

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя
(повне найменування вищого навчального закладу)
Факультет комп'ютерно-інформаційних систем і програмної інженерії
(назва факультету)
Кафедра комп'ютерних наук
(повна назва кафедри)

КВАЛІФІКАЦІЙНА РОБОТА
на здобуття освітнього ступеня

Магістр

(назва освітнього ступеня)

на тему: Методи ідентифікації паління на основі аналітичного
опрацювання візуальних даних засобами
тривимірних згорткових нейронних мереж

Виконала: студентка 6 курсу, групи СНмз-61
спеціальності

122 Комп'ютерні науки

(шифр і назва спеціальності)

Михайлів М.В.

(підпис)

(прізвище та ініціали)

Керівник

доц. Марценко С.В.

(підпис)

(прізвище та ініціали)

Нормоконтроль

доц. Мацюк О.В.

(підпис)

(прізвище та ініціали)

Завідувач кафедри

доц. Боднарчук І.О.

(підпис)

(прізвище та ініціали)

Рецензент

Бойко І.В.

(підпис)

(прізвище та ініціали)

м. Тернопіль – 2021

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії

(повна назва факультету)

Кафедра комп'ютерних наук

(повна назва кафедри)

ЗАТВЕРДЖУЮ

Завідувач кафедри

доц. Боднарчук І.О.

(підпис)

(прізвище та ініціали)

« »

20__ р.

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

на здобуття освітнього ступеня

магістр

(назва освітнього ступеня)

за спеціальністю

122 Комп'ютерні науки

(шифр і назва спеціальності)

студенту

Михайліву Миколі Васильовичу

1. Тема роботи Методи ідентифікації паління на основі аналітичного опрацювання візуальних даних засобами тривимірних згорткових нейронних мереж

Керівник роботи

Марценко Сергій Володимирович, к.т.н., доцент

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом по університету від « 28 » жовтня 2021 року № 4/7-907

2. Термін подання студентом роботи 20.12.2021

3. Вихідні дані до роботи

наукові літературні джерела

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)

1 Аналіз предметної області. 2 Теоретична частина.

3. Практична частина. 4. Охорона праці та безпека в надзвичайних ситуаціях

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)

1. Тема роботи. 2. Актуальність. 3. Мета, задачі, об'єкт, предмет дослідження.

4. Завдання розпізнавання паління. 5. Методи розпізнавання паління. 6. Традиційні методи МН та НМГН. 7. Виявлення ознак паління. 8. Попередня обробка відеопослідовності.

9. Вибір архітектури НМ. 10. Аргументи на користь ResNet. 11. Тривимірні ЗНМ

12. Навчання НМ. 13. Алгоритм процесу розпізнавання паління з відеопослідовності.

14. Експериментальні дослідження. 15. Приклади кліпів класу «паління». 16. Приклади кліпів, які не містять паління. 17. Експеримент. 18. Точність розпізнавання моделі.

19. Основні результати дослідження

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Охорона праці	Приймак М.В., професор	01.11.21	07.12.21
Безпека в НС	Клепчик В.М.	01.11.21	09.12.21

7. Дата видачі завдання _____ 2021 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Термін виконання етапів роботи	Примітка
1	Затвердження теми кваліфікаційної роботи	28.10.21	Виконано
2	Аналіз літературних джерел	29.10.21-18.11.21	Виконано
3	Обґрунтування актуальності дослідження	18.11-21.11.21	Виконано
4	Аналіз предмету дослідження та предметної області	21.11-26.11.21	Виконано
5	Проведення дослідження методів та засобів аналітичного опрацювання даних	22.11-30.11.21	Виконано
6	Оформлення розділу «Аналіз предметної області»	20.11-26.11.21	Виконано
7	Оформлення розділу «Теоретична частина»	27.11-02.12.21	Виконано
8	Оформлення розділу «Практична частина»	03.12-10.12.21	Виконано
9	Оформлення розділу «Охорона праці та безпека в надзвичайних ситуаціях»	26.11-12.12.21	Виконано
10	Нормоконтроль	11.12-15.12.21	Виконано
11	Попередній захист роботи	16.12.21	Виконано
12	Захист кваліфікаційної роботи	21.12.21	

Студент _____
(підпис)Михайлів М.В.

(прізвище та ініціали)Керівник роботи _____
(підпис)Марценко С.В.

(прізвище та ініціали)

АНОТАЦІЯ

Методи ідентифікації паління на основі аналітичного опрацювання візуальних даних засобами тривимірних згорткових нейронних мереж // Кваліфікаційна робота освітнього рівня «Магістр» // Михайлів Микола Васильович // Тернопільський національний технічний університет імені Івана Пулюя, факультет комп'ютерно-інформаційних систем та програмної інженерії, кафедра комп'ютерних наук, група СНмз–61 // Тернопіль, 2021 // С. – 71, рис. – 18, табл.– 6, слайдів – 19, додат. – 2, бібліогр. – 52.

Ключові слова: ВІДЕОРЯД, ЗГОРТКОВА НЕЙРОННА МЕРЕЖА, МАШИННЕ НАВЧАННЯ, ПАЛІННЯ, РОЗПІЗНАВАННЯ ОБРАЗІВ

Кваліфікаційна робота присвячена дослідженню методів ідентифікації паління на основі аналітичного опрацювання візуальних даних засобами тривимірних згорткових нейронних мереж.

Розглянуто процес розпізнавання образів з використанням машинного навчання як основного методу розпізнавання паління. Проведено аналіз предметної області. Виявлено ознаки паління за візуальними даними. Досліджено переваги застосування нейронних мереж глибокого навчання перед традиційними методами машинного навчання. Описано передобробку відеопослідовності. Вибрано архітектуру згорткових нейронних мереж, яка згодом перетворена на тривимірну. Описано навчання нейронної мережі та її оптимізація.

Розроблено та описано алгоритм розпізнавання паління на послідовності відео. Проведено навчання та тестування тривимірної згорткової нейронної мережі на наборі даних HMDB51. Точність розпізнавання паління за візуальними даними розрахована з застосуванням різних метрик.

ANNOTATION

Smoking identification methods based on analytical processing of visual data by means of three-dimensional convolutional neural networks // Master thesis // Mykhayliv Mykola // Ternopil Ivan Pul'uj National Technical University, Faculty of Computer Information Systems and Software Engineering, Department of Computer Science // Ternopil, 2021 // P. - 71, Fig. - 18, Table – 6, Slide - 19, References - 52.

Keywords: VIDEO SEQUENCE, CONVOLUTIONAL NEURAL NETWORK, MACHINE LEARNING, SMOKING, PATTERN RECOGNITION

This thesis deals with the research of smoking identification methods on the basis of analytical processing of visual data by means of three - dimensional convolutional neural networks.

The process of pattern recognition using machine learning as the main method of smoking recognition is considered. The analysis of the subject area is carried out. Signs of smoking were revealed according to visual data. The advantages of using neural networks of deep learning over traditional methods of machine learning are investigated. Video sequence preprocessing is described. The architecture of convolutional neural networks is chosen, which is later transformed into three-dimensional. The training of the neural network and its optimization are described.

A algorithm for smoking detection on a video sequence has been developed and described. Training and testing of a three-dimensional convolutional neural network on the HMDB51 data set were performed. The accuracy of visual recognition of smoking by visual data is calculated using different metrics.

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ СКОРОЧЕНЬ І ТЕРМІНІВ

CNN (Convolutional Neural Network) – ЗНМ (згортова нейронна мережа)

DL – Deep Learning.

FAR – False Acceptance Rate.

FRR – False Rejection Rate.

GPU – Graphics Processing Unit.

КМ – K-Means.

OpenCV – Open Source Computer Vision Library.

SVM - Support Vector Machine.

TR – True Recognition.

БД – база даних.

ВООЗ – Всесвітня організація охорони здоров'я.

КЗ – комп'ютерний зір.

МН – машинне навчання.

НМ – нейронна мережа.

НМГН – нейронна мережа глибокого навчання.

ОС – операційна система.

ПЗ – програмне забезпечення.

СГС – стохастичний градієнтний спуск.

ШІ – Штучний Інтелект.

ШНМ – Штучна нейронна мережа.

ЗМІСТ

Вступ.....	9
1 Аналіз предметної області.....	11
1.1 Постановка проблеми	11
1.2 Методи розпізнавання паління	13
1.3 МН	14
1.4 НМ	16
1.5 Традиційні методи МН та НМГН.....	22
1.6 Виявлення ознак паління.....	23
1.7 Висновки до першого розділу.....	24
2 Теоретична частина.....	26
2.1 Попередня обробка відеопослідовності.....	26
2.2 Вибір архітектури НМ.....	28
2.2.1 AlexNet.....	28
2.2.2 VGGNet.....	30
2.2.3 ResNet.....	31
2.3 Тривимірні ЗНМ.....	35
2.4 Модифікована архітектура НМ	38
2.5 Навчання НМ.....	39
2.6 Оптимізація.....	40
2.6.1 СГС	40
2.6.2 Пакетна нормалізація.....	43
2.7 Висновки до другого розділу	44
3 Практична частина	45
3.1 Алгоритм процесу розпізнавання паління з відеопослідовності	45
3.2 Експериментальні дослідження.....	51
3.2.1 Набір даних.....	51
3.2.2 Експеримент	53

3.3 Висновки до третього розділу	56
4 Охорона праці та безпека в надзвичайних ситуаціях.....	58
4.1 Режими праці і відпочинку при роботі з ЕОМ.....	58
4.2 Вплив електромагнітного імпульсу (ЕМІ) ядерного вибуху на елементи виробництва та заходи захисту.....	60
4.4 Висновки до четвертого розділу.....	64
Висновки	65
Перелік джерел.....	66
Додатки	

ВСТУП

Актуальність теми. За останнє десятиліття розпізнавання людської діяльності пройшло великий шлях і відіграло важливу роль у галузі обчислень. Така зростаюча популярність може бути пояснена безліччю реальних додатків, що застосовують розпізнавання діяльності, в основному пов'язаних із проблемами, орієнтованими на людину, такими як, охорона здоров'я та дотримання правопорядку.

Розпізнавання людських дій, а особливо такого складної дії, як паління, в реальному середовищі застосовують додатки в самих різних галузях, включаючи інтелектуальне відеоспостереження, аналіз поведінки співробітників та покупців. Тим не менш, точне розпізнавання дій є дуже складним завданням через фони, оклюзії та варіації кутів огляду. Спеціально для розпізнавання людських дій різні класи дій можуть виглядати зовсім по-різному з точки зору їх зовнішнього вигляду та моделей руху. В даний час робляться численні спроби використання методів інтелектуального аналізу даних та МН для точного виявлення дій людини у системах інтелектуального відеоспостереження.

Тому тема дослідження методів ідентифікації паління на основі аналітичного опрацювання візуальних даних є актуальною.

Мета дослідження: розпізнати процес паління тютюнових виробів за візуальними даними із застосуванням НМ.

Для досягнення вказаної мети, в роботі поставлено та розв'язано **наступні задачі:**

- виділити особливості процесу паління на послідовності відео;
- розробити алгоритм розпізнавання паління;
- програмно реалізувати розроблений алгоритм;
- провести експериментальні дослідження.

Об'єкт дослідження: процес ідентифікації паління на відеопослідовності.

Предмет дослідження: методи розпізнавання образів за відеорядом.

Наукова новизна роботи:

- проаналізовано загальні ознаки, що дозволяють розпізнавати паління на відео;
- вибрано архітектуру, яка згодом перетворена на тривимірну згорткову НМГН шляхом зміни форми вхідного об'єму для включення просторово-часової інформації та використання тривимірних ядер усередині архітектури;
- описано процес навчання та оптимізації НМ, що виконується за допомогою СГС з імпульсом та пакетної нормалізації після кожного згорткового шару;
- запропоновано алгоритм розпізнавання паління на відеопослідовності;
- проведено навчання та тестування тривимірної ЗНМ на наборі даних.

Практичне значення одержаних результатів. Впровадження розробки дозволить достатньо точно виконати візуальну ідентифікацію процесу паління на відеопослідовності.

Апробація. Окремі результати роботи були представлені на ІХ науково-технічній конференції «Інформаційні моделі, системи та технології» Тернопільського національного технічного університету імені Івана Пулюя (08-09 грудня 2021 р.) у вигляді опублікованих тез [21].

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Постановка проблеми

Про шкоду паління тютюну. Щороку через вживання тютюну та тютюнових виробів понад п'ять мільйонів людей гине – це число перевищує кількість смертей від таких серйозних захворювань, як ВІЛ чи СНІД, туберкульоз та малярія. Що ж станеться, якщо збережеться сьогоднішня тенденція – число смертельних випадків, що фіксуються щорічно, пов'язаних із вживанням тютюну перевищити вісім мільйонів людей вже до 2030 року, а загальна кількість таких випадків смертності в 21-му столітті може сягнути навіть мільярда. Такі значні цифри змушують задуматися. Безумовно, вживання тютюну є загрозою, причому однією з найбільш небезпечних і значних для здоров'я людей, коли-небудь існували у світі. За офіційними даними кожні кілька секунд вмирає одна людина. Що це спричиняє? Багато людей з раннього віку починають вживати тютюнові вироби. До того ж, половина нинішніх споживачів тютюну врешті-решт помре від будь-якої хвороби, пов'язаної із пристрастю до цього небезпечного легалізованого наркотику.

У медичній спільноті вже давно намагаються вжити заходів щодо зниження споживання тютюнової продукції. Звернемося до документів, а саме до доповіді ВООЗ 2020 р. [1]. Відстежуючи тютюнову епідемію та надаючи урядам та іншим зацікавленим сторонам інформацію, необхідну їм для підготовки своїх заходів щодо скорочення кількості осіб, котрі палять, ВООЗ закликає вжити серйозних заходів для боротьби з тютюновою залежністю. Із принципів статті 8 Рамкової конвенції ВООЗ [2] чітко зрозуміло, що безпечної величини впливу диму від паління тютюну не має і що тільки створення навколишнього середовища абсолютно позбавленого тютюнового диму, є єдиною можливим шляхом захисту населення від згубних наслідків вдихання навіть вторинного тютюнового диму. Адже щороку фіксується понад 600 000

смертельних випадків саме через вдихання вторинного тютюнового диму. За даними дослідників, у тютюновому диму міститься значна кількість токсичних хімічних речовин, з яких, як відомо, 250 шкідливих і більше 50 містять канцерогени, 6 компонентів впливають на здатність до дітонародження.

Завдання розпізнавання паління. Включає ідентифікацію різних дій з відеокліпів, тобто послідовності кадрів, де дія може виконуватися або не виконуватися протягом усієї тривалості відео. Це виглядає як природне розширення завдань класифікації зображень на кілька кадрів, а потім агрегування прогнозів кожному за кадром. Незважаючи на успіх архітектур глибокого навчання у класифікації зображень, прогрес в архітектурах для класифікації відео та навчання представленню був повільнішим.

Що зробило це завдання складним?

1. Величезні обчислювальні витрати: проста згортова двовимірна мережа для класифікації 101 класу має загалом 5 мільйонів параметрів, тоді як та ж архітектура при додаванні тимчасової компоненти та перетворення на тривимірну структуру призводить до 33 мільйонів параметрів. Тренування тривимірної ЗНМ на UCF101 займає від 3 до 4 днів, а на Sports-1M – близько двох місяців, що робить складний пошук у архітектурі трудомістким процесом.

2. Захоплення тимчасового контексту. Для покращення результатів розпізнавання дій на відеопослідовності, додають захоплення просторово-часового контексту між кадрами Крім того, отримана просторова інформація має компенсуватися рухом камери. Навіть наявності сильного просторового виявлення об'єктів недостатньо, оскільки інформація про рух також несе у собі дрібніші деталі. Існує локальний, а також глобальний контекст з інформацією про рух, котрий необхідно збирати для надійних передбачень.

3. Проектування класифікаційних архітектур. Проектування архітектур, які можуть збирати просторово-тимчасову інформацію, включає безліч варіантів, які нетривіальні та коштовні для оцінки. Наприклад, деякі можливі стратегії можуть бути: одна мережа для збирання просторово-часової

інформації або дві окремих для кожної просторової та тимчасової; злиття прогнозів між кількома кліпами; наскрізне навчання або виділення функцій та класифікації окремо.

1.2 Методи розпізнавання паління

Розпізнавання образів є основним результатом роботи алгоритмів машинного та глибокого навчання. Коли людина дивиться на фотографію або відео, вона може легко визначити людей, об'єкти, дії та візуальні деталі. Саме мета розпізнавання образів полягає у тому, щоб навчити комп'ютер видобуванню значущих ознак із зображення або відео і використовувати для того, щоб досягти високого рівня розуміння того, що містить зображення.

