

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА
Факультет інформаційних технологій
Кафедра інтелектуальних технологій**

**КВАЛІФІКАЦІЙНА РОБОТА
на здобуття освітнього ступеня «магістр»
НА ТЕМУ:
Система розпізнавання частково закритих облич
з використанням нейронних мереж**

Галузь знань: 12 «Інформаційні технології»
Спеціальність: 122 «Комп'ютерні науки»
Освітньо-наукова програма «Технології штучного інтелекту»

Виконала:
студентка 2 курсу магістратури, групи ТШІ-21

Абрамова Марія Вікторівна
(ПІБ)
Самохвалов Юрій Якович
(ПІБ)
д.т.н, професор
(науковий ступінь, вчене звання)

Науковий керівник:

Засвідчую, що в цій кваліфікаційній роботі
немає запозичень з праць інших авторів без
відповідних посилань

Студентка)

підпис

Кваліфікаційна робота допущена до захисту
рішенням кафедри *інтелектуальних технологій*

Протокол № ____ від «____» травня 2022 р.

Зав. кафедри _____ доц. Іларіонов О.Є.
підпис

Київ 2022

РЕФЕРАТ

Кваліфікаційна робота складається зі вступу, 4 розділів, висновків, списку використаної літератури із 43 джерел та 1 додатку. Загальний обсяг роботи 74 сторінки. Робота містить 4 таблиці та 34 рисунки.

Актуальність теми. За останні роки системи розпізнавання обличчя на основі технологій штучного інтелекту, досягли високого ступеня досконалості. Проте досі існує ряд чинників, які негативно впливають на результат розпізнавання. Одним з таких чинників є часткове закриття обличчя сторонніми елементами, що в реальних умовах зустрічається дуже часто. Найбільш суттєво впливає закриття нижньої частини обличчя або навколо окулярної зони. Закриття нижньої частини обличчя в реальних умовах часто пов'язане з використанням засобів індивідуального захисту, що стало масовим явищем в період пандемії COVID-19. Масове носіння масок стало перешкодою для систем ідентифікації. Крім того, маска може слугувати засобом маскування з метою приховування особистості, що дає широкі можливості для здійснення кримінальних правопорушення, терористичних та диверсійних атак. З початку пандемії було розроблено багато ефективних систем детекції облич, проте ефективність ідентифікації ще має місце для покращення та дослідження. Тому ідентифікація частково закритих обличч людей є важливою та актуальною темою.

Мета роботи. Метою кваліфікаційної роботи є проведення дослідження існуючих методів та технологій ідентифікації обличч та створення системи для розпізнавання осіб з частково закритими обличчями, а саме засобами індивідуального захисту, з можливістю подальшого використання в сфері безпеки.

Об’єкт дослідження. Процес розпізнавання частково закритих облич людей.

Предмет дослідження. Предметом дослідження є методи та технології розпізнавання людей з частково закритими обличчями.

Результат роботи. Досліджено існуючі методи розпізнавання облич, розроблено систему ідентифікації осіб із закритою нижньою частиною обличчя, при наявності зображень даної особи в базі даних. Досягнута висока точність розпізнавання, мінімізовано хибно-позитивні та хибно-негативні результати розпізнавання.

Апробація результатів дослідження. Основні положення кваліфікаційної роботи були представлені на науково-практичній конференції «Інноваційні наукові дослідження в умовах світових змін» (м. Вінниця, 29 – 30 квітня 2022 р) та опубліковані у збірнику матеріалів цієї конференції.

Ключові слова: Штучний інтелект. Методи машинного навчання. Нейронні мережі. Глибинне навчання. Розпізнавання облич. Ідентифікація особи. Згортова нейронна мережа. Ембедінг. Python. PyTorch.

Моделі: CNN, DCNN, ResNet, InceptionResNet, MTCNN, FaceNet.

ABSTRACT

The qualification work consists of an introduction, 4 chapters, a conclusion, a list of references (43 sources) and 1 application. The total volume of work is 74 pages. The work contains 4 tables and 34 figures.

The actuality of the theme. In recent years, face recognition systems based on artificial intelligence technologies have reached a high level of perfection. However, there are still a number of factors that negatively affect the result of recognition. One of them is the partial occlusion of the face by some elements, this situation is very common in the wild. The occlusion of the lower part of the face or the ocular area has the most significant effect. Blocking of the lower part of the face in natural conditions is often associated with the use of personal protective equipment, which became a mass phenomenon during the pandemic COVID-19. Regular wearing of masks has become an obstacle for identification systems. In addition, the mask can serve as a means of disguise to conceal the identity, which provides additional opportunities for criminal offences, terrorist attacks and sabotage. Many effective facial detection systems have been developed since the pandemic, but the effectiveness of identification still needs to be improved and researched. Therefore, the identification of partially occluded faces is an important and relevant topic.

The purpose of the qualification work is to research the existing methods and technologies of face identification and create a system for recognizing persons with partially occluded faces, namely personal protective equipment, with the possibility of further use in the security sphere.

The object of the research is the process of face recognition with partial occlusions.

The subject of the research is the recognition methods and technologies of people with partially occluded faces.

Results of the work. The existing face recognition methods have been studied, and a system of identification of persons with an occluded lower part of the face has been developed. It identifies a person in case the presence of images of this person is in the database. High recognition accuracy is achieved, and false-positive and false-negative recognition results are minimized.

Publications. The main results of the qualification work were presented as a thesis at the scientific-practical conference "Innovative research in the context of world change" (Vinnytsia, April 29 - 30, 2022) and published in the materials of this conference.

Keywords: Artificial Intelligence. Methods of machine learning. Neural networks. Deep learning. Face recognition. Face identification. Convolutional neural network. Embedding. Python. PyTorch.

Models: CNN, DCNN, ResNet, InceptionResNet, MTCNN, FaceNet.

ЗМІСТ

ЗМІСТ	6
ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ	8
ВСТУП	9
РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ. ПОСТАНОВКА ЗАВДАННЯ КВАЛІФІКАЦІЙНОЇ РОБОТИ	11
1.1 Опис проблеми, що вирішується, та її актуальність	11
1.2 Аналіз стану вирішення проблеми.	13
1.2.1 Аналіз існуючих технологій розпізнавання облич	13
1.2.2 Аналіз існуючих технологій частково закритих облич	17
1.3 Аналіз методів та технологій в системах розпізнавання облич	23
1.3.1 Методи на основі комплексного підходу	24
1.3.2 Методи на основі локального підходу	25
1.3.3 Технології глибинного навчання	26
1.4 Постановка завдання розпізнавання частково закритих облич	27
1.5 Висновки до першого розділу	28
РОЗДІЛ 2. ДОСЛІДЖЕННЯ МЕТОДІВ ТА ТЕХНОЛОГІЙ РОЗПІЗНАВАННЯ	30
2.1 Методи та технології розв’язання завдань дослідження	30
2.1.1 Нейронна мережа ResNet	32
2.1.2 Нейрона мережа Inception	33
2.1.3 Нейрона мережа Inception-ResNet	34
2.1.4 Багатозадачні каскадні згорткові нейронні мережі	35
2.1.5 Система розпізнавання облич FaceNet	38

	7
2.1.6 Технологія трансферного навчання	40
2.1.7 Метод вирівнювання зображень	41
2.2 Висновки до другого розділу	42
РОЗДІЛ 3. РОЗРОБКА СИСТЕМИ РОЗПІЗНАВАННЯ ЧАСТКОВО ЗАКРИТИХ ОБЛИЧ	43
3.1 Обґрунтування та вибір початкових даних для розробки програмного продукту	43
3.2 Розробка підходу до розпізнавання частково закритих облич	45
3.3 Процес навчання нейронної мережі	50
3.3.1 Програмна реалізація навчання нейронної мережі	50
3.3.2 Вибір моделі нейронної мережі	51
3.4 Функції та структура системи розпізнавання	53
3.4.1 Функції системи розпізнавання	53
3.4.2 Структура системи розпізнавання	56
3.5 Висновки до третього розділу	58
РОЗДІЛ 4. ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ ОТРИМАНИХ РЕЗУЛЬТАТІВ	60
4.1 Характеристика оцінок ефективності результатів розпізнавання частково закритих облич	60
4.2 Опис результатів роботи на основі контрольного прикладу	62
4.3 Оцінка ефективності результатів дослідження	63
4.4 Висновки до четвертого розділу	65
ВИСНОВКИ	66
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	68
Додаток А: Приклади зображень з датасетів	72

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

ЗІЗ	засоби індивідуального захисту
ШІ	штучний інтелект
CNN / ЗНМ	згорткова нейронна мережа
COVID-19, ковід, коронавірус	COronaVirus Disease 2019, коронавірусна хвороба 2019
CPU	центральний процесор
DCNN	глибинна згорткова нейронна мережа
FTE	Failure to Enroll
GCP	google cloud platform
GPU	графічний процесор
MTCNN	багатозадачні каскадні згорткові нейронні мережі
NIST	Національний інститут стандартів і технології
PCA	метод головних компонент
VM	віртуальна машина

ВСТУП

Обличчя є одними з найважливіших джерел інформації для людського ока. Саме завдяки зору ми ідентифікуємо та отримуємо перші дані про людину. Отримується інформація щодо багатьох атрибутів, таких як: особистість, емоції, стать та вік. Тому здатність розпізнавати обличчя є вкрай необхідною для соціального сприйняття та взаємодії [1]. Для моделювання функцій людського ока використовуються методи комп'ютерного зору.

За останні роки системи розпізнавання обличчя на основі технологій штучного інтелекту, досягли високого ступеня досконалості, однак існує ряд чинників, які негативно впливають на результат розпізнавання. Одним з таких чинників є часткове закриття обличчя сторонніми елементами, що в реальних умовах зустрічається дуже часто.

Викликом останніх двох років, принаймні до 24 лютого 2022 року, для України та всього світу стала пандемія вірусу COVID-19. Через пандемію в багатьох країнах було запроваджено сурові карантинні обмеження. Україна не стала виключенням. Одним з обмежень, що стало невід'ємною частиною пандемії 2020 - 2022 років є обов'язковість носіння захисних масок в публічних місцях (масковий режим), що ставить перед собою на меті зменшення швидкості розповсюдження вірусу COVID-19. Проте маска сама по собі може слугувати засобом маскуванню з метою приховування особистості, що дає широкі можливості для здійснення кримінальних правопорушень, терористичних та диверсійних атак. Тому ідентифікація частково закритих облич людей є важливою та актуальною темою для вирішення безпекових питань, зокрема в період війни так і в умовах пандемії.

Мета роботи. Метою кваліфікаційної роботи є проведення дослідження існуючих методів та технологій ідентифікації облич та створення системи для розпізнавання осіб з частково закритими обличчями, а саме засобами індивідуального захисту, з можливістю подальшого використання в сфері безпеки.

Для досягнення мети необхідно розв'язати наступні задачі:

- дослідження існуючих підходів, методів, виявлення їх недоліків, місць для вдосконалення;
- розробка підходу до розпізнавання частково закритих облич, використовуючи існуючі методи та технології;
- проектування та реалізація системи ідентифікації;
- оцінка ефективності роботи системи.

Об'єкт дослідження. Процес розпізнавання частково закритих облич людей.

Предмет дослідження. Предметом дослідження є методи та технології ідентифікації людей з частково закритими обличчями.

Практичне значення. Результати кваліфікаційної роботи можуть використані для ідентифікації людей у сферах безпеки, як в державному так і приватному секторі.

Апробація результатів дослідження. Основні положення кваліфікаційної роботи були представлені на науково-практичній конференції «Інноваційні наукові дослідження в умовах світових змін» (м. Вінниця, 29 – 30 квітня 2022 р) та опубліковані у збірнику матеріалів цієї конференції.

Структура роботи складається зі вступу, 4 розділів, 15 підрозділів, висновку, списку використаних джерел із 43 найменувань, 1 додатку. Загальний обсяг роботи 74 сторінки.

РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ. ПОСТАНОВКА ЗАВДАННЯ КВАЛІФІКАЦІЙНОЇ РОБОТИ

1.1 Опис проблеми, що вирішується, та її актуальність

Технології штучного інтелекту (ШІ), а саме стрімкий розвиток глибоких нейронних мереж, вплинули на те, що результати досліджень в сфері розпізнавання образів вийшли на новий рівень точності та ефективності.

Крім цього, технологія розпізнавання облич привернула величезну увагу протягом останніх трьох десятиліть, оскільки вона вважається доволі популярною задачею в сфері аналізу зображень і розпізнавання образів. Хоча існуючі системи досягли високого ступеня досконалості, їх продуктивність обмежена певними умовами [2]. Наприклад, ідентифікація зображень облич, які отримані у певному середовищі (при незручному освітленні, поставі або виразі обличчя, у випадку часткової оклюзії, маскуванню або руху камери) створює проблеми для аналізу. Тому, існуючі технології ще далекі від можливостей людського ока.

Традиційно системи розпізнавання облич представлені системами розпізнавання незакритих облич, тобто на фотографії обличчя обов'язково є очі, ніс і рот. Однак, існує ряд обставин, коли обличчя закриваються масками, наприклад, під час пандемії, в медичних установах, в умовах надмірного забруднення чи у різноманітних наукових лабораторій тощо. Як показали дослідження, ця перешкода може бути достатньою для зниження точності роботи деяких алгоритмів розпізнавання облич [3], емоцій [4, 5]. Під час пандемії COVID - 19 науковці отримали нагоду зібрати достатню кількість зображень людей з закритими обличчям та вирішити цю проблему раз і назавжди. На рис 1.1 можна побачити статистичні дані, що демонструють

відсоток людей, що одягали маску у публічних місцях, станом на літо 2020 року [6]. Маючи таку велику кількість закритих облич задача ідентифікації людини у натовпі є майже нездійсненною. Слід зазначити, що подібні технології не мають бути застосовані з метою придушення прав і свобод людини. Але можуть також бути корисним при розпізнаванні терористів та військових злочинців. Подібні технології вже були використані для встановлення осіб причетних до військових злочинів у Бучі в березні 2022 року.

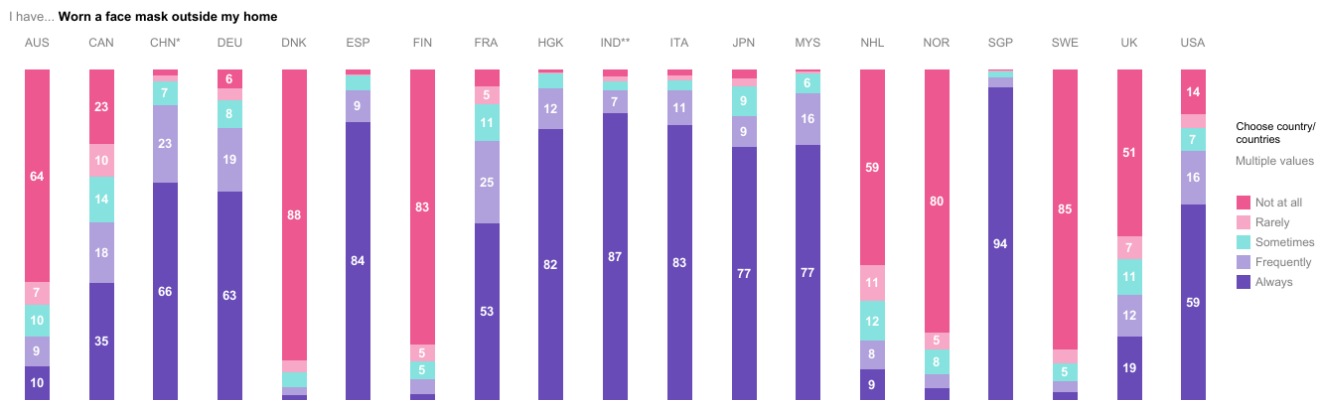


Рис 1.1 Чи одягає людина маску в публічному місці

На рис 1.2 наведено очікувані темпи зростання ринку розпізнавання облич в Північній Америці [7]. Темпи зростання світового ринку розпізнавання облич становлять 15,4%, з оціночною вартістю 12,1 мільярда доларів до 2028 року. Подібні прогнози демонструють продовження поступового росту запиту на технології розпізнавання облич в найближчі роки.

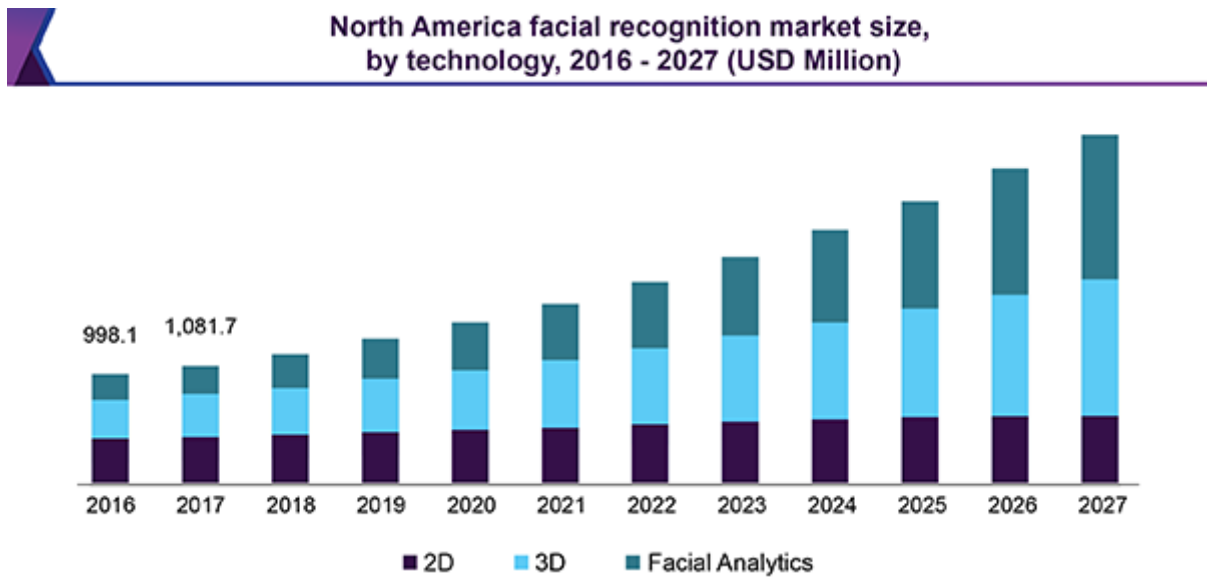


Рис 1.2 Прогноз зростання ринку розпізнавання облич

Таким чином, існує ринковий запит на розробку методів та технологій розпізнавання облич, а ідентифікація частково закритих облич людей є важливою та актуальною темою.

