навченої згорткової нейронної мережі на цільовому наборі даних може значно підвищити продуктивність, одночасно зменшуючи вимоги до цільових позначених даних [35, 40, 41, 42].

Передавальне навчання передбачає використання знань, отриманих в одному завданні чи домені, для покращення навчання в іншому пов'язаному завданні чи домені [43]. Замість навчання моделі з нуля попередньо навчена модель використовується як відправна точка. Захоплюючи загальні закономірності та характеристики з величезного набору даних, попередньо навчена модель діє як база знань, забезпечуючи міцну основу для ефективного вирішення нових схожих проблем.

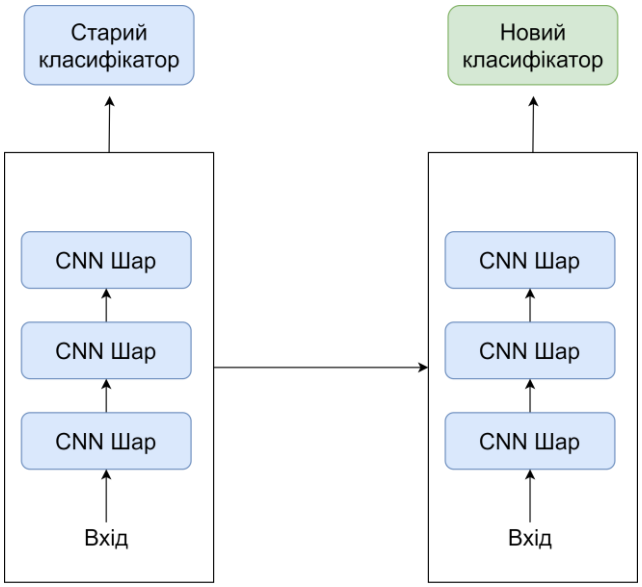

Рисунок 2.2 – Схема тонкого настроювання адаптація попередньо підготовлених моделей

Тонке настроювання – це процес взяття попередньо навченої моделі та її адаптації до нового завдання шляхом подальшого навчання на спеціальному наборі даних для завдання (рис 2.2.). Мета полягає в тому, щоб удосконалити вивчені уявлення попередньо підготовленої моделі, дозволяючи їй краще відповідати конкретним нюансам і характеристикам цільового завдання. Під час тонкого ннастроювання вагові коефіцієнти попередньо навченої моделі коригуються на основі нового набору даних, дозволяючи їй вивчати шаблони, що стосуються конкретного завдання, зберігаючи при цьому загальні знання [44].

Існує кілька варіантів реалізації ідеї тонкого настроювання глибоких мереж на практиці. Природний підхід оптимізувати всі параметри глибокої мережі з використанням цільових даних навчання (після їх ініціалізації параметрами попередньо навченої моделі). Однак, якщо цільовий набір даних невеликий, а кількість параметрів величезна, точне налаштування всієї мережі може призвести до перенавчання [45]. В якості альтернативи останні кілька шарів глибокої мережі можуть бути точно налаштовані, заморожуючи параметри інших початкових шарів до їх попередньо навчених значень [46, 47]. Це пов'язано з поєднанням обмежених даних навчання цільової задачі та

емпіричних даних у тому, що початкові рівні вивчають низькорівневі функції, які можна безпосередньо використовувати у різних завданнях комп'ютерного зору. Однак кількість початкових шарів, які необхідно заморозити під час тонкого налаштування, залишається вибором вручну, оптимізація якого може виявитися неефективною, особливо для мереж з сотнями або тисячами шарів. Крім того, емпірично було помічено, що сучасні успішні багатоколійні глибокі архітектури, такі як ResNets [36], поводяться як ансамблі дрібних мереж [48]. Незрозуміло, чи є обмеження точного налаштування останніми суміжними шарами найкращим варіантом, оскільки ефект ансамблю зменшує припущення про те, що ранні або середні рівні повинні використовуватися спільно із загальними функціями низького або середнього рівня.

Поточні методи також використовують глобальну стратегію точного налаштування, тобто одне й те саме рішення про те, які параметри заморозити, а які точно налаштувати, приймається для всіх прикладів у цільовому завданні. Передбачається, що таке рішення є оптимальним для всього розподілу цільових даних, що може бути невірним, особливо у разі недостатності цільових навчальних даних. Наприклад, певні класи в цільовій задачі можуть мати більшу схожість з вихідним завданням, і маршрутизація цих цільових прикладів через початкові попередньо навчені параметри (під час виведення) може бути кращим вибором з точки зору точності [42].

Можна виділити такі сприятливі причини використання передавального навчання:

а) обмеженість даних: трансферне навчання корисне, коли доступний набір даних для цільового завдання невеликий. Попередньо навчені моделі можуть використовувати знання з великих та різноманітних наборів даних для гарного узагальнення нових завдань з обмеженими даними;

б) ефективність навчання. Навчання моделей з нуля може зайняти багато часу та витрат у обчислювальному відношенні. Трансферне навчання дозволяє використовувати попередньо навчені моделі, скорочуючи час навчання та вимоги до ресурсів;

г) підвищення продуктивності. Попередньо навчені моделі вже засвоїли корисні функції та закономірності з великих наборів даних. Точне налаштування цих моделей на даних для конкретних завдань допомагає досягти вищої продуктивності порівняно з навчанням із нуля;

ґ) адаптація предметної галузі. Трансферне навчання особливо корисне, коли вихідна предметна область (попередня підготовка) та цільова область (тонка настройка) мають деякі подібності. Це дозволяє моделям адаптуватися та добре працювати в нових галузях.

Використання передавального навчання буде доречним при: нестачі даних (якщо у вас невеликий набір даних для цільової задачі, трансферне навчання може забезпечити значний приріст продуктивності за рахунок використання попередньо вивчених моделей), схожих завданнях (передавальне навчання працює найкраще, коли вихідне та цільове завдання пов'язані. Якщо завдання мають загальні характеристики чи шаблони, попередньо навчені моделі можуть ефективно передавати знання) та при обмеженні за часом та ресурсами (в такому випадку, використання передавального навчання дозволяє отримати вигоду з вивчених уявлень попередньо навченої моделі та знизити навантаження на навчання).

Підходи та кроки до використання трансферного навчання:

а) вибір попередньо навченої моделі. Вибір заздалегідь навченої моделі, яка відповідає проблемній області та задачі. Треба зважати на такі фактори, як архітектура (наприклад, VGG, ResNet, BERT) та набір даних, на якому була попередньо навчена модель;

б) заморозити початкові шари: заморозити початкові шари попередньо навченої моделі, щоб зберегти вивчені уявлення. Ці рівні фіксують загальні функції, які, ймовірно, можна застосувати до нового завдання;

в) замінити або додати шари, специфічні для завдання. Потрібно змінити архітектуру попередньо навченої моделі відповідно до конкретних вимог завдання. Є можливість замінити остаточні шари класифікації або додати нові шари поверх попередньо вивченої моделі;

г) підготовити набір даних. Підготовка набору даних для конкретного завдання, організувавши його у відповідні піднабори для навчання, перевірки та тестування. Потрібно переконатися, що набір даних позначений і сумісний із вхідним форматом, який очікувано попередньо вивченою моделлю;

ґ) навчання та точне налаштування. Спочатку потрібно навчити модифіковану модель за допомогою заморожених шарів, використовуючи набір даних для конкретного завдання. Цей крок дозволяє доданим шарам адаптуватися до нового завдання, зберігаючи при цьому підготовлені знання. Згодом виконується точне налаштування всієї моделі, розблокувавши попередньо навчені шари та продовживши навчання на наборі даних для конкретного завдання;

д) оцінка та ітерація. Наступним кроком треба оцінити продуктивність точно налаштованої моделі на наборі перевірок. У разі потреби виконати ітерацію та додаткове налаштування, коригуючи гіперпараметри або змінюючи архітектуру;

е) тестування та розгортання. Якщо ви задоволені продуктивністю моделі, оцініть її на окремому наборі тестових даних, щоб оцінити її здатність до узагальнення. Нарешті, розгорніть модель для прогнозування нових, невідомих даних.

