Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет *Інфокомунікаціїі*
(повна назва)

Кафедра *Інформаційно-мережної інженерії*
(повна назва)

# КВАЛІФІКАЦІЙНА РОБОТА
# Пояснювальна записка

Рівень вищої освіти *другий (магістерський)*

*Дослідження методів розпізнавання обличчя у масках*

(тема)

Виконав:
студент 2 курсу, групи *ІМІм-20-2*
*Рєзнік Я.В.*

Спеціальності *172 Телекомунікації та радіотехніка*
(код і повна назва спеціальності)
Тип програми *Освітньо-професійна*
(освітньо-професійна або освітньо-наукова)
Освітня програма *Інформаційно-мережна інженерія*
(повна назва освітньої програми)
Керівник *проф. Безрук В.М.*
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____ *Безрук В.М.*
(підпис) (прізвище, ініціали)

2022 р.

Не містить відомостей, заборонених до відкритого публікування

Студент _____ *Рєзнік Я.В.*
( підпис ) (прізвище та ініціали)

Керівник _____ *Безрук В.М.*
( підпис ) (прізвище та ініціали)

Харківський національний університет радіоелектроніки

Факультет *Інфокомунікацій*
(повна назва)

Кафедра *Інформаційно-мережної інженерії*
(повна назва)

Рівень вищої освіти *другий (магістерський)*

Спеціальність *172 Телекомунікації та радіотехніка*
(код і повна назва)

Тип програми *Освітньо-професійна*
(освітньо-професійна або освітньо-наукова)

Освітня програма *Інформаційно-мережна інженерія*
(повна назва)

ЗАТВЕРДЖУЮ:
Зав. кафедри ІМІ_____
(підпис)
“____” _____2022 року

# ЗАВДАННЯ
## НА КВАЛІФІКАЦІЙНУ РОБОТУ

Студентові *Рєзніку Ярославу Володимировичу*
(прізвище, ім’я, по батькові)

1. Тема роботи *Дослідження методів розпізнавання обличчя у масках*

затверджені наказом університету від 08 листопада 2022 року № 1674 *Ст*

2. Термін подання студентом роботи до екзаменаційної комісії *13 травня 2022 р.*

3. Вихідні дані до роботи *Розпізнавання облич в масках, технології до використання: OpenCV, Keras, TensorFlow, Caffe-based detector та MobileNet. Нейромережі: згорткові, глибокого навчання. Генеративно-змагальні нейронні мережі.*

4. Перелік питань, що потрібно опрацювати в роботі
*Вступ*
*1. Аналіз предметної області*
*2. Розпізнавання маскованих облич*
*3. Аналіз технологій розпізнавання масок на обличчі*
*Висновки*

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри) *Слайди у форматі Power Point (назва, мета і задачі роботи, стан та історія розвитку предметної області, технології OpenCV, Tensorflow, Caffe-based detector, MS-COCO, MobileNetV2, Keras, моделі машинного навчання, генеративно-змагальні мережі, алгоритм розпізнавання маскованих облич, висновки)*

## КАЛЕНДАРНИЙ ПЛАН

| № з/п | Назва етапів атестаційної роботи | Строк виконання етапів роботи | Примітка |
|---|---|---|---|
| 1 | *Ознайомлення із завданням. Уточнення ТЗ* | *14.03.22* | *виконано* |
| 2 | *Підбір літератури за темою роботи* | *15.03-20.03.22* | *виконано* |
| 3 | *Виконання розділу 1* | *21.03-28.03.22* | *виконано* |
| 4 | *Виконання розділу 2* | *29.03-20.04.22* | *виконано* |
| 5 | *Виконання розділу 3* | *21.04-01.05.22* | *виконано* |
| 6 | *Оформлення пояснювальної записки* | *02.05-09.05.22* | *виконано* |
| 7 | *Оформлення презентаційного матеріалу,* | *10.05-11.05.22* | *виконано* |
| 8 | *підготовка до захисту у ЕК* | *12.05-15.05.22* | *виконано* |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

Дата видачі завдання *14.03.2022 р.*

Студент _____ *Рєзнік Я.В.*
( підпис ) (прізвище та ініціали)

Керівник роботи _____ *Безрук В.М.*
( підпис ) (прізвище та ініціали)

# РЕФЕРАТ

Пояснювальна записка: 63 с., 26 рис., 8 джерел, 2 додатки.

Об'єкт дослідження – методи розпізнавання маскованих облич.

Мета роботи – дослідження методів розпізнавання маскованих облич та інструментів розпізнавання захисних масок на обличчі.

Результати – в роботі були досліджені методи розпізнавання облич, прикритих маскою, проаналізований кожний з етапів цього процесу та розглянуті існуючі методики досягнення цих задач. Окремо розглянуті інструменти для розпізнавання масок на обличчі, проаналізована їхня ефективність.

РОЗПІЗНАВАННЯ ОБЛИЧ, ШТУЧНИЙ ІНТЕЛЕКТ, ЗАХИСНІ МАСКИ, ВИДИМІ ОБЛАСТІ, НЕЙРОННА МЕРЕЖА, МОДЕЛЬ, ОЗНАКА, ТЕХНОЛОГІЯ.

# THE ABSTRACT

Explanatory note: 63 p., 20 fig., 8 sources, 2 app.

The object of study is methods of recognition of masked faces.

The purpose of the work is to study the methods of recognizing masked faces and tools for recognizing protective masks on the face.

Results - the methods of recognition of masked faces were studied, each of the stages of this process was analyzed and the existing methods of achieving these tasks were considered. The tools for face mask recognition are considered separately, their efficiency is analyzed.

FACE RECOGNITION, ARTIFICIAL INTELLIGENCE, PROTECTIVE MASKS, VISIBLE AREAS, NEURAL NETWORK, MODEL, SIGN, TECHNOLOGY.

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ ..... 8
ВСТУП ..... 9
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ..... 10
1.1 Комп’ютерний зір ..... 10
1.2 Розпізнавання облич ..... 12
1.3 Штучний інтелект ..... 16
2 РОЗПІЗНАВАННЯ МАСКОВАНИХ ОБЛИЧ ..... 18
2.1 Принцип розробки MFR ..... 19
2.2 Моделі глибокого навчання ..... 21
2.3 Згорткові нейронні мережі ..... 21
2.3.1 Автокодери ..... 23
2.3.2 Генеративні змагальні мережі ..... 23
2.3.3 Глибока мережа довіри ..... 24
2.3.4 Вичерпне навчання з підкріпленням ..... 24
2.3.5 Конкретні глибокі мережі MFR ..... 24
2.4 Виділення ознак ..... 25
2.5 Виявлення маски ..... 27
2.6 Демаскування обличчя ..... 29
2.7 Відновлення обличчя ..... 30
2.8 Розпізнавання та відповідність облич ..... 31
2.9 Розпізнавання облич у масці ..... 34
3 АНАЛІЗ ТЕХНОЛОГІЙ РОЗПІЗНАВАННЯ МАСОК НА ОБЛИЧЧІ ..... 37
3.1 OpenCV ..... 37
3.1.1 Історія ..... 37
3.1.2 Застосування ..... 38
3.2 Caffe-based detector ..... 39
3.3 Keras ..... 40
3.4 TensorFlow ..... 41
3.5. MobileNetV2 ..... 42
3.5.1 MobileNetV1 ..... 43

3.5.2 MobileNetV2 ..... 43
3.5.3 Загальна архітектура ..... 43
3.5.4 Дослідження абляції ..... 45
3.5.5 Результати експерименту ..... 46
3.5.6 Виявлення об'єктів за допомогою MS COCO ..... 48
3.5.7 Семантична сегментація за допомогою PASCAL VOC 2012 ..... 49
ВИСНОВКИ ..... 50
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ ..... 52
ДОДАТОК А СЛАЙДИ ПРЕЗЕНТАЦІЇ ..... **Ошибка! Закладка не определена.**
ДОДАТОК Б ПУБЛІКАЦІЯ ЗА ТЕМАТИКОЮ РОБОТИ **Ошибка! Закладка не определена.**

# ПЕРЕЛІК СКОРОЧЕНЬ

FR (Face Recognition) — Розпізнавання облич.
MFR (Masked Face Recognition) — Розпізнавання маскованих облич
AI (Artifical Intelligence) – Штучний інтелект

# ВСТУП

Метою даної роботи є дослідження методів розпізнавання маскованих облич та інструментів розпізнавання захисних масок на обличчі.

Оскільки COVID-19 може поширюватися через контактні та забруднені поверхні, традиційні біометричні системи, засновані на паролях або відбитках пальців, більше не є безпечними. На цьому тлі, розпізнавання обличчя стає найбезпечнішим інструментом у забезпеченні громадської безпеки, оскільки позбавляє від необхідності торкатися будь-якого предмету.

Дослідження коронавірусу показали, що носіння маски як здоровими, так і хворими людьми значно знижує передачу вірусу. З іншого боку, носіння маски може призводити до таких соціальних проблем:

1) Шахраї та злодії використовують маски для крадіжки та вчинення злочинів, не будучи ідентифікованими.

2) Коли велика частина обличчя закрита маскою, контроль публічного доступу та аутентифікація обличчя стають надзвичайно складними завданнями.

3) Існуючі системи розпізнавання облич неефективні під час носіння маски, оскільки все зображення обличчя отримати неможливо.

4) Через ці проблеми маски для обличчя значно ускладнюють існуючі алгоритми розпізнавання облич.

Таким чином, перед науковою та інженерною спільнотою, постає виклик з удосконалення існуючих методів розпізнавання обличчя. У даній роботі були проаналізовані основні підходи для виконання цього завдання.

# 1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

## 1.1 Комп’ютерний зір

Комп’ютерний зір — це розділ штучного інтелекту (ШІ), який дозволяє комп’ютерам і системам отримувати корисну інформацію з цифрових фотографій, відео та інших візуальних даних, а також виконувати дії або давати рекомендації на основі цих даних.

Людський зір подібний до комп’ютерного, але люди мають перевагу. Людський зір отримує користь від навколишнього життя, розрізняючи об’єкти один від одного, наскільки вони віддалені, чи рухаються вони, і чи є щось невірне із образом [1].

Комп’ютерний зір навчає комп’ютери виконувати ці завдання, але використовуючи камери, дані та алгоритми, а не сітківку, зорові нерви та зорову кору, він має впоратись із цим за набагато менший час [1]. Оскільки система, навчена перевіряти предмети або контролювати виробничий актив, може оцінювати тисячі продуктів або процесів за хвилину, виявляючи несправності або проблеми, невидимі для людей, вона може швидко перевершити людські можливості.

Ринок комп’ютерного зору розширюється, його застосування охоплюють від енергетики та комунальних послуг до промислових та автомобільних. За оцінками, до 2022 року він досягне 48,6 млрд доларів США [1].

Комп’ютерний зір багато в чому покладається на дані. Він багаторазово аналізує дані, доки не зможе розрізнити відмінності та в кінцевому підсумку розпізнати зображення. Наприклад, щоб навчити комп’ютер розпізнавати автомобільні шини, йому потрібно представити величезну кількість зображень шин і пов’язаних з ними об’єктів, щоб він міг зрозуміти відмінності та розпізнати шину, особливо бездефектну [1].

Для досягнення цього результату використовуються глибоке навчання, машинне навчання та згорткові нейронні мережі (CNN) [1].

Використовуючи алгоритмічні моделі, машинне навчання дає змогу комп’ютеру тренуватися щодо контексту візуального введення. Якщо в модель надано достатню кількість даних, комп’ютер буде «дивитися» на дані та

навчиться розрізняти зображення. Алгоритми дозволяють машині самонавчатися, на відміну від програмування на розпізнавання зображення.

З 1950-х років вчені та інженери намагалися розробити методи, що дозволяють машинам бачити і розуміти візуальні дані. У 1959 році нейрофізіологи вперше спробували співвіднести реакцію в мозку кішки, демонструючи тварині різноманітні зображення. Вони помітили, що спочатку він реагував на гострі краї або лінії, що, з наукової точки зору, вказує на те, що обробка зображення починається з простих форм, таких як прямі краї [1].