1.2 Аналіз стану вирішення проблеми.

1.2.1 Аналіз існуючих технологій розпізнавання облич

Розпізнавання частково закритих облич, є крайовим випадком розпізнавання облич. Тому спочатку необхідно розглянути загальний стан вирішення проблеми ідентифікації облич. Ключові етапи розвитку технології розпізнавання облич зображено на рис 1.3 [1], розглянемо більш детально кожен з цих етапів.

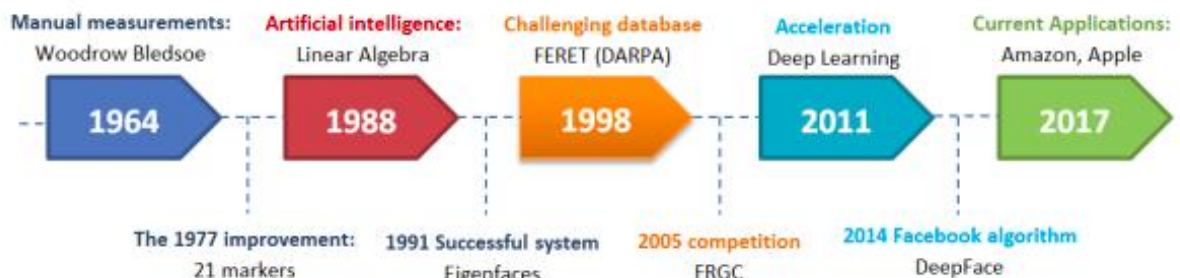


Рис 1.3 Розвиток розпізнавання облич

Піонерами в напрямку технології розпізнавання облич вважають Вуді

Бледсо, Хелена Чан Вульфа і Чарльза Біссона. В 1960-х вони працювали над використанням комп'ютера для розпізнавання людських облич. Їхній ранній проект розпізнавання облич, що побачив світ у **1965**, під лейблом компанії Panoramic Research Incorporated (Panoramic), отримав назву - man-machine (людина - машина), оскільки координати рис обличчя на фотографії повинна була встановити людина, перш ніж вони могли бути використані комп'ютером для розпізнавання. На графічному планшеті людині потрібно було точно визначити координати рис обличчя, таких як центри зіниць, ніс, краї брів тощо. Система людина - машина могла обробляти близько 40 зображень на годину. Комп'ютер автоматично вираховував відстані для кожної фотографії, тобто обчислював відстані між заданими координатами рис обличчя та повертав результат як можливу відповідність [8].

У 1967 Вуді Бледсо взяв участі у іншому дослідженні, метою якого було допомогти правоохоронним органам швидко просіювати бази даних арештованих у пошуках співпадінь. Остання версія алгоритму використовувала 22 виміри. Коли Вуді Бледсо і його колега Пітер Харт попросили трьох людей зіставити підгрупи зі 100 осіб, навіть найшвидшому з них знадобилося шість годин. Але їх алгоритм виконав аналогічне завдання приблизно за три хвилини, досягши стократного скорочення часу. Люди краще розуміли повороти голови та краще працювали з поганою якістю фотографій. Слід зазначити, що цей алгоритм, на відміну від попередніх, розумів "вікові зміни" та значно перевищував людей при аналізі цього фактору, що було неймовірним проривом для того часу.

У **1970** році нині відомий японський вчений Такео Канаде, якому на той момент було 25 років, продемонстрував систему розпізнавання обличчя, яка визначала 21 анатомічну особливість, такі як підборіддя, очі, ніс, та обчислювала співвідношення відстані між рисами обличчя без людського

втручання. В 1977 році науковець видав першу докладну книгу про технологію розпізнавання обличчя “Computer recognition of human faces” [9].

У 1988 році Лоуренс Сирович і Майкл Кірбі з Університету Брауна почали застосовувати лінійну алгебру до проблеми розпізнавання облич. Система, яка стала відома як **Eigenface**, що має у своєму ядрі використання PCA сприяла стрімкому прогресу у сфері розпізнавання облич [10]. У 1991 році Метью Терк і Алекс Пентленд продовжили роботу над розвитком та вдосконаленням цього підходу. Вони розробили комп'ютерну систему, яка працює майже в реальному часі, що може знаходити та відстежувати голову людини, а потім розпізнавати людину, порівнюючи характеристики обличчя з характеристиками вже відомих людей [11]. Було продемонстровано спосіб обчислення власних векторів коваріаційної матриці, таким чином щоб комп'ютери того часу могли виконувати розкладання на великій кількості зображень.

У 1993 в США Агентство передових оборонних дослідницьких проєктів (DARPA) і Національний інститут стандартів і технологій (NIST) розгорнули програму технології розпізнавання облич (FERET). Головною задачею програми було об'єднати існуючі дослідницькі групи та проаналізувати існуючі технології розпізнавання. Важливим етапом цієї програми було створення першого великого набору даних облич. До цього моменту набори даних були дуже обмеженими, в середньому використовувалося зображення 50 осіб, в таких випадках складно казати про репрезентативність вибірки, наприклад, для задачі верифікації при знятті готівки в банкоматі. В результаті процесу збору даних, що тривав три роки, було зібрано 14 126 зображень, 1199 різних людей. З'явилася відкрита база зображень, що дало можливість порівнювати результати різних досліджень [12].

У **2006** проведено Face Recognition Grand Challenge, основною метою якого було просування та розвиток технології розпізнавання облич, протягом FRGC було порівняно останні алгоритми розпізнавання обличчя. У тестах були використані зображення обличчя високої роздільної здатності, 3D-сканування обличчя та зображення райдужної оболонки ока тощо. Результати показали, що нові алгоритми були в 10 разів точнішими, ніж алгоритми розпізнавання обличчя 2002 року і в 100 разів точнішими, ніж алгоритми 1995 року. Таким чином, було продемонстровано значний прогрес у технології розпізнавання обличчя за останнє десятиліття.

2014 відзначився появою Deepface - нового алгоритму розпізнавання облич від Facebook. Це була глибинна згорткова нейронна мережа, що містила більше 2х мільйонів параметрів. Розпізнавання було реалізовано як для 2D так і для 3D зображень [13]. На той момент компанія стверджувала, що її метод наближається до продуктивності людського ока. Найкраща точність на наборі 2D даних LFW [14] становила 97%.

Вдосконалення технології розпізнавання обличчя спонукало до численних інвестицій у комерційні, промислові, юридичні та державні додатки. Серед них варто відзначити наступні продукти:

- Face ID - 3D сканер обличчя, представлений компанією Apple у 2017 році, що можна використовувати для біометричної аутентифікації, оплати, розпізнавання емоцій.
- Skynet System - найбільша у світі система відеоспостереження, створена урядом Китайської Народної Республіки. Вона використовує технології ідентифікації облич, що дозволяє дуже швидко ідентифікувати будь-якого громадянина материкового Китаю, заявлена точність сягає 99.8 %. Станом на 2019 рік система Skynet налічує 200 мільйонів камер громадського спостереження в материковому Китаї, що розташовані в 16 провінціях.

- Amazon Rekognition - продукт компанії Amazon з вбудованим класифікатором. Ця функція дозволяє виявити наявність засобів індивідуального захисту, що є актуальним рішенням не тільки на період пандемії, а й в роботі підприємств хімічної промисловості, закладів охорони здоров'я, харчовій промисловості тощо.

1.2.2 Аналіз існуючих технологій частково закритих облич

В 2019 році, зв'язку з появою нового вірусу COVID-19 та його швидким розповсюдженням, людство було вимушено пристосовуватися до нових викликів та загроз. COVID-19 - це інфекційна хвороба, яка вперше була виявлена у грудні 2019 року в місті Ухань, материковий Китай. Коронавірус поширюється в людській популяції виключно від людини до людини за допомогою переважну повітряно-крапельного механізму передачі вірусу. В зв'язку з цим основними методами попередження розповсюдження захворювання стали соціальне дистанціювання та носіння масок.

З'явилося чимало нових тем для досліджень в різних наукових сферах, наприклад розуміння впливу масок для обличчя на різні аспекти соціальної взаємодії. Для сфери розпізнавання та ідентифікації облич ця пандемія також стала певним викликом, оскільки одне лише носіння засобів індивідуального захисту пов'язано з приховуванням певної частини обличчя людини, що досі було складним випадком для задачі ідентифікації облич та розпізнавання емоцій. Крім того регулярне носіння масок стало масовим явищем, по всій земній кулі.

Проблема розпізнавання частково закритих не є новою, скоріше випадок, коли обличчя закрито маскою або шарфом певною мірою сприймався як виняток. Розробка розпізнавання таких випадків проводиться окремо від розпізнавання облич не закритих ніякими перешкодами. Наприклад, згадана

вище технологія DeepFace тренувалася та тестувалася на наборі даних, що не містить подібних зображень [13].

Оскільки у кваліфікаційній роботі проводиться розпізнавання саме частково закритих облич, слід пояснити термін “частково закриті обличчя”. Нижче наведена загальна класифікація оклюзій (перекриттів) облич за кількістю, характером, позицією, що наведена в дослідженні “Facial Expression Analysis under Partial Occlusion” [4]:

1. Різний тип. Тип оклюзії може значно відрізнятися в залежності від ситуації, де знаходиться обличчя. Якщо немає попередніх знань про те, який тип оклюзії відбудеться в конкретному контексті, системи розпізнавання повинні враховувати всі можливі типи, що є дуже складним і технічно нездійсненним, принаймні наразі.
2. Змішаний тип. На обличчі можуть існувати кілька типів оклюзії. Існування та поширеність змішаної оклюзії збільшує труднощі в обробці оклюзії та дослідженні впливу окремої оклюзії, оскільки розділення змішаної оклюзії досі залишається невивченою областю.
3. Нефіксоване розташування. Більшість типів оклюзії, як правило, не фіксуються на певному місці обличчя. Хоча місце розташування є приблизно передбачуваним для деяких систематичних перекриттів обличчя, таких як окуляри та капелюх, все одно важко передбачити їх точне розташування. Також складно передбачити розташування для тимчасової оклюзії, наприклад, прикриття рукою.
4. Різна тривалість. Різні типи оклюзії можуть існувати протягом різної тривалості. Передбачається, що оклюзія через рух руки по обличчю триватиме лише кілька секунд, тоді як оклюзія сонцезахисних окулярів часто існує протягом загальної тривалості відеоданих.

5. Непередбачувана властивість. Через мінливість об'єктів, що призводять до оклюзії, візуальні властивості (наприклад, форма, зовнішній вигляд і розмір) отриманої оклюзії часто є непередбачуваними. Наприклад, можуть бути великі відмінності у візуальних властивостях між сонцезахисними окулярами, окулярами для корекції зору та захисними окулярами.
6. Місцевий вплив. На відміну від варіацій пози та освітлення, які часто впливають на цілісну область обличчя, більшість типів оклюзії впливають лише на невелику частину обличчя. Цю властивість можна розглядати як перевагу, оскільки ефект оклюзії можна компенсувати за допомогою незакритих частин обличчя.

В кваліфікаційній роботі будуть розглядатися зображення людей з закритою нижньою частиною обличчя, при цьому датасет містить як зімітовані закриття облич, так і реальні зображення. На рис 1.4 це категорії (b), (d). За наведеною вище класифікацією, даний тип закриття облич слід віднести до оклюзій місцевого впливу, оскільки закритою буде зберігатися певна частина обличчя.

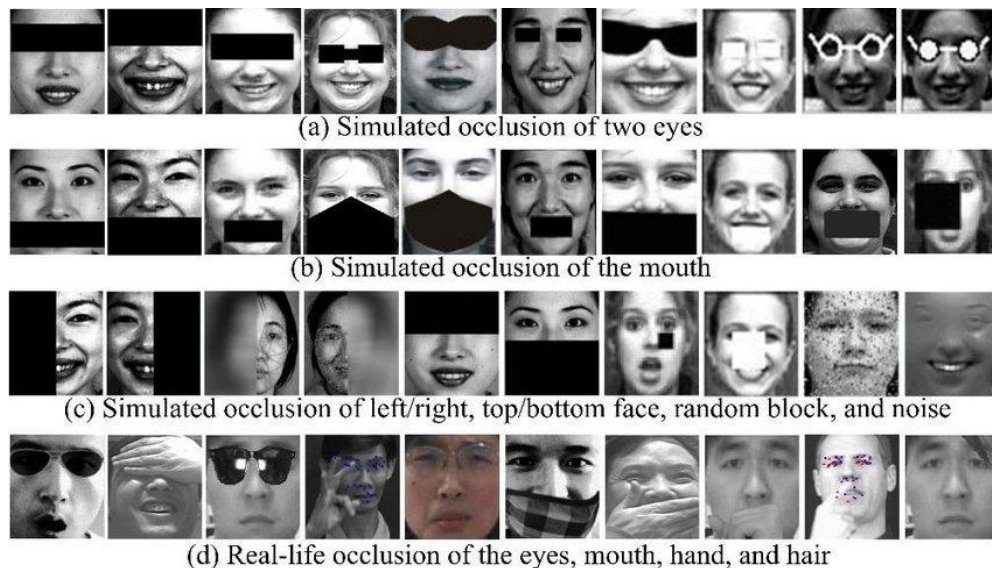


Рис 1.4 Категорії закритих облич.

У **2015** було представлено новий датасет, що став своєрідним "викликом" для алгоритмів розпізнавання. Дослідники стверджували, що існує розрив між поточною продуктивністю розпізнавання обличчя та реальністю. Було створено датсет під назвою WIDER FACE1, який у був 10 разів більший за наявні набори даних. Унікальністю цього набору була наявність надзвичайно складних ситуацій для розпізнавання, а саме: велика кількість людей в кадрі, закриття обличч предметами, різноманітність поз та ракурсів, тощо [15]. Ще через два роки було зібрано датасет Mafa (MAsked FAcEs), з фокусом на обличчя, що закриті маскою. Було запропоновано підхід, що перевершив існуючі методи на 15% при роботі з замаскованими обличчями [16]. В Додатку А наведено приклади зображення з цих датасетів.

У **2017** році для розпізнавання частково закритих обличч, були використані наступні методи: новостворений Face Attention Network (далі FAN) [17] та модифіковані Single Stage Headless Face Detector [18] та Single Shot Scale-invariant Face Detector [19]. Точність детекції обличч на обох датасетах з частково закритими обличчями сягала близько 90%.

Розглянемо архітектуру FAN. Автори адаптували архітектуру RetinaNet [20] для виявлення обличчя, використовуючи принцип «розділяй і володарюй». ResNet-50 [21] використовується в якості основи для створення ієрархії пірамід функцій з багатою семантичною інформацією. Наступний етап - U-подібна структура двох підмереж: одна для класифікації, а інша для регресії. Підмережа класифікації застосовує чотири шари згортки (3×3) кожен з 256 фільтрами, за якими слідує шар згортки (3×3) з фільтрами $K \times A$, де K - кількість класів, а A - кількість якорів. Підмережа регресії ідентична класифікаційній підмережі, за винятком того, що вона закінчується фільтрами згортки $4 \times A$ з лінійною активацією. На рис 1.5 наведена схематична архітектура даної мережі.

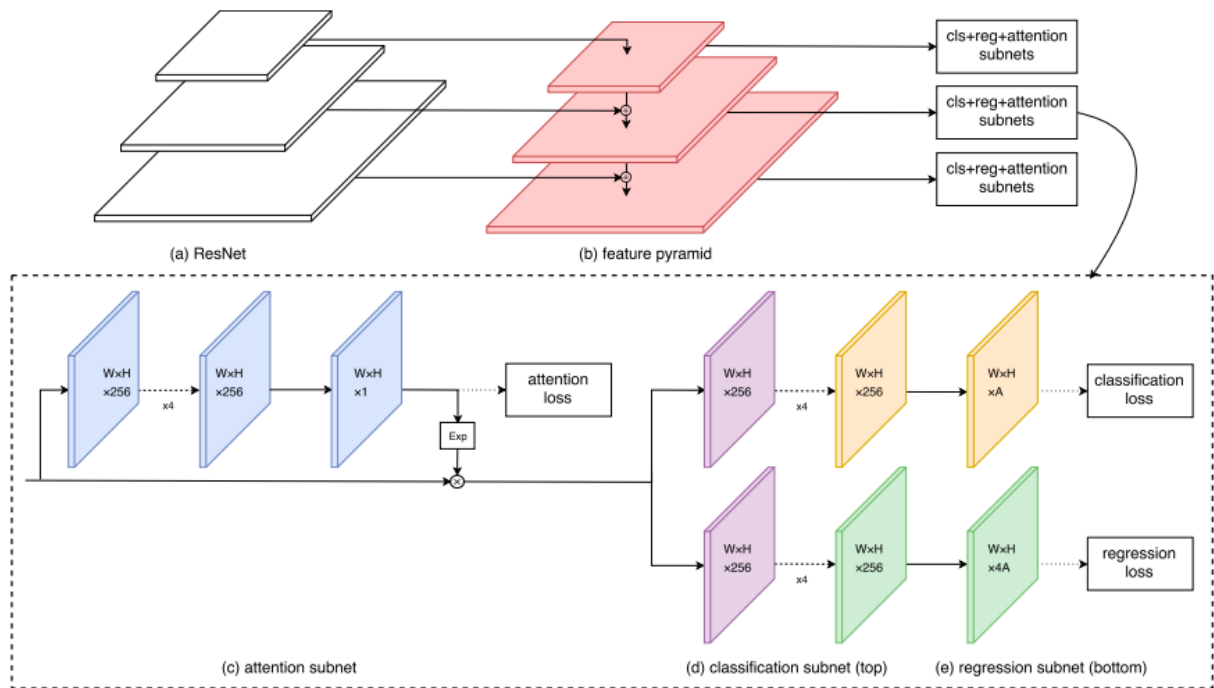


Рис 1.5 Схематичне зображення архітектури FAN [17]

У 2020 році командою дослідників NIST було проведено порівняльний аналіз алгоритмів, що були розроблені у 2018 - 2019 [3]. Поширена вимога носіння захисних масок у громадських місцях, в період пандемії COVID-19, викликала потребу зрозуміти, як технологія розпізнавання обличч справляється з закритими обличчями, часто з видимою лише навколо окулярну зоною і вище. Було використано два великими набори даних з фотографіями, зібраних в урядових програмах США. Разом ці набори даних дозволили нам обробити загалом 6,2 мільйона зображень 1 мільйона людей за допомогою 89 алгоритмів.