Далі розглянемо переваги та недоліки тонкого настроювання як підходу передавального навчання. До переваг можна віднести:

– скорочення часу навчання та вимог до ресурсів. Трансферне навчання усуває необхідність навчання моделей з нуля, заощаджуючи значні обчислювальні ресурси та час;

– покращена продуктивність за обмежених даних. Попередньо навчені моделі, навчені великих наборах даних, фіксують загальні закономірності. Точне налаштування дозволяє цим моделям адаптуватися до конкретних завдань навіть за обмеженої кількості розмічених даних, що призводить до підвищення продуктивності;

– узагальнення та переносимість. Трансферне навчання дозволяє моделям добре узагальнювати пов'язані завдання або області, використовуючи отримані знання з одного завдання для іншого;

– доступна сучасна продуктивність. Попередньо навчені моделі, у тому числі випущені дослідницькою спільнотою, забезпечують доступ до найсучаснішої продуктивності, не вимагаючи великих знань чи обчислювальних ресурсів.

Недоліками ж будуть:

– невідповідність домену. Попередньо навчені моделі не завжди можуть ідеально узгоджуватися з цільовим завданням або доменом, що потенційно може призвести до неоптимальної продуктивності, якщо значні відмінності;

– переоснащення: точне налаштування невеликого набору даних для конкретного завдання може збільшити ризик перенавчання, коли модель не може узагальнити далеко за межі навчальних даних;

– обмежена інтерпретованість. Попередньо навчені моделі можуть бути складними і не піддаються інтерпретації через їх розмір і глибину, що ускладнює розуміння та налагодження їхньої внутрішньої роботи [44].

Трансферне навчання та тонке налаштування стали незамінними інструментами для практиків машинного навчання, пропонуючи численні переваги та дозволяючи прориви в різних сферах. Використовуючи потужність попередньо підготовлених моделей, дослідники можуть використовувати величезні обсяги знань і досягати найсучаснішої продуктивності зі скороченим часом навчання та ресурсами. У зв’язку з постійним прогресом у галузі, трансферне навчання та тонке налаштування продовжують розширювати межі того, що можливо в машинному навчанні та штучному інтелекті, штовхаючи нас до більш розумних та ефективних систем.

2.4 Архітектура нейронної мережі SlowFast

У сучасній галузі глибокого навчання та комп'ютерного зору, архітектури нейронних мереж відіграють важливу роль у забезпеченні ефективного аналізу та обробки відеоданих. Однією з інноваційних архітектур, що заслуговує на увагу, є повільний потік (англ. SlowFast). Ця архітектура була розроблена для вирішення складних завдань відеорозпізнавання, де потрібна як висока просторова, так і тимчасова роздільна здатність. Два потоки, повільний (Slow) та швидкий (Fast), інтегруються в єдину структуру, забезпечуючи баланс між точністю та ефективністю в обробці відеопотоку.

Мережі SlowFast можна описати як однопотокову архітектуру, яка працює з двома різними частотами кадрів, але ми використовуємо концепцію шляхів, щоб відобразити аналогію з біологічними Parvo- та Magnocellular. Наша загальна архітектура має повільний шлях і швидкий шлях , які об’єднані бічними з’єднаннями з мережею SlowFast. Таблиця 2.1 ілюструє таку концепцію.

Таблиця 2.1 – Приклад екземпляра мережі SlowFast

Етап	Повільний потік	Швидкий потік	Вихідний розмір $T \times S^2$
Необроблений кліп	–	–	64×224^2
Шар даних	крок $16,1^2$	крок $2,1^2$	$Slow: 4 \times 224^2$ $Fast: 32 \times 224^2$
$conv_1$	$1 \times 7^2, 64$ *крок* $1,2^2$	$5 \times 7^2, 8$ *крок* $1,2^2$	$Slow: 4 \times 122^2$ $Fast: 32 \times 122^2$
$pool_1$	1×3^2 max *крок* $1,2^2$	1×3^2 max *крок* $1,2^2$	$Slow: 4 \times 56^2$ $Fast: 32 \times 56^2$

res_2	$\begin{bmatrix} 3\times1^2, 64 \\ 1\times3^2, 64 \\ 1\times1^2, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 3\times1^2, 8 \\ 1\times3^2, 8 \\ 1\times1^2, 32 \end{bmatrix}\times3$	$Slow: 4\times56^2$ $Fast: 32\times56^2$
res_3	$\begin{bmatrix} 3\times1^2, 128 \\ 1\times3^2, 128 \\ 1\times1^2, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 3\times1^2, 16 \\ 1\times3^2, 16 \\ 1\times1^2, 64 \end{bmatrix}\times3$	$Slow: 4\times28^2$ $Fast: 32\times28^2$
res_4	$\begin{bmatrix} 3\times1^2, 256 \\ 1\times3^2, 256 \\ 1\times1^2, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 3\times1^2, 32 \\ 1\times3^2, 32 \\ 1\times1^2, 128 \end{bmatrix}\times6$	$Slow: 4\times14^2$ $Fast: 32\times14^2$
res_5	$\begin{bmatrix} 3\times1^2, 512 \\ 1\times3^2, 512 \\ 1\times1^2, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 3\times1^2, 64 \\ 1\times3^2, 64 \\ 1\times1^2, 256 \end{bmatrix}\times3$	$Slow: 4\times7^2$ $Fast: 32\times7^2$
глобальний середній пул, concat, fc			#classes

Зауважимо що розміри ядер позначаються як $\{T\times S^2, C\}$ для тимчасових, просторові та розміри каналів. Кроки позначаються як $\{$*часовий крок*, *просторовий крок*$^2\}$. Тут коефіцієнт швидкості $\alpha=8$, а коефіцієнт каналу $\beta=1/8$, $\tau=16$. Зелені кольори позначають вищу тимчасову роздільну здатність, а червоні кольори позначають менше каналів для швидкого шляху. Невироджені часові фільтри підкреслені. Залишкові блоки показані дужками. Основою є ResNet-50.

Повільний потік охоплює широкий контекст і фокусується на просторовій інформації, що дозволяє мережі краще розуміти зміст кожного кадру. Він може бути будь-якою згортковою моделлю, яка працює з кліпом відео як просторово-часовим обсягом. Ключовою концепцією нашого повільного шляху є великий часовий крок τ на вхідних кадрах, тобто він обробляє лише один із τ кадрів. Типове значення τ, становить 16 – ця швидкість оновлення становить приблизно 2 кадри в секунду для відео зі швидкістю 30 кадрів/с. Позначаючи

кількість кадрів, відібраних повільним шляхом, як T, необроблена довжина кліпу становить $T \times \tau$ кадрів. Паралельно з повільним шляхом, швидкий шлях є іншою згортковою моделлю з наступними властивостями:

а) висока частота кадрі. Мета полягає в тому, щоб мати точне представлення вздовж тимчасового виміру. Наш швидкий шлях працює з невеликим часовим кроком τ/α, де $\alpha > 1$ – це співвідношення частоти кадрів між швидким і повільним шляхами. Обидва шляхи працюють на одному необробленому кліпі, тому швидкий шлях відбирає кадри αT, у α разів щільніші, ніж повільний шлях. Типове значення $\alpha = 8$ у наших експериментах. Наявність α є ключем до концепції SlowFast (таблиця 2.1, вісь часу). Це чітко вказує на те, що два шляхи працюють на різних часових швидкостях, і, таким чином, керує експертизою двох підмереж, що створюють екземпляри двох шляхів;

б) особливості високої тимчасової роздільної здатності. Швидкий шлях не тільки має високу роздільну здатність вхідного сигналу, але й використовує функції високої роздільної здатності в усій мережевій ієрархії. У наших екземплярах ми не використовуємо рівні тимчасового зменшення дискретизації (ані часове об’єднання, ані часові згортки) протягом усього швидкого шляху до глобального рівня об’єднання перед класифікацією. Таким чином, наші тензори ознак завжди мають αT-кадри вздовж часового виміру, зберігаючи часову точність, наскільки це можливо;

в) низька пропускна здатність каналу. Швидкий шлях також відрізняється від існуючих моделей тим, що він може використовувати значно меншу пропускну здатність каналу для досягнення високої точності для моделі SlowFast. Це робить його легким.