Сьогодні інтерес до систем, що використовують КЗ, не стає слабшим і не тільки через інтелектуальні причини, пов'язані з фундаментальними дослідженнями, а й через можливість таких автоматизованих систем різко збільшувати виробничий потенціал організацій. Як правило, найважливішим компонентом практичного візуально керованого агента є його модуль розпізнавання дій та об'єктів. Інформація про довкілля може надаватися системою КЗ, що діє як датчик зору і надає інформацію високого рівня про навколишнє середовище та роботу. ШІ та КЗ поділяють інші теми, такі як розпізнавання образів та методи навчання. Отже, КЗ іноді розглядається як частина області ШІ чи галузі комп'ютерних наук загалом.

Основні підходи у розпізнаванні образів: МН; НМГН.

Розпізнавання дій – це процес аналізу стану, переміщення та взаємодії об'єктів у послідовності зображень, таких як відео та їх класифікації в контексті навколишніх кадрів для інтерпретації чи прогнозування руху об'єкта.

НМГН дозволяють вирішувати завдання у сфері аналізу зображень і відео, особливо їх часто застосовують при вирішенні завдань, пов'язаних із класифікацією дій людини.

Система розпізнавання паління за візуальними даними дозволить здійснювати контроль за дотриманням правил техніки пожежної безпеки, а також згодом дозволить своєчасно реагувати при виявленні порушень, пов'язаних з палінням у місцях, які для цього непризначені.

1.3 МН

МН [3] - це одна з найбільш швидкозростаючих технічних областей на сьогоднішній день, що лежить на перетині інформатики та статистики, а також в основі ШІ та науки про дані.

Недавній прогрес у МН був обумовлений як розробкою нових алгоритмів, так і постійним зростанням загальнодоступності онлайн-даних та недорогих обчислень. Використання методів МН можна знайти в науці, техніці та комерції, це дозволяє приймати обґрунтовані рішення у багатьох сферах життя, включаючи охорону здоров'я [4], виробництво [5], освіту, фінансове моделювання, поліцейську діяльність та маркетинг. МН [6] має широкий спектр застосувань, включаючи робототехніку та комп'ютерні ігри, обробку природної мови, медичну діагностику, класифікацію ДНК, пошукові системи, виявлення спаму та шахрайства, а також розпізнавання мови, рукописного тексту та розпізнавання образів у КЗ.

Продуктивність простих алгоритмів МН залежить від подання даних. Наприклад, коли логістична регресія використовується для рекомендації кесаревого розтину [7], система ШІ безпосередньо не обстежує пацієнта. Натомість лікар повідомляє системі кілька частин відповідної інформації, такої як наявність чи відсутність рубця на матці. Кожна частина інформації включена в представлення пацієнта, називається функцією. Логістична регресія дізнається, як кожна з цих особливостей пацієнта співвідноситься з різними наслідками. Тим не менш, вона не може впливати на те, як функції визначаються яким-небудь чином. Якби логістичній регресії було дано МРТ-

сканування пацієнта, а не формалізований звіт лікаря, вона б не змогла дати корисні прогнози.

Залежність продуктивності від представлення даних є загальним явищем, яке проявляється в інформатиці і навіть у повсякденному житті. У галузі ІТ такі операції, як пошук у колекції даних, володіють експоненційною зростаючою швидкістю виконання, якщо колекція структурована та розумно проіндексована. Люди можуть легко виконувати арифметику на арабських цифрах, але арифметика на римських цифрах набагато складніша. Не дивно, що вибір представлення дуже впливає на продуктивність алгоритмів МН. На рис. 1.1 [8] зображено приклад різних представлень: необхідно розділити дві категорії даних, малюючи лінію між ними на діаграмі розсіювання.

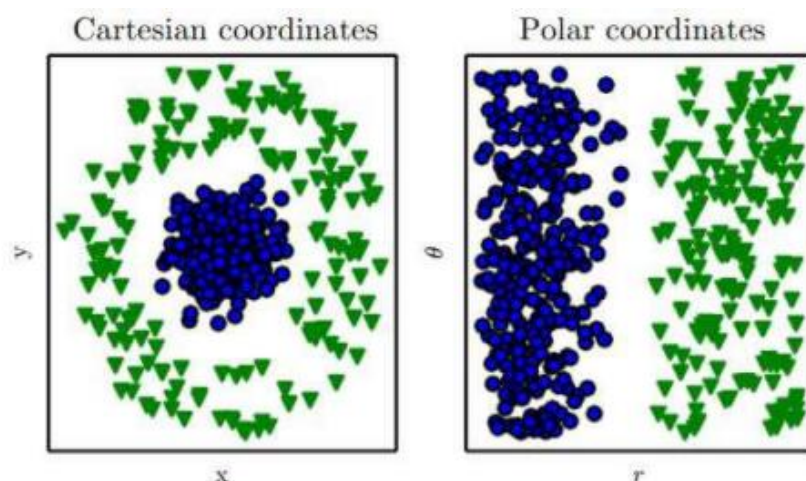


Рисунок 1.1 – Приклад різного представлення даних

На графіку зліва представлені деякі дані, при використанні декартових координат, розв'язання задачі неможливо. На графіку праворуч представлено дані з полярними координатами, і завдання стає простим для вирішення.

Одним із вирішень цієї проблеми є використання МН для виявлення не тільки відображення від подання до висновку, а й самого представлення. Цей підхід відомий як репрезентативне навчання [9, 10]. Вони також дозволяють системам ШІ швидко адаптуватися до нових завдань із мінімальним втручанням

людини. Алгоритм навчання подання може виявити гарний набір функцій для простого завдання за лічені хвилини або складного завдання за години чи місяці. Мануальне проектування функцій для складного завдання потребує великих людських зусиль та часу; воно може зайняти десятиліття для всього співтовариства дослідників [11].

Навчання - це процес, за допомогою якого модель вивчає або розпізнає закономірності даних для створення відповідних прогнозів. Крім репрезентативного, існує ще кілька видів навчання: з учителем; без вчителя; із підкріпленням.

1.4 НМ

НМ – це модель, чия багаторівнева структура є подібною до мережевої структури нейронів у людському мозку із шарами зв'язаних вузлів. Така мережа може навчатися на отриманих даних, тому її можна навчити розпізнавати шаблони, здійснювати класифікацію даних та прогнозування майбутніх подій [22].

Спостереження за поведінкою природних нейронів призвело до розкриття його основних принципів роботи та цікавих властивостей. Вперше математичний опис штучного нейрона було зроблено Маккаллохом і Піттсом в 1943 [23]. Розенблатт у 1957 році удосконалив свою ідею створення сучасного штучного нейрона, званого простим персептроном [24].

Варто навести основні компонент НМ.

Вхідні значення. Представляють різні мережеві входи, які є незалежними змінними. Кожен із цих входів множиться на вагу з'єднання чи синапс.

Вагові коефіцієнти або синапси. Зв'язок між нейронами, який має певне значення. Чим вище значення, тим більша вага та тим вищий пріоритет даного нейрона. Крім того, в математиці та програмуванні, ваги розглядаються в матричному форматі. На рис. 1.2 зображена проста НМ та відображення

першого прихованого шару на матрицю ваги 4×3 . Підсвічені ваги в кожному рядку повторюються, і лише один із них буде помножено на відповідний вхід [25]. Також біля вхідного шару є 3 нейрони, а у наступного шару (прихований шар) – 4. Далі відбувається створення матриці з 3 рядків та 4 стовпців та вставка значення кожної ваги до неї.

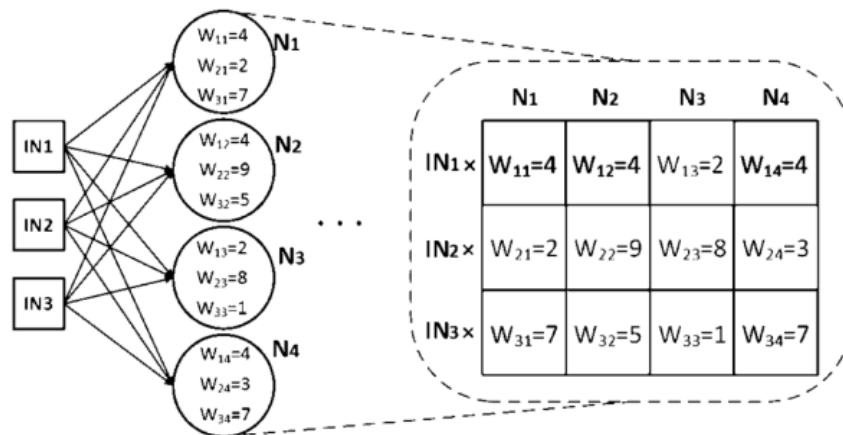


Рисунок 1.2 – Приклад простої НМ

Зміщення. Можливість зміни розташування функції через виникнення даних шумів чи важко прогнозованих параметрів. Значення зміщення дозволяє зміщувати функцію активації вгору або вниз. На рис. 1.3 представлена обчислювальна робота, що проводиться в НМ з обліком зміщень.

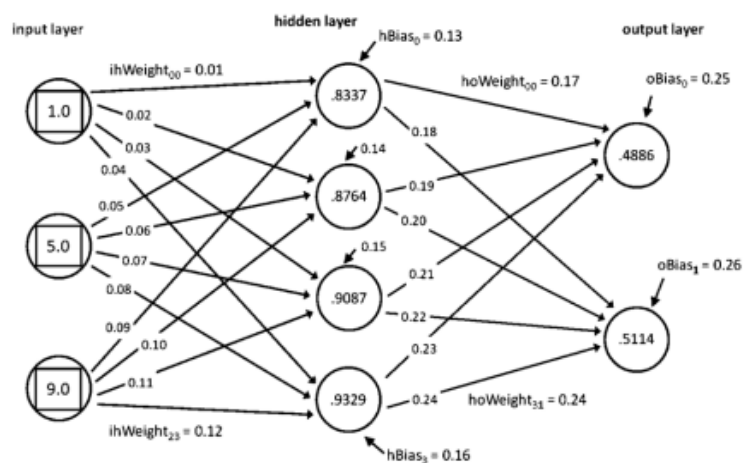


Рисунок 1.3 – Демонстрація впливу зміщень на обчислення

Функція активації. Визначає властивості штучного нейрона, важлива для ШНМ, щоб вчитися розуміти щось справді складне. Їх основна мета – перетворити вхідний сигнал вузла ШНМ у вихідний сигнал. Цей вихідний сигнал використовується як вхід для наступного шару у стеку. Функція активації вирішує, чи слід активувати нейрон, обчислюючи зважену суму та додатково додаючи до неї зсув. Сенс полягає в тому, щоб запровадити нелінійність у вихід нейрона.

Багатошарова НМ складається з великої кількості шарів (нейронів), з'єднаних разом. Шари в мережі зазвичай поділяються на три класу: вхідні шари, які одержують інформацію для обробки; вихідні шари, у яких перебувають результати обробки; шари між ними, відомі як приховані шари. ШНМ прямого зв'язку дозволяють сигналам проходити лише в один бік, від входу до виходу. По-перше, мережа навчається на наборі парних даних для визначення відображення вводу-виводу. Ваги зв'язків між нейронами фіксуються, і мережа використовується для визначення класифікацій нового набору даних.

НМГН – це модель ШН, що імітує роботу людського мозку при обробці даних та створення шаблонів для використання при прийнятті рішень. Глибоке навчання – найбільш захоплююча та потужна гілка МН. Термін «глибокий» зазвичай відноситься до прихованих шарів в НМ. Традиційні НМ містять лише 2-3 прихованих шари, в той час як глибокі сітки можуть мати до 150 шарів.

НМГН використовуються в області цифрової обробки зображень для вирішення складних завдань: зафарбовування, класифікація, сегментація, виявлення зображень та розпізнавання дій. Методи, що використовуються НМГН, здебільшого покращують продуктивність прогнозування з використанням великих даних та обчислювальних ресурсів і розширюють межі можливого. Проблеми, які вважалися нерозв'язними, тепер вирішуються з надлюдською точністю. Класифікація зображень є яскравим прикладом цього. Після того, як Крижевський, Суцкевер та Хінтон випустили свою статтю в 2012

році [26], НМГН з того часу домінують в цій галузі, завдяки кращій продуктивності у порівнянні із традиційними методами. Такі мережі, зазвичай, краще працюють з великою кількістю зображень, а власне використання графічного процесора дозволяє зменшити час, котрий потрібний для навчання моделі.

ЗНМ. Моделі глибокого навчання, такі як ЗНМ, застосовуються для автоматичного вивчення властивостей об'єкту для його ідентифікації. ЗНМ [27] є спеціалізованим видом НМ для обробки даних, що мають відому топологію, подібну до сітки. Приклади включають дані тимчасового ряду, які можна розглядати як одновимірну сітку, що бере вибірки через регулярні інтервали часу, та дані зображення, які можна розглядати як двовимірну сітку пікселів. ЗНМ мережі були надзвичайно успішні у практичному застосуванні. Власне сама назва ЗНМ говорить, що в мережі має місце математична операція, звана згорткою. ЗНМ – це просто НМ, які застосовують згортку, а не загальне множення матриць, принаймні на одному з їхніх рівнів.

Приклад роботи ЗНМ з декількома згортковими шарами, в яких фільтри застосовуються до кожного навчального зображення з різною роздільною здатністю, а вихідні дані кожного згорнутого зображення служать вхідними даними для наступного шару, продемонстровано на рис. 1.4.

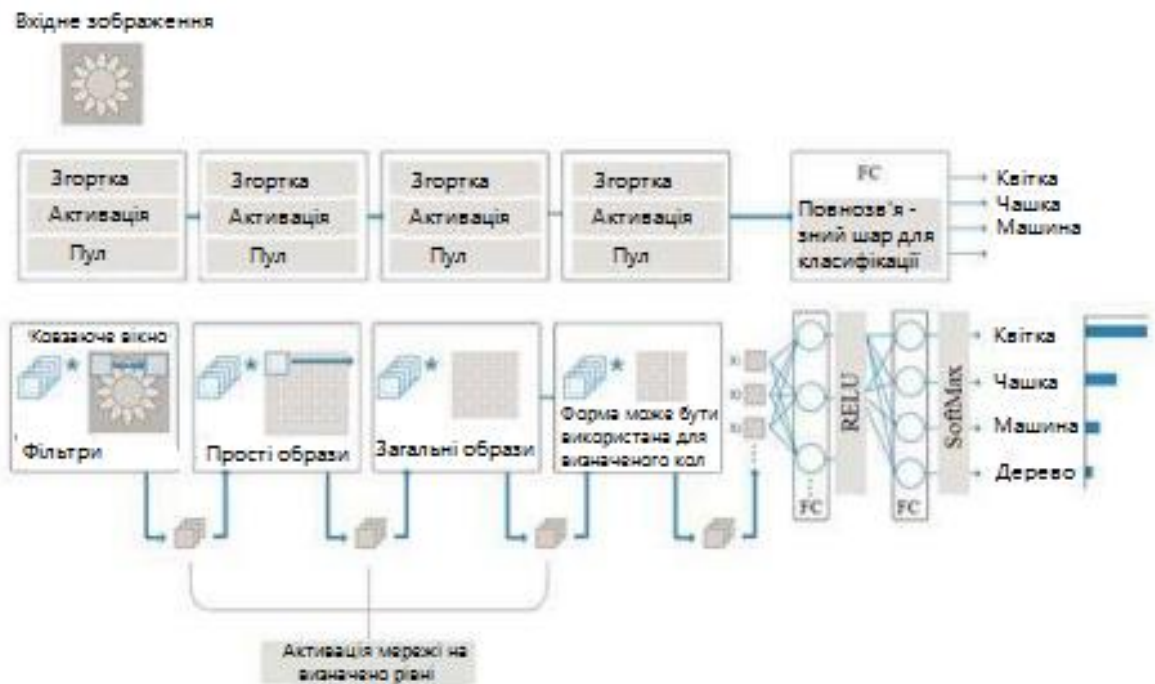


Рисунок 1.4 – Приклад мережі з кількома шарами згортки

ЗНМ виключають необхідність мануального вилучення функцій, тому не потрібно визначати функції, що використовуються для класифікації зображень. ЗНМ працює, витягуючи ознаки безпосередньо з зображень. Таке автоматичне вилучення ознак робить моделі глибокого навчання дуже точними для завдань КЗ, таких як класифікація об'єктів та розпізнавання дій.

Шари ЗНМ. Як і інші НМ, будова ЗНМ така: вхідний шар, вихідний шар та множини прихованих проміжних шарів [28]. Ці шари виконують операції, що змінюють дані з метою вивчення особливостей, які характерні для цих даних. Виділяють три найбільш поширені рівні: згортка, активація (ReLU) та субдискретизація.