Приблизно в той же час була створена перша технологія сканування зображень для комп'ютерів, що дозволило комп'ютерам оцифрувати та отримувати зображення [1]. Коли в 1963 році комп'ютери змогли перетворити двовимірні зображення в тривимірні форми, була досягнута нова віха. ШІ виник як сфера досліджень у 1960-х роках, що також поклало початок спробам ШІ вирішити проблему зору людини.

У 1974 році була введена технологія оптичного розпізнавання символів (OCR), яка могла розпізнавати текст, надрукований будь-яким шрифтом або гарнітурою [1]. Використовуючи нейронні мережі, інтелектуальне розпізнавання символів (ICR) може аналогічним чином декодувати рукописний текст.

Відтоді OCR та ICR увійшли в обробку документів та рахунків, розпізнавання номерних знаків транспортних засобів, мобільні платежі, машинний переклад та низку інших поширених програм.

Девід Марр, невролог, продемонстрував у 1982 році, що зір функціонує ієрархічно, і розробив методи для роботів розпізнавання країв, кутів, кривих та інших фундаментальних структур [1]. Одночасно вчений-комп'ютерник Куніхіко Фукусіма створив мережу клітин, що розпізнають образи. Мережа Неокогнітрон складалася із згорткових шарів у нейронній мережі.

У 2000 році ідентифікація об'єктів стала основним напрямком дослідження, а в 2001 році з'явилися перші програми для розпізнавання обличчя в режимі реального часу. У 2000-х виникла стандартизація того, як позначаються та анотуються набори візуальних даних. Набір даних ImageNet був опублікований у 2010 році [1]. Він включає мільйони фотографій із мітками, що охоплюють тисячу класів елементів, і служить основою для сучасних CNN і моделей глибокого навчання. У 2012 році команда з Університету Торонто брала участь у конкурсі ідентифікації зображень за

допомогою кадрів CNN. Модель AlexNet різко знизила частоту помилок розпізнавання зображень [1]. Після такого розвитку подій рівень помилок знизився до кількох відсотків.

## 1.2 Розпізнавання облич

Система розпізнавання облич — це технологія, яка звіряє людське обличчя з цифрового зображення або відеокадру з базою даних облич [2]. Зазвичай вона використовується для автентифікації користувачів за допомогою служб ідентифікації та працює шляхом точного визначення та вимірювання рис обличчя на певному зображенні.

Розробка подібних систем почалася в 1960-х роках як комп'ютерного додатка. Системи розпізнавання облич в останні роки все частіше використовуються у смартфонах та в інших видах технологій, таких як робототехніка [1]. Системи розпізнавання облич класифікуються як біометричні, оскільки вони передбачають вимірювання фізіологічних характеристик людини. Хоча системи розпізнавання обличчя як біометрична технологія мають нижчу точність, ніж розпізнавання райдужної оболонки ока та розпізнавання відбитків пальців, вони широко використовуються завдяки безконтактному процесу.

Розпізнавання облич можна розглядати як підмножину виявлення об’єктного класу. Метою виявлення об’єктного класу є визначення місцезнаходження та розміру всіх елементів зображення, які належать до цього класу[2]. Верхня частина тулуба, люди та автомобілі – все це приклади є прикладами об’єктних класів.

Виявлення фронтальних людських облич є центром уваги алгоритмів виявлення облич. Зображення відповідає зображенню в базі даних. Будь-які модифікації ознак обличчя бази даних зроблять процедуру зіставлення марною [2].

На основі генетичного алгоритму та технології eigen-face надійний метод виявлення обличчя: Для початку всі ділянки долини на зображенні рівня сірого перевіряються на ймовірне розташування людського ока. Потім генетичний алгоритм використовується для побудови всіх можливих зон обличчя, включаючи брови, райдужну оболонку, ніздрі та куточки рота [2].

Далі, кожен можливий кандидат на обличчя нормалізується, щоб зменшити вплив освітлення, а також ефекти викликані рухом голови. Значення придатності кожного кандидата розраховується шляхом проектування його на власні грані [2]. Усі кандидати на обличчя з високим значенням придатності вибираються для подальшої перевірки після ряду ітерацій. На цьому етапі вимірюється симетрія обличчя, і для кожного кандидата на обличчя підтверджується наявність різних рис обличчя.

Переваги і недоліки

У 2006 році продуктивність найновіших алгоритмів розпізнавання обличчя була перевірена в рамках Face Recognition Grand Challenge (FRGC). В експериментах використовувалися зображення обличчя з високою роздільною здатністю, 3-D сканування обличчя та фотографії райдужної оболонки ока. Результати показали, що нові алгоритми в десять разів точніші за алгоритми розпізнавання обличчя 2002 року і в сто разів точніші за алгоритми 1995 року. Деякі комп’ютери перевершували людей у розпізнаванні облич і навіть розрізняють однояйцевих близнюків [2].

Однією з істотних переваг системи розпізнавання обличчя є те, що вона може здійснювати масову ідентифікацію без участі досліджуваного. Правильно розроблені системи, встановлені в аеропортах, мультиплексах та інших громадських місцях, можуть ідентифікувати людей серед натовпу, навіть якщо перехожі не здогадуються про пристрій [2]. Однак у порівнянні з іншими біометричними підходами розпізнавання облич може бути менш надійним та ефективним [2]. Оскільки фотографії облич можуть сильно відрізнятися, стандарти якості мають вирішальне значення в системах розпізнавання облич. Під час зйомки обличчя такі фактори, як освітлення, емоції, положення та шум, можуть впливати на ефективність систем розпізнавання обличчя [3]. Розпізнавання обличчя має найвищий рівень помилкового прийняття та відхилення серед будь-якої біометричної системи, що викликає занепокоєння щодо її корисності для безпеки на залізницях та аеропортах.

У 2008 році Ральф Гросс, дослідник Інституту робототехніки Карнегі-Меллона, описав одну проблему, пов'язану з кутом огляду обличчя: «Розпізнавання обличчя стало дуже відмінним на повній передній панелі та на 20 градусах, але як тільки ви головою щодо профілю, виникли проблеми [3].

Окрім коливань положення, фотографії обличчя з низькою роздільною здатністю також важко розрізнити [2]. Це одна з найбільш значущих проблем

розпізнавання облич у системах спостереження.

Розпізнавання обличчя менш ефективне, коли емоції обличчя змінюються. Величезна посмішка може зробити систему менш ефективною. Наприклад, у 2009 році Канада обмежила використання нейтральних виразів обличчя в паспортних зображеннях.

Існує також невідповідність у наборах даних, які використовуються дослідниками. Дослідники можуть використовувати що завгодно: від кількох людей до сотень предметів і сотень фотографій до тисяч зображень [1]. Для дослідників дуже важливо ділитися один з одним наборами даних, які вони використовували, або створити принаймні стандартний набір даних.

Критики технології стверджують, що програма Лондонського району Ньюхем не виявила жодного злочинця станом на 2004 рік, незважаючи на те, що кілька правопорушників у базі даних системи живуть у районі, і система функціонує вже кілька років [2]. «Наскільки відомо поліції, технологія автоматичного розпізнавання обличчя Ньюхема ніколи не виявляла живу мету». Ці докази, здається, суперечать твердженням про те, що система була відповідальна за 34% зниження рівня злочинності (тому вона була поширена і в Бірмінгемі).

Експеримент, проведений у 2002 році Департаментом поліції Тампи, мав такі ж незадовільні результати [3].

Система в Бостонському аеропорту Логан була закрита в 2003 році після того, як після дворічного тестового періоду не вдалося знайти жодних збігів.

У 2014 році Facebook повідомила, що в стандартизованому тесті на розпізнавання обличчя з двома варіантами їхня онлайн-система отримала 97,25 відсотка точності в порівнянні з критерієм людини в 97,5 відсотка [2].

Правозахисні групи і захисники конфіденційності, такі як Electronic Frontier Foundation, Big Brother Watch і ACLU, стурбовані тим, що використання технологій спостереження ставить під загрозу конфіденційність.

Розпізнавання обличчя може використовуватися для виявлення додаткових особистих даних, пов’язаних з особою, таких як облікові записи в соціальних мережах, діяльність в Інтернеті та звички людини під час подорожей. Крім того, якщо люди не маскують свої обличчя, вони мають мінімальні можливості для запобігання або порушення моніторингу розпізнавання облич [3]. Це глибоко змінює динаміку повсякденної конфіденційності, дозволяючи будь-якому рекламодавцю, державній установі

чи випадковому незнайомцю таємно збирати конфіденційну інформацію будь-яких людей, виявлених за допомогою технології розпізнавання обличчя. Споживачі можуть не розуміти або не знати про те, як використовуються їхні дані, що не дасть їм погодитися на поширення їхньої особистої інформації.

У липні 2015 року Бюро підзвітності уряду Сполучених Штатів опублікувало звіт підкомітету з питань конфіденційності, технологій та закону Сенату Сполучених Штатів Америки і відповідного федерального законодавства [3]. У ньому зазначається, що раніше досліджувалися проблеми з технологією розпізнавання обличчя, і вони вказують на необхідність для Сполучених Штатів оновити свої правила конфіденційності, щоб федеральний закон завжди відповідав ефекту нових технологій. Згідно з документом, різні корпоративні, урядові та комерційні групи створюють або прийняли «добровільні норми конфіденційності». Вимоги зацікавлених сторін відрізнялися, але основною метою було отримати дозвіл та проінформувати людей про плановане використання технології розпізнавання облич. У документі стверджується, що добровільні правила конфіденційності допомогли полегшити проблеми з конфіденційністю, які виникають, коли люди не поінформовані про те, як використовуються їхні персональні дані.

Коли російський стартап NtechLab представив технологію розпізнавання обличчя FindFace у 2016 році, це викликало проблему конфіденційності в міжнародних ЗМІ, пообіцявши, що російські користувачі можуть робити зображення незнайомих людей на вулиці та підключати їх до профілю в соцмережі на сайті соціальних мереж Вконтактє (VK) [2].

У грудні 2017 року Facebook запустив нову функцію, яка сповіщає користувачів, коли хтось завантажує зображення, на якому Facebook знаходить чиєсь обличчя, навіть якщо ця людина не була позначена. Незважаючи на попередні резонанси, Facebook намагався інтерпретувати нові можливості в позитивному ключі. Роб Шерман, директор із конфіденційності Facebook, описав новий інструмент як надання користувачам більшого контролю над їхніми онлайн-зображеннями. «Ми розглядали це як надзвичайно потужну функцію», — пояснює він. «Можливо, там є зображення, про які ви не знаєте». Декілька групових позовів були подані проти DeepFace від Facebook відповідно до Закону про конфіденційність біометричної інформації, у яких стверджується, що Facebook збирає та зберігає дані про розпізнавання облич своїх користувачів без їхньої інформованої згоди, що прямо порушує Закон про

конфіденційність біометричної інформації 2008 року (BIPA). Остання справа була відхилена в січні 2016 року через відсутність юрисдикції суду [2]. У Сполучених Штатах такі компанії, як Clearview AI, використовують Першу поправку до Конституції Сполучених Штатів для очищення даних облікових записів користувачів на сайтах соціальних мереж, щоб створити системи розпізнавання облич [3].

У 2019 році газета Financial Times повідомила, що програмне забезпечення для розпізнавання облич використовується в лондонському районі Кінгс-Кросс [2].

У районі лондонського залізничного вокзалу Кінгс-Крос є магазини, офіси, штаб-квартира Google у Великобританії та частина коледжу Святого Мартіна [3]. "Сканування облич людей, коли вони законно займаються своїм повсякденним життям, щоб ідентифікувати їх, є можливою небезпекою для конфіденційності, яка має насторожити нас усіх", - йдеться в повідомленні Управління інформаційного комісара Великобританії. Елізабет Денхем, Комісар з питань інформації Великобританії, почала розслідування щодо використання системи розпізнавання облич Кінгс-Кросс, якою керує компанія Argent [2]. У вересні 2019 року Арджент заявив, що технології розпізнавання облич більше не будуть застосовуватися на Кінгс-Кросс. Арджент заявив, що зловмисне програмне забезпечення було встановлено на двох камерах, які знімали пішохідну дорогу, що проходить через комплекс у період з травня 2016 року по березень 2018 року [2]. У жовтні 2019 року заступник мера Лондона Софі Лінден заявила у звіті, що столична поліція надала зображення семи осіб до Argent для включення до їхньої системи розпізнавання облич Кінгс-Кросс у таємному контракті.