Тобто, швидкий шлях – це згортка, аналогічна повільному шляху, але має співвідношення $\beta(\beta < 1)$ каналів повільного шляху. Типове значення $\beta = 1/8$ у наших експериментах. Зауважте, що обчислення (операції з плаваючими числами або FLOP) загального рівня часто є квадратичними з точки зору коефіцієнта масштабування каналу. Саме це робить швидкий шлях більш

ефективним з точки зору обчислень, ніж повільний шлях. У наших екземплярах швидкий шлях зазвичай займає ~20% від загального обсягу обчислень. Цікаво, що дані свідчать про те, що приблизно 15-20% клітин сітківки в зоровій системі приматів є М-клітинами (які чутливі до швидкого руху, але не до кольорів або просторових деталей).

Низьку пропускну здатність каналу також можна інтерпретувати як слабшу здатність представляти просторову семантику. Технічно наш Швидкий шлях не має спеціальної обробки просторового виміру, тому його здатність до просторового моделювання має бути нижчою, ніж повільний шлях через меншу кількість каналів. Хороші результати нашої моделі свідчать про те, що бажаним компромісом для швидкого шляху є послаблення його здатності до просторового моделювання при одночасному посиленні його здатності до часового моделювання [47].

Отже, архітектура SlowFast є важливим кроком у розвитку глибоких нейронних мереж для відеоаналізу. Її здатність ефективно обробляти як просторову, так і тимчасову інформацію робить її досить перспективною для різних програм, включаючи детекцію, класифікацію та трекінг об'єктів у відеопотоці. На основі балансу між двома потоками, SlowFast обіцяє покращені результати на завданнях, пов'язаних з аналізом динамічних відеоданих, та надає дослідникам та практикам ефективний інструмент для вирішення складних завдань відеорозпізнавання у реальному часі.

2.5 Метрики оцінки якості класифікаційної моделі

Показники оцінювання схожі на інструменти вимірювання, які ми використовуємо, щоб зрозуміти, наскільки добре модель машинного навчання виконує свою роботу. Вони допомагають нам порівняти різні моделі та визначити, яка з них найкраще підходить для певного завдання. У світі проблем класифікації є деякі метрики, які часто використовуються, щоб побачити,

наскільки хороша модель, і дуже важливо знати, яка метрика підходить для нашої конкретної проблеми. Коли ми розуміємо деталі кожного показника, стає легше вирішити, який з них відповідає потребам нашого завдання [50].

Метрики оцінки можуть допомогти оцінити продуктивність обраної моделі, контролювати систему машинного навчання у виробництві та керувати моделлю відповідно до заданих потреб. В такому випадку метою є створення та вибір моделі, яка забезпечує високу точність даних поза вибіркою. Дуже важливо використовувати кілька метрик оцінювання для оцінки вашої моделі, оскільки модель може працювати добре, використовуючи одне вимірювання з одного показника оцінки, і може працювати погано, використовуючи інше вимірювання з іншого показника оцінки.

Класифікація полягає у передбаченні міток класів за вхідними даними. У бінарній класифікації є лише два можливих вихідних класи (тобто дихотомія). У багатокласовій класифікації може бути присутнім більше двох можливих класів [51].

Визначимо основну термінологію, пов’язану з проблемою класифікації.

– мітки основної істини: вони відносяться до фактичних міток, які відповідають кожному прикладу в нашому наборі даних. Це основа всіх оцінок і прогнози порівнюються з цими значеннями;

– прогнозовані позначки: це позначки класів, передбачені з використанням моделі машинного навчання для кожного прикладу в нашому наборі даних. Ми порівнюємо такі прогнози з мітками основної істини, використовуючи різні метрики оцінки, щоб розрахувати, чи зможе модель вивчити уявлення в наших даних.

Тепер розглянемо проблему бінарної класифікації для легшого розуміння. Маючи лише два різні класи в нашому наборі даних, порівняння базових міток істинності з прогнозованими мітками може призвести до одного з наступних чотирьох результатів, як показано на рисунку 2.3.

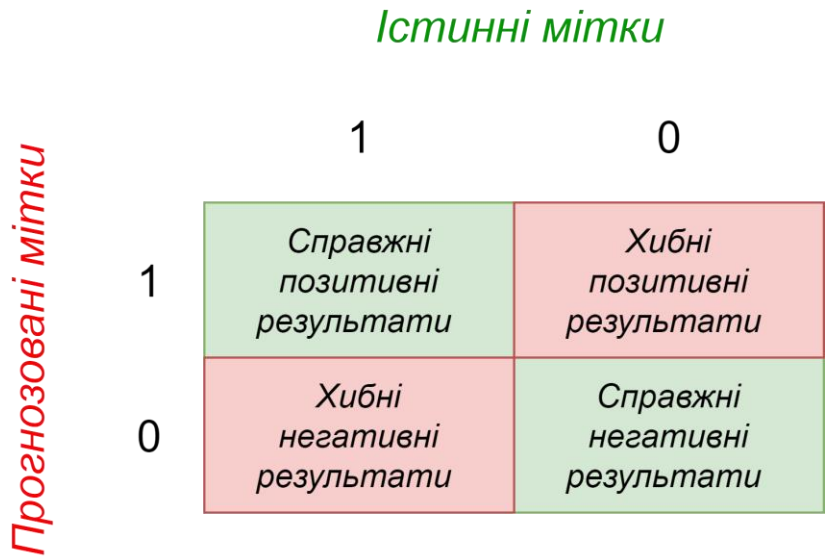

Рисунок 2.3 – Проблема бінарної класифікації

Використовуючи 1 для позначення позитивної мітки та 0 для негативної мітки, передбачення можна віднести до однієї з чотирьох категорій.

Справжні позитивні результати (англ. True Positives): модель прогнозує позитивну позначку класу, коли основна істина також позитивна. Це потрібна поведінка, оскільки модель може успішно передбачити позитивну мітку.

Хибні позитивні результати (англ. False Positives): модель прогнозує позитивну мітку класу, коли основна справжня мітка є негативною. Модель помилково ідентифікує вибірку даних як позитивну.

Хибні негативні результати (англ. False Negatives): модель прогнозує негативну відмітку класу для позитивного прикладу. Модель помилково ідентифікує вибірку даних як негативну.

Позитивні негативні результати (англ. True Negatives): також потрібна поведінка. Модель правильно ідентифікує негативний зразок, передбачаючи 0 для зразка даних, що має позначку істинності 0 [50].

До загальних показників класифікації, які використовуються для оцінки моделей входять: точність (англ. Accuracy), влучність (англ. Precision), повнота (англ. Recall), F1-оцінка (англ. F1-Score) та особливий випадок: F-оцінка з

фактором β та логістична втрата (англ Log Loss). Розберемо кожен з них по черзі:

а) Точність – це найпростіший, але інтуїтивно зрозумілий спосіб оцінки ефективності моделі для проблем класифікації. Він вимірює частку загальних міток, які модель правильно передбачила. Тому точність можна обчислити наступним чином:

$$Accuracy = \frac{TruePositives + TrueNegatives}{TruePos + FalsePos + FalseNeg + TrueNeg}.$$

Точність використовують:

1) для початкової оцінки моделі – зважаючи на свою простоту, точність є широко використовуваним показником. Це є хорошою відправною точкою для перевірки того, чи може модель добре навчатися, перш ніж використовувати метрики, специфічні для нашої проблемної області;

2) зі збалансованими наборами даних – підходить лише для збалансованих наборів даних, де всі мітки класів мають однакові пропорції. Якщо це не так, і одна мітка класу значно перевищує кількість інших, модель все одно може досягти високої точності, завжди прогнозуючи більшість класів. Показник точності однаково штрафує за неправильні прогнози для кожного класу, що робить його непридатним для незбалансованих наборів даних;

3) при неправильній класифікації витрати рівні – підходить для випадків, коли хибно-позитивні та хибно-від’ємні результати однаково погані. Наприклад, для проблеми аналізу настроїв однаково погано класифікувати негативний текст як позитивний або позитивний текст як негативний. Для таких сценаріїв хорошим показником є точність;

б) Показник влучності зосереджується на тому, щоб усі позитивні прогнози були правильними. Він вимірює, яка частка позитивних прогнозів була насправді позитивною. Математично це представляється як

$$Presicion = \frac{TruePositives}{TruePos + FalsePos}.$$

Влучність використовують:

1) при високій ціна хибних спрацьовувань – розглянемо сценарій, у якому ми навчаємо модель виявлення раку. Для нас буде важливіше, щоб ми не помилково класифікували пацієнта, у якого немає раку, тобто хибнопозитивний результат. Ми хочемо бути впевненими, коли робимо позитивний прогноз, оскільки помилкова класифікація людини як хворої на рак може призвести до непотрібного стресу та витрат. Тому ми дуже цінуємо те, що прогнозуємо позитивну мітку лише тоді, коли фактична мітка є позитивною;

2) коли якість важлива кількість – розглянемо інший сценарій, у якому ми створюємо пошукову систему, що містить запити користувачів із набором даних. У таких випадках ми цінуємо, щоб результати пошуку точно відповідали запиту користувача. Ми хочемо повернути який-небудь документ, який має відносин до користувача, тобто. хибне спрацьовування. Тому ми прогнозуємо позитивний результат тільки для документів, які точно відповідають запиту користувача. Ми цінуємо якість, а не кількість, оскільки віддаємо перевагу невеликому кількості тісно пов'язаних результатів замість більшого кількості результатів, які можуть бути релевантними або нерелевантними для користувача. Для таких сценаріїв нам потрібна висока влучність;

в) Повнота, також відоме як чутливість (англ. Sensitivity), вимірює, наскільки добре модель може запам’ятати позитивні мітки в наборі даних. Воно вимірює, яку частку позитивних міток у нашому наборі даних модель передбачає як позитивну:

$$Recall = \frac{TruePos}{TruePos + FalseNeg}.$$

Чим більше значення тим краще модель запам’ятовує, які зразки даних мають позитивні позначки.

Повнота використовують при високій вартості хибнонегативних результатів, тобто коли пропуск позитивної мітки може мати серйозні наслідки. Розглянемо сценарій, у якому ми використовуємо модель машинного навчання для виявлення шахрайства з кредитними картками. У таких випадках важливо раннє виявлення проблем. Ми не хочемо пропустити шахрайську транзакцію, оскільки це може збільшити збитки. Таким чином, ми цінуємо повноту над точністю, коли неправильну класифікацію транзакції як шахрайську можна легко перевірити, і ми можемо дозволити собі кілька хибних спрацьовувань над хибно-негативними [50].

г) F1-Оцінка. Це єдиний показник, який поєднує влучність і повноту. Чим вищий показник F1, тим краща продуктивність нашої моделі. Діапазон для оцінки F1 становить [0,1]. Оцінка F1 – це середньозважене значення точності та запам’ятовування. Класифікатор отримає високу оцінку F, лише якщо і влучність, і повнота високі. Цей показник надає перевагу лише тим класифікаторам, які мають однакову влучність і повноту.

$$F1 = \frac{2}{\frac{1}{precision} + \frac{1}{recall}};$$

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall}.$$

Оцінка F1 є узагальненим випадком загальної оцінки F. Загальна оцінка F має коефіцієнт β, який визначає, наскільки влучність/повнота впливає на оцінку:

$\beta < 1$: оцінка, орієнтована на влучність;

$\beta > 1$: оцінка, орієнтована на повноту

Оцінка F1 є узагальненим випадком, коли β дорівнює 1, що означає, що влучність і повнота збалансовані [52].

$$F_\beta Score = \frac{\left(1+\beta^2\right)\cdot Precision \cdot Recall}{\left(\beta^2 \cdot Precision\right) + Recall}.$$

F1-Оцінку використовують:

1) при незбалансованому наборі даних – на а відміну від точності, F1-Score підходить для оцінки незбалансованих наборів даних, оскільки ми оцінюємо продуктивність на основі здатності моделі згадувати менший клас, зберігаючи при цьому високу загальну влучність;

2) для знаходження компромісу між точністю та пригадуванням – обидва показники протилежні один одному. Емпірично покращення одного часто може призвести до погіршення іншого. F1-Score допомагає збалансувати обидва показники та є корисним у сценаріях, коли повнота та влучність однаково важливі. Беручи до уваги обидва показники для розрахунку, показник F1 є широко використовуваним показником для оцінки моделей класифікації;

г) Логістична втрата. Логістична втрата (або перехресна ентропійна втрата) є одним із основних показників для оцінки продуктивності проблеми класифікації.

Для окремої вибірки з істинною міткою $y \in \{0,1\}$ та оцінкою ймовірності $p = P(y=1)$ втрата дорівнює:

$$\log loss_{(N=1)} = y\log(p) + (1-y)\log(1-p).$$

Підкреслимо, що вибір відповідних метрик оцінки моделі класифікації є невід'ємним кроком у створенні поінформованих стратегій прийняття рішень. Ретельно вибрані метрики стають надійним інструментом у розумінні того,

наскільки ефективно модель справляється з поставленими завданнями. Правильно налаштовані метрики як надають об'єктивні показники точності передбачень, а й орієнтують нас у тих подальших дій. У світлі постійного вдосконалення моделей та стратегій, уміння адаптувати метрики під конкретні вимоги завдання стає ключовим елементом у прийнятті обґрунтованих рішень та подальшому покращенні результатів класифікації.

Висновки за розділом 2

Даний розділ присвячено вибору та обґрунтуванню методу розв’язання задачі розпізнавання шопліфтингу на відео. Починаючи з основних відомостей про використання нейронних мереж у комп’ютерному зорі, було детально розглядає теорію розпізнавання відео та розкрито особливості тонкого налаштування як підходу передавального навчання.

Окремий аспект розділу – архітектура нейронної мережі SlowFast. Було систематично висвітлює основні компоненти цієї архітектури, надано відомості про її принциповий принцип роботи та визначає сфери її застосування.

У завершенні розділу розглядаються метрики оцінки якості класифікаційної моделі. Вказується на важливість використання відповідних метрик та обґрунтовує їхній вибір для конкретного завдання.

У цілому, даний розділ становить цілісний огляд ключових аспектів вибору методу розв’язання задачі розпізнавання відео, починаючи від теоретичних основ до конкретних архітектур та оціночних критеріїв.

3 ПРОГРАМНА РЕАЛІЗАЦІЯ

3.1 Мова програмування Python

Мова програмування Python здобула широкий розпізнавання та популярність завдяки своїй простоті, ефективності та гнучкості. Розроблена Гвідо ван Россумом і вперше випущена у 1991 році, Python став потужним інструментом для створення різноманітних програмних застосунків, від веб-додатків до штучного інтелекту.

Однією з ключових особливостей Python є його читабельний синтаксис, який сприяє швидкому розвитку коду та підтримує зрозумілість для програмістів на різних рівнях навчання. Python використовує динамічну типізацію, що полегшує роботу змінними та сприяє підтримці великих проектів.

Мова програмування Python також славиться своєю великою екосистемою бібліотек та фреймворків, які розширюють її можливості. Бібліотеки, такі як NumPy для роботи з масивами даних, pandas для обробки та аналізу даних, і TensorFlow для розвитку моделей машинного навчання, забезпечують велику гнучкість та швидкість розробки.

Python використовується в різних галузях, включаючи розробку веб-додатків, наукові дослідження, обробку даних, ігор, а також в сфері штучного інтелекту та машинного навчання. Його універсальність робить його ідеальним вибором для різноманітних завдань у сучасному програмуванні.

Також Python є однією з найпопулярніших мов програмування у сфері машинного навчання. Бібліотеки, такі як TensorFlow, PyTorch та Scikit-learn, надають потужні інструменти для розробки та навчання моделей. Його зручний інтерфейс дозволяє дослідникам та розробникам швидко експериментувати, створювати та оптимізувати складні моделі штучного інтелекту.

Мова програмування Python грає ключову роль у сучасній обробці зображень та комп'ютерному зорі. Бібліотеки, такі як OpenCV та scikit-image,

дозволяють легко обробляти та аналізувати зображення. Велика спільнота розробників використовує Python для створення різноманітних застосунків у сферах від розпізнавання обличчя до розпізнавання об'єктів та роботи з відеопотоками.

Ці дві області використання роблять Python важливим інструментом для високотехнологічних галузей, таких як розробка штучного інтелекту та обробка зображень, і вказують на його ключову роль у розвитку новітніх технологій.

Ще важливо зазначити про Anaconda, яка представляє собою комплексну платформу та дистрибутив для роботи з обчисленнями наукових даних та машинним навчанням в середовищі Python. Його включення у себе не лише основної версії мови, але й популярних бібліотек, таких як NumPy та pandas, робить Anaconda ідеальним інструментом для розробників та дослідників. Він спрощує встановлення та управління середовищами, забезпечуючи готовий старт для роботи з великими обсягами даних.