Висновок був невтішний: “Навіть найкращі з 89 протестованих комерційних алгоритмів розпізнавання обличчя мали показники помилок від 5% до 50% при зіставленні цифрових масок для обличчя з фотографіями тієї ж людини без маски”. Слід зауважити, що автори алгоритмів не знали, що їх роботи будуть перевірятися таким чином, та не були попереджені заздалегідь.

На думку дослідників, при створенні наступного покоління алгоритмів, слід зауважити наступні аспекти:

1. Точність алгоритму з замаскованими обличчями значно знизилася. Використовуючи незамасковані зображення, найточніші алгоритми не можуть аутентифікувати людину приблизно в 0,3% випадків. Замасковані зображення підвищили рівень збоїв навіть цих найкращих алгоритмів приблизно до 5%, тоді як багато інших компетентних алгоритмів виходили з ладу від 20% до 50% часу.
2. Замасковані зображення частіше призводили до того, що алгоритми не могли обробити обличчя, що технічно називається «невдачею реєстрації або шаблону». Алгоритми розпізнавання обличчя зазвичай працюють, вимірюючи риси обличчя — наприклад, їх розмір і відстань один від одного — а потім порівнюючи ці вимірювання з результатами на іншій фотографії. Невдача реєстрації шаблону означає, що алгоритм не може виділити риси обличчя достатньо добре, щоб зробити ефективне порівняння.
3. Чим більшу частину носа покриває маска, тим нижча точність алгоритму. У дослідженні розглядалися три рівні охоплення носа — низький, середній і високий — виявляючи, що точність погіршується з більшим закриттям носа.
4. Форма і колір маски мають значення. Частота помилок алгоритму, як правило, була нижчою з круглими масками. Чорні маски також погіршили роботу алгоритму в порівнянні з синіми, хоча через обмеження часу та ресурсів команда не змогла повністю перевірити ефект кольору.

Таким чином, в результаті аналізу стану проблеми можна стверджувати, що протягом 30 років методи та технології ідентифікації облич розвивалися та

досягли певного ступеня досконалості. Проте перекриття облич негативно впливає на точність ідентифікації.

1.3 Аналіз методів та технологій в системах розпізнавання облич

Класична система двовимірного розпізнавання облич працює з зображеннями або відео. Система спочатку повинна виявити обличчя у вхідному зображенні/відео та відділити його від області виявлення. Далі відбувається попередня обробка зображення згідно заданої конфігурації, наприклад адаптація до потенційних змін освітлення. Наступним етапом є виділення ознак та завершальним етапом є ідентифікація особистості, що використовує класифікацію за раніше виділеними ознаками.

Залежно від характеру використовуваних методів сегментації та класифікації виділяються наступні підходи до розпізнавання облич [5]:

- комплексний підхід
- локальний підхід
- технології на базі глибинного навчання

У роботі "Past, Present, and Future of Face Recognition: A Review" [1] методи з локальним підходом підрозділяються на геометричні та текстурні.

Таблиця 1

Комплексні	Локальні (геометричні)	Локальні (текстурні)	Глибинне навчання
PCA	EGM	LBP	DeepFace
LDA	EBGM	LDP	VGGNet
ICA	MEGM	BSIF	GoogleNet
DWT		HOG	FaceNet

У випадку комплексного підходу аналізується все зображення обличчя, а дескриптори локальних ознак передбачають використання локальних компонентів обличчя окремо та подальшого їх об'єднання в комбіновану гістограму. В свою чергу глибини навчання передбачає використання згорткових нейронних мереж для аналізу зображень. Розглянемо деякі методи, наведені у Таблиці 1.

1.3.1 Методи на основі комплексного підходу

- **Метод головних компонентів (PCA)**, відомий як *eigenfaces* [11]. Основною метою цього методу є зменшення великої розмірності шляхом втрати незначної кількості даних. На даний момент PCA використовується для вираження великого одновимірного вектора пікселів, побудованих із двовимірного зображення обличчя, у компактні основні компоненти простору зображень.
- **Лінійний дискримінантний аналіз Фішера (LDA)**, відомий як *fisherfaces*, і **незалежний компонентний аналіз (ICA)** є найбільш поширеними лінійними методами, які використовуються для систем розпізнавання обличчя. LDA шукає проєкційні осі, на яких точки даних різних класів знаходяться далеко одна від одної, в той момент, коли точки даних одного класу повинні бути близькими одна до одної. На відміну від PCA, який кодує інформацію в ортогональному лінійному просторі, LDA кодує дискримінаційну інформацію в лінійно відокремленому просторі, використовуючи базиси, які не обов'язково є ортогональними [22].

Спільним недоліком для розглянутих методів є сильна чутливість до змін контексту та невідповідності. З цієї причини в більшості випадків обличчя доводиться вирізати з зображення, необхідно провести попередню детекцію та

сегментацію обличчя. З іншого боку, оскільки набір даних розглядається як єдина матриця, необхідно забезпечити геометричну узгодженість у всіх екземплярах обличчя. Всі зображення обличчя повинні бути ретельно зіставлені в стандартній системі відліку. Незначна помилка в орієнтації обличчя може спричинити значні помилки класифікації обличчя. Тобто є чуттєвість до зсувів і поворотів облич.

1.3.2 Методи на основі локального підходу

- Однією з поширених форм розпізнавання обличчя на основі орієнтирів є **еластичне зіставлення графіків (EGM)** [23]. Даний алгоритм використовує візуальні характеристики засновані на вейвлетах Габора, які, як виявилось, є якісною моделлю ранньої візуальної обробки в мозку, точніше, простих клітинах первинної зорової кори. Крім цього сам алгоритм є алгоритмічною версією динамічного узгодження зв'язків (DLM) [23], що є моделлю інваріантного розпізнавання об'єктів у мозку. Візуальні об'єкти в EGM представлені у зваженого графу, де вершини представляють локальні текстури на основі вейвлетів Габора, а ребра представляють відстані між розташуваннями вузлів на зображенні.
- **Дискретне вейвлет-перетворення (DWT)** [24] використовується в основному для стиснення зображень та вибору ознак. Одним з популярних вейвлет - перетворень є перетворення Хаара, що є основою для методу Віюлі Джонса.
- **Локальні двійкові шаблони (LBP)** [25], локальні спрямовані шаблони (LDP) і алгоритми вилучення ознак з інваріантним масштабом (SIFT) були успішно використані для виділення локальних ознак [5].
- **Гістограма напрямлених градієнтів (HOG)** - також найчастіше використовується для виділення локальних ознак. При розпізнаванні

зображень найчастіше використовується для детекції обличчя, є одним з найбільш точних та часто використовуваних методів для цієї задачі.

На даний момент, більшість методів з цієї групи використовуються у комбінації з іншими методами в якості допоміжних у рішеннях задач ідентифікації та класифікації. Такий підхід збільшує точність розпізнавання.

1.3.3 Технології глибинного навчання

Глибинне навчання, належність до класу машинного навчання, використовує послідовні приховані шари для різних рівнів обробки інформації.

Розглянемо мережу VGGNet. VGG означає Visual Geometry Group. Це стандартна архітектура глибинної згорткової нейронної мережі (DCNN) з кількома шарами. «Глибинний» відноситься до кількості шарів з VGG-16 або VGG-19, що складається з 16 і 19 згорткових шарів, тобто глибина даних мереж 16 і 18 відповідно [26]. Архітектура VGG є основою інноваційних моделей розпізнавання образів. Розроблена як глибока нейронна мережа, VGGNet також перевершує baseline для багатьох завдань і наборів даних. Більше того, зараз це одна з найпопулярніших архітектур розпізнавання зображень.

Проте класичне використання мереж - класифікаторів, таких як VGG16, не підходять для вирішення задачі ідентифікації, через велику кількість класів. Крім того їх збалансованість у цій задачі не гарантується. Також існує проблема з додаванням класів. В класичному варіанті при додаванні нового класу необхідно перенавчити модель, що є дорогавартісним процесом в випадку великих навчальних вибірок та великої кількості класів. Тому використовується підхід із створенням ембедінг векторів. Проте існують випадки, такі як часткове закриття облич, коли існуючі методи мають низку точність ідентифікації.

Аналіз показує, що існуючі методи поодиночі не дозволяють вирішити задачу ідентифікації частково закритих облич, але при компіляції даних

технологій, зважаючи на їх переваги та недоліки, можна досягти більшої точності розпізнавання ніж при використанні лише одного з цих методів.

1.4 Постановка завдання розпізнавання частково закритих облич

Завдання дослідження, полягає у тому, щоб розробити підхід до розпізнавання частково закритих облич, використовуючи існуючі методи та технології, враховуючи їх недоліки, що не дають вирішити цю задачу в повному обсязі. На базі даного підходу розробити систему розпізнавання.

На вхід системи подається зображення людини або декількох людей у масці. Необхідно розпізнати їх обличчя та, при наявності даної особи у наборі даних, встановити особу за допомогою технологій штучного інтелекту.

Для створення системи ідентифікації необхідно розробити кілька підсистем, що вирішують наступні завдання розпізнавання:

- детекція обличчя - пошук обличчя на зображенні і виділення контурів обличчя.
- пошук даної особи у базі даних і ідентифікація.

Окрім безпосереднього розпізнавання необхідно реалізувати підсистему вводу - виводу даних.

Основні етапи, які необхідно виконати для розробки технологій розпізнавання та створення системи розпізнавання:

1. дослідження існуючих підходів, методів, виявлення їх недоліків, місць для вдосконалення;
2. пошук публічно доступних баз даних зображень та обробка цих зображень з метою приведення до одного спільного формату та створення нового набору даних;
3. програмна реалізація та порівняння роботи існуючих методів на створеному наборі даних, аналіз результатів;

4. реалізація та тестування підходу до розпізнавання частково закритих облич;
5. проектування системи ідентифікації;
6. розробка підсистем детекції та ідентифікації облич;
7. створення системи розпізнавання облич;
8. оцінка ефективності роботи системи.

1.5 Висновки до першого розділу

Задача розпізнавання частково закритих облич полягає у тому, що маючи зображення людини або декількох людей з частково закритими обличчями розпізнати їх обличчя та, при наявності цієї особи у базі даних, ідентифікувати людину.

Завдання дослідження, полягає у тому, щоб розробити підхід до розпізнавання частково закритих облич, використовуючи існуючі методи та технології. На базі даного підходу розробити систему розпізнавання.

Під частково закритими обличчями розуміються зображення облич людей з закритою нижньою частиною обличчя.

В результаті аналізу проблеми було зроблено висновок, що тема розпізнавання закритих облич є актуальною. При розробці системи слід звернути увагу на ймовірний вплив колір, форми маски, положення та повороту голови на результат розпізнавання, забезпечити сегментацію обличчя людини від інших елементів зображення.

Розглянуто різні підходи до розпізнавання облич, однак зроблено висновок, що існуючі методи поодиночі не дозволяють вирішити задачу ідентифікації частково закритих облич. Проте при об'єднанні даних технологій та методів можна побудувати підхід до розпізнавання частково закритих облич. Оскільки методи на основі глобального підходу мають ряд недоліків та гірше справляються з ідентифікацією не закритих маскою облич в порівнянні з

глибинними згортокими мережами. Було вирішено віддати перевагу використанню технологій глибинного навчання. Серед методів на основі локального підходу є методи, що використовуються для детекції облич. Їх необхідно розглянути при виборі технології для вирішення задачі детекції частково закритих облич.

Таким чином, було проаналізовано актуальність, стан проблеми та методи і технології розпізнавання облич.

РОЗДІЛ 2. ДОСЛІДЖЕННЯ МЕТОДІВ ТА ТЕХНОЛОГІЙ РОЗПІЗНАВАННЯ

2.1 Методи та технології розв’язання завдань дослідження

Глибинне навчання є частиною сімейства методів машинного навчання. Це набір методів, що використовує багат шарові штучні нейронні мережі. Глибинні нейронні мережі можна розділити на три основні класи залежно від підходу та архітектури [1] :

- З безконтрольним навчанням або генеративні (автокодер (AE), машина Больцмана (BM) рекурентна нейронна мережа (RNN) та мережа суми добутку (SPN));
- Контрольовані або дискримінаційні (згорткові нейронні мережі (CNN));
- Гібридні нейронні мережі.

В даній роботі будуть використані нейронні мережі, в основі яких лежить CNN, що є загальновживаним стандартом для розпізнавання та класифікація зображень. Згорткова нейронна мережа (CNN/ЗНМ) складається з шарів входу та виходу, а також із декількох прихованих шарів (Рис 2.1), такими як шари згортки, агрегування (пулінгу) та повнозв'язний шари [27].

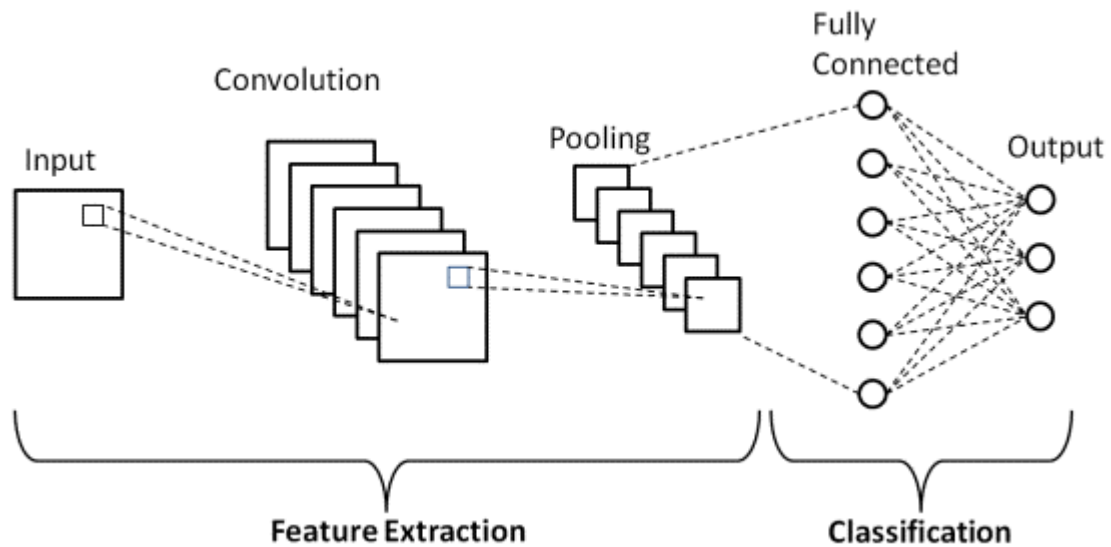


Рис 2.1. Схематичне представлення CNN

- Convolutional - згортковий шар, дозволяє об'єднувати значення розташованих поруч пікселів, шляхом накладання на кожну зону зображення матриці-фільтра, тобто перемноження їх значень та сумування всіх значень результуючої матриці і виділяти більш узагальнені ознаки зображення
- Pooling - шар агрегування, також застосовується для отримання узагальненої характеристики групи пікселів, для цього один піксель з невеликої групи обирають по заданному правилу, наприклад – максимальний і використовують його для представлення всієї групи. При цій операції зменшення розмірності даних отриманого на вхід із згорткового шару.
- Fully connected - повнозв'язний шар, який додається після комбінації згорткового та шару агрегування. У випадку задачі класифікації, результат - N-просторовий вектор, де N - число класів, з яких програма вибирає потрібний. Кожне число в цьому N-просторовому векторі представляє ймовірність належності вхідного зображення або сегменту зображення певному класу.

Саме згорткова структура нейронної мережі дає можливість працювати з зображеннями більш ефективно. Зараз глибинні згорткові нейронні мережі (DCNN) застосовуються для вирішення великого спектру завдань пов'язаних з фотографіями. З їх допомогою виконують такі завдання, як класифікація, пошук об'єктів на зображенні, знаходження меж цих об'єктів, побудова 3D-моделі по фотографії.

Відомою проблемою, що виникає при збільшенні глибини нейронної мережі є зникнення градієнту, що призводить до збільшення рівня похибки в процесі навчання та тестування. Для вирішення цієї проблеми та проблеми ефективного використання ресурсів при навчанні використовуються архітектурні модифікації на базі DCNN.