Поліція Південного Уельсу неодноразово тестувала автоматичне розпізнавання облич у період з 2017 по 2019 роки [2].

## 1.3 Штучний інтелект

ШІ – це здатність цифрового комп’ютера або робота, керованого комп’ютером, виконувати завдання, зазвичай пов’язані з розумними істотами. Цей термін часто застосовують до проекту розробки систем з когнітивними процесами, схожими на людину, такими як здатність міркувати, відкривати сенс, узагальнювати або вчитися на минулому досвіді [1]. З моменту розробки

цифрового комп’ютера в 1940-х роках було продемонстровано, що комп’ютери можна запрограмувати для виконання надзвичайно складних завдань з великим вправністю, наприклад, відкриття доказів математичних теорем або гри в шахи.

Незважаючи на постійний прогрес у швидкості обробки комп'ютера та об'єму пам'яті, жодна програма ще не може зрівнятися з людською гнучкістю в ширших сферах або в задачах, які вимагають широких повсякденних знань. З іншого боку, деякі програми перевершили рівень продуктивності людських експертів і професіоналів у виконанні конкретних завдань, в результаті чого штучний інтелект у цьому обмеженому сенсі можна знайти в таких різноманітних програмах, як медична діагностика, комп’ютерні пошукові системи та розпізнавання голосу чи рукопису [1].

Вся поведінка людини, крім найпростішої, приписується інтелекту, тоді як навіть найскладніша поведінка комах ніколи не приписується інтелекту.

До прикладу, розглянемо поведінку оси-копача Sphex ichneumoneus. Коли самка оси повертається до своєї нори з їжею, вона спочатку відкладає її на поріг, потім перевіряє, чи немає в її норі зловмисників, і, нарешті, якщо все добре, вона несе їжу всередину [1]. Якщо їжу відсунути на кілька дюймів від входу в її нору, поки вона всередині, вона повторить всю процедуру стільки разів, скільки їжа буде зміщена. Інтелект, якого явно не вистачає Сфексу, повинен включати здатність пристосовуватися до нових обставин.

Всі методи і технології описані у даній роботі, які використовуються для розпізнавання маскованих облич, базуються на використанні штучного інтелекту [1].

# 2 РОЗПІЗНАВАННЯ МАСКОВАНИХ ОБЛИЧ

Системи розпізнавання облич (FR) зазвичай покладаються на основні риси обличчя, такі як очі, ніс і рот, тобто незакриті обличчя. Однак у різних ситуаціях та обставинах людям необхідно бути у масках з частково або повністю або закритим обличчям [4]. Типовими прикладами є пандемії, лабораторії, медичні процедури та надмірне забруднення. За даними Всесвітньої організації охорони здоров'я (ВООЗ) та Центрів з контролю та профілактики захворювань, найкращий спосіб захистити людей від вірусу COVID-19 та уникнути поширення або зараження захворювання – це практика дотримання соціальної дистанції та носити захисні маски [4]. Як результат, кожна країна у світі зобов'язує людей носити захисні маски для обличчя в громадських місцях, що вимагає дослідження ефективності систем розпізнавання облич при роботі з обличчями в масках.

Однак впровадження таких заходів з епідеміологічної безпеки створює серйозний виклик для існуючих систем безпеки та аутентифікації, які покладаються на FR [4]. Більшість останніх алгоритмів було розроблено, щоб визначити, чи є обличчя прихованим, тобто для виявлення замаскованих облич. Хоча захист публічного здоров'я є обов'язковим, існує нагальна потреба в автентифікації осіб, які носять маски, не вимагаючи їх знімати.

У багатьох місцях, таких як контроль доступу до приміщень та прикордонні контрольно-пропускні пункти, необхідно показувати обличчя перед камерою, що створює проблему для розпізнавання обличчя, оскільки для виявлення та розпізнавання цього обличчя потрібні закриті його ділянки [4].

Крім того, численні організації вже розробили та розгорнули необхідні набори даних для розпізнавання обличчя як засобу аутентифікації або ідентифікації [4]. Аутентифікація на обличчі, також відома як відповідність один до одного, — це процедура встановлення особи, яка підтверджує, що вона є тим, за кого себе видає. Під час безпекової аутентифікації фіксується зображення обличчя, з якого генерується біометричний шаблон і порівнюється з наявним особистим підписом. На відміну від цього, ідентифікація обличчя, також відома як відповідність один до багатьох, є типом біометричного

розпізнавання, який ідентифікує людину шляхом порівняння та аналізу індивідуальних малюнків із великою базою даних відомих облич [4].

На жаль, закриті обличчя ускладнюють точне розпізнавання суб'єктів, ставить під загрозу дійсність існуючих наборів даних і робить такі внутрішні системи FR непрацездатними [4].

## 2.1 Принцип розробки MFR

Як показано на малюнку 2, цей розділ описує, як системи MFR зазвичай розробляються через серію складних етапів. Загальна методологія в першу чергу базується на широко прийнятих моделях глибокого навчання для вивчення відмінних характеристик облич у масках. Як свідчить цей конвеєр, для розробки остаточної системи розпізнавання зазвичай робиться ряд важливих кроків, що описані в наступних підрозділах [4].

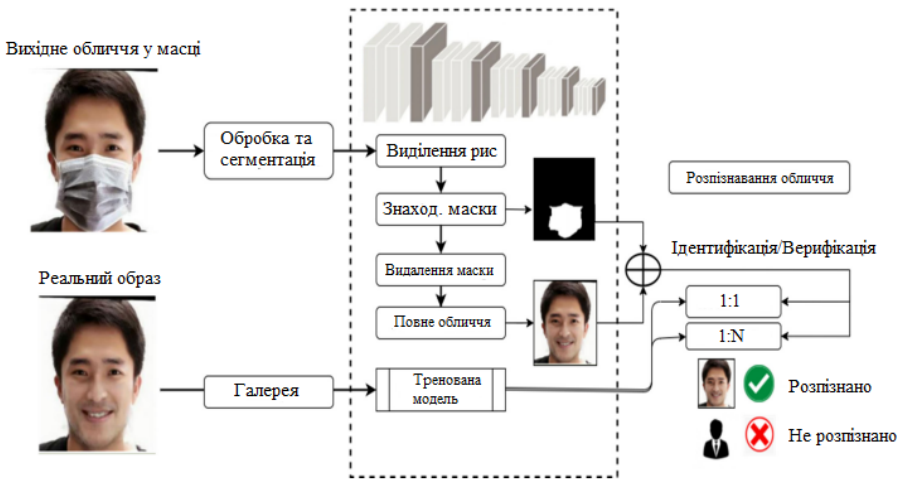


Рисунок 2.1 Схематичне зображення процесу розпізнавання облич у масці

Спочатку збирається колекція зображень оригінальних замаскованих і реальних облич. Як правило, це включає їх організацію в каталоги по категоріям для навчання моделей, перевірки та тестування. Потім виконуються деякі операції попередньої обробки, такі як збільшення даних і сегментація зображення [4]. Потім набір важливих рис обличчя витягується за допомогою однієї або кількох моделей глибокого навчання, які зазвичай попередньо тренуються на зображеннях загального призначення та точно налаштовуються

на новій колекції, тобто обличчя в масках. Такі характеристики повинні бути достатньо дискримінаційними, щоб точно визначити маски для обличчя. Потім застосовується процедура демаскування обличчя, щоб відновити замасковане обличчя та повернути оцінку вихідного обличчя. Нарешті, передбачуване обличчя порівнюється з оригінальними обличчями, що відповідають дійсності, щоб визначити, чи була ідентифікована особа [4].

Попередня обробка зображень

На продуктивність систем розпізнавання обличчя (FR) з масками чи без них значно впливає характер зображень обличчя, які використовуються під час навчання, перевірки та тестування. Недостатньо загальнодоступних пар зображень обличчя з об’єктами маски та без них, щоб адекватно тренувати систему MFR для прогресування. Отже, це посилює необхідність збагачувати тестовий стенд додатковими синтетичними зображеннями, що містять різні типи масок для обличчя, і покращувати можливості узагальнення моделей глибокого навчання [5]. MaskTheFace, MaskedFace-Net, глибока згорткова нейронна мережа (DCNN), CYCLE-GAN, Identity Aware Mask GAN (IAMGAN) та starGAN є одними з найбільш популярних методів синтезу масок на обличчі.

Зображення також ретельно обробляється за допомогою розширення даних, що дозволяє застосовувати різноманітні операції для збільшення кількості та різноманітності зображень, включаючи обрізання, перегортання, обертання та вирівнювання [4]. Інші процеси покращення, такі як зміна масштабу зображення, сегментація, видалення шуму та згладжування, також використовуються для покращення якості представлення зображення. Крім того, можна виконати коригування зображення для покращення його різкості, при цьому дисперсія Лапласа є однією з найбільш широко використовуваних методик.

Кілька методів миттєво або семантично сегментують зображення на локальні частини, а потім представляють їх упорядкованою властивістю частин обличчя або набором дискримінаційних компонентів, щоб створити більш точні образи зображення [4]. Тим не менш, деякі методи передають зображення до існуючого інструменту для виявлення орієнтирів на обличчі, тоді як інші представляють вхідне нерухоме зображення за допомогою загального дескриптора, такого як регуляризація низького рангу і розріджене представлення [4].

## 2.2 Моделі глибокого навчання

Численні добре відомі методи, такі як LBP, SIFT та Gabor, були запропоновані та намагаються розпізнати людські обличчя за локальними або глобальними рисами, створеними вручну [5].

Однак здатність цих цілісних підходів підтримувати неконтрольовані зміни обличчя, які відхиляються від початкових припущень обмежена. Пізніше були введені неглибокі представлення зображень, такі як дескриптори словника на основі навчання, щоб вирішити проблеми відмінності та компактності попередніх методів. Незважаючи на те, що було досягнуто підвищення точності, ці неглибокі зображення продовжують демонструвати низьку стійкість до реальних додатків і нестабільність проти змін у зовнішньому вигляді обличчя [5].

Після 2010 року було швидко розроблено і впроваджено кілька глибокорівневих методів глибокого навчання для вилучення ознак і перетворення зображень. З часом вони продемонстрували перевагу у вивченні кількох рівнів образу обличчя, які відповідають різним рівням абстракції, демонструючи надійну стійкість до змін зображення обличчя, таких як освітлення, вираз, поза та одягнена маска. Поєднуючи низькорівневі та високорівневі абстракції, моделі глибокого навчання здатні представляти та розпізнавати ідентичність обличчя з високою відмінністю [5].

## 2.3 Згорткові нейронні мережі

Згортка нейронна мережа (CNN) є однією з найефективніших нейронних мереж, що демонструє свою перевагу в різноманітних застосунках, таких як класифікація зображень, розпізнавання, пошук і виявлення об'єктів. CNN зазвичай складаються з вхідного шару, згорткового, шару піддискретизації, повністю підключеного та вихідного шарів для контролю ступеня зсуву, масштабу та спотворення. Ці рівні включають вхідний, згортковий, підвибірковий та повністю зв’язаний [5]. Вони можуть ефективно вивчати внутрішньокласні відмінності, такі як освітлення, поза, вираз обличчя та вік, на

основі даних про навчання. Численні великомасштабні набори даних обличчя були використані для навчання моделей на основі CNN.

AlexNet є однією з найпопулярніших попередньо підготовлених архітектур, яка була успішно застосована до завдань FR. Завдяки наявності інтегрованих графічних процесорів (GPU), AlexNet зміг скоротити час навчання та частоту помилок для великих наборів даних. VGG16 і VGG19 також широко використовуються архітектури на основі CNN, які були реалізовані в різних програмах комп'ютерного зору, таких як розпізнавання обличчя [5]. Як правило, моделі на основі VGG пропонують функції або уявлення на основі згортки . Незважаючи на їх чудову досягнуту точність, час їх навчання та складність є проблематичними.