Завдяки вбудованим засобам Anaconda, таким як Jupyter, розробники отримують зручний інтерфейс для інтерактивної роботи з кодом, візуалізації даних та проведення наукових обчислень. Jupyter - це веб-сервіс, що ідеально поєднується з Anaconda, дозволяючи створювати та редагувати документи з інтерактивним кодом та графікою. Це зробило Jupyter невід'ємною частиною наукового програмування та аналізу даних.

Для розробників Python, які шукають високопродуктивне інтегроване середовище розробки (IDE), PyCharm є ідеальним вибором. Його розширені можливості редагування коду, відлагодження та підтримка великих проектів надають розробникам зручність та продуктивність, сприяючи високоякісній розробці програм.

Усі ці інструменти - Anaconda, Jupyter та PyCharm - взаємодіють узгоджено, створюючи потужну інфраструктуру для розробки, аналізу та візуалізації даних. Такий інтегрований підхід сприяє легкості та ефективності програмістів та дослідників у світі наукового програмування.

Python здобув популярність завдяки своїй простоті, читабельності та

розширюваності. Заслуговуючи на визнання від спільноти розробників, він продовжує залишатися однією з найбільш улюблених мов програмування для різноманітних проектів. З постійними оновленнями та розвитком, Python лишається важливою складовою сучасного програмування.

3.2 Алгоритм розв’язання задачі з побудови нейронної моделі для розпізнавання шопліфтингу за допомогою нейронної мережі SlowFast

Побудуємо алгоритм розв’язання задачі розпізнавання шопліфтингу за допомогою нейромережі SlowFast. Можна виділити п’ять основних етапи алгоритму.

На першому етапі збирається вибірка для подальшого використання в навчанні моделі та її тестуванні. В нашому випадку набір відео з зафіксованими випадками шопліфтингу з камер спостереження в магазинах роздрібної торгівлі. Набір було створено власноруч, тобто знята достатня кількість відео на камерах відеоспостереження з використанням послуг акторів.

На другому етапі проводиться робота над даними:

а) відеофрагменти з кожною людиною вирізаються з необробленого відео за допомогою програмного коду, який використовує наступні алгоритми комп’ютерного зору: YOLO-NAS для виявлення людини на зображенні, DeepSort для відстеження конкретної людини на послідовності кадрів;

б) розмірність кадрів приводиться до одного значення, тобто векторне представлення кожного зображення кожного відеофрагменту набору даних становиться рівним (256, 256, 3), де 256 – висота та ширина кадру, а 3 – кольорова модель RGB;

в) вирізані відеофрагменти розбиваються на відеофрагменти меншої тривалості (3-5 секунд), маркуються по наступим класам: «крдіжка» – 1, «не крадіжка» – 0, далі відеофрагменти проходять наступну обробку: кількість кадрів

в секунду рівномірно зменшується з 15 до 5 за допомогою алгоритму UniformTemporalSubsample з фреймворку PyTorch, набір даних ділиться на тренувальну, валідаційну та тестову вибірки;

г) проводиться штучне збільшення кількості даних в тренувальному наборі даних, застосовуються такі техніки як: поворот зображення на випадковий кут від -10 до 10 градусів, горизонтальне віддзеркалення, розмивання Гауса, відкидання пікселів, адаптивне вирівнювання гістограми, еластична трансформація;

Третім етапом є точне налаштування (навчання за допомогою підготовлених даних) попередньо навченої моделі SlowFast на наборі Kinetics-400 [51].

На четвертому етапі проводиться оцінка якості моделі та налаштування гіперпараметрів.

П’ятий етап, тобто останній, є введення моделі в продукт, це включає застосування таких технік як квантування (quantization) та обрізання (pruning) задля оптимізації моделі для використання на пристроях з малою обчислювальною потужністю.

У результаті відпрацювання даного алгоритму буде побудована та підготовлена для використання модель , яка зможе якісно визначати магазинну крадіжку на відео.

3.3 Опис програми

Програма для вирішення задачі розпізнавання шопліфтингу при відеоспостереженні за допомогою SlowFast написана мовою програмування Python. Програмний код включає в себе деяку кількість .py файлів, які використовуються як модулі. Архітектура та ваги базової моделі загружаються з відкритого джерела за допомогою бібліотеки PyTorchVideo, потім ваги деяких шарів заморожуються для запобігання їх зміни при навчанні, на кінець моделі вбудовуються нові шари, такі як: лінійний, функція активації ReLU, шар Drop-

out (використовується для запобігання перенавчанню), ще один лінійний, та функція активації Sigmoid, яка виводить ймовірність класу 1 – крадіжка.

Програма містить дві основні функції: функція тренування, та функція тестування. Функція тренування має низку параметрів, таких як:

– device – пристрій, на якому модель повинна бути навчена, в нас 'cuda' для GPU;

– model – модель машинного навчання, яку слід навчати;

– criterion – функція втрат, використовувана для навчання – бінарна перехресна ентропія;

– optimizer – оптимізаційний алгоритм для оновлення ваг моделі – різновид стохастичного градієнтного спуску – Adam;

– lr_scheduler – планувальник швидкості навчання для регулювання швидкості навчання під час навчання;

– classification_dataloader_train – завантажувач даних для навчального набору;

– classification_dataloader_val – завантажувач даних для перевірочного набору;

– best_epoch – епоха, з якої повинно початися або продовжитися навчання;

– num_epoch – загальна кількість епох для навчання;

– best_val_epoch_accuracy – найкраща точність перевірки, отримана до цього моменту;

– checkpoint_dir – каталог для збереження контрольних точок моделі;

– saving_dir_experiments – каталог для збереження даних, пов'язаних з експериментом;

– logger – об’єкт журналювання для запису деталей навчання;

– epoch_start_unfreeze – епоха, з якої можна розморозити шари моделі;

– block_start_unfreeze – блок, з якого слід розморозити шари.

При вказані необхідних параметрів та запуску функції тренування відбувається ініціація процесу тренування моделі та вся потрібна інформація виводиться в процесі виконання програми.

Функція тестування має наступні параметри:

– device – пристрій, на якому модель повинна бути навчена, в нас 'cuda' для GPU;

– model – модель машинного навчання, яку слід навчати;

– classification_dataloader – завантажувач даних для тестового набору;

– path_save – шлях для збереження результатів тестування (зображення та текстовий звіт);

– class2label – словник, що відображає класи в мітки;

Отже, мета програми – отримати класифікатор, який говорив би нам, з деякою ймовірністю, де нормальна поведінка покупця, а де крадіжка.

Основний код моделі містить два етапи тренування моделі та тестування. На етапі тренування відбувається наступне:

а) встановлюється початкове значення для генератора випадкових чисел (seed) для забезпечення відтворюваності результатів;

б) ініціалізується змінна для збереження контрольної точки та інші параметри для кращого епохи та точності на валідації;

в) завантажуються шляхи до CSV-файлів навчального, валідаційного та тестового наборів даних з конфігурації;

г) створюються директорії для збереження результатів експерименту та контрольних точок моделі;

ґ) ініціалізується логер для журналювання результатів експерименту;

д) створюються завантажувачі даних для тренувального, валідаційного та тестового наборів;

е) визначається пристрій (GPU або CPU) для використання моделі під час тренування;

є) завантажується модель для тренування та встановлюється оптимізатор, критерій втрат та планувальник швидкості навчання;

ж) запускається функція тренування моделі (train_model), яка проводить тренування протягом заданої кількості епох, використовуючи навантажувачі даних та інші параметри;

з) після завершення тренування результати оновлюються, і в разі необхідності зберігається контрольна точка моделі.

На етапі тестування відбувається наступне:

а) перевіряється, чи необхідно виконати тестування;

б) шукається остання контрольна точка моделі для використання під час тестування;

в) якщо контрольна точка знайдена, вона завантажується, та модель оновлюється;

г) збираються відповідності класів на мітки для використання під час тестування;

ґ) виконується модель для тренувального, валідаційного та тестового наборів даних за допомогою зазначеної контрольної точки.