Згортковий шар. У цьому шарі карти об'єктів з попередніх шарів об'єднані з ядрами, що навчаються. Вихідні дані ядер проходять через лінійну або нелінійну функцію активації, таку як, сигмоїдальна, гіперболічна дотична, Softmax, випрямлена лінійна та одинична функції, щоб сформувати вихідні карти показників. Кожна з вихідних карт об'єктів можуть бути об'єднані з більш, ніж однією вхідною картою об'єктів. Загалом, ми маємо формулу (1.1):

$$x_j^l = f \left(\sum_{i \in M_j} x_i^{l-1} \cdot k_{ij}^l + b_j^l \right), \quad (1.1)$$

де x_j^l - вихідний сигнал поточного рівня; x_i^{l-1} - вихідний сигнал попереднього рівня; k_{ij}^l - ядро для поточного рівня; b_j^l - зміщення для поточного рівня; M_j - вибір вхідних карт.

Для кожної вихідної карти додається адитивний зсув. Однак вхідні картки будуть згорнуті з різними ядрами для створення відповідних вихідних карток. Вихідні карти нарешті проходять лінійну чи нелінійну функцію активації.

Шар активації. Забезпечує більш швидке та ефективне тренування, перетворюючи негативні значення на нуль і підтримуючи позитивні значення. Це називається активацією, оскільки тільки активовані функції переносяться на наступний рівень.

Шар субдискретизації. Субдискретизація чи об'єднання (pool) спрощує вивід, виконуючи нелінійну знижувальну дискретизацію, зменшуючи кількість параметрів, які має вивчити мережу. У статтях з НМ використовується термін – пул. Шар підвибірки виконує операцію субдискретизації на вхідних картках. Це також називають об'єднуючим шаром. У цьому шарі кількість вхідних та вихідних карток об'єктів не змінюється. Наприклад, якщо є N вхідних карт, то буде рівно N вихідних карт. Через операцію понижаючої вибірки розмір кожного вимірювання вихідних карт буде зменшений залежно від розміру маски понижаючої вибірки. Наприклад, якщо використовується ядро з дискретизацією 2×2 , то кожен вихідний вимір буде половиною відповідного вхідного виміру для всіх зображень. Ця операція може бути сформульована у виді (1.2):

$$x_j^l = \text{down}(x_j^{l-1}), \quad (1.2)$$

де $down(\cdot)$ - представляє функцію підвибірки.

На цьому рівні в основному виконуються два типи операцій субдискретизації: середня чи максимальна. У разі підходу середнього пулу функція зазвичай підсумовує $N \times N$ патчів карт об'єктів з попереднього шару та вибирає середнє значення. З іншого боку, у разі максимального пулу найбільше значення вибирається з $N \times N$ патчів карток об'єктів, тому розміри вихідної картки зменшуються в n разів.

1.5 Традиційні методи МН та НМГН

Швидкий розвиток НМГН та покращення можливостей пристроїв, включаючи обчислювальну потужність, об'єм пам'яті, енергоспоживання, дозвіл датчика зображення та оптику, підвищили продуктивність та економічну ефективність, що ще більше прискорило поширення додатків з урахуванням КЗ. Порівняно з традиційними методами КЗ, такі НМ дають змогу інженерам КЗ отримувати більшу точність у таких завданнях, як класифікація зображень, семантична сегментація, виявлення об'єктів та одночасна локалізація, зіставлення та розпізнавання дій.

Оскільки НМГН вчать, а не програмуються, програми, які використовують цей підхід, часто вимагають менше експертного аналізу та тонкого налаштування і використовують величезний обсяг відеоданих, доступних у сучасних системах. НМГН також забезпечують чудову гнучкість, оскільки моделі та структури ЗНМ можна перенавчити з використанням налаштованого набору даних для будь-якого варіанта використання, на відміну від алгоритмів МН, які, як правило, більш специфічні для конкретної галузі.

На рис. 1.5 представлено основні відмінності роботи алгоритмів МН від алгоритмів НМГН.

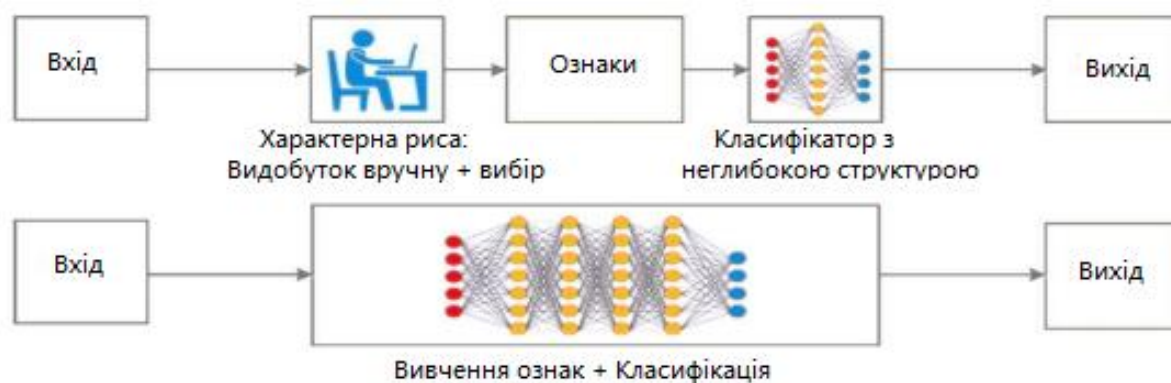


Рисунок 1.5 – Схеми роботи методів

а) МН; б) методів НМГН

Розвиток ЗНМ має величезний вплив на області КЗ в останні роки і є відповідальним за великий стрибок у можливостях розпізнавати об'єкти [29]. Цей сплеск було досягнуто завдяки збільшенню обчислювальної потужності та обсягу даних, доступних для навчання НМ. Нещодавній вибух та широке поширення різних архітектур глибоких НМ для КЗ очевидний, оскільки основний документ ImageNet Classification with Deep Convolutional Neural Networks цитувався понад 3000 разів [30].

1.6 Виявлення ознак паління

В результаті проведеного аналізу предметної області отримано перелік ознак, які дозволяють розпізнавати паління на відеопослідовності.

У таблиці 1.1 представлені ознаки паління, які застосовуються в проаналізованих роботах, а також загальні ознаки, що дозволяють розпізнавати паління.

Таблиця 1.1 – Ознаки паління

Підхід	Особа	Цигарка	Дим	Рука	Поза
[12]	+	+			
[13]	+	+			
[14]		+	+		
[15]		+	+	+	
[16]	+				
[17]			+		
[18]			+		
[19]					+

Майже всі наведені вище методи ґрунтуються на розпізнаванні обличчя, руки, диму, форми сигарет. Проте ключовою проблемою для аналізу події є нерегулярна форма паління. Тобто різні способи того, як люди тримають сигарету, різноманітні види тютюнових виробів, різнопланова зйомка. Також існує низка факторів, які можуть вплинути на точність розпізнавання під час використання даних методів. До таких факторів відносяться: розташування камер відеоспостереження, на яких не завжди вдається розрізнити обличчя, а тим більше дим і сигарету; слабка освітленість; паління без рук тощо. До того ж, майже у всіх описаних раніше методах, розпізнавання процесу паління відбувається без урахування тимчасової інформації. то призводить до виникнення складнощів, оскільки за одним кадром проблематично зробити висновок про наявність паління у відеопотоці.

1.7 Висновки до першого розділу

У першому розділі введено поняття розпізнавання образів, МН як основного методу розпізнавання паління. Також розглянуто основні типи

навчання та завдання, які вони дозволяють вирішити. Розглянуті методи МН розпізнавання паління, які у свою чергу поділяються на традиційні методи МН та НМ. Проведено аналіз предметної області. Виявлено ознаки паління за візуальними даними, котрі використані у різних роботах. Представлено основні поняття ЗНМ. Досліджено переваги застосування НМГН перед традиційними методами МН.

2 ТЕОРЕТИЧНА ЧАСТИНА

2.1 Попередня обробка відеопослідовності

Оскільки відео, отримані в реальних умовах зйомки, можуть містити різні артефакти, що виникають внаслідок шумів обладнання або через складні погодні умови, якість роботи НМ може суттєво погіршуватися [31]. Для того, щоб отримати правильні прогнози із застосуванням НМГН, потрібно попередньо обробити дані, усунувши ці шуми.

Попередня обробка зображень включає широкий спектр операцій, які можуть розглядатися як самоціль, але в даному разі їх призначення – спрощення чи поліпшення наступного аналізу. Попередня обробка підвищує якість зображення, усуваючи ненавмисні спотворення або покращуючи деякі функції зображення, які важливі для подальшої обробки та створення більш відповідного зображення, ніж оригінал для конкретного завдання. Дії, що можуть бути виконані на цифрових зображеннях, включають точкові, локальні чи «сусідні» та глобальні операції. Точкові операції перетворюють пікселі незалежно від сусідніх пікселів. Значення сірого кольору для вихідного зображення у конкретному пікселі залежить тільки від значення сірого для того ж пікселя у вхідному зображенні. Вони відтворюють пікселі в одному зображенні, щоб сформувати інше, використовуючи одну функцію відображення [21].

Точкові операції не враховують просторову організацію зображення, яка формує фундаментальний характер зображень на відміну інших типів даних. Приклади цих операцій включають розтягування контрасту, сегментацію на основі значення сірого та вирівнювання гистограми [32], також середнє віднімання та масштабування за деякими факторами. У контексті глибокого навчання та класифікації відео розглянемо такі методи попередньої обробки [33]: середнє віднімання; масштабування.

У роботі використовується модуль, що дозволяє здійснити попередню обробку відеопослідовності. OpenCV – бібліотека з відкритим вихідним кодом, що має програмну ліцензію університету Берклі, яка включає кілька сотень алгоритмів комп'ютерного зору [34]. Нова НМГН застосовує модуль OpenCV, містить дві функції, які можна використовувати для попередньої обробки кадрів та підготовки їх до класифікації за допомогою моделей глибокого навчання. Тому середнє віднімання розглядається як метод, що використовується для покращення роботи ЗНМ. Перш ніж приступити до тренування НМГН, необхідно обчислити середню інтенсивність кольору для кожного з червоного, зеленого та синього каналів за всіма зображеннями в навчальному наборі. У результаті отримаємо такі змінні μ_R , μ_G , μ_B .

Зазвичай результуючі значення являють собою три кортежі, що складаються із середнього значення червоного, зеленого та синього каналів відповідно. Наприклад, середні значення для навчального набору ImageNet є такими: $R = 103,93$, $G = 116,77$ та $B = 123,68$. Для набору даних, котрий застосовується у роботі, використовуються такі значення: $R = 114.7748$, $G = 107.7354$, $B = 99.4750$.

На рис. 2.1 показаний приклад візуалізації середнього віднімання, де середнє значення RGB (у центрі) було обчислено з набору даних зображень, шляхом віднімання з вихідного зображення (ліворуч), що привело до вихідного зображення (праворуч).

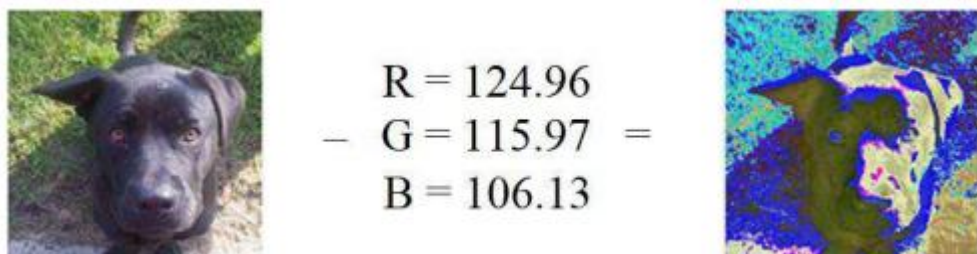


Рисунок 2.1 – Візуалізації середнього віднімання

Однак у деяких випадках середні значення червоного, зеленого та синього можуть бути розраховані по каналах, а не за кольорами, що призводить до матриці $M \times N$. У цьому випадку обчислюється матриця $M \times N$ для кожного каналу, потім віднімається з вхідного зображення під час навчання або тестування. Обидва методи є формами середнього віднімання, але версія, у якій застосовується інтенсивність кольору, використовується частіше, особливо для великих наборів даних.

Перед тим як передати відео через мережу, віднімаємо середнє, тобто μ від кожного каналу вхідного відеопотоку. Також враховуємо коефіцієнт масштабування σ , який додає нормалізації:

$$R = \frac{R - \mu_R}{\sigma}, \quad G = \frac{G - \mu_G}{\sigma}, \quad B = \frac{B - \mu_B}{\sigma}, \quad (2.1)$$

де R , G , B – червоний, зелений, синій канали відповідно; μ - середня інтенсивність кольору; σ - коефіцієнт масштабування

Значення σ може бути стандартним відхиленням по навчальному набору, тим самим перетворюючи крок попередньої обробки на стандартну оцінку. Однак σ може бути встановлена вручну, щоб масштабувати простір вхідного зображення до певного діапазону.

2.2 Вибір архітектури НМ

Основні архітектури, що застосовуються для класифікації зображень та відео: AlexNet; VGG; ResNet.

Далі розглянемо архітектури, наведені вище.

2.2.1 AlexNet

У 2012 році Алекс Крижевський випустив AlexNet [26] і з великим відривом за відсотком точності виграв складний конкурс ImageNet. 30 вересня

2012 року AlexNet брав участь у конкурсі з великомасштабного візуального розпізнавання ImageNet. Мережа досягла топ-5 помилок 15,3%, більш ніж на 10,8 процентних пункти нижче, ніж у учасника, котрий зайняв друге місце. Основним результатом оригінальної статті було те, що глибина моделі була важливою для її високої продуктивності, яка була обчислювально дорогою, але стала можливою через використання графічних процесорів (GPU) під час навчання. AlexNet містив вісім шарів: перші п'ять були згортковими шарами, деякі з них йшли шарами з максимальним об'єднанням, а останні три були повнозв'язними шарами [26]. Алекс використав функцію активації ненасиченого ReLU, яка показала покращення тренувальних показників у порівнянні з tanh та сигмоїдними функціями.

Рис. 2.2 ілюструє архітектуру AlexNet, у якій чітко показується розмежування обов'язків між двома графічними процесорами. Один графічний процесор запускає частини шару зверху малюнку, в той час як інший запускає частини шару внизу. Графічні процесори спілкуються лише на певних рівнях. Мережевий вхід має розмір 150 528, а число нейронів в шарах мережі, що залишилися, становить 253 440-186 624-64 896-64 896-43 264- 4096-4096-1000.

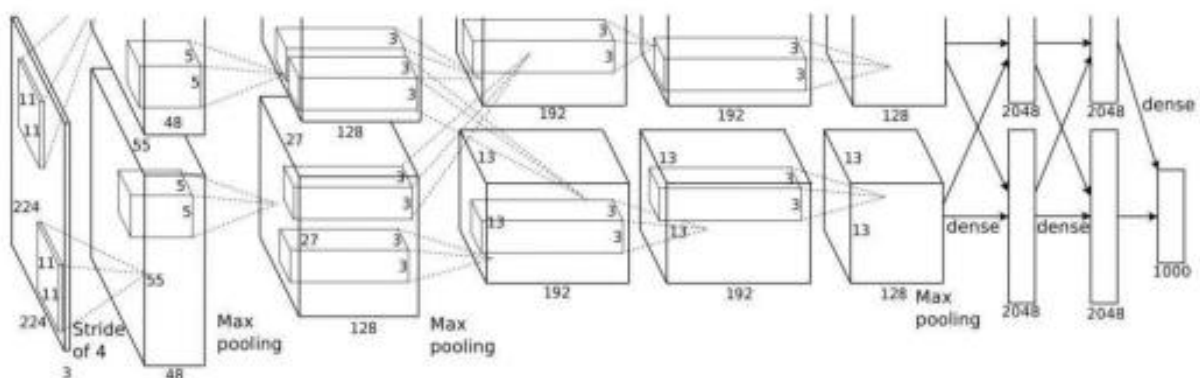


Рисунок 2.2 – Архітектура AlexNet

Дозволяє отримати точність розпізнавання паління на відеопослідовності 56,25%, оскільки паління є складною дією та для отримання вищої точності

необхідно витягувати низькорівневі ознаки, які дозволяють точніше ідентифікувати процес паління на послідовності відео.

2.2.2 VGGNet

Мережі VGG [35] з Оксфорда були першими, які використовували набагато менші фільтри 3×3 у кожному згортковому шарі, також об'єднували їх у вигляді послідовності згортки. Великою перевагою VGG стало розуміння того, що множинна згортка 3×3 в послідовності може емулювати ефект великих рецептивних полів, наприклад, 5×5 та 7×7 .

У мережах VGG використовуються кілька згорткових рівнів 3×3 для представлення складних функцій. На рис. 2.3 блоки 3, 4, 5 VGG-E: фільтри 256×256 та 512×512 3×3 використовуються кілька разів поспіль для вилучення та комбінації складних функцій. Це фактично схоже на великі класифікатори 512×512 з 3 шарами, які є згортковими, та становить множину параметрів. Але навчання цих мереж було важким, і їх доводилося розбивати на дрібніші мережі з додаванням шарів по одному. Все це через відсутність надійних способів упорядкування моделі або будь-якого обмеження величезного простору пошуку, що забезпечується великою кількістю параметрів.