Складність завдання розпізнавання зображень з часом зросла, що вимагало використання більш глибоких нейронних мереж. Однак додавання більше шарів у мережу ускладнює її навчання; як наслідок, зазвичай спостерігається зниження точності [4, 5]. Залишкова мережа (ResNet) була розроблена для подолання цієї труднощі шляхом накладання додаткових шарів для досягнення вищої продуктивності та точності. Inception та його варіанти також є популярними архітектурами на основі CNN; їхня новизна полягає у використанні модулів або блоків для побудови мереж зі згортковими шарами на відміну від стекування згорткових шарів. Xception є екстремальною версією Inception, в якій роздільні згортки по глибині замінюють модулі Inception. У таблиці 1 наведено підсумок основних характеристик поширених моделей на основі CNN, що використовуються в області MFR [5].

| Модель | Версії | Параметри тренування | Згорткові шари | Інші шари | Усього шарів |
|---|---|---|---|---|---|
| AlexNet | - | 62 M | 5 | 3 | 8 |
| VGG | VGG16 | 138 M | 13 | 3 | 16 |
| | VGG19 | 143 M | 16 | 3 | 19 |
| ResNet | ResNet50 | 25 M | 48 | 2 | 50 |
| | ResNet101 | 44 M | 99 | 2 | 101 |
| MobileNet | MobileNet | 13 M | 28 | 2 | 30 |
| | MobileNet-v2 | 3.5 M | - | - | 53 |
| Inception | GoogleNet | 7 M | 22 | 5 | 27 |
| | InceptionV2 | 56 M | 22 | 26 | 48 |
| | InceptionV3 | 24 M | 22 | 26 | 48 |
| | InceptionV4 | 43 M | - | - | 164 |
| | Inception-ResNet-V2 | 56 M | - | - | 164 |
| Xception | - | 23 M | 36 | 35 | 71 |

Рисунок 2.2 – Характеристики попередньо навчених моделей на базі CNN

### 2.3.1 Автокодери

Autoencoder — це добре відома глибока нейронна мережа, яка забезпечує неконтрольовану парадигму на основі навчання функцій для ефективного кодування та декодування даних. Завдяки його здатності автоматично вивчати надійні функції з величезної кількості нерозмічених даних, значні зусилля дослідження були присвячені кодуванню вхідних даних у низьковимірний простір із значущими та дискримінаційними представленнями, яке виконується декодером. Потім декодер повертає процес, використовуючи зворотне поширення під час навчання, щоб генерувати ключові характеристики на етапі кодування. Автокодери, такі як LSTM-автокодери, двоканальні SSDA (DC-SSDA), декорумповані автокодери та варіаційні автокодери на основі 3D орієнтирів, ефективно використовувалися для задачі OFR [5].

### 2.3.2 Генеративні змагальні мережі

Генеративні змагальні мережі (GAN) використовуються для автоматичного виявлення та вивчення регулярних шаблонів із вхідних даних за відсутності обширних навчальних даних [5].

GAN складається з двох нейронних мереж: генератора і дискримінатора. Генератор створює нові функції, використовуючи випадкові значення із заданого розподілу як шумові дані.

Дискримінатор — це двійковий класифікатор, який визначає автентичність згенерованих ознак. GAN називаються змагальними через їхнє змагальне навчання, в якому генератор і дискримінатор змагаються за оптимізацію протилежних функцій програшу в мінімаксній грі (тобто грі з нульовою сумою) [4]. Загальні проблеми FR, такі як синтез обличчя, перехресне розпізнавання обличчя, інваріантне розпізнавання обличчя та розпізнавання облич без макіяжу, були вирішені за допомогою GAN, факт чого важливо відзначити.

### 2.3.3 Глибока мережа довіри

Глибока мережа довіри (DBN) — це набір кількох прихованих одиниць різних шарів, які з'єднані внутрішньо без з'єднання одиниць в одному шарі [5]. Зазвичай вона складається з серії обмежених машин Больцмана (RBM) або автокодерів, у яких кожен прихований підрівень діє як видимий шар для наступного прихованого підшару, а останній шар є шаром softmax, який використовується в процесі класифікації. DBN також використовувалися у FR та OFR.

### 2.3.4 Вичерпне навчання з підкріпленням

Навчання з підкріпленням відбувається у місцевому середовищі; отже, він імітує процес прийняття рішень людиною, дозволяючи агенту вибирати дію на основі свого досвіду [5]. Агент — це сутність, здатна сприймати навколишнє середовище за допомогою датчиків і діяти на нього за допомогою виконавчих механізмів. У методах глибокого FR, таких як увага і marginaware, успішно застосовується комбінація глибокого навчання та навчання з підкріпленням.

### 2.3.5 Конкретні глибокі мережі MFR

Численні архітектури глибокого навчання були розроблені або налаштовані спеціально для задач FR або OFR, і вони внесли помітний внесок у підвищення продуктивності [5].

FaceNet відображає зображення в евклідовому просторі за допомогою глибоких нейронних мереж, які потім вбудовують обличчя на основі втрати триплету. В евклідовому просторі відстань між зображеннями, що належать одній людині, буде невеликою, тоді як відстань між зображеннями, що належать різним людям, буде великою. Ця функція дозволяє FaceNet

виконувати різні завдання, включаючи виявлення облич, розпізнавання та групування [5]. SphereFace— це додаткова добре відома система FR, яка надає геометричну інтерпретацію та дає змогу CNN вивчати ознаки кутової дискримінації, що робить її ефективною для навчання представлення обличчя. ArcFace також є ефективною мережею FR, заснованою на навчанні подібностям, яка замінює втрату кутового запасу на втрату softmax. Використовуючи косинусну подібність, вона обчислює найкоротшу відстань між двома зображеннями [5].

На основі косинусної втрати великої межі, також було запропонувано MFCosface як алгоритм MFR. Він ефективно долає проблему низького рівня розпізнавання, спричиненого оклюзіями маски, шляхом визначення основних рис обличчя замаскованих облич. Крім того, MFCosface покладається на великі втрати косинуса [5]. Він оптимізує відображення рис обличчя, додаючи до моделі механізм уваги. VGGFace— це система розпізнавання обличчя, яка використовує для розпізнавання нейронну мережу глибокої згортки на основі VGG-Very-Deep16 CNN. Також включено детектор обличчя та локалізатор на основі каскадної моделі деталей, що деформуються. DeepID був розроблений для виявлення дискримінаційного глибокого представлення обличчя шляхом класифікації великомасштабних зображень обличчя на велику кількість ідентичностей, тобто ідентифікації обличчя. Проте досліджені образи облич ставляться під сумнів через значні внутрішньоособистісні варіації, які були зменшені багатьма варіантами DeepID, включаючи ідентифікацію спільного обличчя, представлену в DeepID + 2 [5].

## 2.4 Виділення ознак

Вилучення ознак є важливим кроком у процесі розпізнавання обличчя, метою якого є вилучення набору дискримінаційних ознак, які можна використовувати для представлення та вивчення ключових атрибутів обличчя, таких як очі, рот, ніс і текстура. Оклюзії та маски ускладнюють процес, і існуючі системи розпізнавання облич необхідно модифікувати, щоб отримати репрезентативні, але надійні дані про риси обличчя. Підходи до виділення ознак для розпізнавання замаскованих облич можна розділити на неглибокий і глибокий методи представлення.

Неглибоке виділення ознак — це класичний метод, який чітко формулює набір функцій, створених вручну з неглибоким вивченням або оптимізацією. Деякі методи покладаються на створені вручну низькорівневі функції, щоб знайти приховані частини обличчя та виключити їх з розпізнавання дескрипторів, які представляють цілісне навчання, локальні особливості та підходи до неглибокого навчання, включають LBP, SIFT, HOG та кодові книги [4].

Вони досягли помітної точності та стійкості проти багатьох змін обличчя, таких як освітлення, спорідненість, поворот, масштаб і трансляція в завданнях розпізнавання обличчя без закритого типу [4]. Під час роботи із закритими обличчями, такими як маски для обличчя, продуктивність неглибоких рис погіршилася, і глибокі уявлення, отримані моделями глибокого навчання, значно перевершили їх.

Глибоке навчання було використано для розробки та тестування різноманітних методів виділення рис з облич [5]. Лі та ін. припустили, що особливості замаскованих облич часто включають інформацію, пов’язану з областю маски, яку слід моделювати окремо, тому вони вивчили два центри для кожного класу замість одного, один для зображень повного обличчя та один для зображень замаскованих облич. Пісня та ін. запропонував багатоетапну стратегію навчання масок, засновану насамперед на CNN, з метою виявлення та видалення пошкоджених функцій з розпізнавання. Використовуючи додаткову підмережу для отримання важливих ділянок обличчя, багато інших методів, що враховують увагу та контекст, витягують особливості зображення.

Розпізнавання, реконструкція та розпізнавання замаскованих облич також використовували уявлення графічних зображень із згортковими мережами глибоких графів (GCN) [5]. Використовуючи просторові або спектральні фільтри, створені для спільної або фіксованої структури графіка, GCN продемонстрували високі можливості в навчанні та обробці зображень облич. Однак через велику кількість шарів GCN та високу обчислювальну складність навчання графічних представлень часто обмежене. Закрите або замасковане тривимірне розпізнавання облич також було досліджено за допомогою 3D просторових функцій. Методи 3D розпізнавання облич розроблені для імітації реального бачення та розуміння рис обличчя людини, і, таким чином, можуть допомогти покращити роботу існуючих систем двовимірного розпізнавання.

Багато змін обличчя, такі як зміни освітлення, вирази обличчя та напрямок обличчя, не впливають на тривимірні риси обличчя.

## 2.5 Виявлення маски

Маски для обличчя, які бувають різних стилів, розмірів, текстур і кольорів, останнім часом стали одними з найпоширеніших речей, які частково закривають обличчя. Це підкреслює важливість правильного тренування моделей глибокого навчання для виявлення масок. У задачі виявлення маски налаштовується та досліджується більшість існуючих методів виявлення, які зазвичай вводяться для виявлення об’єктів. У сфері виявлення об’єктів, де для класифікації пропозицій об’єктів використовується глибока ConvNet, регіони з функціями CNN (R-CNN) отримали широке поширення . R-CNN витягує тисячі областей обличчя із закритих облич, передаючи їх у мережу CNN і використовуючи алгоритм вибіркового пошуку для створення вектора ознак для кожної області. Після цього машина опорного вектора класифікує наявність об’єкта в цій пропонованій області обличчя-кандидата на основі вилученої ознаки (SVM). Щоб підвищити продуктивність шляхом трансформації архітектури R-CNN, були представлені Fast R-CNN і Faster R-CNN. Однак ці методи мають суттєві недоліки, наприклад, процес навчання є багатоетапним конвеєром, який інтенсивний як у просторі, так і в часі [4]. Крім того, R-CNN виконує прямий перехід ConvNet для кожної пропозиції об’єкта повільно, без спільного доступу.

Дослідниками запропоновано контекстну увагу R-CNN як структуру виявлення маски обличчя. Витягуючи відмінні риси, ця структура використовується для збільшення внутрішньокласної відстані та зменшення міжкласової відстані.

У результаті більше досліджень було зосереджено на використанні глибоких мереж на основі сегментації для виявлення масок. Повністю згорткова нейронна мережа (FCN) — це архітектура семантичної сегментації, яка в основному використовується з автокодером на основі CNN, який не містить жодних щільних шарів. Це більш просунута версія популярного модуля класифікації, який замінює повністю зв’язані шари на 1:1 згортку [4].

U-Net спочатку був розроблений для сегментації біомедичних зображень, але зараз широко використовується в різноманітних програмах комп'ютерного зору, включаючи виявлення обличчя.

Він має кодер, який використовує шари згорткового та максимального об'єднання для захоплення контексту зображення, і декодер, який використовує транспоновані згортки для підвищення вибірки закодованих даних. Карти характеристик кодера потім об'єднуються з картами об'єктів декодера. Це допомагає вивчати контекстну інформацію (відносини між пікселями зображення) [5].

Крім того, в літературі були запропоновані інші ефективні методи MFR або OFR.