д) результати виконання зберігаються у відповідні файли (зображення та текстовий звіт);

е) обчислюється та виводиться звіт про точність та класифікацію;

є) будується та зберігається матриця плутанини у вигляді зображення;

ж) завершується завантаження результатів, якщо вказані відповідні параметри.

Для визначення якості роботи класифікатора, ми використовували такі метрики як точність, влучність, повнота, F - оцінка, матриця невідповідностей та значення бінарної перехресної ентропії як функції втрат на кожній епосі при навчанні та валідації.

Висновки за розділом 3

На основі аналізу пунктів цього розділу можна зробити наступні висновки:

– мова програмування Python є потужним інструментом для вирішення задач машинного навчання та обробки зображень;

– алгоритм розв’язання задачі розпізнавання шопліфтингу за допомогою нейромережі SlowFast є ефективним та дозволяє отримати точну модель;

– програма для реалізації алгоритму написана на мові програмування Python і містить всі необхідні функції для навчання та тестування моделі.

4 РЕЗУЛЬТАТИ ОБЧИСЛЮВАЛЬНОГО ЕКСПЕРИМЕНТУ ТА ЇХ АНАЛІЗ

4.1 Навчання моделі SlowFast та оцінка якості класифікації

При навчанні моделі було виконано один експеримент, який полягає в навчанні попередньо навченої нейронної мережі SlowFast для класифікації відео за активностями на них. Для навчання було використано знятий, проанотований та оброблений набір даних.

Для навчання моделі обирались наступні параметри при конфігурації:

– num_classes = 1 – значення, що подається як аргумент для функції активації сигмоїди, яка стоїть на виході моделі та дає змогу отримати для певного екземпляру з вибірки ймовірність того, що на відео відбувається магазинна крадіжка;

– num_epoch = 10 – кількість епох для навчання;

– frames_per_second = 5 – кількість кадрів в секунду;

– name_torchvideo_model = slowfast_r50 – назва попередньо навченої моделі SlowFast на основі ResNet:

– learning_rate = 0.001 – значення початкового темпу навчання для процесу тренування;

– scheduler_step_size = 3 – розмір кроку для планувальника темпу навчання;

– scheduler_gamma = 0.1 – коефіцієнт, на який темп навчання зменшується після кожного кроку;

– n_nodes = 128 – кількість вузлів у конкретному шарі моделі;

– dropout = 0.3 – рівень випадкового вимикання, який є технікою регуляризації для запобігання перенавчанню.

При використанні GPU з відеокарти NVIDIA RTX4090 з 24-ма гігабайтами відео пам’яті та 16384-ох ядер CUDA тривалість навчання зайняла приблизно добу.

Процес навчання зайняв 10 епох, на перших трьох значення темпу навчання було рівне 0.001, на 3-5 епохах воно змінилось до 0.0001, на 6-8 – 0.00001, на 9 – 0.000001.

На рисунку 4.1 можна побачити зміну значень тренувальної та валідаційної помилки протягом епох.

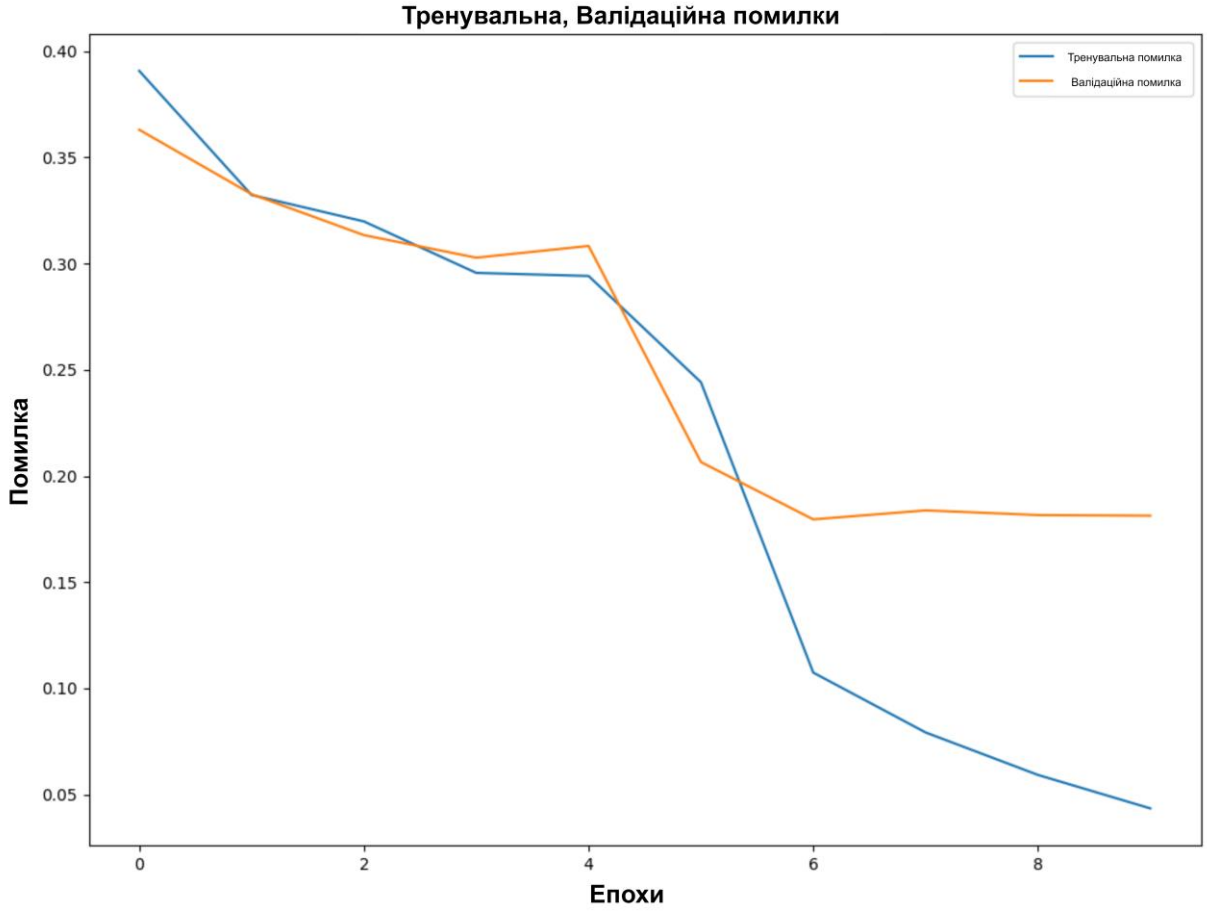

Рисунок 4.1 – Зміна значень тренувальної та валідаційної помилок протягом епох

На рисунку 4.2 можна побачити зміну значень тренувальної та валідаційної точності протягом епох.

З рисунків 4.1 та 4.2 можна зробити висновки, що найкраща модель була на епохах 5 та 6. Ваги моделі на 6-ій епосі були збережені як найкращі. Можна побачити, що з 4-ої по 5-ту епохи відбувся скачок точності та помилки, що каже знаходження моделлю якихось важливих закономірностей в даних. Також

важливо зазначити, що після 6-ої епохи модель почала перенавчатись, тобто сильно підлаштовуватись під тренувальні дані.