2.1.1 Нейронна мережа ResNet

У 2015 році дослідниками Microsoft Research представлено нову архітектуру нейронної мережі під назвою ResNet [21] або залишкова нейронна мережа [21]. В основі цієї архітектури використовується залишковий (residual) блок, що вирішує проблему зникаючого/вибухаючого градієнту. У цій мережі використано техніку, яка називається пропуском з'єднань, тобто пропускає навчання з кількох шарів і підключається безпосередньо до виходу з останнього пропущеного. Схематично цей крок зображено на рис 2.2.

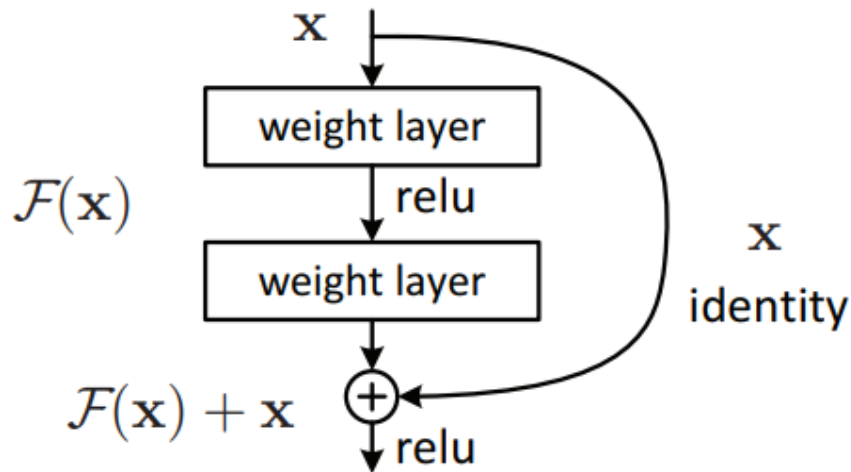


Рис 2.2. Пропуск з'єднань в ResNet

2.1.2 Нейрона мережа Inception

Ідея мережі Inception полягає у переході до більш розрідженої архітектури, зменшення глибини мережі, зберігаючи її ефективність при розпізнаванні. Відомо, що ключові частини зображення можуть мати надзвичайно великі варіації в розмірі. Через цю величезну різницю в розташуванні інформації вибрати правильний розмір ядра для операції згортки - важко. Більше ядро є кращим для інформації, яка поширюється більш глобально, а менше ядро є кращим для інформації, яка поширюється більш локально.

Для вирішення цієї проблеми було запропоновано модуль ("наївний" початковий модуль, модуль скорочення, рис 2.3), що складається з шару, де на одному рівні одночасно розміщувалися фільтри різних розмірів [28].

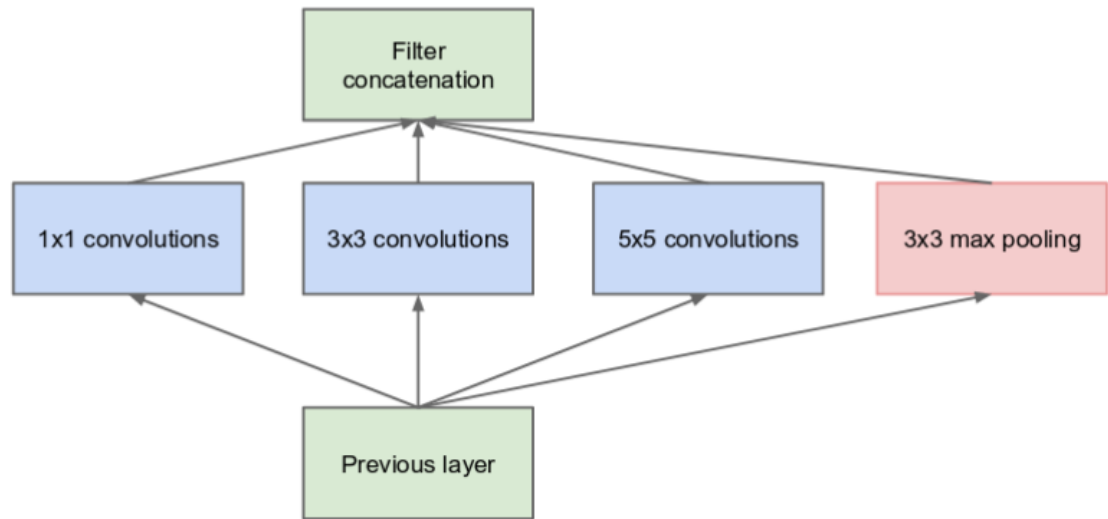


Рис 2.3. “Наївна” версія модулю скорочення Inception V1

2.1.3 Нейрона мережа Inception-ResNet

Inception-ResNet - остання версія мережі Inception, що поєднує у собі переваги Inception та ResNet. Існує дві підверсії Inception ResNet, а саме v1 і v2 [29]. Обидві підверсії мають однакову структуру для модулів A, B, C (на рис 2.4 inception-resnet-a, inception-resnet-b, inception-resnet-c) і блоків скорочення (reduction). Єдина відмінність між версіями - це налаштування гіпер параметрів.

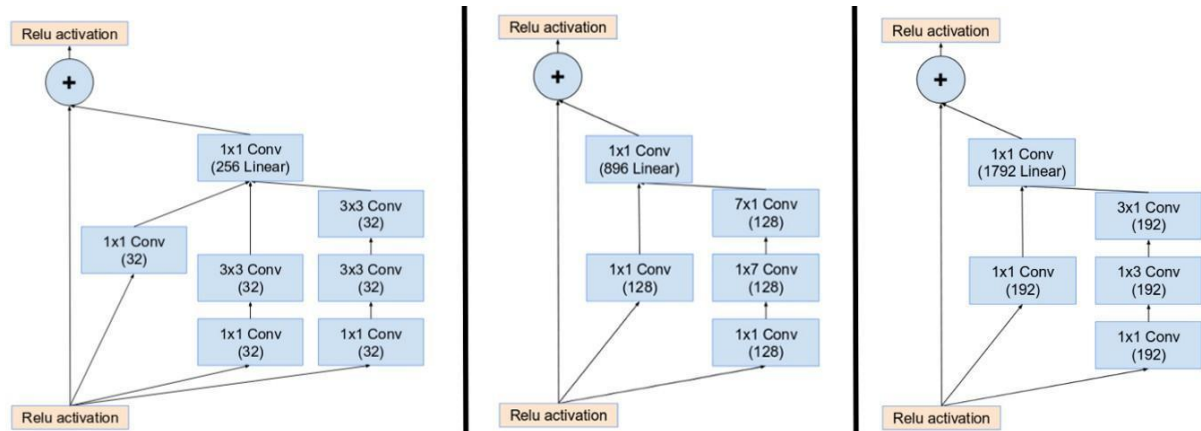


Рис 2.4. Модулі inception-resnet-a, inception-resnet-b, inception-resnet-c

Шар агрегації (pooling), що був наявний в InceptionV3 був замінений залишковим з'єднанням, а також додатковою згорткою 1x1 перед об'єднанням.

Архітектура Inception-ResNet містить два типи блоків скорочення, схематично зображених на рис 2.5.

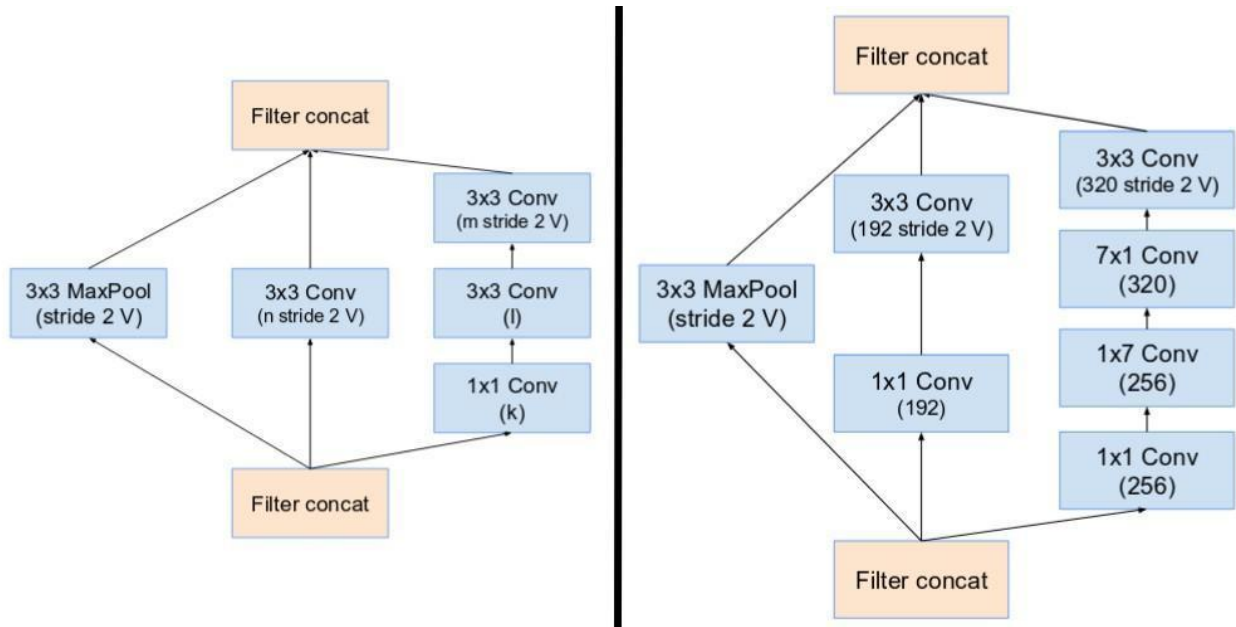


Рис 2.5. Блоки скорочення InceptionResNet

2.1.4 Багатозадачні каскадні згорткові нейронні мережі

Багатозадачні каскадні згорткові мережі (MTCNN) — це фреймворк, розроблений як рішення як для задачі виявлення обличчя, так і для знаходження ключових точок на обличчі. Процес складається з трьох етапів згорткових мереж, які здатні розпізнавати обличчя та орієнтири, такі як очі, ніс і рот [30].

На першому етапі використовується неглибока CNN для швидкого створення вікон-кандидатів. На другому етапі уточнюються запропоновані вікна-кандидати за допомогою більш складної CNN. І нарешті, на третьому етапі маємо ще одну CNN, складнішу за інші, для подальшого уточнення результату та виведення орієнтирів обличчя.

MTCNN складається з трьох нейронних мереж (рис 2.6):

1. Мережа пропозиції (P-Net)

На першому етапі використовується повністю згорткова мережею (FCN). Різниця між CNN і FCN полягає в тому, що повністю згортка мережа не

використовує повнозв'язний шар як частину архітектури. Ця мережа пропозицій використовується для отримання вікон-кандидатів та їх векторів, що задають координати обмежувальної рамки. Після отримання цих координат, виконується об'єднання регіонів, що накладаються. Кінцевим результатом цього етапу є згенерований набір областей (вікон-кандидатів) після об'єднання.

2. Уточнювальна мережа (R-Net)

Усі кандидати з P-Net надходять у Refine Network. R-Net додатково зменшує кількість вікон, використовує алгоритм немаксимального придушення (NMS) для того, щоб обрати один елемент з множини кандидатів, що перекривають одне одного. R-Net виводить:

- вектор ймовірності належності двом класам (обличчя, не обличчя)
- вектор із 4 елементів, координат вікна, де знаходиться обличчя
- вектор із 10 елементів для локалізації обличчя ключових точок обличчя

3. Вихідна мережа (O-Net)

Цей етап подібний до попереднього, але ця вихідна мережа має на меті більш детально вивести по дві координати для кожної точки (ліве око, ніс, праве око, лівий кут рота, правий кут рота).

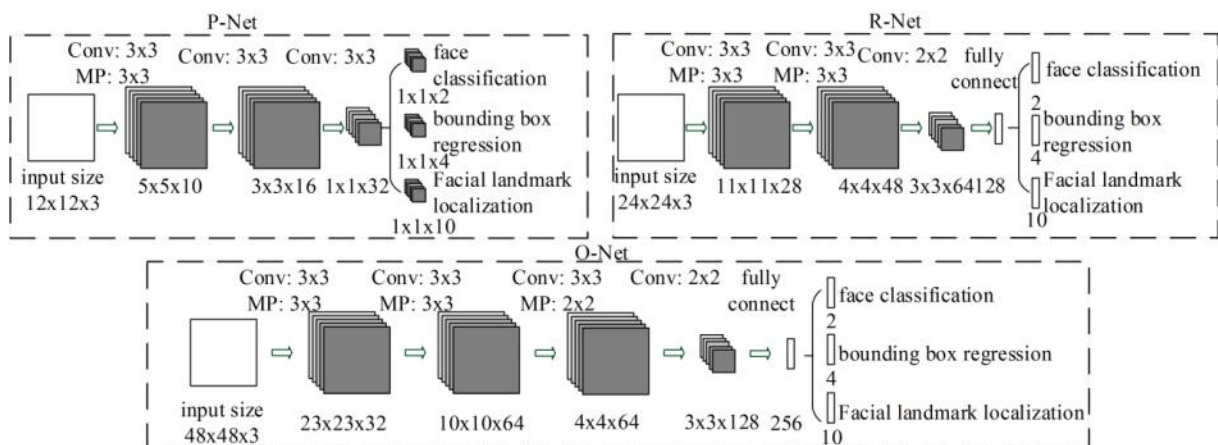


Рис 2.6. Архітектура MTCNN

На кожному з етапів вирішується одне з наступних завдань:

- 1) Класифікація облич - це проблема бінарної класифікації, яка використовує кросс ентропію як функцію втрат :

$$L_i^{det} = -(y_i^{det} \log(p_i) + (1 - y_i^{det})(1 - \log(p_i))) \quad (1)$$

- 2) Лінійна регресія для координат вік обличчя

$$L_i^{box} = \|\hat{y}_i^{box} - y_i^{box}\|_2^2$$

- 3) Лінійна регресія для координат ключових точок

$$L_i^{landmark} = \|\hat{y}_i^{landmark} - y_i^{landmark}\|_2^2$$

Цей метод було обрано, через стійкість до поворотів та нахилів голови (така проблема характерна для каскаду Хаара). На результат детекції не впливає розмір зображення (цей недолік характерен для програмної реалізації гістограми напрямлених градієнтів для пакету Dlib). Програмна реалізація мовою python є найшвидшою в порівнянні з іншими методами детекції, що є критичним для роботи системи розпізнавання. На рис 2.7 наведено порівняння швидкості роботи пакету tensorflow MTCNN, з аналогічними програмними рішеннями. Та головне те, що даний підхід продемонстрував стійкість до перекриттів нижньої частини обличчя. В роботі [31] продемонстровано, що даний метод є більш стійким до перекриттів зони губ та носа в порівнянні з алгоритмом HOG.

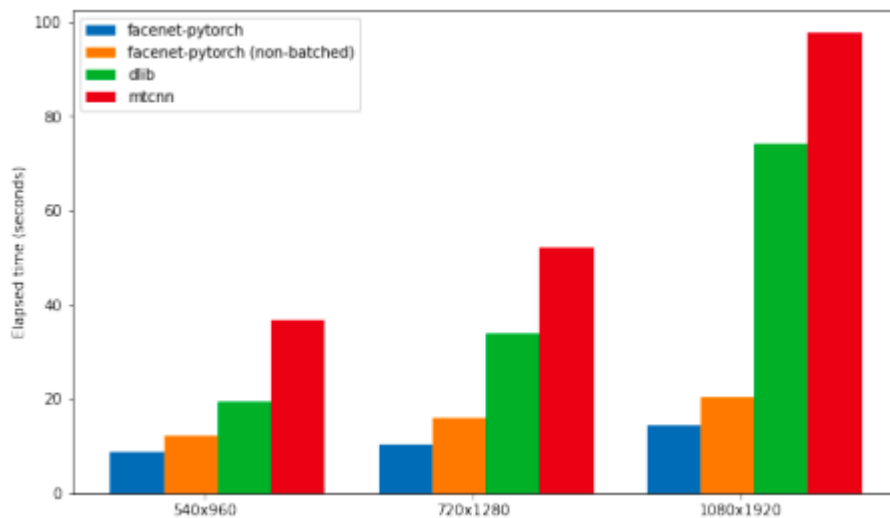


Рис 2.7. Швидкість роботи різних пакетів для детекції обличчя

2.1.5 Система розпізнавання обличчя FaceNet

FaceNet — це система розпізнавання облич, описана Флоріаном Шроффом у статті «FaceNet: A unified embedding for face recognition and clustering» [32].

FaceNet (рис 2.8) приймає зображення обличчя людини як вхідні дані та виводить вектор із 128 чисел, які представляють ключові ознаки обличчя. У машинному навчанні цей вектор називається вбудовуванням (embedding). Тобто на виході маємо відображення зображень облич у компактний евклідовий простір, де відстані безпосередньо відповідають мірі подібності обличчя.

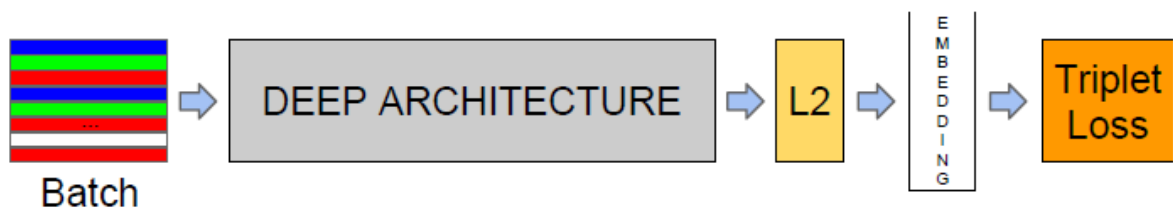


Рис 2.8 Схематичне зображення FaceNet[32]

В основі FaceNet - глибока згортова нейронна мережа, навчену через функцію втрат (а саме триплетну функцію втрат) , яка спонукає вектори однієї ідентичності ставати більш схожими (відстань зменшується), тоді як очікується,

що вектори для різних людей стануть менш схожими (відстань зростає). При цьому існує декілька варіантів DCNN, які можна використовувати.