Мережа уваги до обличчя (FAN) була введена як одноразовий детектор облич, який використовує мережу пірамід функцій для вирішення проблем оклюзії та помилкових спрацьовувань для облич різного масштабу. Інші запропонували використовувати попередньо навчені CNN для виділення кандидатських областей обличчя та представлення їх за допомогою високорозмірних дескрипторів для виявлення замаскованих облич [5]. Дескриптори обличчя потім перетворюються у вектори ваги за допомогою модуля локального лінійного вбудовування, який відновлює всі відсутні ознаки обличчя в замаскованих областях. Нарешті, зважені вектори використовуються як вхідні дані в завданнях класифікації та регресії для визначення реальних ділянок обличчя. Лін та ін. запропонував модифікований LeNet (MLeNet), який зменшує переобладнання та покращує продуктивність виявлення замаскованих облич за допомогою невеликої кількості навчальних зображень за рахунок збільшення кількості одиниць у вихідному шарі та карт ознак з меншим розміром фільтра. Використовуючи кілька фільтрів, представили багатографічні функції GCN для виявлення масок для обличчя. Вони витягли та вивчили ключові риси обличчя, використовуючи вбудовану геометричну інформацію, розраховану за допомогою графіків відстані та кореляції [4].

Негі та ін. запропонували моделі глибокого навчання на основі CNN і VGG16 і комбіновані запобіжні заходи на основі штучного інтелекту для виявлення масок для обличчя в наборі даних Simulated Masked Face Dataset (SMFD).

Peng та ін. представив використання глибоких мереж на основі злиття локальних функцій у нелінійному просторі для виявлення замаскованого

обличчя. Багато інших досліджень на основі виявлення спиралися на традиційні локальні та глобальні риси обличчя, засновані на ключових рисах обличчя, таких як ніс і рот.

Чен та ін. нещодавно запропонували концепцію помічника маски для обличчя (FMA) як методу виявлення обличчя на основі мобільного мікроскопа. Вони використовували глобально та локально узгоджене завершення зображення (GLCM), щоб виділити риси текстури та вибрати контраст, кореляцію, енергію та однорідність як риси обличчя. Щоб задовольнити нижчі обчислювальні вимоги для вбудованих систем, Fan et al. запропонували одноразовий легкий детектор маски для обличчя на основі глибокого навчання. Вони розробили SL-FMDet (одноразовий легкий детектор маски для обличчя), який добре працював завдяки низьким вимогам до обладнання. Легкий хребет обмежував вилучення функцій, що було значною перешкодою [5].

## 2.6 Демаскування обличчя

У літературі використовувалися різні підходи до видалення об'єктів, включаючи маску. Далі представлено кілька поширених методів для алгоритмів видалення об’єктів, заснованих на навчанні та не на основі навчання.

Шетті та ін. запропонували модель на основі GAN, яка отримує вхідне зображення та автоматично видаляє цільовий об’єкт для підходів, заснованих на навчанні [5]. Лі та ін. та Iizuka et al. запропонували дві моделі для вивчення глобальної когерентності та завершення пошкодженої області шляхом видалення цільового об'єкта та реконструкції пошкодженої частини за допомогою налаштування GAN. Щоб видалити об’єкти із зображень обличчя, Khan et al. використовував підхід на основі GAN від грубого до точного [4].

Бутрос та ін. запропонував модель вбудовування демаскування (EUM) для видалення маски, яка використовує вбудовування ознак, витягнуте з замаскованого обличчя, як вхідні дані. Він створює нову вбудовану функцію з унікальними властивостями, подібними до немаскованого обличчя тієї ж самої особистості. Дін та ін. використовував GAN з двома дискримінаторами для автоматичного видалення маски.

Criminisi та ін. запропонували модель, яка видаляє небажану частину зображення, створює нову область, яка відповідає відсутній області, а потім синтетично відповідає тому, що залишилося від зображення. Щоб обчислити

порядок заповнення, Ван запропонував регуляризований коефіцієнт, який регулює криву функції пріоритету виправлення [5].

Щоб видалити окуляри з зображень обличчя, Park et al. використовували реконструкцію аналізу головних компонент (PCA) та рекурсивну компенсацію помилок. Хейз та ін. запропонував алгоритм завершення зображення, який шукає подібну інформацію у великій базі даних зображень і вбудовує її в пошкоджений піксель вхідного зразка [4].

## 2.7 Відновлення обличчя

Після демаскування обличчя будь-які відсутні частини слід оцінити та відновити, щоб можна було використати процес узгодження ідентичності для розпізнавання особи.

Класифікація на основі розрідженого представлення (SRC) для надійного OFR є однією з перших робіт у реконструкції зображень. Різні варіанти SRC були введені для конкретних проблем FR, таких як розширений SRC (ESRC) для FR з недостатньою вибіркою і групове розріджене кодування (GSC) для покращення дискримінаційної здатності реконструкції обличчя [4]. Запропоновано багато інших методів реконструкції відсутніх частин оклюзованих граней. Юань та ін. об’єднали розріджене представлення та фільтрацію частинок та дослідили результати. Cen та ін. також запропонував схему класифікації для надійного OFR на основі представлення глибинного словника. Методи реконструкції зображень, з іншого боку, виявили ряд добре відомих недоліків, у тому числі потребу в надповному словнику, велике збільшення зображень у галереї, що призвело до проблеми складності, та обмеження можливості узагальнення.

Щоб відновити відсутню частину зображення обличчя, були використані методи глибокого навчання. Методи на основі GAN із глобальними та локальними дискримінаторами використовувалися для вирішення завдання реконструкції обличчя в останні роки [5].

Щоб обчислити відсутні пікселі та регіони, Yeh et al. використовували дані на основі семантичного зображення. Однак вони не в змозі зберегти ідентичність обличчя. В результаті Zhao et al. розробили модель, засновану на втраті ідентичності та дискримінаторі пози під час навчання мережі, щоб отримати відсутні частини пікселя під різними позами голови, намагаючись

зберегти ідентичність. Дуан та ін. запропонував генератор фронтального огляду з кількома оклюзіями, мережу підсилення з багатьма входами та дискримінатор із кількома входами як наскрізну мережу BoostGAN. До цього методу входить мережевий ансамбль деокклюзії та фронталізації обличчя від грубої до тонкої. Для малювання зображень Yu et al. запропонував підхід на основі GAN від грубого до точного з новим контекстним модулем уваги [5].

Завдяки підходу до перекладу зображення в зображення, Дін та ін. використовували зображення на основі GAN для завершення зображення. GAN були використані Duan et al., щоб одночасно виконувати завдання фронталізації та завершення обличчя. Вони запропонували модель уваги, засновану на оклюзованих масках, і представили двоступеневу генеративну змагальну мережу (TSGAN). Крім того, Luo et al. використовували GAN, щоб представити структуру EyesGAN, яка в основному використовується для побудови обличчя навколо очей [4].

Ма та ін. представили метод навчання та збереження обличчя завершеної мережі (LP-FCN) для аналізу зображень обличчя та вилучення функцій збереження ідентичності обличчя (FIP) одночасно. Цей метод заснований насамперед на CNN, який був навчений трансформувати функції FIP. Ці характеристики об’єднуються і подаються в декодер, який генерує ціле зображення. На малюнку 3 зображено два нещодавно запропоновані методи демаскування облич і відновлення відсутніх рис обличчя [5].

## 2.8 Розпізнавання та відповідність облич

Відповідність облич за глибокими ознаками для FR і MFR можна розглядати як проблему перевірки або ідентифікації обличчя. Під час фази навчання та перевірки набір зображень ідентифікованих суб’єктів надходить у систему для виконання цього завдання. На етапі тестування системі представляється невідомий предмет і просять прийняти рішення про розпізнавання. Відповідна функція втрат повинна бути реалізована та використана для ефективного вивчення набору глибоких функцій або дескрипторів. Спільнота MFR використовує один із двох поширених підходів зіставлення: 1-до-1 і 1-до-N [5].

В обох підходах використовуються звичайні міри відстані, такі як L2 і косинус на основі Евкліда. Перевірка обличчя використовує процедуру зіставлення подібності 1 до 1, яка застосовується між колекцією реальних зображень і тестовим зображенням, щоб визначити, чи два зображення відносяться до однієї і тієї ж людини, тоді як ідентифікація обличчя використовує процедуру 1 до -N відповідності схожості, щоб дослідити обличчя конкретної особи в масці [4, 6].

Багато методів, таких як метричне навчання та розріджені уявлення, були введені для покращення рівня розрізнення глибоких ознак з метою зробити процес зіставлення обличчя більш точним та ефективним. Моделі softmax на основі втрат і триплетних моделей широко використовуються в моделях глибокого навчання для збігу ідентичностей обличчя. Моделі, засновані на втратах Softmax, використовують функцію softmax для навчання мультикласового класифікатора для одного класу для кожної особистості в наборі навчальних даних. З іншого боку, моделі, засновані на триплетних втратах, характеризуються негайним вивченням вбудовування шляхом зіставлення результатів різних вхідних даних, щоб мінімізувати відстань між класами і таким чином максимізувати відстань між класами. З іншого боку, моделі на основі втрат Softmax і триплетних втрат страждають від оклюзії маски [6].

Останнім часом у літературі було опубліковано декілька наукових праць щодо вирішення завдань МФР. Наприклад, високу продуктивність FR було продемонстровано за допомогою методів на основі GAN для демаскування облич перед подачею їх у модель розпізнавання обличчя, вилучення рис лише з верхньої частини обличчя та навчання розпізнавання обличчя мережі з поєднанням замаскованих і немаскованих облич. Анвар та ін. поєднав набір даних VGG2 з доповненими замаскованими обличчями та навчив модель за допомогою оригінального конвеєра FaceNet, що дозволило моделі визначати, чи носить обличчя маску чи ні, на основі верхньої половини рис обличчя. Для MFR Montero et al. представив повний навчальний конвеєр для моделей розпізнавання обличчя на основі ArcFace. Ген і його колеги змогли визначити два центри для кожної особи, які відповідають зображенням анфас і обличчу в масці [6].

Афзал та ін. запропонував метод застосування вилучення ознак, обчислення глибини та формулювання тривимірного зображення, який є

обчислювально ефективним. SIFT використовувався для щільного зображення рис обличчя. Потім для обчислення глибини зображення використовується багатовимірний розподіл Гаусса [6]. Нарешті, вони визначили форму, використовуючи техніку затінення, засновану на законі відбиття Ламберта, що дозволило їм відновити дрібні деталі, такі як ямочки та зморшки. Дін та ін. запропонували техніку деокклюзії обличчя для зображень обличчя, яка вимагала від користувача вибору, який об'єкт видалити. Поєднавши ваніль і часткові згортки в одній мережі, вони змогли створити контент, який був добре інтегрований і не мав візуальних артефактів. Вони також створили великий синтетичний парний набір даних із закритим обличчям, використовуючи відкрито доступні набори даних CelebA та CelebA-HQ, щоб вирішити проблему недостатності даних. Вони виявили, що модель, навчена на синтетичному наборі даних із закритим обличчям, ефективно видаляє об'єкти, які не є обличчями, і надає структурно та правдоподібно вміст обличчя у складних зображеннях реального світу.

Ван та ін. запропонував MaskNet, модуль глибокого навчання для вивчення особливостей зображення з незвичайною точністю, ігноруючи ті, які деформовані оклюзіями. Його можна використовувати в різноманітних архітектурах CNN, які потребують меншої кількості персональних ідентифікаційних міток і обчислень. Ефективність MaskNet була продемонстрована за допомогою реальних і синтетичних зображень закритого обличчя [6]. Ця мережа була навчена на CASIA-Webface, точно налаштована на наборі даних AR і, нарешті, протестована на наборі даних LFW. Сонг та ін. запропонували структуру парної диференціальної сіамської мережі (PDSN) для глибоких моделей CNN, щоб знайти еквівалентність між закритими лицьовими частинами та елементами ознак. На наборах даних обличчя з реальними та синтетичними оклюзіями була оцінена продуктивність системи.

Розпізнавання облич за допомогою масок оклюзії, запропоноване Кью та ін., є методом розпізнавання обличчя з оклюзіями на основі однієї наскрізної глибокої нейронної мережі (FROM) [6].

Він використовується для вивчення точних масок функцій, а потім використання глибоких CNN для пошуку пошкоджених функцій і очищення їх за допомогою масок, що вивчаються динамічно. Крім того, автори ефективно тренують FROM, створюючи масиви зображення облич. Вони розглянули

різноманітні набори даних із замаскованими або закритими обличчями, включаючи LFW, Megaface challenge 1, RMF2 та AR.