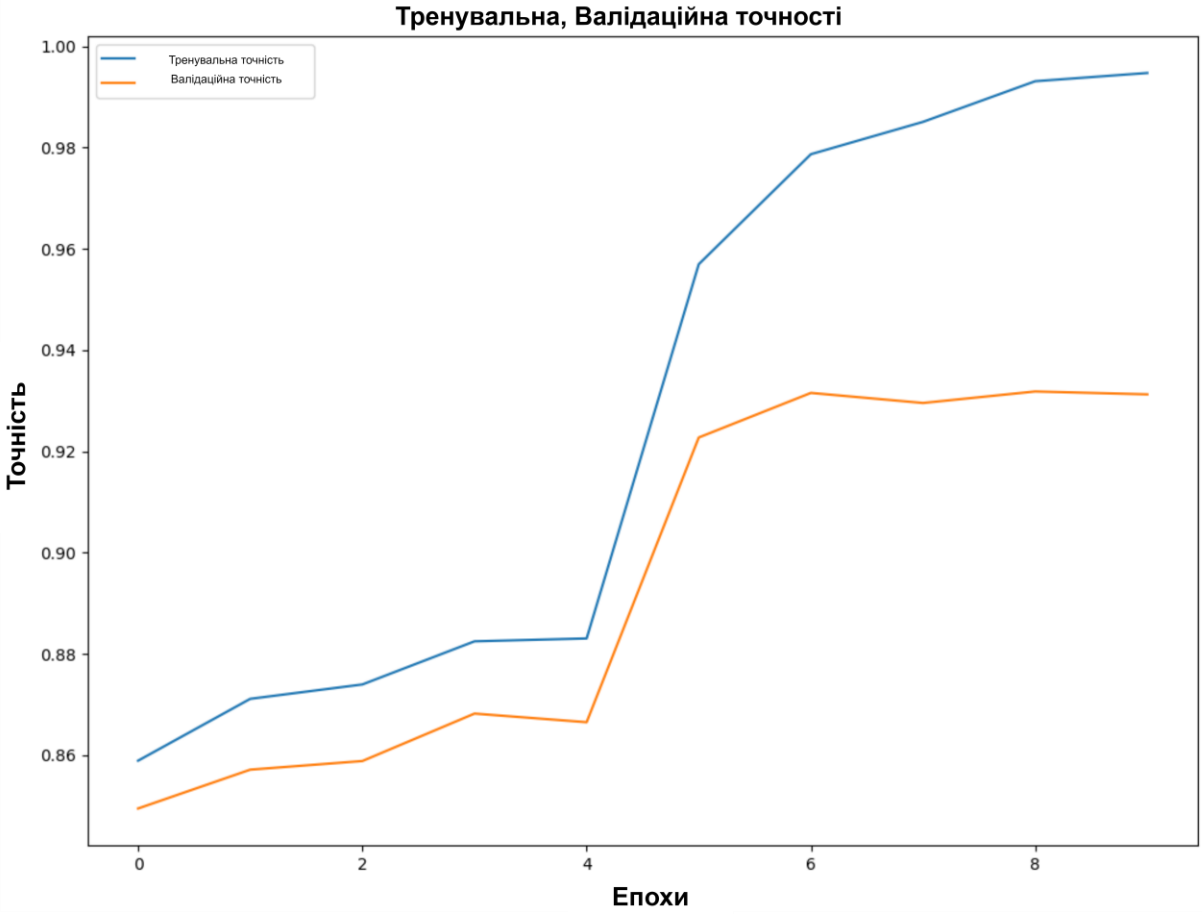

Рисунок 4.2 – Зміна значень тренувальної та валідаційної точності протягом епох

В таблицях 4.1 – 4.3 представлена інформація по метрикам на тренувальному, валідаційному та тестовому наборах даних.

З таблиці 4.1 можна сказати, що загалом модель добре справляється з тренувальним набором, маючи точність 99%. Метрики влучність та повнота обох класів високі, що свідчить про гарний баланс між правильним визначенням позитивних випадків і виявленням всіх позитивних випадків. F - оцінка, яка поєднує влучність та повноту, також висока для обох класів.

Таблиця 4.1 – Значення метрик на тренувальному наборі даних

	Влучність	Повнота	F - оцінка	Кількість екземплярів
Не шопліфтинг	1.00	0.99	0.99	7042
Шопліфтинг	0.99	1.00	0.99	7042

З таблиці 4.2 можна сказати, що результати моделі на валідаційному наборі трошки нижчі порівняно з тренувальним набором. В той час як влучність для обох класів є прийнятною, повторюваність для «Не Шопліфтингу» нижча, що свідчить про те, що модель не так ефективно виявляє всі випадки «Не Шопліфтингу». F - оцінка є в цілому прийнятним, але не таким високим, як на тренувальному наборі.

Таблиця 4.2 – Значення метрик на валідаційному наборі даних

	Влучність	Повнота	F - оцінка	Кількість екземплярів
Не шопліфтинг	0.95	0.91	0.93	1760
Шопліфтинг	0.92	0.95	0.93	1761

З таблиці 4.3 можна сказати, що результати моделі на тестовому наборі відповідають валідаційному. Влучність для обох класів задовільна, але повнота для «Не Шопліфтингу» відносно нижча. F - оцінка схожа на валідаційну, що свідчить про стабільність результатів.

Таблиця 4.3 – Значення метрик на тестовому наборі даних

	Влучність	Повнота	F - оцінка	Кількість екземплярів
Не шопліфтинг	0.96	0.90	0.93	979
Шопліфтинг	0.91	0.96	0.93	978

В таблицях 4.4 – 4.6 представлені матриці невідповідностей на тренувальному, валідаційному та тестовому наборах даних.

Таблиця 4.4 – Матриця невідповідностей на тренувальному наборі даних

Реальні значення	Не шопліфтинг	6970	72
	Шопліфтинг	25	7017
		Не шопліфтинг	Шопліфтинг
		Спрогнозовані значення	

Таблиця 4.5 – Матриця невідповідностей на валідаційному наборі даних

Реальні значення	Не шопліфтинг	1609	151
	Шопліфтинг	89	1672
		Не шопліфтинг	Шопліфтинг
		Спрогнозовані значення	

Таблиця 4.6 – Матриця невідповідностей на тестовому наборі даних

Реальні значення	Не шопліфтинг	883	96
	Шопліфтинг	39	939
		Не шопліфтинг	Шопліфтинг
		Спрогнозовані значення	

Висновки за розділом 4

В даному розділі було проведено навчання моделі SlowFast для класифікації відео за активностями, зокрема виявлення крадіжок у магазинах.

Параметри навчання включали кількість епох (10), кількість кадрів в секунду (5), назву попередньо навченої моделі SlowFast (slowfast_r50), та інші. Процес тривав приблизно добу з використанням GPU NVIDIA RTX4090.

З результатів на тренувальному, валідаційному і тестовому наборах даних можна зробити наступні висновок, що загалом, модель показала гарні результати на тренувальному наборі, але на валідаційному та тестовому є ознаки меншої ефективності, особливо у виявленні «Не Шопліфтингу». Це може свідчити про необхідність подальших покращень, наприклад, оптимізації параметрів моделі або додаткової настройки для покращення її загальної здатності до узагальнення. Загальна точність на валідаційному та тестовому наборах становить близько 93%, що є прийнятним, але не таким високим, як точність на тренувальному наборі.

ВИСНОВКИ

В ході даної кваліфікаційної роботи було досліджено метод виявлення шопліфтингу серед покупців в магазинах роздрібної за допомогою класифікатора, заснованого на попередньо навченій нейронній мережі SlowFast на основі ResNet побудованої з деякими змінами та навченої на власному наборі даних.

Було розглянуто та досліджено задачу розпізнавання відео, а саме: проведено теоретичне вивчення та пояснення методів, які використовуються для конкретної задачі, аналіз та підбір найбільш підходящого алгоритму для вирішення задачі, також виконано навчання класифікатора відео, та проведена робота з аналізу різноманітних метрик для оцінки якості роботи класифікаційної моделі нейронної мережі.

З отриманих результатів, можна зробити висновок, що даний метод відео класифікації має досить високу точність – 93%. Розроблена програма, тобто навчена модель з підібраними вагами надалі може бути використана для розпізнавання шопліфтингу в магазинах роздрібної торгівлі.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Recognition of Shoplifting Activities in CCTV Footage Using the Combined CNN-RNN Model / L. Kirichenko, O. Pichugina, B. Sydorenko, S. Yakovlev. *Progress in Polish Artificial Intelligence Research 4.* 2023. Pp. 61-66.

2. Сидоренко Б. Ю. Розпізнавання крадіжок у магазинах на записах камер відеоспостереження за допомогою комбінованої моделі CNN RNN. *27-й Міжнародний молодіжний форум «Радіоелектроніка та молодь у XXI столітті»* : зб. матеріалів форуму (м. Харків, 10-12 квітня 2023 р.). Т. 7. Харків : ХНУРЕ, 2023. С. 193–194.

3. Smith B.T. Differential Shoplifting Risks of Fast-Moving Consumer Goods. Rutgers. The State University of New Jersey, 2013. 135 p.

4. Beck A., Chapman P., Peacock C. A practical way to shrink shrinkage. *ECR Journal*. 2002, Vol. 2, № 2, P. 59–63.

5. Warr M. Fear of crime in the United States: Avenues for research and policy. *Criminal Justice 2000: Measurement and Analysis of Crime and Justice*. 2000. Vol. 4, № 4, P. 451–489.