На рис. 2.3 показано найбільш поширену реалізацію VGGNet- VGGNet-16.

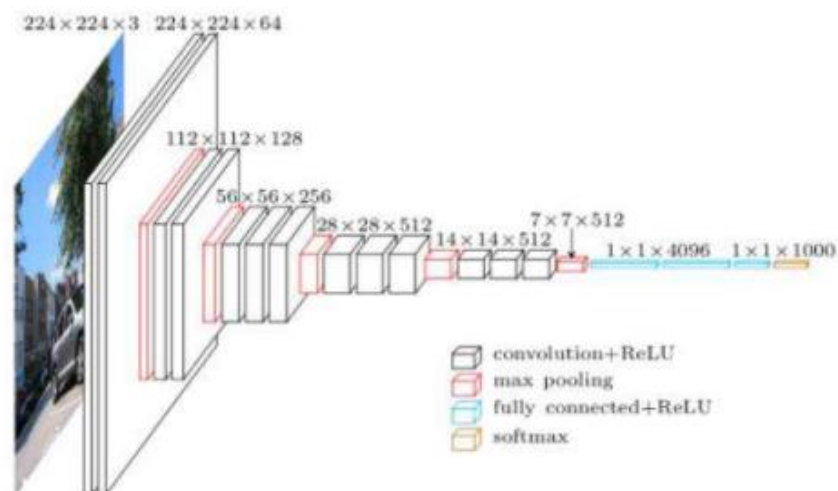


Рисунок 2.3 – Архітектура VGG16

VGGNet має дві відмінні особливості, які ускладнюють її роботу: повільна швидкість тренування суттєво збільшує час, витрачений на навчання; загасаючий градієнт заважає оновленню ваги.

Через свою глибину та кількість повністю підключених вузлів VGG перевищує 533 МБ для VGG16 та 574 МБ для VGG19. Це робить розгортання VGG довгим та ресурсозатратним. Для навчання НМ знадобилося понад 100 епох. Середня точність розпізнавання паління за візуальними даними становила 68.37%, що дозволяє зробити висновок про те, що архітектура VGG не підходить для вирішення поставленого завдання.

2.2.3 ResNet

ResNet або Residual Network [36] - залишкова НМГН, розроблена компанією Microsoft. ResNet дозволяє тренувати до сотень або навіть тисяч шарів і при цьому досягати якісних результатів.

Залишкові НМ використовують пропущені блоки або ярлики для переходу через деякі шари. Типові моделі ResNet реалізовані з використанням дво- або тришарових перепусток, які містять нелінійності (ReLU) та пакетну нормалізацію між ними.

Однією з причин пропуску шарів є вирішення проблеми зникнення градієнтів шляхом повторного використання активацій з попереднього шару, поки сусідній шар не впізнає свої ваги. Під час навчання ваги адаптуються для приглушення рівня вище за напрямом та посилюють пропущений раніше рівень. У найпростішому випадку адаптуються тільки вагові коефіцієнти для з'єднання сусіднього рівня, без явних вагових коефіцієнтів для висхідного рівня. Це працює найкраще, коли один нелінійний шар переступає або всі проміжні шари є лінійними. Якщо ні, то для пропущеного з'єднання слід вивчити явну матрицю ваги. Залишкові модулі необхідні для спрощення навчання глибокої НМ. На відміну від традиційних послідовних мережевих архітектур, таких як AlexNet і VGG, ResNet натомість є формою «екзотичної архітектури», яка спирається на

мікроархітектурні модулі, також звані «мережеві архітектури у мережі». Термін мікроархітектура відноситься до набору «будівельних блоків», що використовуються для побудови мережі. Набір будівельних блоків мікроархітектури разом із стандартними рівнями згортковим і пулом, що призводить до макроархітектури, тобто до самої кінцевої мережі.

ResNet вводить швидкі з'єднання, які пропускають сигнал від одного рівня до наступного. З'єднання проходять через градієнтні потоки мереж від пізніших рівнів до більш ранніх рівнів та полегшують навчання дуже глибоких мереж. На малюнку 16 показаний залишковий блок, який є елементом ResNet. З'єднання обходять сигнал від «вершини» блоку до "хвоста". Архітектура ResNet стала оригінальною роботою, котра демонструє, що надзвичайно глибокі мережі можна навчати з використанням стандартного СГС та функції ініціалізації з використанням залишкових модулів, як показано на рис. 2.4.

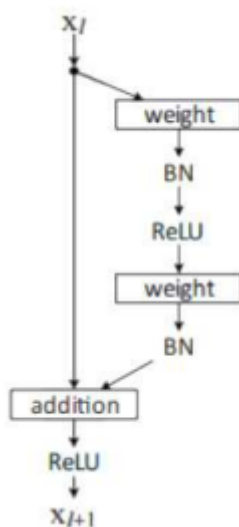


Рисунок 2.4 – Залишковий модуль у ResNet

Для отримання додаткової точності можна оновити залишковий модуль, що використовувався для зіставлення ідентифікаторів, що наочно демонструє рис. 2.5, на якому зліва показаний оригінальний залишковий модуль, а праворуч – оновлений залишковий модуль з використанням попередньої активації [36].

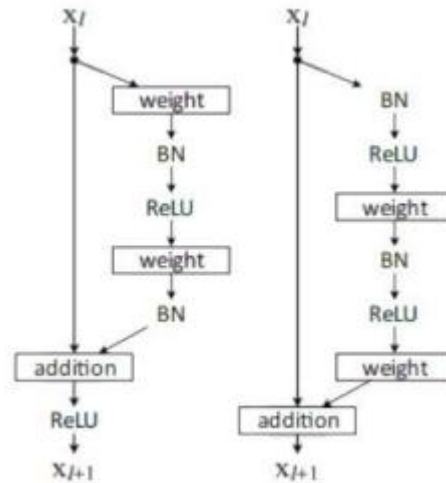


Рисунок 2.5 – Оригінальний та оновлений залишкові модулі

Вихід залишкового рівня може бути визначений на основі вихідних даних $(l-1)^{th}$, які надходять від попереднього рівня, визначеного як x_{l-1} ; $F(x_{l-1})$ – це вихідний сигнал після виконання різних операцій, таких як згортка з різними розмірами фільтрів, нормалізація партії (BN), за якою слідує функція активації, така як ReLU на x_{l-1} . Остаточний результат залишкової одиниці дорівнює x_l , який можна визначити за допомогою рівняння (2.2):

$$x_l = F(x_{l-1}) + x_{l-1}; \quad (2.2)$$

Також варто відзначити, незважаючи на те, що ResNet набагато глибше, ніж VGG16 і VGG19, розмір моделі насправді істотно менше через використання глобального середнього пулу, а не повністю підключених шарів – це зменшує розмір моделі до 83.26 МБ ResNet34. ResNet є однією з небагатьох архітектур, яка перевершила людину. На малюнку 18 показана історія успіху згорткових НМГН у задачі класифікації ImageNet у 2012 році [37]. ResNet-152 показує лише 3,57% помилок, що краще, ніж помилка, котру зробив в цій задачі людина, яка становить 5%.

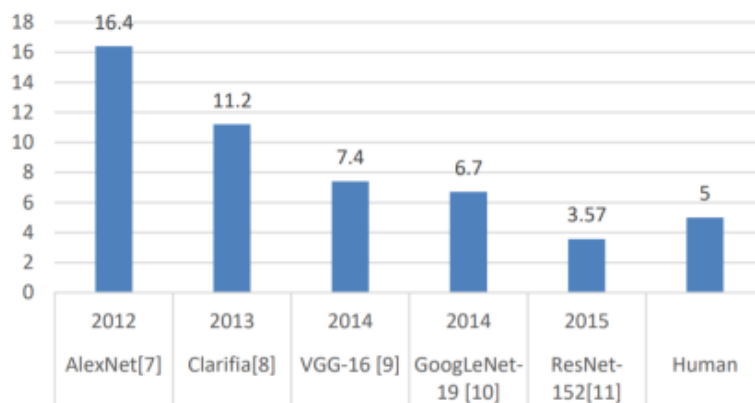


Рисунок 2.6 – Порівняння точності розпізнавання об'єктів НМ та людини

ResNet-34 складається з двох згорткових рівнів, і за кожним згортковим рівнем слідує пакетна нормалізація і ReLU. Швидкий перехід дозволяє з'єднувати верхню частину блоку з шаром відразу після останнього згорткового шару у блоці. У своєму дослідженні Kaiming He показав, що така попередня активація сприяє оптимізації тренувань та знижує перенавчання [38]. ResNet забезпечує з'єднання швидкого доступу, які дозволяють сигналу обходити один рівень і переходити до наступного рівня послідовності. Оскільки ці з'єднання проходять через градієнтні потоки мереж від пізніших рівнів до ранніх рівнів, вони дозволяють полегшити навчання НМГН. Проведений експеримент показує, що ця архітектура дозволяє отримати точність розпізнавання паління на послідовності відео 82,3%, з цього випливає висновок про те, що ResNet може бути використана для вирішення поставленого завдання. Результати експерименту представлені у порівняльній таблиці 2.1.

Таблиця 2.1 - Порівняльні характеристики архітектур НМ

Назва	Кількість епох	Отримана точність, %	Кількість параметрів, М	Розмір файлу, МВ
AlexNet	20	56,25	62	233,1
VGGNet	100	68,37	138	533
ResNet	50	81,1	21	84,24

На основі результатів дослідження, наведених у таблиці 2.1, можна зробити висновок, що найбільш схожою архітектурою для розпізнавання паління за візуальними даними є ResNet, оскільки отримана точність становить 81%, що у свою чергу перевищує точність, отриману під час тестування AlexNet і VGGNet.

ResNet досить добре справляється з розпізнаванням паління, однак такий підхід не враховує інформацію руху, закодовану в кількох суміжних кадрах. Щоб ефективно включити цю інформацію про рух у відеоаналіз, необхідно виконати тривимірну згортку в згорткових шарах ЗНМ, щоб зафіксувати розрізняючі ознаки як у просторовому, і у часовому вимірі.

2.3 Тривимірні ЗНМ

У більшості прикладів розпізнавання дій використовується простий підхід, який полягає в тому, щоб розглядати відеокадри як нерухомі зображення і застосовувати ЗНМ для розпізнавання дій на рівні окремих кадрів. Справді, цей підхід був використаний для аналізу відео ембріонів, що розвиваються [39].

У двовимірних ЗНМ двовимірна згортка виконується на згорткових шарах для видалення об'єктів з локального околу на картах об'єктів у попередньому шарі. Потім застосовується адитивне усунення, і результат передається через сигмоподібну функцію. Формально вартість одиниці в позиції (x,y) у j -й карті об'єктів на i -му шарі, позначеній як v_{ij}^{xy} , визначається (2.3):

$$v_{ij}^{xy} = \tanh(b_{ij} + \sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} w_{ijm}^{pq} v_{(i-1)m}^{(x+p)(y+q)} + b_{ij}), \quad (2.3)$$

де \tanh – це функція гіперболічного тангенсу; b_{ij} – це зміщення для карти об'єктів; m – індекс для набору карток об'єктів у 1-му шарі i , підключеного до

поточної картки об'єктів; w_{ijm}^{pq} – значення позиції p , ядра q , що з k -й картою ознак; P_i і Q_i - висота і ширина ядра відповідно.

У шарах підвибірки роздільна здатність карт об'єктів зменшується за рахунок об'єднання локального околу на картах об'єктів у попередньому шарі, цим самим підвищуючи інваріантність до спотворень на входах. Архітектура ЗНМ може бути побудована шляхом об'єднання декількох шарів згортки та підвибірки почергово. Параметри ЗНМ, такі як зміщення b_{ij} та вага ядра w_{ijm}^{pq} , зазвичай вивчаються з використанням контрольованих або неконтрольованих підходів.

Тривимірна карта активації, отримана під час згортки тривимірної ЗНМ, необхідна для аналізу даних, де важливий тимчасовий або об'ємний контекст. Здатність аналізувати ряд кадрів або зображень у контексті призвела до використання 3D ЗНМ як інструменту для розпізнавання дій.

Виконано перетворення ResNet для обробки тривимірними вхідними даними, а не двовимірними, проводилося з метою включення тимчасового компонента для розпізнавання активності, у випадку паління. ResNet одна з небагатьох архітектур, яка має можливість виконувати класифікацію відео шляхом зміни форми вхідного об'єму даних для включення просторово-часової інформації та використання тривимірних ядер усередині архітектури [40].

Завдання розпізнавання паління включає ідентифікацію різних дій із відеокліпів, у яких паління може виконуватися або не виконуватися протягом усієї тривалості відео. Це виглядає як природне розширення завдань класифікації зображень на кілька кадрів, а потім агрегування прогнозів кожного кадру. Незважаючи на успіх архітектур глибокого навчання в класифікації зображень, прогреси в архітектурах для класифікації відео були більш повільними.

В аналізі відео необхідно захоплювати інформацію про рух, закодовану у кількох суміжних кадрах. З цією метою пропонується виконати тривимірні

згортки в ЗНМ, щоб обчислити елементи як з просторових, так і тимчасових вимірів. Тривимірна згортка досягається шляхом згортки тривимірного ядра кубом, утвореним шляхом об'єднання кількох суміжних кадрів. За допомогою цієї конструкції карти ознак у шарі згортки з'єднуються з кількома суміжними кадрами у попередньому шарі, тим самим захоплюючи інформацію про рух. Формальне значення у позиції (x,y,z) на j -й карті об'єктів у i -му шарі визначається за [41] як (2.4):

$$v_{ij}^{xyz} = \sigma(\sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} \sum_{r=0}^{R_i-1} w_{ijm}^{pqr} v_{(i-1)m}^{(x+p)(y+q)(z+r)} + b_{ij}), \quad (2.4)$$

де v_{ij}^{xyz} - значення в точці (x,y,z) на j -й карті об'єктів i -го шару в мережі; w_{ijm}^{pqr} - це значення точки (p,q,r) на m -й карті об'єктів, яка є виходом верхнього рівня, з'єданого тим самим ядром згортки; P_i та Q_i - висота та ширина 3D-фільтра відповідно; R_i - глибина 3D-фільтра за тимчасовим виміром; b_{ij} - позначає зміщення; $\sigma(\cdot)$ - сигмоїдна функція активації.

На рис. 2.7 показаний приклад 3D згорток. У 3D-згортці те саме 3D-ядро застосовується до 3D-кубів, що перекриваються, у вхідному відео для вилучення рухомих об'єктів.

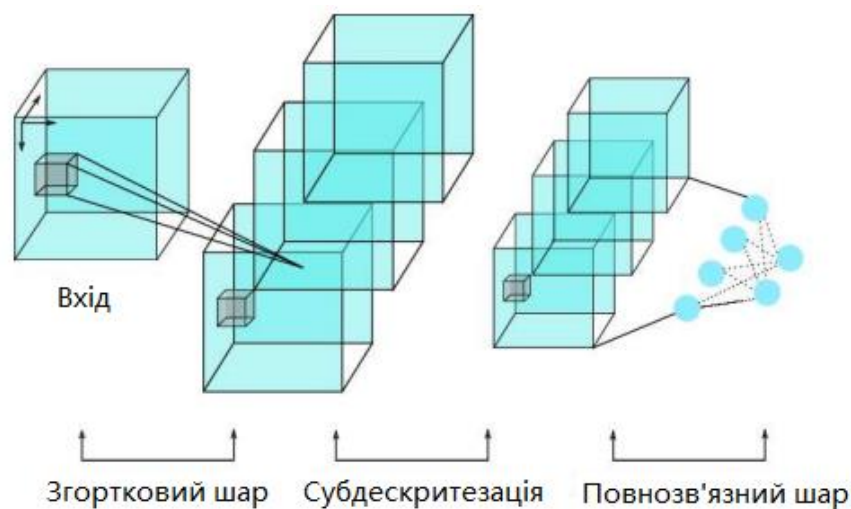


Рисунок 2.7 – 3D згортка

2.4 Модифікована архітектура НМ

Для модифікації вибрано ResNet. ResNet є повністю згортковою, тому її застосовують для вилучення просторово-тимчасового об'єму, на відміну від багатьох архітектур з повнозв'язними шарами, включаючи AlexNet та VGG-16, які містять кілька рівнів максимального пулу, здатні пошкодити оцінку дій, архітектура ResNet містить лише один рівень пула відразу після conv1. Зменшена кількість об'єднуючих шарів робить ResNet значно кращим для розпізнавання паління за візуальними даними, оскільки для розпізнавання цього процесу необхідно зберегти просторові деталі.

Зокрема, використовується 34-шарова ResNet, що показує обчислювальну ефективність з хорошими характеристиками для класифікації [41]. Для того, щоб використовувати ResNet для оцінки багатокадрового оптичного потоку, необхідно провести розширення даної архітектури, замінюючи всі $k \times k$ двовимірні згорткові ядра додатковим тимчасовим виміром $k \times k \times 3$, як описано у статті [42]. Шари пулу в декодері розширюються аналогічно.