Триплетна функція втрат

На відміну від навчання з іншими функціями втрати, де зазвичай потрібні лише дані та результат навчання, застосування триплетної функції вимагає подальшого процесу порівняння з вихідним набором даних. Дана функція втрат була представлена у статті “Deep Metric Learning With Tuplet Margin Loss”, авторами Б. Ю та Д. Тао, у 2019 році [33].

Нехай є зображення (anchor), необхідно вибрати з набору даних позитивне (positive) та негативне (negative) зображення, де позитивне з того самого класу, що й дане зображення, а негативне – з іншого класу [32].

Функція втрат мінімізує відстань між i позитивним, обидва з яких мають однакову ідентичність, і максимізує відстань між даним та негативним, зображенням іншої людини.

Нехай $\square(\square)$ - обмежена для існування на d -вимірній гіперсфері, тобто $\|\square(\square)\|^2 = 1$.

Як показано на рис 2.9, зображення $\square_{\square}$ (якір) конкретної людини ближче до всіх інших зображень $\square_{\square}$ (позитивних) цієї ж людини, ніж до будь-якого зображення $\square_{\square}$ (негативного) будь-якої іншої особи. Маємо таку закономірність:

$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2 ,$$

$$\forall (f(x_i^a), f(x_i^p), f(x_i^n)) \in \mathcal{T} .$$

де $\alpha=0,2$ є інтервалом, який існує між позитивними та негативними парами. При цьому $\mathcal{T}$ — множина всіх можливих триплетів у навчальній множині, потужність якої N .

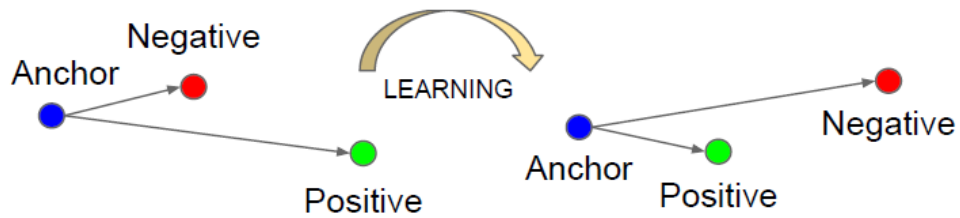


Рис 2.9 Триплетна функція витрат

L - функція втрат, що мінімізується:

$$\sum_i^N \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+$$

Вибір триплетів для функції втрат

Триплети можна обрати на різних етапах навчання. Наприклад, пара вибирається на початку всього процесу навчання, повторно вибирається на початку кожної епохи (автономний триплетний вибір) або повторно вибирається для кожного пакету (онлайн триплетний вибір). Оскільки параметри моделі постійно змінюються протягом навчального процесу, спосіб вибору пари також може змінюватися.

Таким чином, частота оновлення пар також важлива для результату навчання.

2.1.6 Технологія трансферного навчання

Трансферне навчання використовується для вирішення задачі на зберіганні та передачі знань, отриманих під час розв'язання однієї задачі, та застосуванні їх до іншої, але схожої задачі [34]. В випадку навчання нейронних мереж для навчання береться вже навчена модель на деяких даних таких же або близьких за значенням до тих, на яких повинна працювати навчена мережа. Тобто цей підхід полягає в тому, що ваги з уже навченої моделі використовуються для ініціалізації ваг в використаній моделі. Також при навчанні найчастіше змінюються і калібрують лише ваги декількох останніх шарів [35].

Трансферне навчання працює, оскільки на перших шарах згорткова нейронна мережа вчиться виділяти найпростіші ознаки, такі як кути, точки, межі, на наступних згортальних рівнях - більш складні ознаки. Останніми шарами DCNN часто є повнозв'язні шари, які відповідають, наприклад, за класифікацію зображень. Тому якщо мережа була натренована на близькому класі зображень, то і ознаки будуть аналогічні, починаючи від найпростіших і закінчуючи більш складними.

2.1.7 Метод вирівнювання зображень

Під вирівнюванням в даному випадку розуміється приведення зображень облич до певного положення на площині. Розробники FaceNet зазначають, що після додавання етапу вирівнювання обличчя, точність розпізнавання збільшилася майже на 1%. Тобто необхідно виконати якесь афіне перетворення. Існує багато методів, які можуть виконати подібні повороти. Більшість з них використовують ключові точки обличчя, проте в випадку частково закритих облич кількість таких точок суттєво обмежена. Можна було б використовувати координати контуру маски, але при такому підході необхідно врахувати форму маски, що ускладнює задачу. Тому пропонується метод з використанням координат зіниць. Оскільки MTCNN повертає вектор, що містить координати зіниць, можна використати ці значення наступним чином.

Нехай маємо координати зіниць:

(x_1, y_1, x_2, y_2) - координати лівої зіниці

(x_3, y_3, x_4, y_4) - координати правої зіниці

Необхідно врахувати кут повороту зображення, маємо наступне рівняння:

$$\theta = \arctan\left(\frac{y_2 - y_1}{x_2 - x_1}\right) - \arctan\left(\frac{y_4 - y_3}{x_4 - x_3}\right)$$

Результат роботи даного методу зображено на рис 2.10 [36].

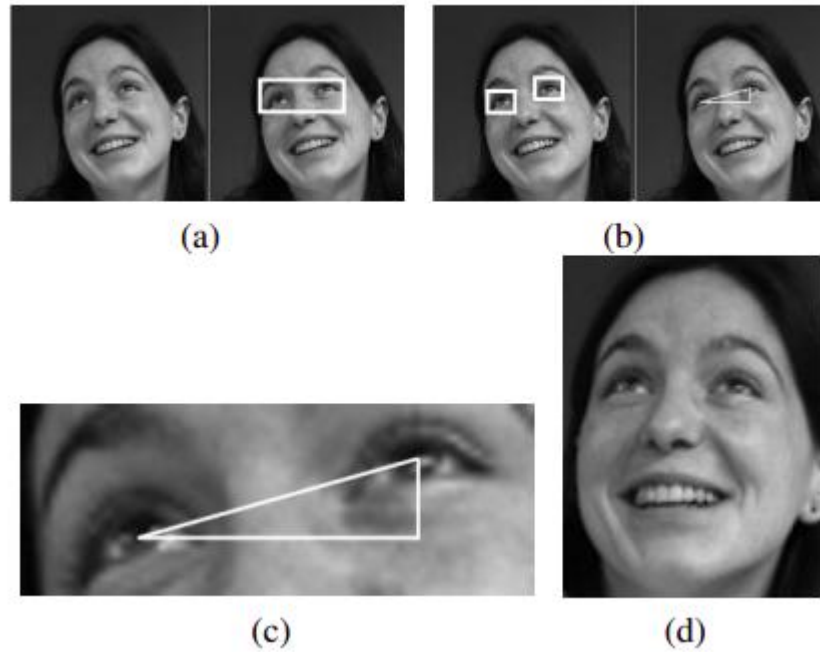


Рис 2.10 Поворот зображення за координатами очей (а) - пошук очей на зображенні, (b) - знаходження координат зіниць (с) - обчислення кута повороту (d) - результат вирівнювання

2.2 Висновки до другого розділу

Для задачі розпізнавання облич згорткові нейронні мережі мають значні переваги над іншими методами. Однак не всі архітектури ЗНМ підходять для вирішення задачі ідентифікації. За критерієм точністю ідентифікації незакритих облич, обрано систему FaceNet, що використовує технологію перетворення растрового зображення у векторне представлення. Крім того, в даній системі можна замінити наступні структурні: архітектура DCNN, функція втрат, розмір вихідного вектору. Гнучкість системи дозволить провести велику кількість експериментів комбінуючі та замінюючи дані блоки та обрати найкращий підхід.

Для детекції облич буде використовуватися MTCNN оскільки ця технологія демонструє стійкість до поворотів, зсувів та перекриттів нижньої частини обличчя.

РОЗДІЛ 3. РОЗРОБКА СИСТЕМИ РОЗПІЗНАВАННЯ ЧАСТКОВО ЗАКРИТИХ ОБЛИЧ

3.1 Обґрунтування та вибір початкових даних для розробки програмного продукту

На початкових етапах розробки, було вирішено використати готовий набір даних RMFD. Оскільки на даний момент це найбільший датасет з замаскованими обличчями (90000 зображень 525 людей), в теорії даний датасет підходив для використання, бо містив велику кількість загалом, близько 5 різних фотографій на людину зроблених у різних ситуаціях, зображення людей у масках та без, при різному освітленні та різних умовах зйомки тощо . Проте під час практичного використання виявилось, що датасет містить суттєву кількість помилково ідентифікованих облич. Було прийнято рішення не використовувати його для навчання та тестування, оскільки даний датасет необхідно заново розмітити. Однак, слід зазначити, що використання реальних фотографій людей у масках може суттєво підвищити якість ідентифікації.

Більшість датасетів, що є у відкритому доступі містять лише одне зображення на людину, що є недостатнім для вирішення задачі ідентифікації. Або загальна кількість зображень є недостатньою для продуктивного навчання моделі. Тому, було прийнято рішення згенерувати набір даних із штучно доданими зображеннями масок. За основу було взято набори даних CASIA [2] та VGGFace2 [3]. На рис 3.1 зображено згенерований набір для зображення з датасету CASIA в результаті накладання медичних масок різних відтінків, чорних тканевих масок та респіраторів N95 і KN95. Накладання маски здійснювалося за допомогою пакету MaskTheFace [37]. Додавання різних типів масок має на меті зробити модель більш стійкою, вона буде менш чутливою до кольору або типу маски, яку носить людина на зображенні.

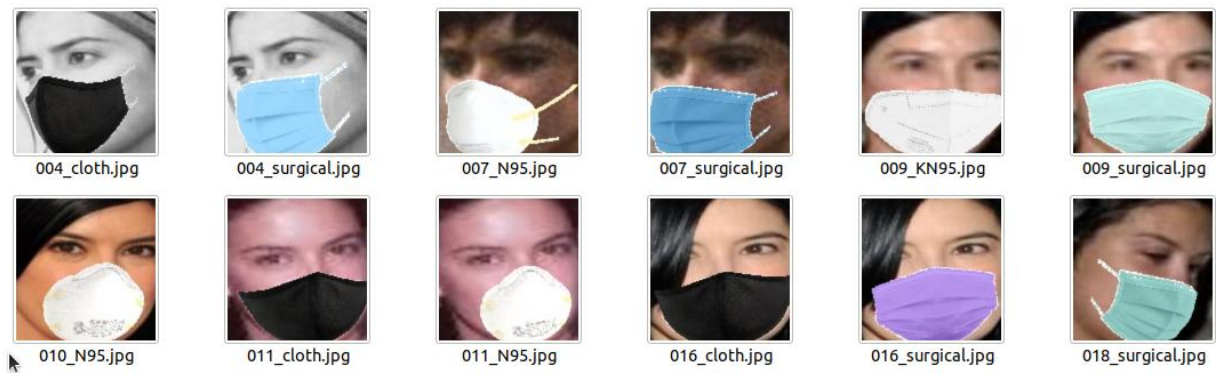


Рис 3.1. Приклад згенерованих зображень з CASIA_masked

В таблиці 2 наведено кількість даних в початкових та у згенерованих наборах даних. Таким чином, маємо базу зображень для формування тестової та навчальної вибірок. Як можна побачити з таблиці кількість згенерованих зображень непропорційно відрізняється від розміру початкових наборі даних. Це пов'язано з тим, що деякі зображення облич були занадто низької якості або неможливо було знайти певні точки обличчя (правий або лівий кут губ та ніс), що унеможливило коректне відпрацювання пакету, тому їх довелося вилучити з набору даних.

Таблиця 2

Набір даних	Кількість осіб	Кількість зображень	Назва згенерованого набору	Кількість зображень
CASIA	10,575	494,414	CASIA_masked	492,832
VGGFace2	9131	3,310000	VGGFace2_masked	2,024,897
LFW	5749	13233	LFW_masked	64,500

Отже, для проведення дослідження було виконано етап генерації нового набору даних, що необхідний для тренування та тестування нейронної мережі, на основі існуючих рішень, враховуючи недоліки попередніх досліджень.

3.2 Розробка підходу до розпізнавання частково закритих облич

На базі існуючих методів, розробимо підхід до розпізнавання частково закритих облич.

В якості основи буде використовуватися DCNN. Для процесу навчання, валідації і тестування моделі необхідно розділити дані на три вибірки. Розподіл здійснено наступним чином (рис 3.2).

Рис 3.2 Розподіл даних

Автори DeepFace [13] наводять послідовність необхідну для рішення задачі ідентифікації. Ця послідовність зображена на рис 3.3 і її буде взято за основу.

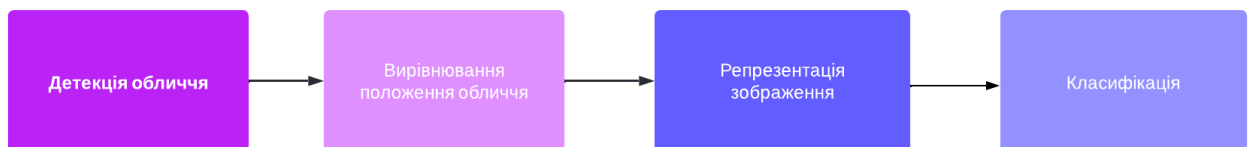


Рис 3.3 Процес розпізнавання

Для детекції обличчя було обрано фреймворк MTCNN, рис 3.4. В якості програмної реалізації використовується бібліотека deepface, що містить вбудовану реалізацію вирівнювання обличчя за положенням очей, метод

описано в пункті 2.1.7. Оскільки навколокулярна зона в даній задачі вважається видимою, то такий підхід до вирівнювання можна застосовувати.

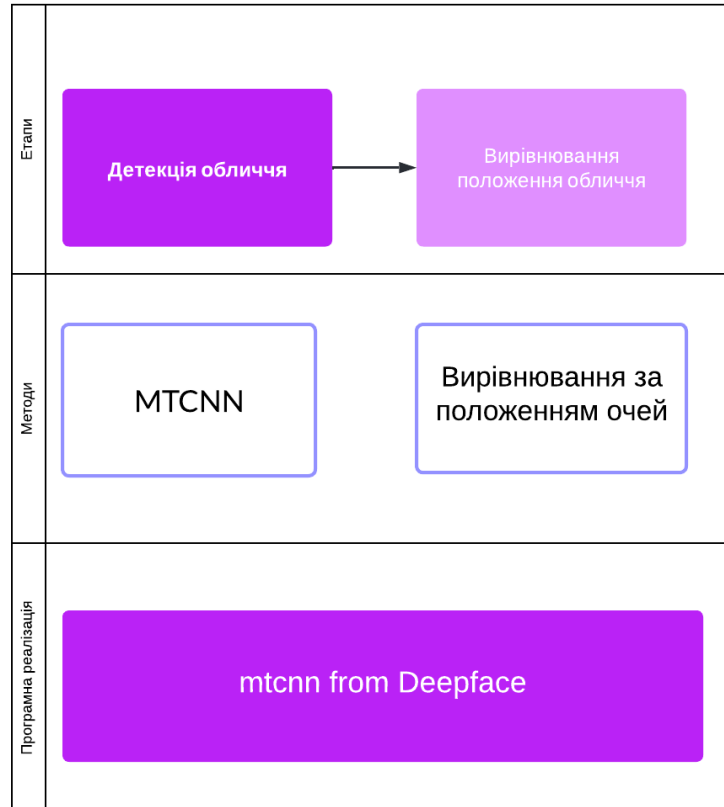


Рис 3.4 Етап детекції та вирівнювання обличчя

На рис 3.5 наведено приклад детекції та вирівнювання для обличчя без маски. Зображення (1) - вхідне, (2) - після детекції, (3) - після вирівнювання.



Рис 3.5 Приклад детекції та вирівнювання

Для ідентифікації необхідно виконати репрезентацію та класифікацію зображень рис 3.6.