Щоб виділити різні місцеві особливості, Wang et al. запропонували попарні самоконтрастні моделі уваги (PSCA). Запропонована втрата розрідженості уваги (ASL) збільшує розріджені реакції на картах уваги, зменшуючи фокус на відволіканих ділянках і збільшуючи увагу на дискримінаційних частинах обличчя. Вони перевірили продуктивність розпізнавання на наборах даних LFW, VGGFace2, MS-Celeb-1M та RMFRD.

Бісвас та ін. запропонував одноразовий метод перцептивного хешування структури сусідства з домінантною частотою (OSF-DNS). За допомогою цього методу покращено завдання верифікації прихованого обличчя та класифікації обличчя. Перевірка закритого обличчя має переваги від можливості зіставляти закриті обличчя з їх неоклюдованими аналогами [6]. Крім того, для класифікації обличчя корисним є отримання ідентичності закритого обличчя за допомогою класифікатора, навченого з незакритими обличчями та перцептивними хеш-кодами як векторами ознак.

## 2.9 Розпізнавання облич у масці

Дін запропонував метод автоматичного видалення об'єктів маски з обличчя та синтезу пошкоджених ділянок із збереженням оригінальної структури обличчя. Вони зберегли консистенцію структури та форми у вилученому обличчі за допомогою двох дискримінаторів, щоб дізнатися загальну структуру обличчя глибоко видаленої ділянки. Для вирішення проблеми недостатності даних використовується синтетичний парний набір даних на основі набору даних CelebA [7]. Їх комбінована модель прямого зв’язку генерує структурно та правдоподібні для сприйняття зображення обличчя, щоб впоратися з реальним зображенням. Чандра та ін. провели порівняльний аналіз чотирьох передових моделей глибокого навчання, а саме VGGFace, FaceNet, OpenFace і DeepFace.

Вони дійшли висновку, що ці моделі мають високу точність у завданні перевірки обличчя.

Монтеро та ін. представили глибоку модель на основі ArcFace зі змінами в магістралі та функцією втрат. Вони створили замасковану версію оригінального набору даних розпізнавання обличчя за допомогою розширення

даних і протестували ResNet-50 на MFR з низькими обчислювальними витратами. Втрата ArcFace потім поєднується з втратою класифікації використання маски, щоб утворити нову функцію під назвою Multi-Task ArcFace (MTArcFace) [7].

Харірі запропонував функції на основі глибокого навчання для відкидання замаскованих областей у MFR.

Вони використовували попередньо навчені глибокі CNN, щоб вибрати найкращі риси із захоплених регіонів, якими були переважно очі та області чола. Парадигма «мішок функцій» потім була застосована до карт ознак останнього згорткового шару для квантування представлення. Вони також використовували набір даних RMFRD, який використовує три попередньо навчені глибокі CNN — VGG-16, AlexNet і ResNet-50 — для вибору глибоких функцій із захоплених регіонів. Махарані та ін. представили підхід MFR, заснований на каскаді Haar і MobileNet для виявлення масок, а потім використали VGG16 і Triplet Loss FaceNet з технікою багатопотокової ідентифікації обличчя [7]. Бутрос та ін. представив модель EUM, яка працювала над головою сучасних моделей розпізнавання обличчя. Вони використовували самообмежений триплет, що дозволило EUM створювати вбудовування на основі немаскованих облич пов’язаних персонажів.

Golwalkar та ін. використовували мережу FaceMaskNet-21, навчені четвірки та глибоке метричне навчання, щоб миттєво ідентифікувати обличчя в масках. Для кожного обличчя в наборі даних, а також для вхідного зображення або відеопотоку в реальному часі було створено 128-d кодування [7].

Вони також використовували функції HOG для більш активного розпізнавання облич, прихованих маскою. Wang та ін. запропонували три набори даних для замаскованих облич: реальний набір даних маскованого розпізнавання облич (RMFRD), набір даних виявлення замаскованих облич (MFDD) і набір даних розпізнавання замаскованих облич (SFRD), щоб виконувати завдання MFR, які вимагають достатньої кількості замаскованих та немасковані зображення. Вони присвоїли вагу важливим рисам, видимим на відкритих ділянках обличчя, таким як деталі очей, лоб і контур обличчя [7].

Анвар та ін. запропонував модель MaskTheFace, яка генерує колекцію замаскованих облич синтетично. MaskTheFace виявляє орієнтири обличчя, щоб визначити основні риси та нахил обличчя. Вони також використовували MaskTheFace для розпізнавання замаскованих облич за допомогою системи

FaceNet, яка додає вбудовування до облич. Для навчання FaceNet вони використали набір даних VGGFace2 і випадково відібрали 42 зображення на людину, щоб створити VGGFace2-mini. Вони створили VGGFace2-mini-SM з нової підмножини, яка включає ті самі зображення, але з масками [7].

WearMask3D — це підхід на основі 3D-моделі, розроблений Хонг та ін., щоб доповнити зображення обличчя в масках у різних позах. Він налаштовує 3D-модель (3DMM) до зображення, а потім створює поверхню 3D-маски, щоб накласти її на модель обличчя. Він застосовує текстуру маски до моделі та перетворює 3D-поверхню на 2D-зображення на основі роздільної здатності та яскравості зображення. Вони також представили міні-набір даних Masked Faces in the Wild (MFW) та оцінили продуктивність моделі на MFW-mini та MFR2.

Мандал та ін. представили фреймворк на основі ResNet-50 для розпізнавання облич у масках.

Вони використовували адаптацію домену для навчання, причому незамасковані обличчя слугували вихідним доменом, а замасковані обличчя — цільовим доменом. Перший сценарій полягав у тому, щоб навчити модель лише на вихідному домені, а потім перевірити її на цільовій. Другий сценарій полягав у навчанні моделі на вихідному домені та частині цільового домену перед тестуванням на решті цільового домену [7].

# 3 АНАЛІЗ ТЕХНОЛОГІЙ РОЗПІЗНАВАННЯ МАСОК НА ОБЛИЧЧІ

## 3.1 OpenCV

OpenCV (англ. Open Source Computer Vision Library, бібліотека комп'ютерного зору з відкритим кодом) — бібліотека алгоритмів комп'ютерного зору, обробки зображень та чисельних алгоритмів загального призначення з відкритим кодом. Реалізована C/C++, також розробляється для Python, Java, Ruby, Matlab, Lua та інших мов. Може вільно використовуватися в академічних та комерційних цілях – поширюється за умов ліцензії BSD [1].

Другий великий апдейт OpenCV було випущено у жовтні 2009 року. OpenCV 2 включає серйозні зміни в інтерфейсі C++, спрямовані на спрощення, поліпшення безпеки, введення нових функцій і збільшення продуктивності (особливо для багатоядерних систем). Офіційні релізи тепер випускаються кожні шість місяців, а технологія ведеться незалежною російською командою за підтримки комерційних корпорацій [1].

### 3.1.1 Історія

Проект OpenCV був офіційно заснований у 1999 році як програма Intel Research для розробки додатків із інтенсивним процесором, як частина послідовності ініціатив, які включали трасування променів у реальному часі та 3D-дисплеї [1]. Ряд спеціалістів з оптимізації з Intel в Росії, а також команда Intel Performance Library Team зробили основний внесок у проект. На початку OpenCV цілі проекту були визначені наступним чином:

Покращення дослідження бачення, надаючи відкритий та оптимізований код для фундаментальної інфраструктури бачення. Більше не буде винаходити велосипед.

Поширення інформації про бачення, надаючи стандартну архітектуру для розробників, роблячи код більш зрозумілим і переносимим.

Розвивиток комерційні додатки на основі бачення, роблячи вільний доступ до портативного, оптимізованого за продуктивністю коду — за ліцензією, яка не вимагає, щоб джерело було відкритим або безкоштовним [1].

Перша альфа-версія OpenCV була представлена громадськості в 2000 році на конференції IEEE з комп'ютерного зору та розпізнавання образів, а потім

п'ять бета-версій у період з 2001 по 2005 рік. У 2006 році була запущена початкова версія 1.0. У жовтні 2008 року була випущена «передрелізна» версія 1.1 [1].

У жовтні 2009 року стала доступною друга основна версія OpenCV. OpenCV 2 забезпечує значні вдосконалення інтерфейсу C++, спрямовані на простіші, безпечніші для типів шаблони, нові методи та більш продуктивні реалізації для існуючих (особливо в багатоядерних системах). Офіційні релізи тепер виходять кожні півроку, а розробкою зараз займається незалежна російська команда, спонсорована комерційними організаціями [1].

Підтримка OpenCV була розпочата у серпні 2012 року неприбутковою організацією OpenCV.org, яка керує сайтом розробників і користувачів.

### 3.1.2 Застосування

Для затвердження загального стандартного інтерфейсу комп'ютерного зору додатків у цій галузі. Для сприяння зростанню числа таких програм та створення нових моделей використання PC [1].

Зробити платформи Intel привабливими для розробників таких додатків за рахунок додаткового прискорення OpenCV за допомогою Intel Performance Libraries (Зараз включають IPP (низькорівневі бібліотеки для обробки сигналів, зображень, а також медіа-кодеки) та MKL (спеціальна версія LAPACK та FFTPack)). OpenCV здатна автоматично виявляти присутність IPP та MKL та використовувати їх для прискорення обробки [1].

Основні модулі

У версії 2.2 бібліотека була реорганізована. Замість універсальних модулів cxcore, cvaux, highGUI та інших було створено кілька компактних модулів з більш вузькою спеціалізацією:

– opencv_core – основна функціональність. Включає базові структури, обчислення (математичні функції, генератори випадкових чисел) і лінійну алгебру, DFT, DCT, введення/виведення для XML і YAML і т. д.

– opencv_imgproc - обробка зображень (фільтрація, геометричні перетворення, перетворення колірних просторів тощо).

– opencv_highgui — простий UI, введення/виведення зображень та відео.

– opencv_ml - моделі машинного навчання (SVM, дерева рішень, навчання зі стимулюванням тощо).

– opencv_features2d - розпізнавання та опис плоских примітивів (SURF(англ.)рус., FAST та інші, включаючи спеціалізований фреймворк).

– opencv_video - аналіз руху та відстеження об'єктів (оптичний потік, шаблони руху, усунення фону).

– opencv_objdetect - Виявлення об'єктів на зображенні (знаходження осіб за допомогою алгоритму Віоли-Джонса, розпізнавання людей HOG і т. д.).

– opencv_calib3d – калібрування камери, пошук стерео-відповідності та елементи обробки тривимірних даних.

– opencv_flann — бібліотека швидкого пошуку найближчих сусідів (FLANN 1.5) та обгортки OpenCV.

– opencv_contrib — супутній код, який ще не готовий для застосування.

– opencv_legacy - застарілий код, збережений для зворотної сумісності.

– opencv_gpu — прискорення деяких функцій OpenCV за рахунок CUDA, створеного за допомогою NVidia.

## 3.2 Caffe-based detector

Caffe — це система глибокого навчання, розроблена для виразності, швидкості та універсальності. Над цим працюють Berkeley AI Research (BAIR) і співробітники спільноти. Янцин Цзя розробив цю ідею в рамках свого навчання в докторантурі в Каліфорнійському університеті в Берклі [4]. Caffe поширюється на умовах ліцензії BSD 2-Clause.

Експресіоністична архітектура сприяє застосуванню та творчості. Моделі та оптимізація налаштовані, а не жорстко закодовані. Встановивши один прапорець, ви можете тренуватися на системі графічного процесора, а потім розгорнути на кластери товарів або мобільні пристрої.

Розширюваний код стимулює активний розвиток. У перший рік Caffe було створено більш ніж 1000 розробниками, і багато суттєвих модифікацій було внесено назад [4]. Завдяки цьому внеску фреймворк відповідає останнім розробкам як у коді, так і в моделях.

Швидкість Caffe робить його ідеальним для наукових досліджень та промислового впровадження. Caffe може обробляти понад 60 мільйонів

фотографій на день на одному GPU NVIDIA K40*. Це 1 мс на зображення для висновку і 4 мс на зображення для навчання, а нові версії бібліотеки та обладнання ще швидше. Caffe, на нашу думку, є однією з найшвидших доступних реалізацій convnet.