У ході модернізації НМ ResNet отримано нову архітектуру, яка надалі застосовується для розпізнавання паління на відеопослідовності. Основна відмінність даної архітектури від ResNet полягає у використанні тривимірних ядер та зміненої операції субдискретизації, за допомогою якої карти ознак у шарі згортки з'єднуються з декількома суміжними кадрами у попередньому шарі, тим самим захоплюючи інформацію про рух. Розміри згорткових ядер становлять $3 \times 3 \times 3$. Мережа використовує 16-кадрові RGB-кліпи як входи. Розміри вхідних кліпів складають $3 \times 16 \times 112 \times 112$. Зниження дискретизації входів виконується періодично з кроком 2. Архітектура НМ, отримана під час модернізації, наведена у таблиці 2.2.

Таблиця 2.2 - Нова архітектура НМ

Назва шару	Функція активації	Ядро	Кількість нейронів
Згортковий шар 1	ReLU	$7 \times 7 \times 7$	64
Згортковий шар 2		$\begin{bmatrix} 3 \times 3 \times 3 \\ 3 \times 3 \times 3 \end{bmatrix} \times 3$	64
Згортковий шар 3		$\begin{bmatrix} 3 \times 3 \times 3 \\ 3 \times 3 \times 3 \end{bmatrix} \times 4$	128
Згортковий шар 4		$\begin{bmatrix} 3 \times 3 \times 3 \\ 3 \times 3 \times 3 \end{bmatrix} \times 6$	256
Згортковий шар 5		$\begin{bmatrix} 3 \times 3 \times 3 \\ 3 \times 3 \times 3 \end{bmatrix} \times 3$	512

У таблиці 2.2 залишкові блоки об'єднані квадратними дужками. Після кожного згорткового шару використовується пакетна нормалізація.

2.5 Навчання НМ

Навчання - це процес, за допомогою якого модель вивчає або розпізнає закономірності даних для створення відповідних прогнозів.

Для навчання НМ дані поділяються на два набори: навчальний та тестовий. Тестовий набір містить уже передбачені значення. Він використовується для перевірки прогнозів, зроблених навчальним набором. Фактично модель навчається на тренувальному наборі та перевіряється на тестовому. Крім того, дйтасет ділиться на позитивні та негативні зразки: позитивні - це відео, що містять паління, негативні – відео, які його не містять. Ефективність моделі оцінюється з урахуванням правильних прогнозів. Навчена та перевірена модель, розроблена для розпізнавання образів з використанням НМГН, називається класифікатором. Далі класифікатор використовується для прогнозування процесів, що характеризуються як паління. Навчання НМ, з

математичної точки зору є багатопараметричним завданням нелінійної оптимізації.

2.6 Оптимізація

2.6.1 СГС

Метод СГС [43] є модифікацію алгоритму градієнтного спуску, оптимізованого за швидкістю збіжності. Різниця між традиційним градієнтним спуском полягає у цьому, що елементи розглядаються окремо. СГС наближає градієнт, використовуючи лише одну точку даних. Таким чином, оцінка градієнта економить багато часу у порівнянні із підсумовуванням за всіма даними. Ефект використання SGD найбільш помітний під час роботи з великими даними. Таким чином, градієнт функції вартості буде розраховуватися не для всіх елементів у вибірці, як це робиться за допомогою традиційного методу градієнтного спуску, а для кожного елемента окремо. Градієнт, розрахований для конкретного елемента, що приймається за апроксимацію реального градієнта. Ваги у моделі перераховуються відповідно до обчисленого градієнта для одного елемента, що призводить до того, що модель коригується при переході від кожного наступного елемента вибірки до наступного.

Для розуміння принципу роботи СГС необхідно звернути увагу, що кожен приклад z представлений парою (x, y) , що складається з довільного вхідного значення x та скалярного вихідного значення y . Функцією втрат є функція $l(\hat{y}, y)$, яка вимірює «вартість» передбачення \hat{y} , коли фактична відповідь дорівнює y . Визначаємо сімейство F функцій $f_w(x)$, параметризованих вектором ваг w . Далі шукаємо функцію $f \in F$, що мінімізує втрати $Q(z, w) = l(f_w(x), y)$, усереднені на прикладах. Важливо пам'ятати, що за усереднення по невідомому розподілу $dP(z)$, доводиться розраховувати середнє значення для вибірки z_1, \dots, z_n :

$$E(f) = \int l(f(x), y) dP(z), \quad E_n(f) = \frac{1}{n} \sum_{i=1}^n l(f(x_i), y_i), \quad (2.5)$$

де $E_n(f)$ – емпіричне значення вимірює ефективність тренувального набору; $E(f)$ – очікуване значення вимірює ефективність узагальнення, тобто очікувану продуктивність у майбутніх прикладах.

Статистична теорія навчання [44] виправдовує мінімізацію емпіричного ризику замість очікуваного ризику, коли обране сімейство F є досить обмежувальним. Спочатку пропонувалося мінімізувати емпіричне значення $E_n(f)$ з використанням градієнтного спуску (GD). Кожна ітерація оновлює ваги w на основі градієнта

$$w_{t+1} = w_t - \gamma \frac{1}{n} \sum_{i=1}^n \nabla_w Q(z_i, w_t), \quad (2.6)$$

де γ - адекватно обрана швидкість навчання.

При достатніх припущеннях про регулярність і в той момент, коли початкова оцінка w_0 досить близька до оптимальної та швидкість навчання досить мала, цей алгоритм досягає лінійної збіжності [45], тобто:

$$-\log \rho \sim t, \quad (2.7)$$

де ρ представляє залишкову помилку error.

З більшості історичних причин лінійна збіжність означає, що залишкова помилка асимптотично зменшується, а квадратична збіжність означає ще швидше асимптотичну збіжність. Обидві швидкості збіжності значно вищі, ніж швидкості збіжності SGD.

Алгоритм СГС є радикальним спрощенням. Замість точного обчислення градієнта $E_n(f_w)$ кожна ітерація оцінює цей градієнт на основі одного випадково вибраного прикладу z_t :

$$w_{t+1} = w_t - \gamma_t \nabla_w Q(z_t, w_t), \quad (2.8)$$

Стохастичний процес $\{w_t, t = 1, \dots\}$ залежить від прикладів, випадково обраних на кожній ітерації. Передбачається, що формула (2.8) поводить ся так само, як очікування (2.5), незважаючи на шум, що вноситься цією спрощеною процедурою.

Оскільки алгоритму не потрібно запам'ятовувати, які приклади взяли під час попередніх ітерацій, він може обробляти їх нальоту в розгорнутій системі. У такій ситуації СГС безпосередньо оптимізує очікуваний ризик, оскільки приклади випадково взяті з основного розподілу істинності.

Імпульс є методом, який допомагає прискорити процес навчання за допомогою SGD. Основна ідея полягає в тому, щоб використовувати ковзне середнє значення градієнта замість використання тільки поточного дійсного значення градієнта. Ми можемо висловити це за допомогою наступного рівняння математично:

$$w_{t+1} = \mu w_t - \gamma_t \nabla_w Q(\theta), \quad \theta = \theta - w_{t+1}. \quad (2.9)$$

На рис. 2.8 показано відмінність реалізації градієнтного спуску з імпульсом та без нього.

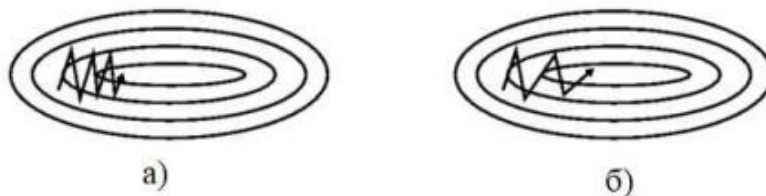


Рисунок 2.8 – СГС без імпульсу (а) та з імпульсом (б)

Доданок імпульсу μ зазвичай встановлюється рівним 0,9. По суті, під час використання імпульсу ми штовхаємо м'яч вниз схилом. Куля накопичує

імпульс, коли він котиться під ухил, стаючи все швидше і швидше на шляху (до тих пір, поки не досягне своєї кінцевої швидкості, якщо є опір повітря, тобто $\mu < 1$). Те саме відбувається з нашими оновленнями параметрів: параметр імпульсу збільшується для вимірювань, градієнти яких вказують в одних і тих самих напрямках, та зменшує оновлення для інших вимірювань, градієнти яких змінюють напрямки. В результаті ми отримуємо швидшу збіжність та зменшуємо коливання.

2.6.2 Пакетна нормалізація

Допомагає прискорити процеси НМГН за рахунок зменшення внутрішньої коваріації шляхом усунення вхідних вибірок. Це означає, що вхідні дані лінійно перетворені, щоб мати нульове середнє значення та одиничну дисперсію.

Також шляхом нормалізації активацій нормалізація партії допомагає стабілізувати розподіл внутрішніх активацій у процесі навчання моделі. Пакетна нормалізація також дозволяє використовувати значно більш високі швидкості навчання і знижує чутливість до ініціалізації. Такі ефекти допомагають прискорити навчання. Batchnorm [46] у спішно використовувався для створення сучасних архітектур, таких як залишкові мережі ResNet. Batchnorm працює над міні-пакетами під час навчання стохастичним градієнтом і використовує середнє значення та дисперсію міні-пакет для нормалізації активацій. Зокрема, розглянемо конкретний вузол у глибокій мережі, створюючи скалярне значення для кожного вхідного прикладу. Враховуючи міні-пакет β з m прикладів, розглянемо значення цього вузла x_1, \dots, x_m . Тоді Batchnorm набуває форми

$$\hat{x}_i \leftarrow \frac{x_i - \mu_\beta}{\sqrt{\sigma_\beta^2 + \epsilon}} \quad (2.10)$$

де μ_β – середнє значення вибірки x_1, \dots, x_m ; σ_β^2 – вибіркова дисперсія; ϵ – невелика постійна.

2.7 Висновки до другого розділу

У другому розділі описано передобробку відеопослідовності. Розглянуто архітектуру згорткових НМГН. Вибрано архітектуру, яка згодом перетворена на тривимірну згорткову НМГН шляхом зміни форми вхідного об'єму для включення просторово-часової інформації та використання тривимірних ядер усередині архітектури. Описано навчання НМ та її оптимізація, яка виконується за допомогою СГС з імпульсом та пакетної нормалізації після кожного згорткового шару.

3 ПРАКТИЧНА ЧАСТИНА

Пропонований метод визначає паління у відеоконтенті за допомогою розпізнавання дій, які характерні для людини, що знаходяться в процесі паління. Розпізнавання дій на відеопослідовності вимагає захоплення контексту з усього відео, а не тільки захоплення інформації з кожного кадру.

3.1 Алгоритм процесу розпізнавання паління з відеопослідовності

Паління відноситься до групи атомарних дій, розпізнати які можна лише за наявності певної сукупності ознак. Паління є процесом з такими характерними особливостями.

1. Зміна положення рук:

1.1. Рука піднімається на рівень губ, пауза, опускається вниз, пауза, знову піднімається.

2. Рух губ на сценах близького плану.

3. Запалювання сигарети може здійснюватися за допомогою сірника або запальнички (за наявності у кліпі):

3.1. Нахил голови.

3.2. Використання запальнички:

3.2.1. Однією рукою людина підносить запальничку в обличчя;

3.2.2. Великим пальцем цієї руки проводиться запуск механізму (дія може повторюватися кілька разів);

3.2.3. Інша рука може перешкоджати загасанню сигарети та перекривати огляд для розпізнавання попередніх дій (у такому разі враховує перебування обох рук лише на рівні губ);

3.2.4. Руки опускаються.

3.3. Використання сірників:

3.3.1. Попередньо цигарка затискається між зубами;

3.3.2. Дві руки знаходяться на рівні грудей/ трохи нижче грудей;

3.3.3. Однією рукою відбувається невеликий помах (дія може повторюватись);

3.3.4. Одна рука залишається на рівні грудей, друга змінює положення, переміщуючись вище до підборіддя або губ.

3.3.5. Помах руки, щоб загасити сірник;

3.3.6. Руки опускаються.

4. Стряхування попелу з сигарети (за наявності в кліпі):

4.1. Відведення руки з сигаретою вниз.

4.2. Рух кистю/пальцями.

Виходячи з отриманого опису розроблено алгоритм розпізнавання паління з відеопослідовності. Для навчання НМ використовується ГСП із імпульсом. Довільно генеруються навчальні зразки з відео у тренувальному наборі. Встановлюються тимчасові позиції шляхом рівномірної вибірки. Далі 32-х кадрові кліпи встановлюються навколо зазначених часових позицій. Якщо відео коротше, ніж 32 кадри, то проводиться зациклювання стільки разів, скільки необхідно для досягнення встановленої тривалості. Потім випадково вибираються просторові положення із чотирьох кутів або одного центру. На додаток до позицій також вказуються просторові масштаби кожної вибірки для виконання багатомасштабного кадрівання. Кадр обрізається в просторово-часових положеннях. Розмір кожної вибірки складає 3 канали \times 32 кадри \times 112 пікселів \times 112 пікселів, і кожна вибір каперевернута по горизонталі із ймовірністю 50%. Також виконується віднімання середнього значення нашого дейтасета із зразка для кожного колірною каналу.

Усі створені зразки зберігають ті ж мітки класів, що й їх вихідні відео. У навчанні моделі використовується крос-ентропія як функція втрати. Параметри тренування включають загасання ваг на 0,001 та 0,9 для імпульсу. Швидкість навчання 0,1 та поділена на 10 після насичення втрат під час перевірки. При

виконанні тонкого налаштування на швидкості навчання 0,001 - згасання ваги $1e-5$.

На першому етапі ініціалізується НМ, що відбувається налаштуванням параметрів і після цього на вхід подається відеопослідовність.

Виконується ініціалізація класів, що дозволяє провести класифікацію набору даних: є паління, немає паління. Визначається тривалість вибірки, тобто кількість кадрів для класифікації – 32, та просторові розміри вибірки – 112 x 112. Для створення вхідних кліпів використовується метод ковзного вікна, в якому відкидається тільки найстаріший кадр у списку, звільняючи місце для найновішого кадру. Далі кожне відео розбивається на 32-кадрові кліпи, що не перекриваються. Дана операція відбувається за допомогою введення циклу, який зчитує кадри з відеопотоку, потім здійснює перевірку захоплення кадру. Якщо кадр захоплено, то кожен кліп обрізається навколо центральної позиції з максимальним масштабом, виконується середнє віднімання і новий кадр додається в чергу, інакше відбувається вихід із циклу. Новий цикл дозволяє зробити перевірку заповнення черги. Після закінчення даного циклу створюється blob-об'єкт, реалізований у програмі функцією `blobFromImages`.

Blob-об'єкт чи крапля – це колекція кадрів з однаковими просторовими розмірами, вираженими шириною та висотою, однаковою глибиною, тобто числом каналів, які мають бути попередньо оброблені однаковим чином.

Blob-об'єкт має таку розмірність: (3, 32, 112, 112). Число 3 позначає кількість каналів у вхідних кадрах, 32 – загальна кількість кадрів у «краплі». Наступні числа позначають висоту та ширину відповідно.

Далі, щоб витягти просторово-часові характеристики, кожен екземпляр передається через тривимірну ЗНМ. Розпізнавання паління відбувається шляхом знаходження багаторазового оптичного потоку. Оптичний потік – це рух об'єктів між послідовними кадрами, викликаний відносним рухом між об'єктом та камерою. Проблема оптичного потоку може бути виражена так, як показано рис 3.1.