6. Hayes R. Employee Theft Control. *Prevention Press*. 1993. 95 p.

7. Sennewald C. Shoplifters Versus Retailers: The Rights of Both. *New Century Press.* 2013. P. 1–6.

8. Caime G., Ghone G. Self Help Guide. *Trix Publishing*. 1996. P. 10–13.

9 .Hayes R., Tallman C. Shoplifting. *The Encyclopedia of Criminology and Criminal Justice.* 2014. 5 p.

10. Gill J., Johnson P., Clark. M. Research Methods for Managers. SAGE Publications. 2010. 288 p.

11. Bamfield J. European Retail Theft Barometer: Monitoring the Costs of Shrinkage and Crime for Europe's Retailers. *Nottingham: Centre for Retail Research.* 2004. Vol. 23, № 5. P. 235-241.

12. Blanco C., Grant J., Petry N. Prevalence and correlates of shoplifting in the United States: Results from the national epidemiologic survey on alcohol and related

conditions (NESARC). *American Journal of Psychiatry*. 2008. Vol,165, № 7. P. 905-913.

13. Farrington D. Measuring, explaining and preventing shoplifting: a review of British research. *Security Journal*. 1999. Vol. 12, № 1, P. 9-27.

14. Schneider J. The link between shoplifting and burglary: the booster burglar. *British Journal of Criminology*. 2005. Vol. 45, №3. P. 395-401.

15. Sutton M. Stolen Goods Markets. *US Department of Justice, Office of Community Oriented Policing Services.* 2014. P. 1-5

16. Clarke R. V. Shoplifting. Problem-Oriented Guides for Series. *Department of Justice, Office of Community Oriented Policing Police Services.* 2003. 57 p.

17. Geason S., Wilson P. Preventing Retail Crime. *Australian Institute of Criminology*. 1992. 88 p.

18. Changes in crime rates during the COVID-19 pandemic / Meyer M., Hassafy A., Lewis G., Shrestha P., Haviland A. M., Nagin, D. S. *Statistics &Public Policy*. 2022. Vol. 9, № 1, P. 97-10.

19. Cameron M (1964) The Booster and the Snitch: Department Store Shoplifiting. *Free Press of Glencoe*. 1964. 212 p.

20. Clarke, R. V. Situational Crime Prevention: Successful Case Studies. *Harrow and Heston.* 1997. 8 p.

21. McShane F., Noonan B. Classification of shoplifters by cluster analysis. *International Journal of Offender Therapy and Comparative Criminology*. 2009. Vol. 37, № 1, P. 30–40.

22. Moore R. Shoplifting in middle America: Patterns and motivational correlates. *International Journal of Offender Therapy and Comparative Criminology*. 1984. Vol. 28, № 1, P. 53–64.

23. Klemke L. The Sociology of Shoplifting. *Westport.* 1992. 175 p.

24. Hayes, R. Shop theft: An analysis of apprehended shoplifters. *Security Journal.* 1997. Vol. 7, № 1, P. 11–14.

25. Merton R. Social structure and anomie. *American Sociological Review*.1993. Vol. 3. P. 672–682.

26. Cornish D., Clarke R. Opportunities, precipitators and criminal decisions: A reply to Wortley's critique of situational crime prevention. *New York: Criminal Justice Press,* 2003. Vol. 16. P. 41–96.

27. Smith M., Cornish D. Theory for Practice in Situational Crime Prevention (Crime Prevention Studies). *Criminal Justice Press.* 2004. Vol. 16. P. 13-21.

28. Felson M., Clarke R. Opportunity Makes the Thief. *Policing and Reducing Crime Unit Research*. 1998. Vol. 98. 28 p.

29. Clarke R. V. Hot Products: Understanding, Anticipating and Reducing Demand for Stolen Goods. *Police Research Series.* 1999. Vol. 112. 2 p.

30. Hayes R. Retail Theft Trends Report. *Loss Prevention Solutions*. 1999. P. 1-6.

31. Complete Guide to Neural Network in Computer Vision. URL : https://codedamn.com/news/machine-learning/complete-guide-to-neural-network (дата звернення: 06.01.2024).

32. Deep Learning for Computer Vision. URL : https://www.run.ai/guides/deep-learning-for-computer-vision (дата звернення: 06.01.2024).

33. U-NET: Computer Vision’s neural network. URL : https://datascientest.com/en/u-net-computer-visions-neural-network (дата звернення: 06.01.2024).

34. A Practical Guide to Video Recognition [Overview and Tutorial]. URL : https://www.v7labs.com/blog/video-recognition-overview-and-tutorial (дата звернення: 06.01.2024).

35. Krizhevsky A., Sutskever I., Hinton G. E.. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems.* 2012. Vol. 25, № 2. 9 p.

36. Deep residual learning for image recognition / He K., Zhang X., Ren S., Sun J. *In CVPR*. 2016. 12 p.

37. . Densely connected convolutional networks / Huang G., Liu Z., Maaten L., Weinberger K. Q. *The IEEE on Conference on Computer Vision and Pattern Recognition.* 2017. 9 p.

38. Pan S., Yang Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*. 2010. Vol. 22, № 10. P.1345–1359.

39. Bengio Y. Deep learning of representations for unsupervised and transfer learning. *In ICML Workshop on Unsupervised and Transfer Learning*. 2012. Vol. 27. P. 17-37

40. Kornblith S., Shlens J., Le Q. V. Do better imagenet models transfer better? *The IEEE on Conference on Computer Vision and Pattern Recognition*. 2018. P. 2661-2671.

41. How transferable are features in deep neural networks? / Yosinski J., Clune J., Bengio Y., Lipson H. *In Advances in Neural Information Processing Systems*. 2014. Vol. 24. 9 p.

42. Convolutional neural networks for medical image analysis: Full training or fine tuning? / Tajbakhsh N., Shin J., Gurudu S., Hurst R., Kendall C., Gotway M., Liang J. *IEEE transactions on medical imaging*. 2016, Vol. 35, № 5. P. 1299–1312.

43. SpotTune: Transfer Learning through Adaptive Fine-tuning / Guo Y., Shi H., Kumar A., Grauman, K., Rosing, T., Feris R. *IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2018. P. 4805-4814.

44. Transfer Learning and Fine-tuning. URL : https://medium.com/@khwabkalra1/transfer-learning-and-fine-tuning-f3db7f7c6ef1 (дата звернення: 06.01.2024).

45. How transferable are features in deep neural networks? / Yosinski J., Clune J., Bengio Y., Lipson H. *In Advances in Neural Information Processing Systems*. 2014. Vol. 24. 9 p.

46. Saleh B., Elgammal A. Large-scale classification of fineart paintings: Learning the right metric on the right feature. *Open Journal of Modern Linguistics*. 2015. Vol. 8, № 4.

47. Factors of transferability for a generic convnet representation / Azizpour H., Razavian A., Sullivan J., Maki A., Carlsson S. *IEEE transactions on pattern analysis and machine intelligence*. 2016. Vol. 38, № 9. P. 1790–1802.

48. Veit A., Wilber M., Belongie S.. Residual networks behave like ensembles of relatively shallow networks. *In NeurIPS,* 2016. Vol. 2, № 3. 9 p.

49. SlowFast Networks for Video Recognition / Feichtenhofer C., Fan H., Malik J., He K. *IEEE/CVF International Conference on Computer Vision.* 2018. 4 p.

50. Understanding Classification Metrics: Your Guide to Assessing Model Accuracy. URL : https://www.kdnuggets.com/understanding-classification-metrics-your-guide-to-assessing-model-accuracy (дата звернення: 06.01.2024).

51. Metrics to Evaluate your Classification Model to take the right decisions. URL : https://www.analyticsvidhya.com/blog/2021/07/metrics-to-evaluate-your-classification-model-to-take-the-right-decisions/ (дата звернення: 06.01.2024).

52. Evaluation Metrics for Classification Models. URL : https://medium.com/analytics-vidhya/evaluation-metrics-for-classification-models-e2f0d8009d69 (дата звернення: 06.01.2024).

53. Kinetics 400 Dataset. Papers With Code. URL : https://paperswithcode.com/dataset/kinetics-400-1 (дата звернення: 07.01.2024).