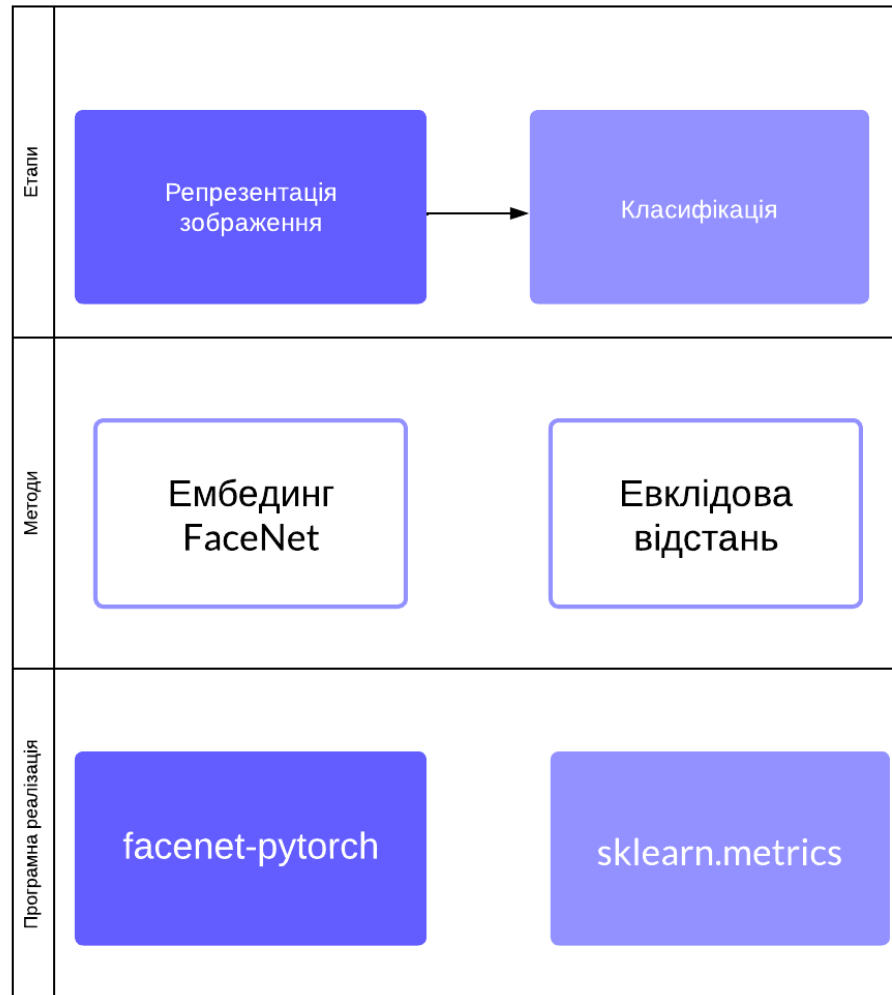


Рис 3.6 Етап репрезентації та класифікації

В якості репрезентації зображення було обрано підхід, оснований на процесі обчислення ембединг вектору. Для реалізації обрано мережу FaceNet, а саме її модифікацію FaceNet - 512, де 512 - розмірність вектору (замість класичної 128). Блок DCNN у FaceNet було замінено на InceptionResNet. InceptionResNet обрано в результаті порівняння результатів навчання серед декількох, а саме:

- InceptionResNet

- EffectiveNet
- Se-ResNetXt

В процесі експериментів використовувалися 3 різні функції активації Triplet Loss, Arc Face. Остаточна версія моделі зображена на рис 3.7.

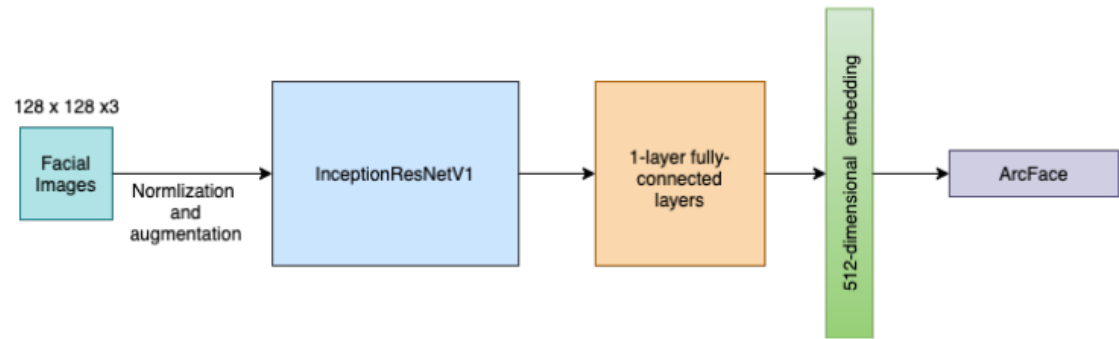


Рис 3.7 Модель для виконання ембедингу

Для програмної реалізації використовується пакет `facenet-pytorch` [38]. Також, для процесу навчання, тестування та візуалізації використовуються бібліотеки:

- `TensorFlow` - це бібліотека з відкритим кодом для чисельних обчислень та масштабованого машинного навчання. `TensorFlow` поєднує в собі безліч моделей машинного навчання та глибокого навчання та алгоритмів. Також міститься зручний інтерфейс роботи з гри.
- `PyTorch` — відкрита бібліотека машинного навчання на основі бібліотеки `Torch`
- Бібліотеки для здійснення маніпуляції з даними, попередньої обробки, математичних операцій: `pandas`, `numpy`, `torch` (а саме `torch.utils.data.DataLoader`, `torch.utils.data.Dataset`)
- Візуалізація даних - `matplotlib`
- Візуалізація прогресу ітераційних процесів - `tqdm`
- Робота з метриками - `sklearn.metrics`

- Для роботи з структурними блоками нейронних мереж використовується пакет *torch.nn*
- Імплементация мережі EfficientNet для PyTorch, містить декілька версій EfficientNet, навчених на датасеті ImageNet - *EfficientNet-PyTorch* [39]

Замість класичної реалізації класифікації, де на виході останнього шару нейронної мережі ми маємо ідентифікатор класу, в задачі ідентифікації останній етап - порівняння значень векторів. Тому, для цього етапу використовують методи для аналізу подібності цих векторів. Найпростіша та класична реалізація, це використання Евклідової відстані або косинусу подібності. Емпіричним шляхом було обрано евклідову відстань, в якості основного методу, обчислюється за формулою:

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

Для обрахунку евклідової відстані та косинуса подібності використовується бібліотека *sklearn.metrics*, пакет *pairwise*. На рис 3.8 візуалізовано приклад порівняння двох зображень без масок з використанням ембедингу та евклідової відстані.

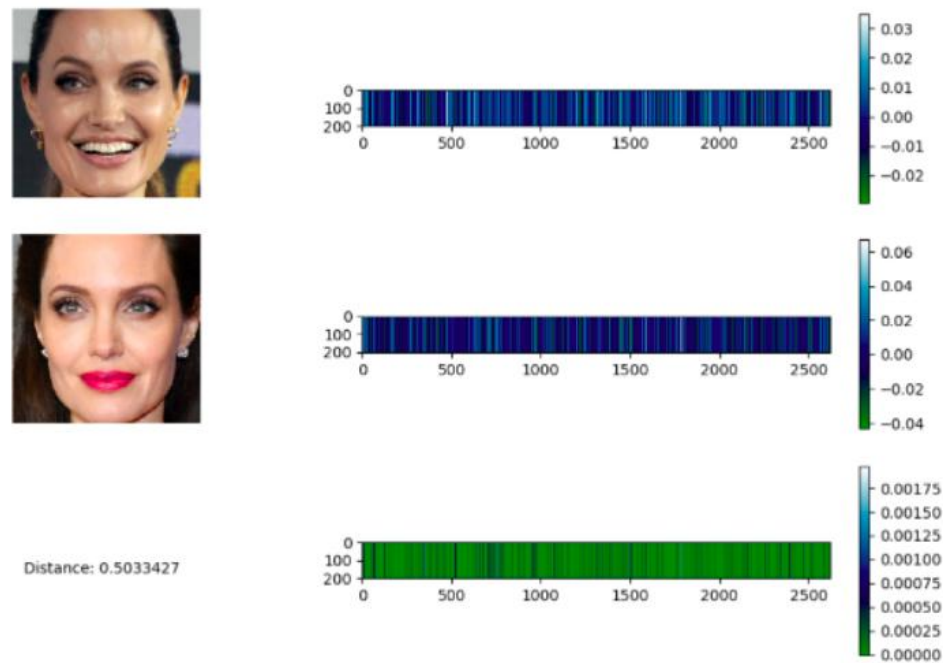


Рис 3.8 Приклад ембедінгу та порівняння зображень

На основі запропоного підходу в процесі створення системи буде розроблено дві підсистеми: детекції та ембедінгу, рис 3.9.

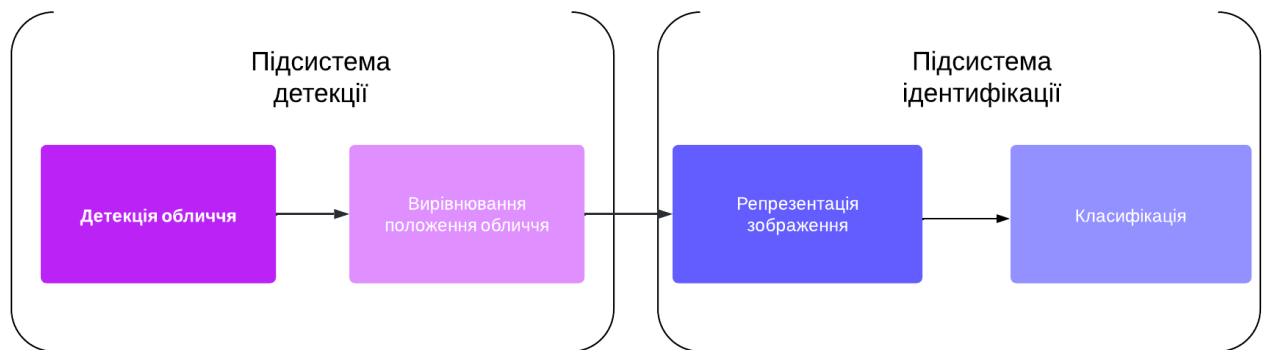


Рис 3.9 Підсистеми системи

3.3 Процес навчання нейронної мережі

3.3.1 Програмна реалізація навчання нейронної мережі

Для створення системи розпізнавання облич частково закритих облич використовувалася мова програмування Python. Основна причина вибору мови полягала у широкому виборі наявних бібліотек та пакетів та інтеграції з GCP.

Оскільки набір даних для тренування мережі був досить об'ємним (більше 12 ГБ), можливостей безкоштовної версії Google Collaboratory було недостатньо.

Під час розробки використовувалися Jupyter Notebook, фактично це інструмент, що дозволяє інтерактивно взаємодіяти з кодом в процесі проведення обчислень. Тому, в якості середовища розробки використовувався веб сервіс JupyterLab[40], оскільки дане середовище дозволяє взаємодіяти з Jupyter ноутбуками та різними типами даних, які були необхідні при розробці, а також є вбудованим в образ Deep Learning VM [41].

Для процесу навчання нейронної мережі необхідно було використати додаткові обчислювальні можливості. Тому, запропоновано використати віртуальне середовище Deep Learning Virtual Machine, компонент Google Cloud Platform (GCP). Ключова відмінність даного середовища від інших типів віртуальних машин (VM), що представлені в GCP, полягає в готових предналаштуваннях, можливості більш розширеної конфігурації CPU і GPU, та в заздалегідь встановлених пакетах для глибокого навчання (TensorFlow, PyTorch, scikit-learn, jupyter). Всі подальші обчислення виконувалися на CPU 4 vCPUs (26 ГБ, n1-highmem-4) або GPU NVIDIA T4. Для полегшення роботи з наборами даних використовувалося віртуальне сховище Cloud Storage та Vertex AI.

3.3.2 Вибір моделі нейронної мережі

В якості DCNN для FaceNet можна обрати різні архітектури, нижче наведено результати точності для різних архітектур, при різних функціях втрат. Triplet Loss - класична функція втрат, яка була використана в оригінальному дослідженні [32], ArcFace за результатами багатьох досліджень - є сучасним state of art [2, 42], тобто функцією, яка немає рівних за ефективністю та точністю. Однак в результаті експериментів, використання TripletLoss дає кращий результат як за точністю, так і за F - мірою.

Таблиця 3

Архітектура	Функція втрат	Трансферне навчання (VggFace)	Точність
InceptionResNetV1	ArcFace	Так	93
InceptionResNetV1	Triplet loss	Так	95
SE-ResNeXt-101	ArcFace	Ні	90
EfficientNet	Triplet loss	Так	91

Найкращий результат за точність показала, InceptionResNetV1 + функція втрат Triplet Loss. На рис 3.10 наведено процес навчання, а саме мінімізація функції втрат.

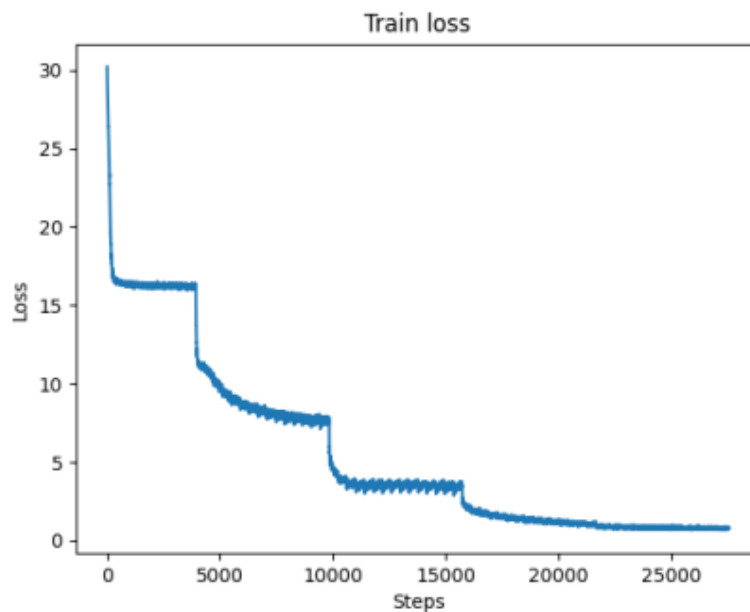


Рис 3.10 Мінімізація функції втрат

Встановлено поріг для значення евклідової відстані. Для тестової вибірки він становить 0,69.

3.4 Функції та структура системи розпізнавання

3.4.1 Функції системи розпізнавання

На рис 3.11 зображено функціональну схему для системи, що буде розроблено на базі підходу, розглянутого в пункті 3.2.

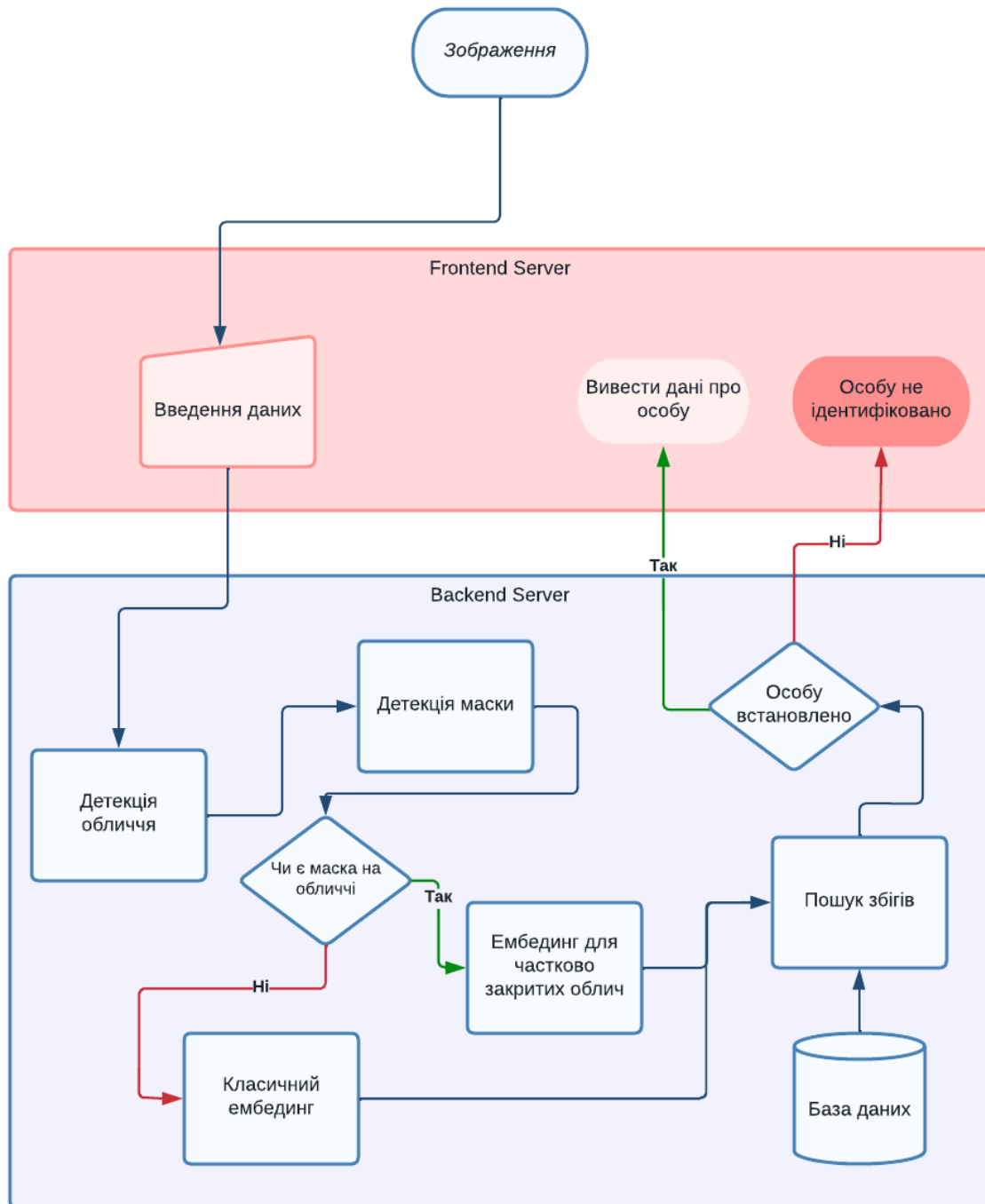


Рис 3.11 Функціональна схема системи розпізнавання

Розглянемо окремо кожну функцію системи.

Введення даних

На вхід системи подається зображення у форматі jpg. Зображення відправляється на сервер, де відбувається аналіз зображення.

Детекція обличчя

Для виявлення облич на зображенні використовується фреймворк MTCNN, який було детально описано в пункті 2.1. Він повертає координати обмежувальної рамки, де на вихідному зображенні знаходиться обличчя. Тобто, якщо на фотографії декілька осіб, буде виявлено контури декількох облич і подальші кроки будуть виконуватися для кожного з них окремо по черзі. Для програмної реалізації використовується бібліотека mtcnn.