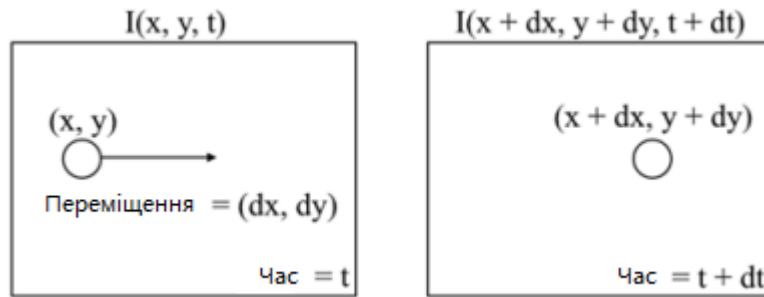


Рисунок 3.1 – Зміна руху точки у кількох кадрах

Щільний оптичний потік можна розглядати як набір полів вектора зміщення d_t між парами послідовних кадрів t та $t + 1$. Через $d_t(u, v)$ вектор зміщення позначається в точці (u, v) у кадрі t переміщує точку відповідну точку в наступному кадрі $t + 1$. Горизонтальні та вертикальні компоненти векторного поля, d_t^x та d_t^y , можна розглядати як канали зображення, добре підходять для розпізнавання з використанням тривимірної ЗНМ. Щоб уявити рух за послідовністю кадрів, ми складаємо канали потоку $d_t^{x,y}$ з L послідовних кадрів, щоб сформувати всього $2L$ вхідних канали. Більш формально, нехай w і h будуть шириною та висотою відео; вхідний об'єм $I_\tau \in \mathbb{R}^{w \times h \times 2L}$ для довільного кадру τ потім будується так:

$$\begin{aligned} I_\tau(u, v, 2k - 1) &= d_{\tau+k-1}^x(u, v), \\ I_\tau(u, v, 2k) &= d_{\tau+k-1}^y(u, v), \quad u = [1; w], v = [1; h], k = [1; L], \end{aligned} \quad (3.1)$$

Для довільної точки (u, v) канали $I_\tau(u, v, c)$, $c = [1; 2L]$ кодують рух у цій точці за послідовністю L кадрів. Далі відбувається одержання траєкторії, тобто альтернативне подання руху, яке замінює оптичний потік, дискретизований в тих самих місцях в декількох кадрах, потоком, дискретизованим уздовж траєкторій руху. В цьому випадку вхідний об'єм, відповідний кадру τ , набуває такого вигляду:

$$I_{\tau}(u, v, 2k - 1) = d_{\tau+k-1}^x(p_k),$$

$$I_{\tau}(u, v, 2k) = d_{\tau+k-1}^y(p_k), \quad u = [1; w], v = [1; h], k = [1; L] \quad (3.2)$$

де p_k - це k -та точка на траєкторії, яка починається в точці (u, v) в системі координат τ і визначається наступним рекурентним співвідношенням:

$$p_1 = (u, v); \quad p_k = p_{k-1} + d_{\tau}(p_{k-1}), k > 1, \quad (3.3)$$

Порівняно з поданням вхідного об'єму, де канали $I_{\tau}(u, v, c)$ зберігають вектори зміщення у місцях (u, v) , вхідний об'єм (3.3) зберігає вектори, вибрані в місцях p_k вздовж траєкторії, як показано на рис. 3.2.

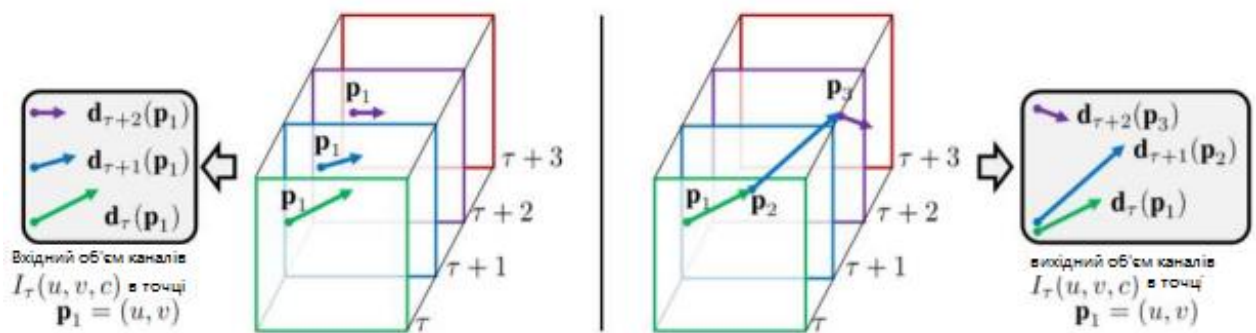


Рисунок 3.2 – Обчислення траєкторії руху.

Далі формується карта рухів. Кожна карта ознак згорткового шару пов'язана з декількома послідовними суміжними кадрами у верхньому шарі. Наступним кроком проводиться оцінка ймовірності наявності паління в кліпах. Мережа «переглядає» послідовність із 32 кадрів, формує траєкторії руху, аналізує подібність із відомим зразком паління, і знаходить ймовірність наявності паління в кожному кадрі, які потім усереднюються по всіх кліпах. Клас, який має максимальну оцінку, позначає дію в даній відеопослідовності. Якщо ймовірність більша або дорівнює 0.5, то паління в даних кадрах розпізнане.

На рис. 3.3 наведено блок-схему розробленого алгоритму розпізнавання паління за візуальними даними.

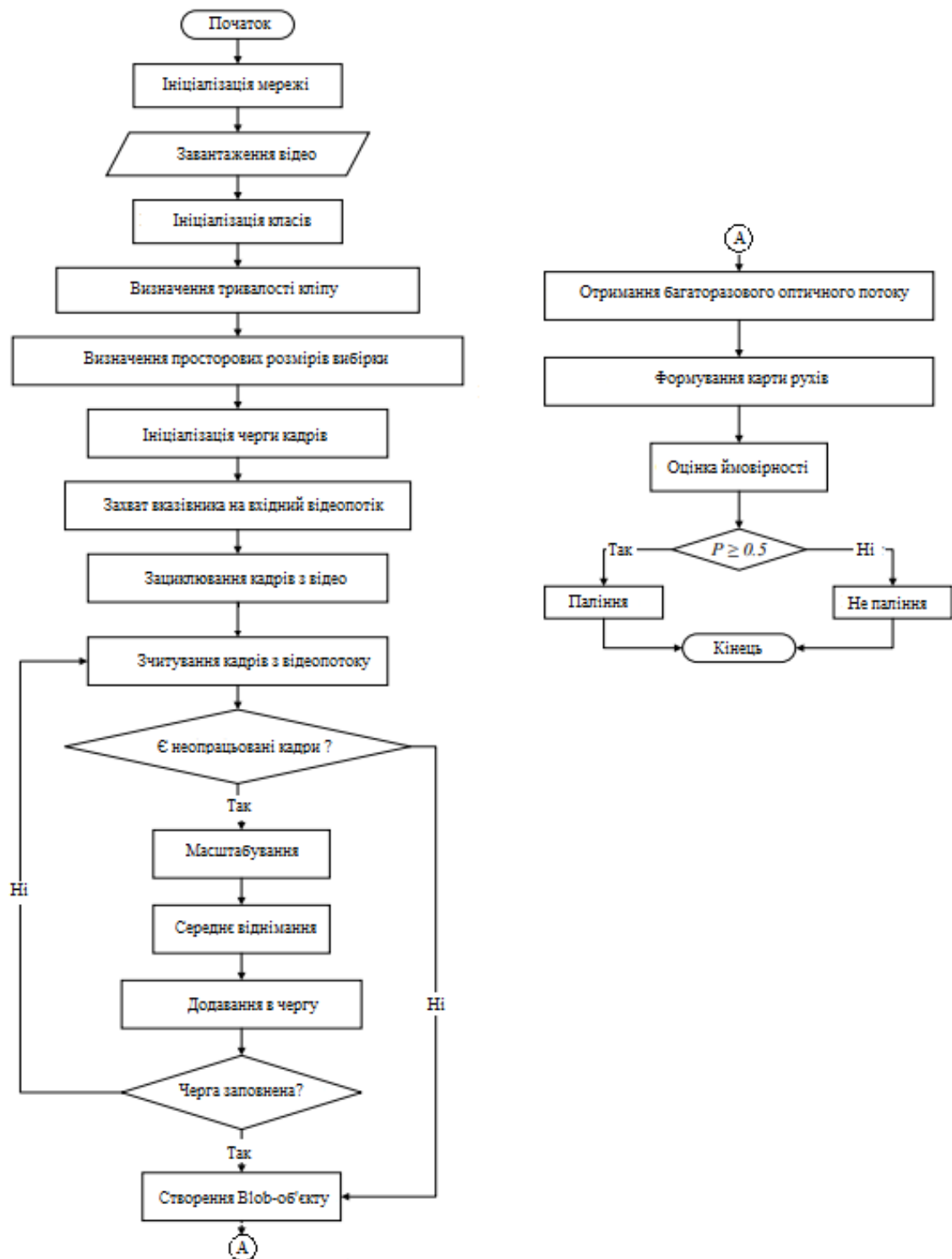


Рисунок 3.3 – Блок-схема алгоритму розпізнавання паління за візуальними даними

3.2 Експериментальні дослідження

3.2.1 Набір даних

У роботі використовується OpenCV [34] для вилучення кадрів із усіх відеокліпів із встановленим розміром кадру 112×112. Далі використовується набір даних HMDB51[47]. Дейтасет HMDB51 включає в себе 51 клас, кожному класі приблизно 100 зразків, що загалом становить 6766 відео. Категорії дій можуть бути згруповані у п'ять типів:

- загальні дії особи: усміхатися, сміятися, жувати, говорити;
- дії з предметними маніпуляціями: курити, їсти, пити;
- загальні рухи тіла: робити колесо, аплодувати, підніматися, підніматися сходами, пірнати, падати на підлогу, прибирати руки назад, робити стійку на руках, стрибати, підтягуватися, віджиматися, бігати, сідати, підніматися з чогось, робити сальто, вставати, повертатися, ходити, робити хвилю;
- рухи тіла при взаємодії з об'єктом: розчісувати волосся, ловити, діставати меч, вести м'яч, грати в гольф, вдаряти по м'ячу, вибирати, наливати, штовхати щось, їздити на велосипеді, їздити на коні, стріляти в м'яч, стріляти із лука, стріляти із рушниці, кидок м'яча;
- рухи тіла для взаємодії з людиною: фехтування, обійми, удар ногою, поцілунок, удар, рукостискання, бій на мечях.

Оскільки основний клас – паління, решта об'єднана у клас "немає паління". Вихідний дейтасет налічує 70 відео, на яких люди різного віку, типів додавання, гендерних ознак, різної расової приналежності, по-різному тримають сигарети, різних форм і видів, знято в процесі паління, і 6766 випадків без паління. Кожен кліп був перевірено принаймні двома спостерігачами для забезпечення узгодженості. то дозволяє розробити більш гнучкі експерименти для оцінки продуктивності моделей комп'ютерного зору з використанням вибраних підмножин бази даних. Додатково для підвищення репрезентативності набору даних, до 70 оригінальним кліпам класу «паління» додаються ще 1500

кліпів, знятих самостійно, а також взятих із YouTube. Це дозволяє зробити навчальний та тестовий набори більш насиченими відеопослідовностями та більше ефективними для подання всіх видів паління.

Для оцінки точності розпізнавання паління на послідовності відео генеруються два окремих набори даних для навчання та тестування. Набори були створені таким чином, щоб гарантувати, що кліпи з одного і того ж відео не використовувалися ні для навчання, ні для тестування. Для кожного класу вибирається 70 навчальних та 30 тестових кліпів. На рис. 3.3 показано приклад кліпів класу «паління».



Рисунок 3.3 – Приклад відео класу «паління»

На рис. 3.4 представлений приклад кліпів, які не містять паління.

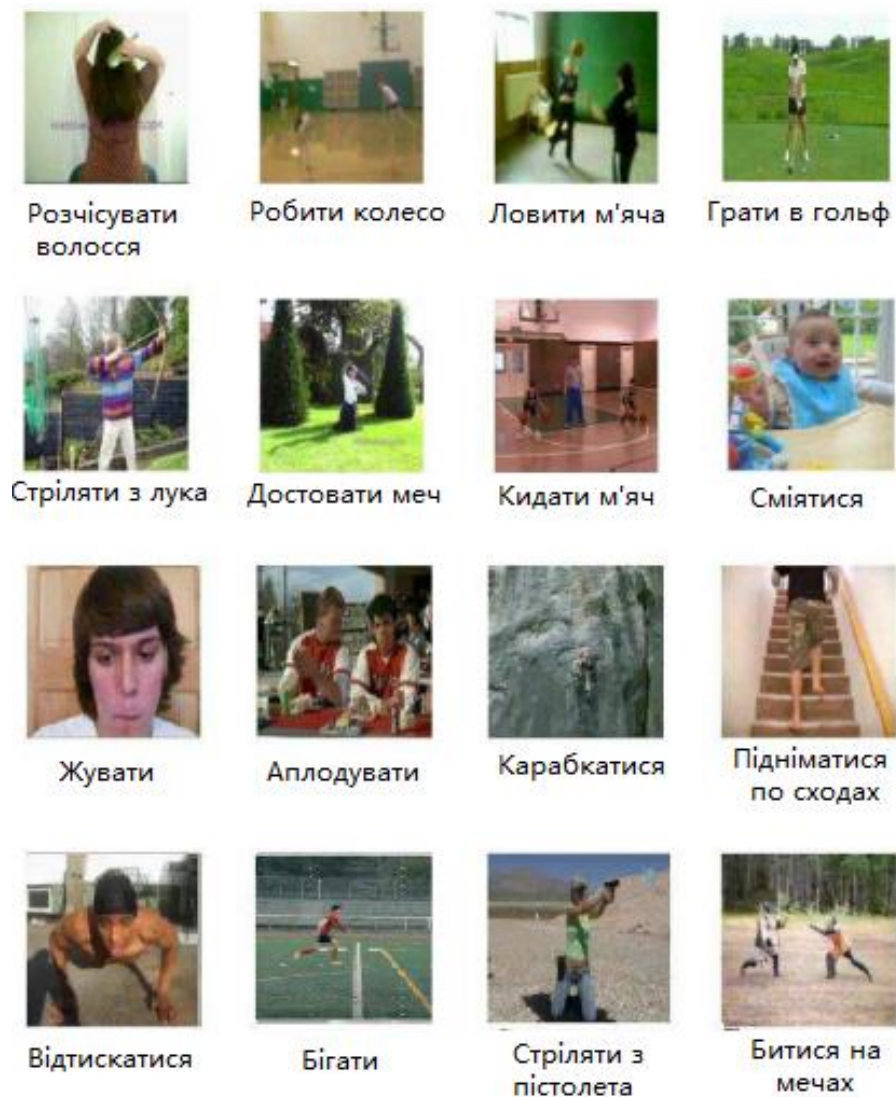


Рисунок 3.4 – Приклад кліпів із класу «не паління»

3.2.2 Експеримент

Для проведення експерименту використовувалося обладнання, інформація про яке наведено у таблиці 2.1. Для забезпечення роботи моделі виявлення паління по відеоданим, необхідні наступні характеристики: мінімум 2 ГБ графічної карти NVIDIA та встановлене ПЗ CUDA і cuDNN. Модель використовує Anaconda та пакети Python, включаючи OpenCV, matplotlib та Pytorch. Експериментальні дослідження проводилися при характеристиках ноутбука, показаних у таблиці 3.1.

Таблиця 3.1 – Конфігурації апаратного забезпечення для експерименту

Апаратне забезпечення	Конфігурації
Процесор	Intel(R) Core(TM) i7-6700HQ
Тактова частота процесора	2.60GHz
Об'єм пам'яті	8 GB
Операційна система	Windows 10
Графічний процесор	NVIDIA GeForce GTX 960M
Виділена пам'ять графічного процесора	2 GB

Оцінка точності моделі проводиться за допомогою метрик True Recognition (TR), False Rejection Rate (FRR) та False Acceptance Rate (FAR). TR дозволяє визначити точність розпізнавання паління, тобто, скільки розпізнано кадрів, що містять паління, дійсно містять дану дію відносно загальної кількості розпізнаних позитивних зразків. FRR, у свою чергу, показує помилково позитивну помилку, а FAR - помилково негативну помилку:

$$\begin{aligned}
 TR &= \frac{N_t}{N_t'}, \\
 FAR &= \frac{N_t - N_t'}{N}, \\
 FRR &= 1 - TR,
 \end{aligned}
 \tag{3.4}$$

де N_t – кількість кадрів із палінням; N_t' – кількість кадрів із палінням, яке розпізнано; N – кількість кадрів у відеопослідовності.

Отримана в результаті модифікації тривимірна згорткова НМГН, в результаті навчання на дейтасеті HMDB51 з розширенням класу «паління» шляхом додавання нових відеокліпів, дозволяє отримати точність розпізнавання, наведену у таблиці 3.2.

Таблиця 3.2 – Точність розпізнавання моделі

Епоха	Втрати при навчанні	Точність при навчанні	Втрати при перевірці	Точність при перевірці
1	1.1552	0.4329	0.7308	0.6699
2	0.9412	0.5801	0.5987	0.7346
3	0.8054	0.6504	0.5181	0.7613
4	0.7215	0.6966	0.4497	0.7984
5	0.6253	0.7572	0.4530	0.7984
...				
46	0.2325	0.9167	0.2024	0.9198
47	0.2284	0.9212	0.2058	0.9280
48	0.2261	0.9212	0.2448	0.9095
49	0.2170	0.9153	0.2259	0.9280
50	0.2109	0.9118	0.2267	0.9125

Графік на рис. 3.5 показує зміну втрати та точності розпізнавання паління на послідовності відео.