Caffe вже керує університетськими дослідницькими проектами, прототипами стартапів і навіть великомасштабними промисловими програмами бачення, голосових та мультимедійних програм [4].

## 3.3 Keras

Keras — відкрита бібліотека, написана мовою Python і забезпечує взаємодію зі штучними нейронними наборами. Вона являє собою надбудову над фреймворком TensorFlow. До версії 2.3 підтримувалися різні версії нейросетевих бібліотек, такі як TensorFlow, Microsoft Cognitive Toolkit, Deeplearning і Theano [5]. Націлена на оперативну роботу з сетями глибинного навчання, при цьому спроектована так, щоб бути компактною, модульною та розширюваною. Вона була створена як частина дослідницьких сил проекту ONEIROS (англ. Open-ended Neuro-Electronic Intelligent Robot Operating System), її основним автором і підтримуючим є Франсуа Шолле (фр. François Chollet), інженер Google.

Планується, що Google підтримуватиме Keras в основній бібліотеці TensorFlow, однак Шолле виділяв Keras у відділену надбудову, так як згідно з концепцією Keras є швидше інтерфейсом, ніж східною системою машинного навчання [5]. Keras надає високоуровневий, більш інтуїтивний набір абстракцій, який робить простим формування нейронних сетей, незалежно від використовуваного в якості вичислительного бэкенда бібліотеки наукового числа.Microsoft працює над доповненням до Keras та низькоуровневих бібліотек CNTK.

Ця бібліотека містить багаточисельні реалізації, широко застосовувані будівельних блоків нейронних сетей, такі як слої, цілі та передавальні функції, оптимізатори та безліч інструментів для просування роботи із зображеннями та текстом. Код, розміщений на GitHub, форуми підтримки включають сторінку питань GitHub, канал Gitter[en] і канал Slack [5].

## 3.4 TensorFlow

TensorFlow — відкрита програмна бібліотека для машинного навчання, розроблена компанія Google для вирішення завдань побудови та тренування нейронної мережі з метою автоматичного навчання та класифікації образів, що досягає якості людського виховання. Прикладається як для досліджень, так і для розробки власних продуктів Google. Основний API для роботи з бібліотекою, реалізованою для Python, також існують реалізації для R, C Sharp, C++, Haskell, Java, Go і Swift [5].

Є продовженням закритого проекту DistBelief. Спочатку TensorFlow був розроблений командою Google Brain для внутрішнього використання в Google, у 2015 році система була переведена у вільний доступ з відкритою ліцензією Apache 2.0 [6].

Закрита система машинного навчання DistBelief розробляла Google Brain для внутрішніх проектів з 2011 року для роботи з нейронними наборами глибокого навчання. Вона стала використовуватися в багатьох дослідницьких і комерційних проектах групи фірм холдинга Alphabet. Після успіху DistBelief, фірма Google вирішила реалізувати проект на новий рівень, а для рефакторингу виділила групу з кількох розробників, в яку вошёл Джефф Дін; метою групи було упрощення та оптимізація кодів бібліотеки, збільшення надії та зручності користування. Нова бібліотека отримала назву TensorFlow. У 2013 році до проекту приєднався Джеффрі Хінтон — навчальний, під керівництвом якого в 2009 році був створений метод узагальненого зворотного поширення помилок та ряд інших покращень, дозволивши істотно вилучити точність нейронної системи (що привело, зокрема, до зниження погрішностей у розпізнаванні речей на 25 %) [5, 6].

TensorFlow 9 листопада 2015 року був відкритий для вільного доступу. TensorFlow є системою машинного навчання Google Brain другого покоління. У той час, як еталонна реалізація працює на єдиних пристроях, TensorFlow може працювати з багатьма паралельними процесорами, як CPU, так і GPU, керуючи архітектурою CUDA для підтримки загального призначення графічних процесорів. TensorFlow доступний для 64-розрядних Linux, macOS, Windows і для мобільних вичислильних платформ, включаючи Android та iOS [6].

Обчислення TensorFlow виражається у вигляді потоків даних через графічний склад. Назва TensorFlow відбувається від операцій із багатомірними масивами даних, які також називаються «тензорами». У червні 2016 року Джефф Дін з Google зазначив, що до TensorFlow зверталися 1500 репозиторіїв на GitHub, і лише 5 з них були від Google [6].

У травні 2016 року Google повідомила про призначення для задач глибинного навчання апаратного ускорителя власної розробки — тензорного процесора (TPU) — спеціалізованої інтегральної схеми, адаптованої під задачі для TensorFlow та забезпечення високої продуктивності в арифметиці зниженої точності (наприклад, для 8-бітної архітектури) і направленою швидше на застосування моделей, чим на їхнє навчання [6].

Повідомлялося, що після використання TPU у власних завданнях Google при роботі з даними вдалося отримати в порядку найкращих показників продуктивності затраченої енергії.

### 3.5 MobileNetV2

MobileNet-v2 — це згортка нейронна мережа, яка має глибину 53 шари. Ви можете завантажити попередньо підготовлену версію мережі, натреновану на більш ніж мільйоні зображень із бази даних ImageNet. Попередньо навчена мережа може класифікувати зображення на 1000 категорій об’єктів, таких як клавіатура, миша, олівець і багато тварин. В результаті мережа навчилася багатим представленням функцій для широкого діапазону зображень. Мережа має розмір вхідного зображення 224x224. Для отримання додаткових відомостей про попередньо підготовлені мережі в MATLAB® див. Попередньо навчені глибокі нейронні мережі [5].

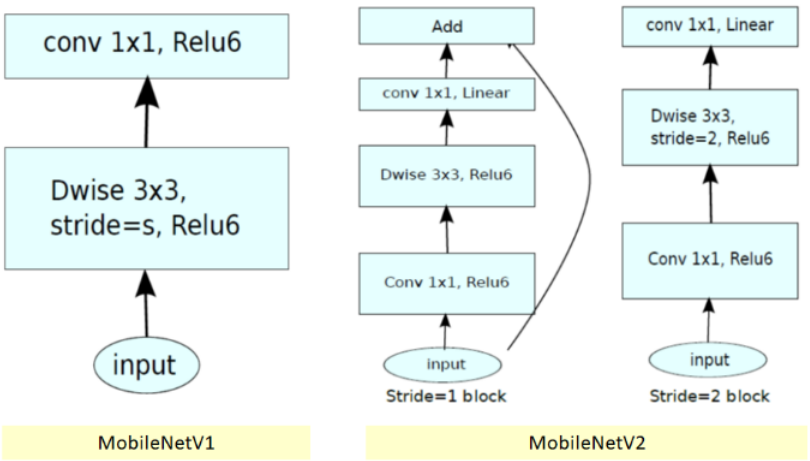

Рисунок 3.1 – Згорткові блоки MobileNet

### 3.5.1 MobileNetV1

У MobileNetV1 є 2 шари.

Перший шар називається глибинною згорткою, він виконує легку фільтрацію, застосовуючи один згортковий фільтр на вхідний канал [5].

Другий шар — це згортка 1×1, яка називається точкової згорткою, яка відповідає за створення нових функцій за допомогою обчислення лінійних комбінацій вхідних каналів.

ReLU6 використовується завдяки своїй надійності під час використання з низькоточними обчисленнями на основі MobileNetV1 [5].

### 3.5.2 MobileNetV2

У MobileNetV2 є два типи блоків. Один – це залишковий блок із кроком 1. Інший – блок із кроком 2 для зменшення розміру.

Є 3 шари для обох типів блоків.

Цього разу перший шар — згортка 1×1 з ReLU6.

Другий шар - це глибинна згортка.

Третій шар - це ще одна згортка 1×1, але без будь-якої нелінійності. Стверджується, що якщо ReLU буде використано знову, глибокі мережі мають потужність лише лінійного класифікатора на ненульовій частині обсягу вихідної області [5].

### 3.5.3 Загальна архітектура

| Input | Operator | $t$ | $c$ | $n$ | $s$ |
|---|---|---|---|---|---|
| $224^2 \times 3$ | conv2d | - | 32 | 1 | 2 |
| $112^2 \times 32$ | bottleneck | 1 | 16 | 1 | 1 |
| $112^2 \times 16$ | bottleneck | 6 | 24 | 2 | 2 |
| $56^2 \times 24$ | bottleneck | 6 | 32 | 3 | 2 |
| $28^2 \times 32$ | bottleneck | 6 | 64 | 4 | 2 |
| $14^2 \times 64$ | bottleneck | 6 | 96 | 3 | 1 |
| $14^2 \times 96$ | bottleneck | 6 | 160 | 3 | 2 |
| $7^2 \times 160$ | bottleneck | 6 | 320 | 1 | 1 |
| $7^2 \times 320$ | conv2d 1x1 | - | 1280 | 1 | 1 |
| $7^2 \times 1280$ | avgpool 7x7 | - | - | 1 | - |
| $1 \times 1 \times 1280$ | conv2d 1x1 | - | k | - | |

MobileNetV2 Overall Architecture

Рисунок 3.2 – Загальна архітектура MobileNetV2

де t: коефіцієнт розширення, c: кількість вихідних каналів, n: число повторень, s: крок. Ядра 3×3 використовуються для просторової згортки [7].

Як правило, первинна мережа (множник ширини 1, 224×224) має обчислювальну вартість 300 мільйонів множин-доданок і використовує 3,4 мільйона параметрів. (Множник ширини введено в MobileNetV1.)

Далі досліджуємо компроміси продуктивності для роздільної здатності введення від 96 до 224 і множників ширини від 0,35 до 1,4 [7].

Вартість обчислень мережі до 585 млн. МДД, а розмір моделі варіюється від 1,7 млн до 6,9 млн параметрів.

Для навчання мережі використовується 16 GPU з розміром пакету 96.

| Size | MobileNetV1 | MobileNetV2 | ShuffleNet (2x,g=3) |
|---|---|---|---|
| 112x112 | 64/1600 | 16/400 | 32/800 |
| 56x56 | 128/800 | 32/200 | 48/300 |
| 28x28 | 256/400 | 64/100 | 400/600K |
| 14x14 | 512/200 | 160/62 | 800/310 |
| 7x7 | 1024/199 | 320/32 | 1600/156 |
| 1x1 | 1024/2 | 1280/2 | 1600/3 |
| **max** | 1600K | **400K** | 600K |

Рисунок 3.3 – Обчислювальна вартість навчання мереж MobileNetV1 та MobileNetV2 у порівнянні з ShuffleNet

### 3.5.4 Дослідження абляції

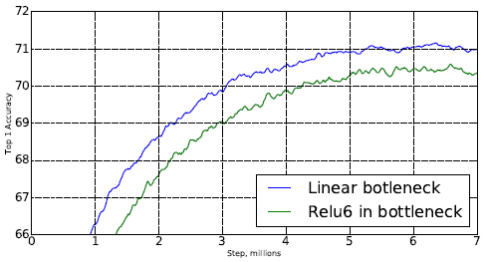


Рисунок 3.4 – Вплив лінійного вузького місця

З видаленням ReLU6 на виході кожного модуля вузького місця підвищується точність [5].

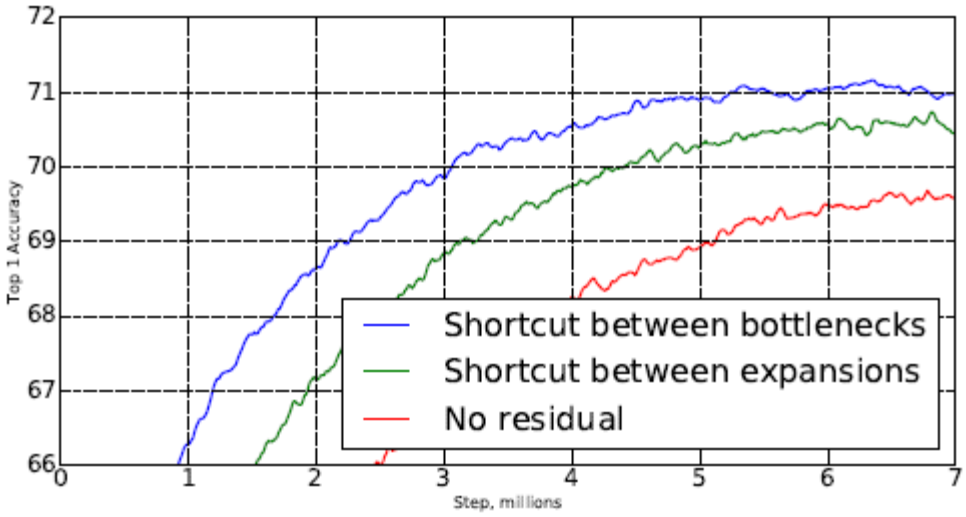


Рисунок 3.5 – Графік впливу ярлика

Завдяки ярлику між вузькими місцями, він перевершує ярлик між розширеннями та той, який не має жодних залишкових зв’язків [8].