Детекція маски

Цей крок представляє собою бінарну класифікацію - “на обличчі є маска” та “маски немає”. Для цього етапу буде використано сторонній пакет FaceMaskDetection [43]. Це рішення створено з використанням архітектури SSD, для тренування були використані датасети MAFA, Wider Face. В таблиці 4 наведено його точність в порівнянні з методами основаними на MobileNet та ResNet-50.

Таблиця 4

	Wider Face + MAFA	MMD + Kaggle + Kyiv Masked	UFMD
Mobile Net V2 - based		92.7	98
ResNet 50 - based	94	91.3	94
SSD - based	96		

Приклад роботи класифікатора рис 3.12, зображення взято з датасету RMFD



Рис 3.12 Приклад класифікації

Таким чином, якщо маска відсутня, то до зображення буде застосовуватися класичний алгоритм створення ембединг вектору, інакше ембединг для частково закритих облич. Дана класифікація допоможе покращити загальну точність системи та відфільтрувати обличчя без масок.

Класичний ембединг

Пакет для створення ембединг вектору для облич без маски, з використанням FaceNet, містить натреновану модель InceptionResNet на базі датасету VGG2, великому наборі даних зображень обличчя з понад 3 000 000 зображень. Попередньо навчена модель досягає 99,65% точності на наборі даних LFW та точності 99,05 на наборі даних CASIA. Дана нейронна мережа також використовувалася при трансферному навчанні - facenet-pytorch [38].

Ембединг для частково закритих облич

У випадку коли обличчя без маски, на даному етапі буде використано підхід запропонований у пункті 3.2.

Пошук збігів

Цей етап полягає в порівнянні отриманого в реальному часі ембедингу з ембедингом зображень облич користувачів у базі даних. Крім цього можлива ситуація, коли в базі ще немає даних про цю особу. Під порівнянням слід розуміти знаходження евклідової відстані рис 3.13.

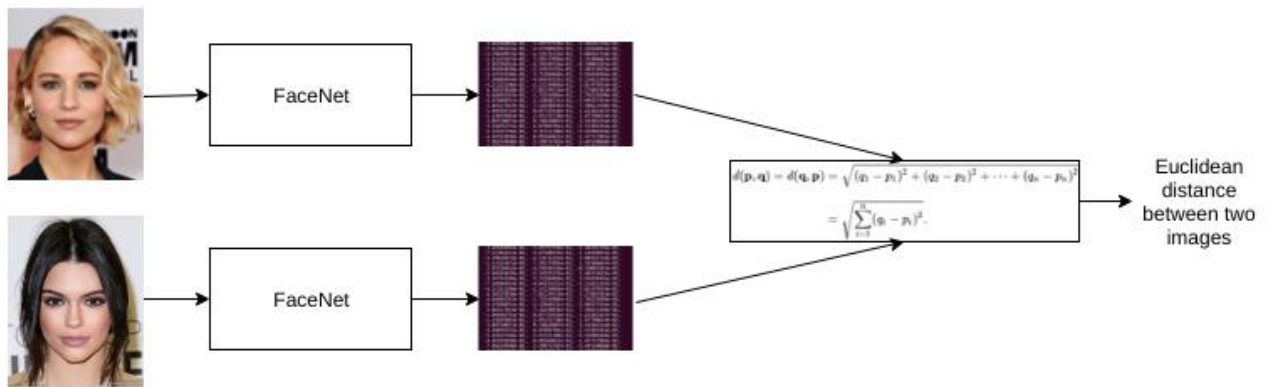


Рис 3.13 Візуалізація роботи FaceNet та порівняння зображень

Особу встановлено/Особу не встановлено

У разі вдалої ідентифікації повертається знайдена інформація про особу, що міститься в базі даних. Планується зберігати вік, стать, ім'я, прізвище, але в подальшому розвитку системи цей перелік можна розширювати. Якщо ідентифікація не вдала - має бути виведено повідомлення про неможливість встановлення особи.

3.4.2 Структура системи розпізнавання

На рис 3.14 зображено структурні елементи системи розпізнавання частково закритих облич

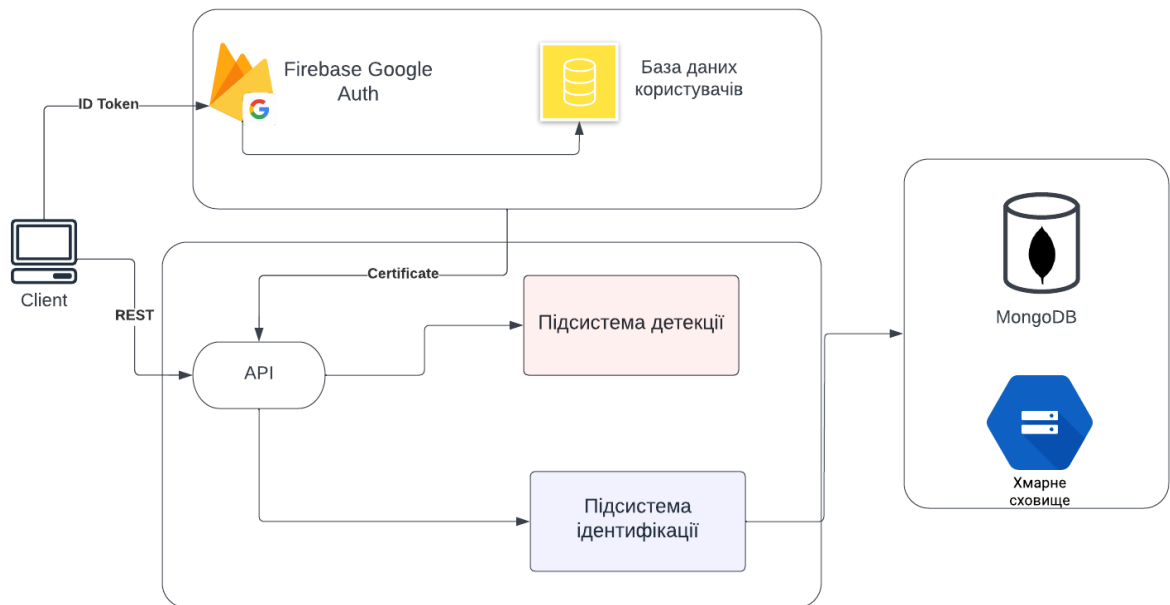


Рис 3.14 Структура системи розпізнавання

Client

Для побудови клієнтської частини буде використано Js бібліотеку React.js. Через клієнтську частину здійснюється взаємодія з користувачами - реєстрація, аутентифікація, отримання зображень, які необхідно ідентифікувати, отримання результатів ідентифікації. Для спілкування з веб сервером використовуються post та get REST запити.

Реєстрація та аутентифікація користувача

Процес реєстрації користувача в системі та аутентифікація здійснюється з використанням сервісу Firebase та протоколу OAuth2. Щоб зареєструвати користувача у системі, необхідно надати облікові дані для автентифікації від користувача, це адреса електронної пошти та пароль користувача або маркером OAuth від провайдера аутентифікації (Facebook, Twitter, Github). В даній системі в якості провайдера використовується Google. Потім ви передаєте ці дані до передаються до Firebase Authentication SDK. Сама перевірка здійснюється на основі JWT токенів. Сервіс дозволяє розділити ролі користувачів, що можна використати для створення багаторівневого доступу до результатів розпізнавання. Також сервіс Firebase має функцію Realtime database, що буде використано для збереження даних про користувачів.

WebServer

Для побудови серверної частини буде використано мову програмування Python, базу даних MongoDB, фреймворк Flask.

API

Забезпечує взаємозв'язок з клієнтом, для цього буде використано мікрофреймворк Flask. З усього функціоналу знадобиться маршрутизація та обробка HTTP методів. Крім цього, Арі модуль відповідає за передачу вхідних даних до підсистем та контролює їх взаємодію. Якщо особу встановлено, повертаються дані про особу та до існуючих зображень особи додається нове

зображення. Якщо особу не встановлено на запит клієнта повертається повідомлення про невдачу ідентифікації, створюється нова сутність в базі даних.

Підсистема детекції

На вхід отримує зображення, виконує детекцію облич, вирівнювання та детекцію маски. На виході отримуємо зображення або чергу з зображень облич, що необхідно ідентифікувати. Для кожного зображення є відповідний клас наявності маски на обличчі.

Підсистема ідентифікації

На вхід отримує зображення обличчя, генерує унікальний для кожної особи ембедінг вектор. Далі відбувається пошук за найбільш подібними значеннями у системі. Якщо особу встановлено, повертається id особи в бд. Підсистема не взаємодіє зі сховищем даних.

Зберігання даних

Для зберігання даних використовуватимуться дві технології - нереляційна база даних MongoDB та хмарне сховище Cloud Storage. В базі даних будуть зберігатися дані осіб, а саме ім'я, прізвище, вік та стать та посилання на зображення, що зберігаються у Cloud Storage.

3.5 Висновки до третього розділу

На підготовчому етапі розробки нейронної мережі було проведено підготовку даних. Використовуючи існуючі набори даних VggFace2, Casia, LFW, згенеровано три набори даних VggFace_masked, Casia_masked, LFW_masked. Процес генерації полягав у накладенні маски на зображення обличчя людини.

Запропоновано підхід до розпізнавання облич на базі існуючих рішень. Для кожного етапу процес розпізнавання облич, а саме детекції, вирівнювання

зображення, репрезентації та класифікації, було визначено методи та технології, а також інструменти для програмної реалізації.

В процесі розробки підсистеми ідентифікації розглянуто декілька архітектур нейронних мереж (InceptionResNet, EfficientNet, SE-ResNeXt), проведено експерименти з різними наборами даних та різними параметрами навчання, та різними варіантами функції втрат. В результаті експериментів, модель InceptionResNet, з триплетною функцією втрат показала найкращий результат за точністю розпізнавання. Точність для облич із масками на тестовому наборі даних складає 95,52%.

На базі даного підходу розроблено систему детекції та ідентифікації. Спроековано та розроблено систему розпізнавання облич.

РОЗДІЛ 4. ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ ОТРИМАНИХ РЕЗУЛЬТАТІВ

4.1 Характеристика оцінок ефективності результатів розпізнавання частково закритих облич

На рис 4.1 наведено матрицю невідповідностей метрик, які використовуються для визначення правильності та точності роботи моделі, в даному випадку класифікатору, а саме:

- TP — істинно-позитивний результат, класифікатор вірно відніс об'єкт до заданого класу;
- TN — істинно-негативний результат, класифікатор вірно підтверджує, що об'єкт не належить до заданого класу;
- FP — хибно-позитивний результат, класифікатор помилко відносить об'єкт до заданого класу;
- FN — хибно-негативний результат, класифікатор хибно стверджує, що об'єкт не належить до розглянутого класу.

		Ground truth		
		+	-	
Predicted	+	True positive (TP)	False positive (FP)	Precision = $TP / (TP + FP)$
	-	False negative (FN)	True negative (TN)	
		Recall = $TP / (TP + FN)$		
				Accuracy = $(TP + TN) / (TP + FP + TN + FN)$

Рис 4.1 Матриця невідповідності метрик

Для оцінки якості моделі, використовуючи дані метрики, необхідно знайти значення наступних показників:

- Accuracy (точність) - оцінка долі правильно класифікованих зображень;
- Recall (повнота) - оцінка частки знайдених об'єктів класу до загальної кількості об'єктів класу. Інакше кажучи, наскільки добре наш класифікатор знаходить об'єкти з класу;
- Precision (також точність) - оцінка частки об'єктів класу серед об'єктів виділених класифікатором;
- F - міра - середнє гармонійне між precision і recall, буде використовуватися для загальної оцінки якості роботи моделі

$$F_1 = \left(\frac{Prec^{-1} + Recall^{-1}}{2} \right)^{-1} = 2 \cdot \frac{Prec \cdot Recall}{Prec + Recall}$$

Використання F - міри обумовлено необхідністю максимізації показників Recall та Precision, при цьому одночасна максимізація обох показників є взаємно виключною, таким чином для оцінки моделі буде використовуватися значення F - міри. В даному випадку використано формулу F1, де рівнозначно враховуються показники як Recall, так і Precision, тобто немає пріоритизації ваги якогось з показників. Відповідно, значення F - міри буде однаково зменшуватися при зменшенні значень і Precision, і Recall.

Таким чином, основні оцінки ефективності методу - F міра та Accuracy.

4.2 Опис результатів роботи на основі контрольного прикладу

Приклад результату ембедінгу для замаскованих зображень особи Geoff Hoop з набору даних LFW.

На вхід підсистеми ідентифікації частково закритих облич отримано зображення, на виході отримуємо ембедінг вектор. В даному випадку було отримано однаковий вектор для всіх трьох зображень, що є надзвичайно добрим результатом. Зазвичай на основі результатів розподілу валідаційної вибірки отримуємо поріг для евклідової відстані. Якщо відстань між векторами більша за цей поріг, тоді зображення належать до різних класів. В розглянутому випадку відстань для трьох зображень є 0 (рис 4.2).

Рис 4.2 Ембединг для зображення Geoff_Hoop

Оскільки евклідова відстань менше за встановлений поріг, можемо зробити висновок, що колір маски, тип маски, поворот голови в контрольному прикладі не вплинули на результат ідентифікації. Якщо розглядати всі 8 зображень Geoff Hoop, то точність для цього прикладу склала 87 %, з 8 зображень, 1 було розпізнано хибно позитивно, що є результатом нижче за середнє, рис 4.3 .

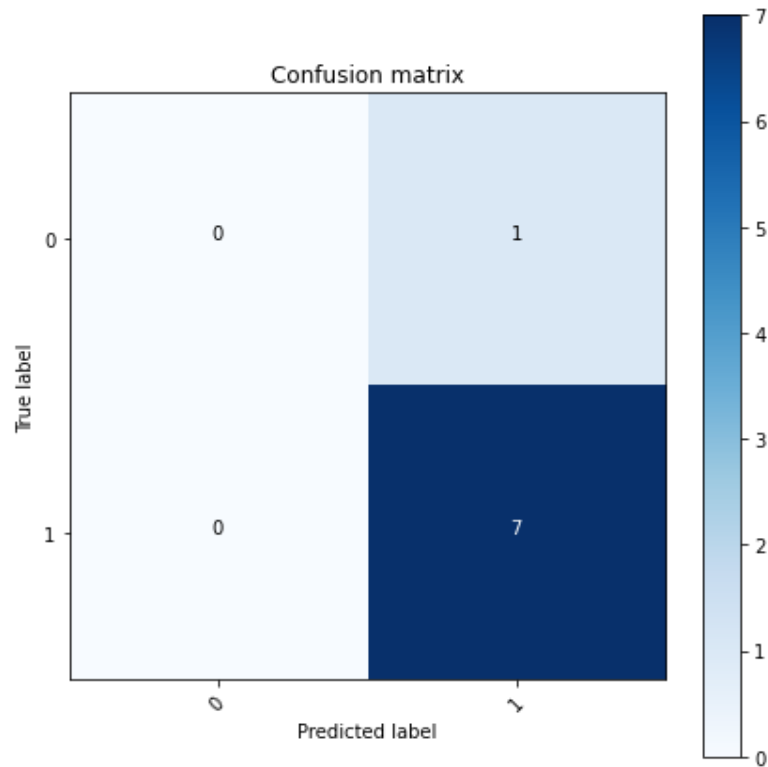


Рис 4.3

4.3 Оцінка ефективності результатів дослідження

Результати тестування етапу ембедінгу FaceNet, розмірність вибірки 648 зображень, датасет LFM, функція втрат Triplet Loss, DCNN - InceptionResNetV1. На рис 4.4 наведено метрики (Precision, Recall, F1-score) для тестової вибірки.

Accuracy: 0.9552469135802469				
	precision	recall	f1-score	support
0	0.93	0.98	0.96	331
1	0.98	0.92	0.95	317
accuracy			0.96	648
macro avg	0.96	0.95	0.96	648
weighted avg	0.96	0.96	0.96	648

Рис 4.4 Precision, Recall, F1-score для тестової вибірки

На рис 4.5 зображено матрицю невідповідностей. Ми можемо побачити, що кількість хибно позитивних випадків складає лише 5 зображень (0,7%), а

хибно негативних відповідно - 24 (3,7%). Таким чином отримуємо, що значення F-міри - 96%.

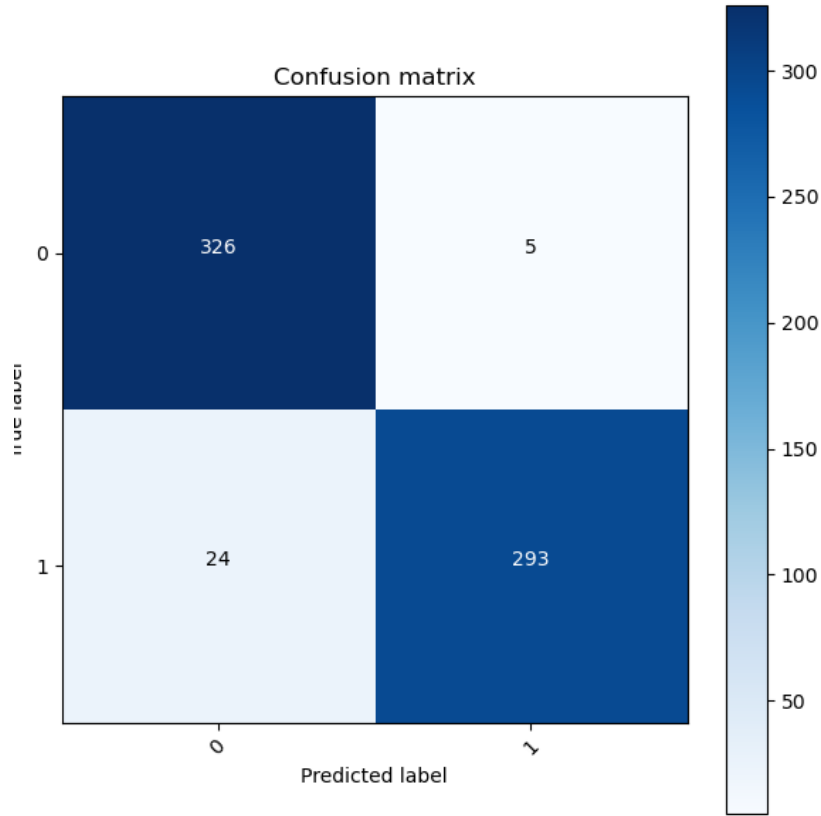


Рис 4.5. Матриця невідповідностей для тестової вибірки

За результатами тестування можна зробити висновки, стосовно роботи системи:

- До переваг можна віднести 95,5% точності ідентифікації обличчя в масці та низькі показники хибно-негативних та хибно-позитивних результатів ідентифікації.
- Детекція обличчя є вдалою при поворотах та наклонах голови. Отримано результат при декількох обличчях на зображенні.