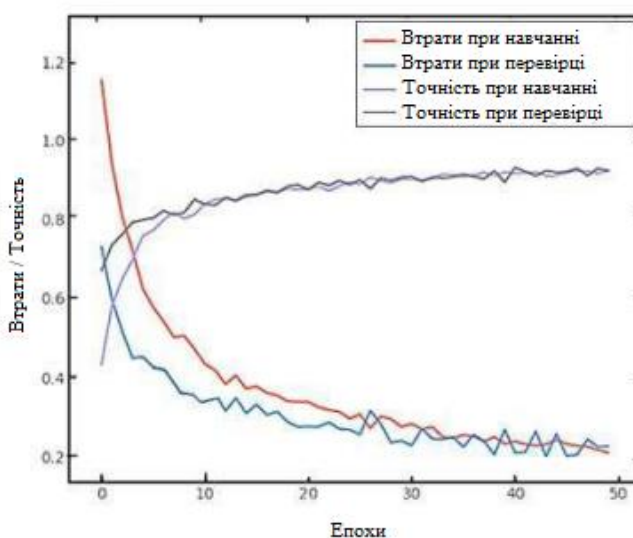


Рисунок 3.5 – Графік зміни втрати та точності під час навчання та перевірки

Таким чином, підсумкова точність на навчальному та тестовому наборах приблизно однакова і дорівнює 91%. У ДОДАТКУ Б представлені результати роботи алгоритму розпізнавання паління на відеопослідовності.

У таблиці 3.3 показані результати обчислень метрик TR, FAR, FRR відеопослідовностей, що брали участь в експерименті, описаних вище.

Таблиця 3.3 – Результати обчислень

Назва	TR (%)	FAR (%)	FRR (%)
Відео 1	87.8	12	12
Відео 3	90.7	9.2	9.1
Відео 5	88.8	11.1	11.1
Відео 7	97.4	2.4	2.5
Відео 8	92.4	7.5	7.5
Відео 10	85.9	14	14
Відео 12	98.2	1.7	1.7
Відео 14	100	0	0
Відео 16	95.4	4.5	4.5
Відео 18	92.5	7.4	7.4
Відео 20	84.3	1.5	1.5

3.3 Висновки до третього розділу

Розроблено та описано алгоритм розпізнавання паління на послідовності відео. Проведено навчання та тестування тривимірної ЗНМ на наборі даних HMDB51. Додатково репрезентативність тестового набору підвищена роликами знятими самостійно та взятими з YouTube. Точність розпізнавання паління за візуальними даними розрахована з застосуванням метрик TR, FAR, FRR. Середня точність склала приблизно 91%.

Також проведено експериментальні дослідження на 20 відеокліпах, серед яких присутні відеоролики, що містять дії, схожі з палінням. З результатів експериментальних досліджень можна зробити висновок, що шаблон паління, представлений у третьому розділі, дозволяє розпізнавати паління на відеопослідовності.

4 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ

4.1 Режими праці і відпочинку при роботі з ЕОМ

Нервово-емоційне напруження, втома очей, гіподинамія, підвищене навантаження на кисті верхніх кінцівок та хребет – усе це негативний вплив на організм людини при роботі з комп'ютером.

В даному випадку для збереження здоров'я працюючих, запобігання професійним захворюванням і підтримки працездатності передбачаються внутрішньозмінні регламентовані перерви для відпочинку [48].

Основним нормативно-правовим документом, який регламентує всі питання, пов'язані із охороною праці, в т.ч. і при роботі з ЕОМ, є [49].

Вимоги при роботі з ЕОМ визначають Державні санітарні правила і норми роботи з візуальними дисплейними терміналами електронно-обчислювальних машин - ДСанПІН 3.3.2.007-98.

Внутрішньозмінні режими праці і відпочинку містять додаткові нетривалі перерви в періоди, що передують появі стомлення і зниження працездатності.

При виконанні робіт, що належать до різних видів трудової діяльності, за основну роботу з візуальними дисплейними терміналами (ВДТ) слід вважати таку, що займає не менше 50% робочого часу. Впродовж робочої зміни мають передбачатися:

- перерви для відпочинку і вживання їжі (обідні перерви);
- перерви для відпочинку і особистих потреб (згідно з трудовими нормами);
- додаткові перерви, що вводяться для окремих професій з урахуванням особливостей трудової діяльності.

Тривалість обідньої перерви визначається чинним законодавством про працю і правилами внутрішнього трудового розпорядку підприємства, організації

чи установи. Як правило, тривалість такої перерви становить 40 – 60 хвилин. Тривалість та кількість інших внутрішньозмінних регламентованих перерв залежить від характеру трудової діяльності, напруженості і важності праці і визначається диференційовано для кожної професії.

За характером трудової діяльності розрізняють три професійні групи, згідно з діючим класифікатором професій [50]:

- розробники програм (інженери-програмісти) виконують роботу переважно з відеотерміналом та документацією при необхідності інтенсивного обміну інформацією з ЕОМ і високою частотою прийняття рішень;

- оператори ЕОМ виконують роботу, пов'язану з обліком інформації, одержаної з ВДТ за попереднім запитом, або тієї, що надходить з нього;

- оператор комп'ютерного набору виконує одноманітні за характером роботи з документацією та клавіатурою і нечастими нетривалими переключеннями погляду на екран дисплея, з введенням даних з високою швидкістю.

Правилами встановлюються такі внутрішньозмінні режими праці та відпочинку при роботі з ЕОМ при 8-годинній денній робочій зміні в залежності від характеру праці:

- для розробників програм із застосуванням ЕОМ слід призначати регламентовану перерву для відпочинку тривалістю 15 хвилин через кожну годину роботи за ВДТ;

- для операторів із застосуванням ЕОМ слід призначати регламентовані перерви для відпочинку тривалістю 15 хвилин через кожні дві години;

- для операторів комп'ютерного набору слід призначати регламентовані перерви для відпочинку тривалістю 10 хвилин після кожної години роботи.

У всіх випадках, коли виробничі обставини не дозволяють застосувати регламентовані перерви, тривалість безперервної роботи з ВДТ не повинна перевищувати 4 години.

При 12-годинній робочій зміні регламентовані перерви повинні встановлюватися в перші 8 годин роботи аналогічно перервам при 8-годинній робочій зміні, а протягом останніх чотирьох годин роботи, незалежно від характеру трудової діяльності, через кожну годину тривалістю 15 хвилин. Необхідно зазначити, що перерви під час роботи не повинні бути строго визначені за часом, а необхідно передбачати певний творчий індивідуальний підхід.

Для запобігання втоми під час деяких перерв доцільно виконувати спеціальні вправи, які наведені у ДСанПН 3.3.2.007-98. Також ці санітарні правила визначають Відстань від екрана до ока фахівців, які працюють за комп'ютером.

Окрім ДНАОП та ДСанНіП, які регламентують вимоги безпеки та санітарно-гігієнічні вимоги до обладнання робочих місць користувачів ВДТ, є ще ціла низка нормативних актів загального призначення, які необхідно враховувати під час організації роботи користувачів ВДТ. Важливим нормативним актом є [51].

Гігієнічна класифікація праці необхідна для оцінки конкретних умов та характеру праці на робочих місцях. На основі такої оцінки приймаються рішення, спрямовані на запобігання або максимальне обмеження впливу несприятливих виробничих факторів.

4.2 Вплив електромагнітного імпульсу (ЕМІ) ядерного вибуху на елементи виробництва та заходи захисту

У воєнний час при застосуванні ядерної зброї проти України на електронно-обчислювальне обладнання в першу чергу буде впливати ЕМІ ядерного вибуху у вигляді короткого імпульсу, який вражає головним чином електричну та електронну апаратуру. ЕМІ виникають в основному в результаті взаємодії гамма-випромінювання з атомами навколишнього середовища. На утворення ЕМІ йде невелика кількість ядерної енергії, але він здатен викликати високі імпульси

струмів та напруг в кабелях повітряних і підземних ліній зв'язку, сигналізації, управління, електропередачі, в антенах радіостанцій. Вплив ЕМІ може привести до згорання чутливих електронних та електричних елементів, зв'язаних з великими антенами чи відкритими дротами, а також до порушень в обчислювальних пристроях. Вплив ЕМІ необхідно враховувати для всіх електричних та електронних систем. Для найбільш важливих приладів треба використовувати засоби захисту і підвищувати їх стійкість до ЕМІ [48].

Особливістю ЕМІ, як вражаючого фактору є його здатність розповсюджуватись на десятки і сотні кілометрів в оточуючому середовищі. Тому ЕМІ може вплинути своєю дією на об'єкти, там де вибухова хвиля, світлове випромінювання, проникаюча радіація втрачають своє значення, як вражаючі фактори. При наземних та низьких повітряних вибухах в лініях зв'язку та електрозабезпечення виникають напруги, які можуть викликати пробій ізоляції провідників та кабелів відносно землі, пробій ізоляції елементів приладів підключених до повітряних і підземних ліній. Степінь враження залежить від наведеного імпульсу напруги чи струму і також електричної міцності обладнання.

Найбільш піддані впливу ЕМІ системи зв'язку, сигналізації, управління. Використані в цих системах кабелі та апаратура мають обмежену електричну міцність не більше 10кВ імпульсної напруги, тоді як наведені імпульси напруги від ЕМІ можуть перевищувати ці значення. Найбільш піддана впливу ЕМІ апаратура виконана на напівпровідниках та інтегральних схемах, працюючих на малих струмах і напругах, і значить відчутних до впливу зовнішніх електричних і магнітних кіл, в тому числі і елементи програмного засобу для управління процесом міграції віртуальних машин в обчислювальній хмарі. ЕМІ пробиває ізоляцію, спалює елементи електричних схем радіоапаратури, викликає коротке замикання в радіопристроях, іонізацію діелектриків, змінює або повністю стирає магнітний запис. Встановлено, що при дії ЕМІ на апаратуру найбільша напруга наводиться на вході. В транзисторах відбувається така залежність: чим більший коефіцієнт підсилення транзистора, тим менша його електрична міцність.

ЕМІ пошкоджує також резистори, викликає іскріння в їх міжконтактних з'єднаннях і деяких областях провідної поверхні. Найбільшу небезпеку ЕМІ представляє для апаратури, яка встановлена в особливо міцних спорудах, які витримують великі тиски ударної хвилі. В цих спорудах апаратура не виходить з ладу від механічних пошкоджень, але ЕМІ може вивести з ладу всю незахищену апаратуру системи зв'язку, сигналізації і керування. Найбільших значень досягають напруги, які наводяться між кабелем і землею. Напруженість електромагнітного поля всередині споруди в деяких випадках недостатня для того, щоб вивести з ладу апаратуру, але такі поля в змозі викликати короткочасний збій роботи радіотехнічних пристроїв.

Розглянемо можливі шляхи рішення задачі захисту від ЕМІ. Ідеальним захистом від ЕМІ виявилось б повне укриття приміщення, в якому розміщена радіоелектронна апаратура, металевим екраном [52]. Водночас зрозуміло, що практично забезпечити такий захист у ряді випадків неможливо, тому що для роботи апаратури часто потрібно забезпечити її електричний зв'язок із зовнішніми пристроями. Тому використовуються менш надійні засоби захисту, такі, як струмопровідні сітки, або плівкові покриття для вікон, щільникові металеві конструкції для повітрезабірників і вентиляційних отворів і контактні пружинні прокладки, розміщені по периметру дверей і люків.

Більш складною технічною проблемою рахується захист від проникнення ЕМІ в апаратуру через різноманітні кабельні входи. Радикальним рішенням даної проблеми міг би стати перехід від електричних мереж зв'язку до практично не схильних до впливу ЕМІ волоконно-оптичних. Проте заміна напівпровідникових приладів у всьому спектрі виконуваних ними функцій електронно-оптичними пристроями можлива тільки у віддаленому майбутньому. Тому в даний час в якості засобів захисту кабельних входів найбільш широко використовуються фільтри, у тому числі волоконні, а також іскрові розрядники, металлоокисні варистори і високошвидкісні зенеровські діоди [52].

Всі ці засоби мають як переваги, так і недоліки. Так, ємнісно-індуктивні фільтри достатньо ефективні для захисту від ЕМІ малої інтенсивності, волоконні фільтри захищають у відносно вузькому діапазоні надвисоких частот. Іскрові розрядники мають значну інерційність й в основному придатні для захисту від перевантажень, що виникають під впливом напруг і струмів, що наводяться в обшивці літака, кожусі апаратури й оплетенні кабеля.

Металоокисні варистори є напівпровідниковими приладами, що різко підвищують свою провідність при високій нарузі. Проте, при застосуванні цих приладів у якості засобів захисту від ЕМІ варто враховувати їх недостатньо високу швидкодію і погіршення характеристик при кількаразовому впливі навантажень. Ці недоліки відсутні у високошвидкісних зенеровських діодах, дія яких заснована на різкій лавиноподібній зміні опору від високого значення практично до нуля, при перевищенні прикладеної до них напруги граничного розміру. Крім того на відміну від варисторів характеристики зенеровських діодів після багатократних впливів високих напруг і переключень режимів не погіршуються.

Найбільш раціональним підходом до проектування засобів захисту від ЕМІ кабельних входів є створення таких роз'ємів у конструкції яких передбачені спеціальні заходи, що забезпечують формування елементів фільтрів і установку вмонтованих зенеровських діодів. Подібне рішення сприяє одержанню дуже малих значень ємності й індуктивності, що необхідно для забезпечення захисту від імпульсів, що мають незначну тривалість і, отже, потужну високочастотну складову. Використання роз'ємів подібної конструкції дозволить вирішити проблему обмеження малогабаритних характеристик пристрою захисту. Складність рішення задачі захисту від ЕМІ і висока вартість розроблених для цього засобів і методів змушують піти по шляху їхнього вибіркового застосування в особливо важливих системах зброї і військової техніки. Такий же шлях обраний і для захисту систем, що мають велику протяжність, керування і зв'язку. Проте основним методом вирішення

проблеми спеціалісти вважають створення так званих розподілених мереж зв'язку [48].

4.3 Висновки до четвертого розділу

В цьому розділі розглянуто важливі питання охорони праці та безпеки в надзвичайних ситуаціях, зокрема описані вимоги до режимів праці і відпочинку при роботі з ЕОМ, а також вплив ЕМІ ядерного вибуху на елементи виробництва та заходи захисту.

ВИСНОВКИ

В кваліфікаційній роботі були отримані наступні результати:

- виконано аналіз існуючих методів виявлення паління за візуальними даними;
- для розпізнавання паління за візуальними даними обрані просторово-часові ознаки, що дозволяють виявляти паління незалежно від відстані до людини, що палить, її пози, статі, віку, видимості сигарети у кадрі, тобто навіть за нерегулярної форми паління;
- модифіковано архітектуру ЗНМ ResNet до тривимірної ЗНМ, що забезпечує роботу з відеопослідовності та облік просторово-часових ознак паління;
- запропоновано алгоритм розпізнавання паління за візуальними даними;
- виконано передобробку кадрів, за допомогою методів середнього віднімання та масштабування для нормалізації оброблюваних відеозображень, з метою підвищення якості роботи алгоритму розпізнавання паління;
- проведені експериментальні дослідження на 20 відеопослідовностях, отримані в реальних умовах зйомки, підтверджують ефективність запропонованого способу розпізнавання паління;
- розпізнавання паління виконується в режимі реального часу;
- середнє значення точності виявлення становило 90%.

ПЕРЕЛІК ДЖЕРЕЛ

1. Всесвітня організація охорони здоров'я. Доповідь ВООЗ про глобальну тютюнову епідемію, 2020 р. – Попередження про небезпеки тютюну. [Електронний ресурс]: – Режим доступу: <https://www.who.int/fctc/ua/> (дата звернення 21.11.2021).
2. Всесвітня організація охорони здоров'я: Стаття 8 Рамкової конвенції Всесвітньої Організації Охорони Здоров'я. [Електронний ресурс] - Режим доступу: https://www.who.int/tobacco/global_report/2013/summary/ru/ (дата звернення 21.11.2021).
3. Jordan, M. I. Machine learning: Trends, perspectives, and prospects // American Association for the Advancement of Science. – 2015. – № 6245. – pp. 255-260/
4. Deo, R. C. Machine learning in medicine // Circulation. – 2015. – R. 132. – №. 20. – p. 1920-1930.
5. Wuest, T. Machine learning in manufacturing: advantages, challenges, and applications // Production & Manufacturing Research. – 2016. – R. 4. – №. 1. – p. 23-45.
6. Hsieh, W.W. Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels: book – Cambridge: Cambridge University Press, 2009. – 349 p.
7. Ranking the risk factors for cesarean: logistic regression analysis of a nationwide study // S. Mor-Yosef, A. Samueloff, B. Modan, D. Navot, J. G. Schenker. – Obstetrics & Gynecology. – 1990. – R. 75. – №. 6. – p. 944-947.
8. Goodfellow I. Deep learning: book – MIT press, 2016.
9. Galton, F. J. Regression towards mediocrity in hereditary stature // Journal of the Anthropological Institute / Institute of Great Britain and Ireland. – 1885. – p. 246–263.