### 3.5.5 Результати експерименту

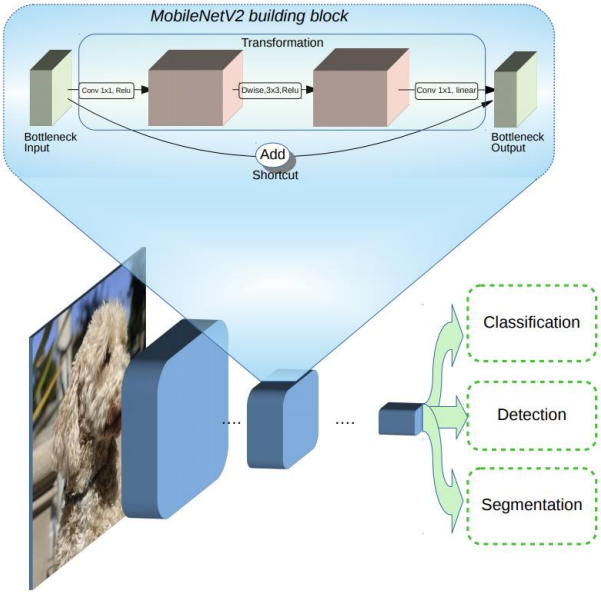


Рисунок 3.6 – MobileNetV2 для класифікації, виявлення та сегментації

| Network | Top 1 | Params | MAdds | CPU |
|---|---|---|---|---|
| MobileNetV1 | 70.6 | 4.2M | 575M | 113ms |
| ShuffleNet (1.5) | 71.5 | **3.4M** | 292M | - |
| ShuffleNet (x2) | 73.7 | 5.4M | 524M | - |
| NasNet-A | 74.0 | 5.3M | 564M | 183ms |
| MobileNetV2 | **72.0** | **3.4M** | **300M** | **75ms** |
| MobileNetV2 (1.4) | **74.7** | 6.9M | 585M | **143ms** |

Рисунок 3.7 – Точність ImageNet Top-1

MobileNetV2 перевершує MobileNetV1 і ShuffleNet (1.5) з порівнянним розміром моделі та вартістю обчислень.

З коефіцієнтом ширини 1,4 MobileNetV2 (1,4) перевершує ShuffleNet (×2), а NASNet — за швидшим часом висновку [8].

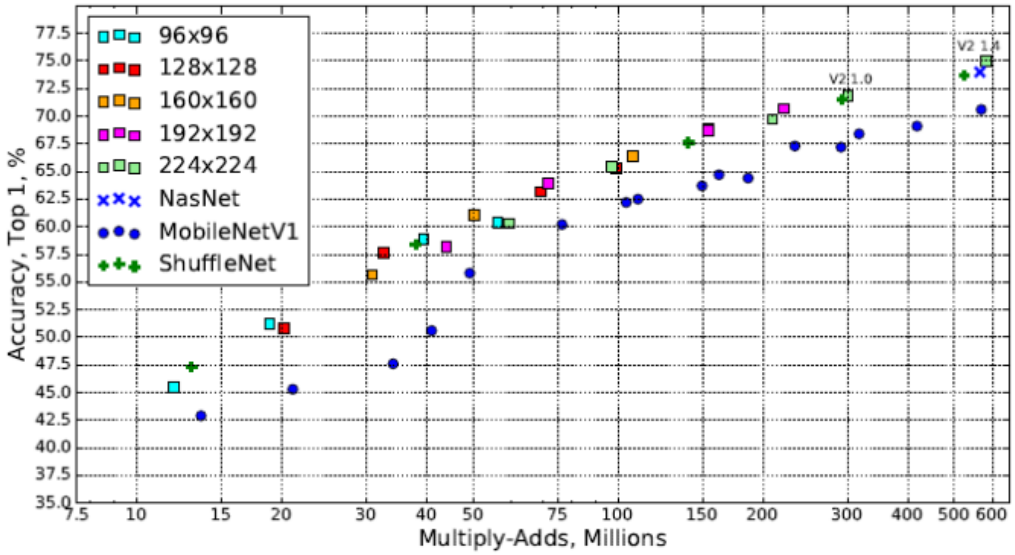


Рисунок 3.8 – Множники ширини для різних систем

Як показано вище, використовуються різні вхідні дозволи та множники ширини. Він стабільно перевершує MobileNetV1 [8].

3.5.6 Виявлення об'єктів за допомогою MS COCO

| | Params | MAdds |
|---|---|---|
| SSD[34] | 14.8M | 1.25B |
| SSDLite | **2.1M** | **0.35B** |

SSDLite

Рисунок 3.9 – Показники ефективності для SSD та SSDLite

По-перше, SSDLite представлено шляхом модифікації звичайних згорток у SSD з роздільними по глибині згортками (MobileNetV1 one) [8].

SSDLite різко знижує кількість параметрів і витрати на обчислення.

| Network | mAP | Params | MAdd | CPU |
|---|---|---|---|---|
| SSD300[34] | 23.2 | 36.1M | 35.2B | - |
| SSD512[34] | 26.8 | 36.1M | 99.5B | - |
| YOLOv2[35] | 21.6 | 50.7M | 17.5B | - |
| MNet V1 + SSDLite | 22.2 | 5.1M | 1.3B | 270ms |
| MNet V2 + SSDLite | 22.1 | **4.3M** | **0.8B** | 200ms |

Рисунок 3.10 – Виявлення об'єктів MS COCO

MobileNetV2 + SSDLite досягає конкурентоспроможної точності зі значно меншою кількістю параметрів і меншою обчислювальною складністю.

І час висновку швидше, ніж у MobileNetV1 [6].

Примітно, що MobileNetV2 + SSDLite в 20 разів ефективніший і в 10 разів менший, але все ще перевершує YOLOv2 на наборі даних COCO.

### 3.5.7 Семантична сегментація за допомогою PASCAL VOC 2012

| Network | OS | ASPP | MF | mIOU | Params | MAdds |
|---|---|---|---|---|---|---|
| MNet V1 | 16 | ✓ | | 75.29 | 11.15M | 14.25B |
| | 8 | ✓ | ✓ | 78.56 | 11.15M | 941.9B |
| MNet V2* | 16 | ✓ | | 75.70 | 4.52M | 5.8B |
| | 8 | ✓ | ✓ | 78.42 | 4.52M | 387B |
| MNet V2* | 16 | | | **75.32** | **2.11M** | **2.75B** |
| | 8 | | ✓ | 77.33 | 2.11M | 152.6B |
| ResNet-101 | 16 | ✓ | | 80.49 | 58.16M | 81.0B |
| | 8 | ✓ | ✓ | 82.70 | 58.16M | 4870.6B |

Рисунок 3.11 – Набір перевірки PASCAL VOC 2012

Тут MobileNetV2 використовується як екстрактор функцій для DeepLabv3.

При відключенні Atrous Spatial Pyramid Pooling (ASPP), а також Multi-Scale і Flipping (MP), а також змінюючи вихідний крок з 8 до 16, отримується mIOU в розмірі 75,32% при дуже низьких розмірах моделі та витратах на обчислення [7].

# ВИСНОВКИ

Технологічний прогрес у всьому світі тісно пов’язаний із багатьма соціальними явищами. Одним із таких чинників, що кинув виклик можливостям людства стала всесвітня пандемія вірусу COVID-19.

Спричинивши значущі трансформації у багатьох сферах людської діяльності, не оминувши і сферу громадської безпеки. Зокрема, поставлена під сумнів ефективність старих методів розпізнавання облич, оскільки тепер у громадських місцях люди змушені носити захисні маски, які собою закривають частину обличчя. Таким чином, стало неможливо достеменно встановлювати особистість людини, яка перебуває в аеропорту, кафе чи просто ходить вулицею.

Для вирішення цих задач існує два основних підходи:

Перший базується на тому, щоб працювати лише з видимими частинами обличчя, зокрема і особливо з очима та зонами навколо них. Прихильники цього методу відзначають, що точки, які розташовані навколо очей людини, несуть дуже великий обсяг інформації не лише про, наприклад, емоційний стан людини, але і дають змогу ідентифікувати особу, оскільки саме вони містять найбільшу долю інформації, яку взагалі можна зчитати з обличчя людини.

Дійсно, зосередження роботи алгоритмів з розпізнавання саме на цих точках обличчя помітно покращує їхню ефективність у випадку роботи з замаскованими обличчями (за різними даними з приблизно 33-35% до майже 60%). Однак, ці показники є все ще недостатніми для здійснення ефективної ідентифікації особи за її обличчям, а робота таких алгоритмів є доволі нестабільною і з цих міркувань в даній роботі не розглядалися.

З іншого ж боку, пропонується підхід, який передбачає відновлення тої частини обличчя, яку не видно за захисною маскою. Для досягнення цієї мети, використовуються технології штучного інтелекту та машинного навчання. Зокрема, технологія GAN дозволяє створювати, або досить правдоподібно відновлювати певні частини зображення.

В першому розділі було проаналізовано предметну область, а також розглянута сама сутність понять, які є ключовими у даній області: Комп’ютерний зір, Розпізнавання облич та Штучний інтелект. Проаналізовано їхню історію, що є наразі дуже важливим, оскільки ці сфери людського життя

отримують в останні роки отримують принципово новий рівень розвитку.

У другому розділі був в цілому та детально описаний процес розпізнавання замаскованого обличчя, докладно описаний кожний з його етапів. Було визначено, що для виконання цих завдань, підходять саме згорткові нейронні мережі, детально розглянуті їхні різновиди та особливості. Розглянуті запропоновані науковою спільнотою підходи до вирішення кожного з завдань у цьому процесі.

Третій розділ був присвячений прикладним інструментам для розпізнавання масок на обличчі. Зокрема, розглянуті бібліотеки, які дають змогу у сукупності, побудувати рішення, яке дозволяє виконати це завдання: OpenCV, Caffe-based detector, Keras та TensorFlow. Цей етап, у процесі розпізнавання замаскованих облич є ключовим, оскільки перш ніж здійснювати ідентифікацію, слід дуже точно визначити, які частини обличчя є видимими, тобто вихідними даними для подальшого процесу розпізнавання, а які закритими, тобто, як у нашому випадку, потребують відновлення свого зображення, це досягається завдяки технологіям штучного інтелекту.

# ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Binford T. O., Tenenbaum J. M. Computer vision. Computer. 1973. Т. 6, № 5. С. 19–24.

2. Saxenna Y. Face Recognition System. International Journal for Research in Applied Science and Engineering Technology. 2020. Т. 8, № 7. С. 1883–1885.

3. Singh S., Prasad S. V. A. V. Techniques and Challenges of Face Recognition: A Critical Review. Procedia Computer Science. 2018. Т. 143. С. 536–543.

4. Masked face recognition: Human versus machine / N. Damer та ін. IET Biometrics. 2022

5. Alyuz N., Gokberk B., Akarun L. 3-D Face Recognition Under Occlusion Using Masked Projection. IEEE Transactions on Information Forensics and Security. 2013. Т. 8, № 5. С. 789–802

6. Masked Face Recognition From Synthesis to Reality / G.-S. J. Hsu та ін. IEEE Access. 2022. Т. 10. С. 37938–37952

7. Towards NIR-VIS Masked Face Recognition / H. Du та ін. IEEE Signal Processing Letters. 2021. Т. 28. С. 768–772.

8. Face Mask Detection. International Journal of Advanced Trends in Computer Science and Engineering. 2021. Т. 10, № 3. С. 1520–1523.