4.4 Висновки до четвертого розділу

Для оцінки ефективності розпізнавання використовується точність та F-міра. При цьому значення F-міри обчислюється як середнє гармонічне між оцінкою частки знайдених об'єктів класу до загальної кількості об'єктів класу та оцінка частки об'єктів класу серед об'єктів виділених класифікатором.

На тестовому прикладі було продемонстровано результати роботи системи. Отримано однаковий ембедінг вектор для трьох зображень, однак точність розпізнавання для 8 зображень була нижчою ніж в середньому по вибірці - 87%. Запропонований підхід на тестовому наборі даних дав результат 95.5 % та F-міра становила 96%.

ВИСНОВКИ

В кваліфікаційній роботі проаналізована проблема розпізнавання частково закритих облич, розглянуто методи та технології розв'язання завдань розпізнавання та ідентифікації облич.

Проведено дослідження існуючих методів та технологій ідентифікації облич та створено систему для розпізнавання осіб з частково закритими обличчями, а саме засобами індивідуального захисту, з можливістю подальшого використання в сфері безпеки.

Для досягнення мети було розв'язано наступні задачі:

- дослідження існуючих підходів, методів, виявлення їх недоліків, місць для вдосконалення;
- розробка підходу до розпізнавання частково закритих облич, використовуючи існуючі методи та технології;
- проектування та реалізація системи розпізнавання;
- оцінка ефективності роботи системи.

Для вирішення завдання ідентифікації були використані технології глибинного навчання, а саме наступні архітектури нейронних мереж: MTCNN, FaceNet.

Проведено попередню обробку наборів даних, та згенеровано новий датасет для навчання та тестування нейронних мереж, використано набори даних VggFace2, Casia, LFW. Для навчання нейронної мережі було обрано хмарне середовище Deeplearning VM з платформи Google Cloud із 8 ГБ CPU та GPU.

Для створення підсистеми детекції облич використано метод MTCNN. З бібліотеки deerface, що містить готову реалізацію вирівнювання облич.

В процесі розробки підсистеми ідентифікації розглянуто декілька архітектур нейронних мереж (InceptionResNet, EfficientNet, SE-ResNeXt), проведено експерименти з різними наборами даних та різними параметрами навчання, та різними варіантами функції втрат. В результаті експериментів, модель InceptionResNet, з триплетною функцією втрат показала найкращий результат за точністю розпізнавання. Точність для облич із масками на тестовому наборі даних складає 95,52%. В результаті аналізу отриманих результатів було розроблено підхід до ідентифікації частково закритих облич, що ліг в основу системи розпізнавання частково закритих облич людей. Даний підхід використовує систему FaceNet, в основі якої мережа InceptionResNetV1. В якості функції витрат - триплетна функцію втрат. Для порівняння ембединг векторів обчислюється евклідова відстань.

В випадку отримання на вхід обличчя без масок використано систему розпізнавання FaceNet, з використанням мережі InceptionResNetV1, натреновану на датасеті VggFaces. Точність на незамаскованому датасеті складає 99%. Для виявлення маски на обличчі людини використано програмне рішення Face Mask Detection.

На основі запропонованих технологій і методів була розроблена система розпізнавання частково закритих облич.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Adjabi I., Ouahabi A., Benzaoui A., Taleb-Ahmed A. (2020), "Past, present, and future of face recognition: A review", *Electronics*, Vol. 9, no. 8, p. 1188.
2. Kortli Y., Jridi M., Falou A.A., Atri M. (2020), "Face recognition systems: A survey.", *Sensors*.
3. Ngan M., Grother P., Hanaoka K. (2020), "Ongoing Face Recognition Vendor Test (FRVT) Part 6A: Face recognition accuracy with masks using pre-COVID-19 algorithms", .
4. Zhang L., Verma B., Tjondronegoro D., Chandran V. (2018), "Facial Expression Analysis under Partial Occlusion", *ACM Computing Surveys*, Vol. 51, no. 2, pp. 1–49.
5. Chengeta K., Viriri S. (2019), "A review of local, holistic and deep learning approaches in facial expressions recognition", In: *2019 Conference on Information Communications Technology and Society (ICTAS)*. IEEE, pp. 1–7.
6. "COVID-19 behaviour tracker | Institute of Global Health Innovation | Imperial College London", available at: <https://www.imperial.ac.uk/centre-for-health-policy/our-work/our-response-to-covid-19/covid-19-behaviour-tracker/> (accessed April 20, 2022).
7. "Facial Recognition Market Size & Trends Report, 2021-2028", available at: <https://www.grandviewresearch.com/industry-analysis/facial-recognition-market> (accessed April 20, 2022).
8. "Woody Bledsoe: His Life and Legacy | AI Magazine", available at: <https://doi.org/10.1609/aimag.v17i1.1207> (accessed April 20, 2022).
9. "Computer Recognition Of Human Faces by Takeo Kanade", available at: https://www.goodreads.com/book/show/213526.Computer_Recognition_Of_Human_Faces (accessed April 20, 2022).
10. Sirovich L., Kirby M. (1987), "Low-dimensional procedure for the characterization of human faces.", *Journal of the Optical Society of America. A, Optics and image science*, Vol. 4, no. 3, pp. 519–524.
11. Turk M., Pentland A. (1991), "Eigenfaces for recognition.", *Journal of Cognitive*

- Neuroscience*, Vol. 3, no. 1, pp. 71–86.
12. "Face Recognition Technology (FERET) | NIST", available at: <https://www.nist.gov/programs-projects/face-recognition-technology-feret> (accessed May 17, 2022).
 13. Taigman Y., Yang M., Ranzato M., Wolf L. (2014), "DeepFace: Closing the Gap to Human-Level Performance in Face Verification", In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 1701–1708.
 14. "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments", .
 15. Yang S., Luo P., Loy C.C., Tang X. (2016), "WIDER FACE: A face detection benchmark", In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 5525–5533.
 16. Ge S., Li J., Ye Q., Luo Z. (2017), "Detecting Masked Faces in the Wild with LLE-CNNs", In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 426–434.
 17. Wang J., Yuan Y., Yu G. (2017), "Face Attention Network: An Effective Face Detector for the Occluded Faces", arXiv.
 18. Najibi M., Samangouei P., Chellappa R., Davis L. (2017), "SSH: Single Stage Headless Face Detector", arXiv.
 19. Zhang S., Zhu X., Lei Z., Shi H., Wang X., Li S.Z. (2017), "SS³FD: Single Shot Scale-invariant Face Detector", arXiv.
 20. Lin T.-Y., Goyal P., Girshick R., He K., Dollár P. (2017), "Focal Loss for Dense Object Detection", arXiv.
 21. He K., Zhang X., Ren S., Sun J. (2015), "Deep Residual Learning for Image Recognition", arXiv.
 22. Belhumeur P.N., Hespanha J.P., Kriegman D.J. (1997), "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection", *IEEE transactions on pattern analysis and machine intelligence*, Vol. 19, no. 7, pp. 711–720.
 23. Lades M., Vorbruggen J.C., Buhmann J., Lange J., von der Malsburg C., Wurtz R.P., Konen W. (1993), "Distortion invariant object recognition in the dynamic link architecture", *IEEE Transactions on Computers*, Vol. 42, no. 3, pp. 300–311.

24. Guetbi C., Kouame D., Ouahabi A., Chemla J.P. (1998), "Methods based on wavelets for time delay estimation of ultrasound signals", In: *1998 IEEE International Conference on Electronics, Circuits and Systems. Surfing the Waves of Science and Technology (Cat. No.98EX196)*. IEEE, pp. 113–116.
25. Ahonen T., Hadid A., Pietikäinen M. (2006), "Face description with local binary patterns: application to face recognition.", *IEEE transactions on pattern analysis and machine intelligence*, Vol. 28, no. 12, pp. 2037–2041.
26. Simonyan K., Zisserman A. (2014), "Very Deep Convolutional Networks for Large-Scale Image Recognition", arXiv.
27. Phung, Rhee (2019), "A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets", *Applied Sciences*, Vol. 9, no. 21, p. 4500.
28. Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. (2014), "Going Deeper with Convolutions", arXiv.
29. "<https://arxiv.org/pdf/1602.07261v2.pdf>", available at: <https://arxiv.org/pdf/1602.07261v2.pdf> (accessed May 4, 2022).
30. Zhang K., Zhang Z., Li Z., Qiao Y. (2016), "Joint face detection and alignment using multitask cascaded convolutional networks", *IEEE signal processing letters*, Vol. 23, no. 10, pp. 1499–1503.
31. Hofer P., Roland M., Schwarz P., Schwaighofer M., Mayrhofer R. (2021), "Importance of different facial parts for face detection networks", In: *2021 IEEE International Workshop on Biometrics and Forensics (IWBF)*. IEEE, pp. 1–6.
32. Schroff F., Kalenichenko D., Philbin J. (2015), "FaceNet: A unified embedding for face recognition and clustering", In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 815–823.
33. Yu B., Tao D. (2019), "Deep metric learning with triplet margin loss", In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, pp. 6489–6498.
34. Yang L., Hanneke S., Carbonell J. (2013), "A theory of transfer learning with applications to active learning", *Machine learning*, Vol. 90, no. 2, pp. 161–189.
35. Pan S.J., Yang Q. (2010), "A survey on transfer learning", *IEEE transactions on knowledge and data engineering*, Vol. 22, no. 10, pp. 1345–1359.

36. Karaaba M.F., Surinta O., Schomaker L., Wiering M.A. (2015), "In-Plane Rotational Alignment of Faces by Eye and Eye-Pair Detection", Unpublished.
37. "GitHub - aqeelanwar/MaskTheFace: Convert face dataset to masked dataset", available at: <https://github.com/aqeelanwar/MaskTheFace> (accessed May 3, 2022).
38. "GitHub - AravindChandradoss/facenet-pytorch: Face Recognition with One Shot (Siamese network) and Model based (PCA) using Pretrained Pytorch face detection and recognition models", available at: <https://github.com/AravindChandradoss/facenet-pytorch> (accessed May 9, 2022).
39. "GitHub - lukemelas/EfficientNet-PyTorch: A PyTorch implementation of EfficientNet and EfficientNetV2 (coming soon!)", available at: <https://github.com/lukemelas/EfficientNet-PyTorch> (accessed May 9, 2022).
40. "Project Jupyter | Home", available at: <https://jupyter.org> (accessed May 9, 2022).
41. "Deep Learning VM Images | Google Cloud", available at: <https://cloud.google.com/deep-learning-vm> (accessed May 9, 2022).
42. Srivastava Y., Murali V., Dubey S.R. (2019), "A Performance Comparison of Loss Functions for Deep Face Recognition", arXiv.
43. "GitHub - AIZOOTech/FaceMaskDetection: 开源人脸口罩检测模型和数据 Detect faces and determine whether people are wearing mask.", available at: <https://github.com/AIZOOTech/FaceMaskDetection> (accessed May 9, 2022).

Додаток А: Приклади зображень з датасетів

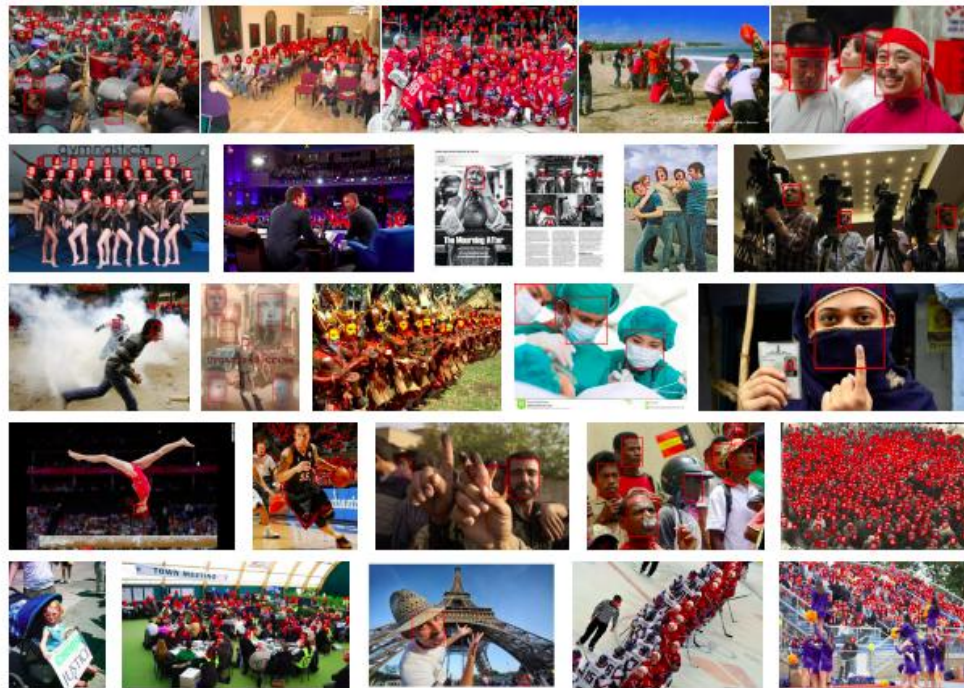


Рис 1. Wider Face [15]

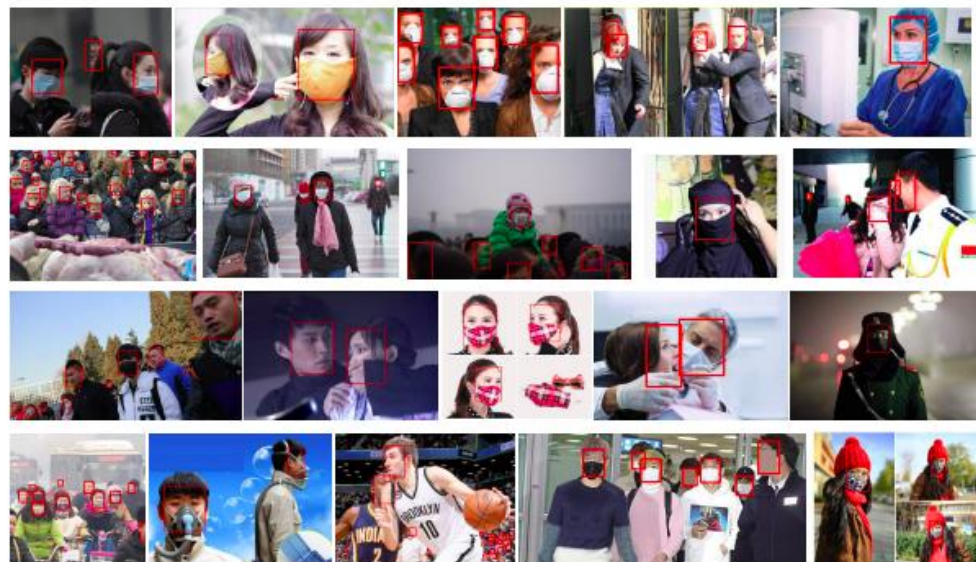


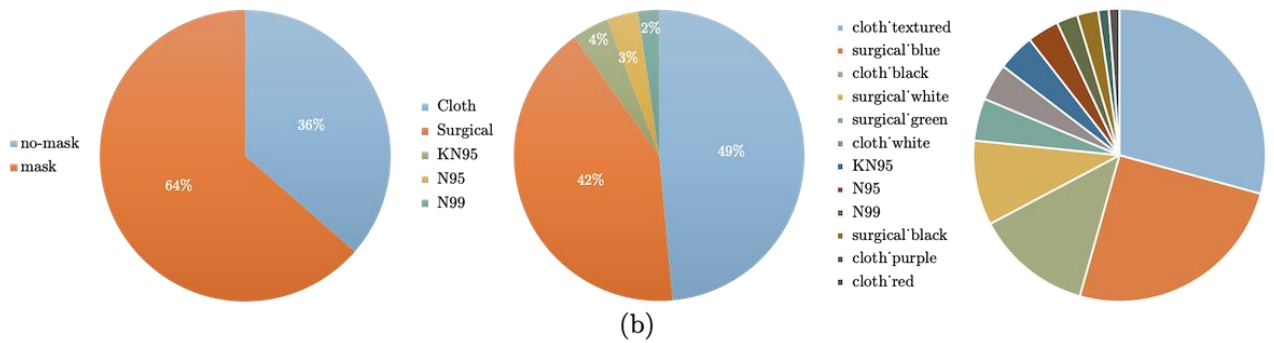
Рис 2 . MAFA [16]



Puc 3. LFW [14]



(a)



Puc 4. MFR2