10. Estivill-Castro V. Why so many clustering algorithms: a position paper //ACM SIGKDD explorations newsletter. – 2002. – R. 4. – №. 1. – p. 65-75.
11. Kaelbling L. P., Littman M. L., Moore A. W. Reinforcement learning: A survey //Journal of artificial intelligence research. – 1996. – R. 4. – p. 237-285.
12. Wu P. Human smoking event detection using visual interaction clues //2010 20th International Conference on Pattern Recognition. – IEEE, 2010. – p. 4344-4347.
13. Smoking-faces: This is a command line tool that analyze videos and detects if a person is smoking. [Электронный ресурс] - Режим доступа: <https://github.com/chok68/smoking-faces/blob/master/smokingdetection.py> (дата звернення 25.11.2021).
14. Odetallah, A. D., Agaian S. S. Human visual system-based smoking event detection //Mobile Multimedia/Image Processing, Security, and Applications 2012. – International Society for Optics and Photonics, 2012. – R. 8406. – p. 840607.
15. Dunne, É. Smoking Detection in Video Footage : ʘʘʘʘʘ. ʘʘʘ. – School of Computer Science and Statistics Smoking Detection in Video Footage. A Dissertation submitted in partial fulfilment of the requirements for the degree of MAI (Computer Engineering). – Submitted to the University of Dublin, Trinity College, 2018. – 43 p.
16. Poonam, G., Shashank B. N., Athri G. R. Development of framework for detecting smoking scene in video clips //Indonesian Journal of Engineering and Computer Science. – 2019. – p. 22-26.
17. Iwamoto, K. et al. Cigarette smoke detection from captured image sequences //Image Processing: Machine Vision Applications III. – International Society for Optics and Photonics, 2010. – R. 7538. – p. 753813.
18. Iwamoto, K. et al. Cigarette Smoke Detection Using Feature Values Based on The Kernel LMS Algorithm // IEICE technical report. Circuits and systems 109(434), 2010. – p. 247-248
19. Bien, T. L., Lin C. H. Detection and recognition of indoor smoking events //Fifth International Conference on Machine Vision (ICMV 2012): Algorithms,

Pattern Recognition, and Basic Technologies. – International Society for Optics and Photonics, 2013. – R. 8784. – p. 878424.

20. Cortes C., Vapnik V. Support-vector networks //Machine learning. – 1995. – R. 20. – №. 3. – p. 273-297. 85

21. Михайлів М.В. Попередня обробка відеозображень з використанням нейронних мереж // Інформаційні моделі, системи та технології: Праці ІХ наук.-техн. конф. (Тернопіль, 08-09 грудня 2021 р.) Тернопіль, 2021. – С. 58.

22. Maglogiannis I. G. (ed.). Emerging artificial intelligence applications in computer engineering: real word ai systems with applications in ehealth, hci, information retrieval and pervasive technologies. – Ios Press, 2007. – p. 160.

23. Cowan J. D. Discussion: McCulloch-Pitts and related neural nets from 1943 to 1989 //Bulletin of mathematical biology. – 1990. – R. 52. – №. 1-2. – p. 73- 97.

24. Rosenblatt F. The perceptron, a perceiving and recognizing automaton Project Para. – Cornell Aeronautical Laboratory, 1957.

25. Yasoubi A., Hojabr R., Modarressi M. Power-efficient accelerator design for neural networks using computation reuse //IEEE Computer Architecture Letters. – 2016. – R. 16. – №. 1. – p. 72-75.

26. Krizhevsky A., Sutskever I., Hinton G. E. Imagenet classification with deep convolutional neural networks //Advances in neural information processing systems. – 2012. – p. 1097-1105.

27. Le Cun Y. et al. Handwritten digit recognition: Applications of neural network chips and automatic learning //IEEE Communications Magazine. – 1989. – R. 27. – №. 11. – p. 41-46.

28. Yamashita R. et al. Convolutional neural networks: an overview and application in radiology //Insights into imaging. – 2018. – R. 9. – №. 4. – p. 611-629.

29. Voulodimos A. et al. Deep learning for computer vision: A brief review //Computational intelligence and neuroscience. – 2018. – R. 2018.

30. Nash W., Drummond T., Birbilis N. A review of deep learning in the study of materials degradation //npj Materials Degradation. – 2018. – R. 2. – №. 1. – p. 1-12.
31. Padmavathi G., Shanmugapriya D., Kalaivani M. Video pre-processing of image information for vehicle identification. – 2011.
32. Marchant J. A. 5.4 Machine Vision in the Agricultural Context. – 2006.
33. BlobFromImage. OpenCV. [Электронный ресурс] - Режим доступа: <https://docs.opencv.org> (дата звернення 29.11.2021).
34. Bradski G., Kaehler A. Learning OpenCV: Computer vision with the OpenCV library. – " O'Reilly Media, Inc.", 2008.
35. Simonyan K., Zisserman A. Very deep convolutional networks for largescale image recognition // arXiv preprint arXiv:1409.1556. – 2014.
36. He K. et al. Deep residual learning for image recognition // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – p. 770-778.
37. Russakovsky O. et al. Imagenet large scale visual recognition challenge // International journal of computer vision. – 2015. – R. 115. – №. 3. – p. 211-252.
38. He K. et al. Identity mappings in deep residual networks // European conference on computer vision. – Springer, Cham, 2016. – p. 630-645.
39. Ning F. et al. Toward automatic phenotyping of developing embryos from videos // IEEE Transactions on Image Processing. – 2005. – R. 14. – №. 9. – p. 1360-1371.
40. Dubey S., Boragule A., Jeon M. 3D ResNet with Ranking Loss Function for Abnormal Activity Detection in Videos // 2019 International Conference on Control, Automation and Information Sciences (ICCAIS). – IEEE, 2019. – p. 1-6.
41. Ji S. et al. 3D convolutional neural networks for human action recognition // IEEE transactions on pattern analysis and machine intelligence. – 2012. – R. 35. – №. 1. – p. 221-231.

42. Yu G., Li T. Recognition of human continuous action with 3D CNN // International Conference on Computer Vision Systems. – Springer, Cham, 2017. – p. 314-322.
43. Bottou L. Large-scale machine learning with stochastic gradient descent // Proceedings of COMPSTAT'2010. – Physica-Verlag HD, 2010. – p. 177-186.
44. Vapnik V. N., Chervonenkis A. Y. On the uniform convergence of relative frequencies of events to their probabilities // Measures of complexity. – Springer, Cham, 2015. – p. 11-30.
45. Dennis Jr J. E. RB Schnabel Numerical methods for unconstrained optimization and nonlinear equations. – 1983.
46. Sari E., Nia V. P. Understanding BatchNorm in Ternary Training // Journal of Computational Vision and Imaging Systems. – 2019. – R. 5. – №. 1. – p. 21-28.
47. Kuehne H. et al. HMDB: a large video database for human motion recognition // 2011 International Conference on Computer Vision. – IEEE, 2011. –pp. 2556-2563.
48. Зеркалов Д.В. Безпека життєдіяльності та основи охорони праці. Навчальний посібник. К.: «Основа». 2016. 267 с.
49. Закон України «Про охорону праці». [Електронний ресурс] – Режим доступу: <https://zakon.rada.gov.ua/laws/show/2694-12> - (дата звертання 06.12.2021).
50. Класифікатор професій ДК 003:2010/ [Електронний ресурс] – Режим доступу: <https://zakon.rada.gov.ua/rada/show/va327609-10> - (дата звертання 07.12.2021).
51. Гігієнічна класифікація умов праці за показниками шкідливості та небезпечності факторів виробничого середовища, важкості та напруженості трудового процесу. – К.: МОЗ України, 1998. – 34 с.
52. Сакевич В.Ф., Поліщук О.В. Цивільна оборона. Теоретичні основи. Навчальний посібник. — Вінниця : ВНТУ, — 2009. — 136 с.

ДОДАТКИ

ДОДАТОК А
Тези конференції

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ТЕРНОПЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ІВАНА ПУЛЮЯ

МАТЕРІАЛИ

ІХ НАУКОВО-ТЕХНІЧНОЇ КОНФЕРЕНЦІЇ
«ІНФОРМАЦІЙНІ МОДЕЛІ,
СИСТЕМИ ТА ТЕХНОЛОГІЇ»



8–9 грудня 2021 року

ТЕРНОПЛЬ
2021

В.О. Колодій, В.Г. Онуцький АНАЛІЗ МЕТОДІВ ДОСЛІДЖЕННЯ ВІБРАЦІЙНОЇ СТІЙКОСТІ ЕЛЕКТРОУСТАНОВОК V.O. Kolodiy, V.G. Onutsky ANALYSIS OF METHODS FOR STUDYING THE VIBRATION RESISTANCE OF ELECTRICAL INSTALLATIONS	49
О.О. Лішчук, Д.А. Радчук, Т.В. Зашчук РОЗУМНІ МІСТА ТА ІНТЕРНЕТ РЕЧЕЙ O.O. Lishchuk, D.A. Radchuk, T.B. Zashchuk SMART CITIES AND THE INTERNET OF THINGS	50
Д.І. Матічук, В.В. Нукитюк ОНЛАЙН-ІНСТРУМЕНТ GOOGLE SHEETS ДЛЯ СИСТЕМАТИЗОВАНИХ КОНСОЛІДОВАНИХ ДАНИХ ВАКЦІНАЦІЇ НЕМОВЛЯТ D. Matsyuk, V. Nykutyuk GOOGLE SHEETS ONLINE TOOL FOR SYSTEMATIZED CONSOLIDATED INFAN VACCINATION DATA	51
М. Мандзій, І. Поліщук, Р. Контсогограда, І. Дедів ЗАДАЧА ОПТИМАЛЬНОГО ВІЯВЛЕННЯ СИГНАЛІВ В СУМІШІ ІЗ ЗАВАДАМИ В ОБЛАСТІ РАДІОТЕХНІКИ M. Mandziy, I. Polishchuk, P. Kotsyograda, I. Dediv THE PROBLEM OF OPTIMAL DETECTION OF SIGNALS IN MIXTURE WITH INTERFERENCES IN THE FIELD OF RADIO ENGINEERING	52
Л. Матітчук, І. Павлов, В. Сташук ТЕОРЕТИЧНЕ ОБГРУНТУВАННЯ МЕТОДУ ВІЯВЛЕННЯ КОМП'ЮТЕРНИХ АТАК L. Matiychuk, I. Pavlov, V. Stashuk THEORETICAL JUSTIFICATION OF THE METHOD OF DETECTION OF COMPUTER ATTACKS	53
Л. Матітчук, І. Павлов, В. Сташук ОЦІНКА ІСНУЮЧИХ СИСТЕМ ВІЯВЛЕННЯ АТАК L. Matiychuk, I. Pavlov, V. Stashuk EVALUATION OF EXISTING ATTACK DETECTION SYSTEMS	55
А.В. Мельничук МЕТОДИ ЗАХИСТУ ІНФОРМАЦІЇ В РАМКАХ ПРЕДМЕТНО- ОРІЄНТОВАНОГО ПРОЄКТУВАННЯ ІНФОРМАЦІЙНИХ СИСТЕМ A.V. Melnychuk INFORMATION PROTECTION METHODS WITHIN DOMAIN-DRIVEN DESIGN OF THE INFORMATION SYSTEM	57
М.В. Михайлів ПОПЕРЕДНЯ ОБРОБКА ВІДЕОЗОБРАЖЕНЬ З ВИКОРИСТАННЯМ НЕЙРОННИХ МЕРЕЖ M.V. Mykhayliv PRE-PROCESSING OF VIDEO IMAGES USING NEURAL NETWORKS	58
О. Данильчів, А. Хом'як, Т. Назаревич ВИКОРИСТАННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ ДОСЛІДЖЕННЯ СТАНУ РОСЛИН В РОЗУМНИХ ТЕПЛИЦЯХ O. Danylytsiv, A. Khomyak, T. Nazarevych THE USE OF NEURAL NETWORKS FOR STUDY THE CONDITION OF PLANTS IN SMART GREENHOUSES	59

УДК 004.852

М.В. Михайлів

(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

ПОПЕРЕДНЯ ОБРОБКА ВІДЕОЗОБРАЖЕНЬ З ВИКОРИСТАННЯМ НЕЙРОННИХ МЕРЕЖ

UDC 004.852

M.V. Mykhayliv

PRE-PROCESSING OF VIDEO IMAGES USING NEURAL NETWORKS

Відео, отримані в реальних умовах зйомки, можуть містити різні артефакти, що виникають внаслідок шумів обладнання або через складні погодні умови, тому якість роботи нейронних мереж може суттєво погіршуватися. Для того, щоб отримати правильні прогнози із застосуванням нейронних мереж глибокого навчання, потрібно попередньо обробити дані, усуваючи ці шуми. Попередня обробка зображень включає широкий спектр операцій, їх призначення – спрощення чи поліпшення наступного аналізу. Попередня обробка підвищує якість зображення, усуваючи ненавмисні спотворення або покращуючи деякі функції зображення, які важливі для подальшої обробки та створення більш відповідного зображення, ніж оригінал для конкретного завдання. Дії, що можуть бути виконані на цифрових зображеннях, включають точкові, локальні чи «сусідні» та глобальні операції. Точкові операції перетворюють пікселі незалежно від сусідніх пікселів. Значення сірого кольору для вихідного зображення у конкретному пікселі залежать тільки від значення сірого для того ж пікселя у вхідному зображенні. Вони відтворюють пікселі в одному зображенні, щоб сформувати інше, використовуючи одну функцію відображення.

Точкові операції не враховують просторову організацію зображення. Приклади цих операцій включають розтягування контрасту, сегментацію на основі значення сірого та вирівнювання гистограми, також середнє піднімання та масштабування за деякими факторами. У контексті глибокого навчання та класифікації відео розглянемо такі методи попередньої обробки: середнє піднімання; масштабування.

OpenCV – бібліотека з відкритим вихідним кодом, що має програмну ліцензію університету Берклі, яка включає кілька сотень алгоритмів комп'ютерного зору. Нейронна мережа застосовує модуль OpenCV, містить дві функції, які можна використовувати для попередньої обробки кадрів та підготовки їх до класифікації за допомогою моделей глибокого навчання. Тому середнє піднімання розглядається як метод, що використовується для покращення роботи згорткових нейронних мереж. Перш ніж приступити до тренування нейронних мереж глибокого навчання, необхідно обчислити середню інтенсивність кольору для кожного з червоного, зеленого та синього каналів за всіма зображеннями в навчальному наборі. Зазвичай результуючі значення являють собою три кортежі, що складаються із середнього значення червоного, зеленого та синього каналів відповідно. Наприклад, середні значення для навчального набору ImageNet є такими: $R = 103.93$, $G = 116.77$ та $B = 123.68$. Для набору даних, котрий застосовується у роботі, використовуються такі значення: $R = 114.7748$, $G = 107.7354$, $B = 99.4750$.

Однак у деяких випадках середні значення кольорів можуть бути розраховані по каналах, а не за кольорами, що призводить до матриці $M \times N$. У цьому випадку обчислюється матриця $M \times N$ для кожного каналу, потім віднімається з вхідного зображення під час навчання або тестування. Обидва методи є формами середнього піднімання, але версія, у якій застосовується інтенсивність кольору, використовується частіше, особливо для великих наборів даних.

Додаток Б

Результати роботи алгоритму розпізнавання паління на відеопослідовності

№ з/п	Характеристики	Приклади кадрів	
1.	Тривалість: 4,47 сек Роздільна здатність: 1280×720 Кількість кадрів:107		
2.	Тривалість: 2,97 сек Роздільна здатність: 1280×720 Кількість кадрів:71		
3.	Тривалість: 15,5 сек Роздільна здатність: 1920×1088 Кількість кадрів:464		
4.	Тривалість: 2,63 сек Роздільна здатність: 1920×1080 Кількість кадрів:78		
5.	Тривалість: 7,93 сек Роздільна здатність: 1280×720 Кількість кадрів:198		
6.	Тривалість: 2,23 сек Роздільна здатність: 1280×720 Кількість кадрів:53		
7.	Тривалість: 4,07 сек Роздільна здатність: 1920×1080 Кількість кадрів:121		

Продовження Додатку Б

8.	Тривалість: 4,87 сек Роздільна здатність: 1280×720 Кількість кадрів:146		
9.	Тривалість: 4,10 сек Роздільна здатність: 406×720 Кількість кадрів:44		
10.	Тривалість: 4,30 сек Роздільна здатність: 1920×1088 Кількість кадрів:128		
11.	Тривалість: 1,53 сек Роздільна здатність: 408×720 Кількість кадрів:107		
12.	Тривалість: 3,90 сек Роздільна здатність: 270×360 Кількість кадрів:117		
13.	Тривалість: 3,00 сек Роздільна здатність: 1080×1920 Кількість кадрів:180		
14.	Тривалість: 17,3 сек Роздільна здатність: 480×360 Кількість кадрів:518		
15.	Тривалість: 5,07 сек Роздільна здатність: 1280×720 Кількість кадрів:151		