

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Національний аерокосмічний університет ім. М. Є. Жуковського
«Харківський авіаційний інститут»

Факультет програмної інженерії та бізнесу

Кафедра інженерії програмного забезпечення

Пояснювальна записка до дипломного проекту

магістра

(освітній ступінь)

на тему «Дослідження застосування штучних нейронних мереж у задачах ідентифікації облич на зображенні»

XAI.603.667п1.121.156339.200

Виконав: студент 6 курсу групи № 667п1
Спеціальність 121 – Інженерія програмного
забезпечення

(код та найменування)

Освітня програма Хмарні обчислення та
Інтернет речей

(найменування)

Костров І.А.

(прізвище й ініціали студента)

Керівник Кіріленко О.Г.

(прізвище та ініціали)

Рецензент Коваленко А.А.

(прізвище та ініціали)

Харків – 2020

Міністерство світи і науки України
Національний аерокосмічний університет ім. М. Є. Жуковського
«Харківський авіаційний інститут»

Факультет програмної інженерії та бізнесу
(повне найменування)

Кафедра інженерії програмного забезпечення
(повне найменування)

Рівень вищої освіти другий (магістерський)

Спеціальність 121 – інженерія програмного забезпечення
(код та найменування)

Освітня програма хмарні обчислення та Інтернет речей
(найменування)

ЗАТВЕРДЖУЮ
Завідувач кафедри

(підпис) (ініціали та прізвище)
“ _____ ” _____ 2020 року

З А В Д А Н Н Я
НА ДИПЛОМНИЙ ПРОЕКТ СТУДЕНТУ

Кострову Іллі Андрійовичу
(прізвище, ім'я, по батькові)

1. Тема дипломного проекту Дослідження застосування штучних нейронних мереж у задачах ідентифікації облич на зображенні

керівник дипломного проекту Кіріленко Олена Георгіївна, к.т.н, професор
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом вищого навчального закладу від “__” ____ 2020 року № __

2. Термін подання студентом роботи _____

3. Вихідні дані до роботи: програмний додаток для детекції та генерації структурної моделі обличчя

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

1 Критичний аналіз існуючих методів, моделей і програмно-апаратних системи для розпізнавання образів на фотозображенні;

2 Розроблення методу та моделі нейронної мережі для детекції та ідентифікації облич на фотозображенні;

3 Розробка програмного додатку для детекції, розпізнавання та генерації структурної моделі облич по фотозображенню

5. Перелік графічного матеріалу

РПЗ – стор. 88, рисунків – 31 шт., таблиць – 2 шт., презентація – 18 слайдів.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1	Кіріленко О.Г., проф. каф. 603		
2	Кіріленко О.Г., проф. каф. 603		
3	Кіріленко О.Г., проф. каф. 603		

8. Нормоконтроль _____ В.А. Постернакова « ____ » _____ 2020 р.
(підпис) (ініціали та прізвище)

7. Дата видачі завдання _____

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проекту	Строк виконання етапів проекту	Примітка
1	Отримання і затвердження теми диплому	03.09.2019	
2	Аналіз предметної області	04.09.2019	
3	Постановка задачі	20.11.2019	
4	Проведення теоретичних досліджень	22.11.2019	
5	Розробка прототипу ПЗ	02.09.2020	
6	Підготовка пояснювальної записки	22.10.2020	
7	Оформлення пояснювальної записки до дипломного проекту	10.11.2020	
8	Передзахист дипломного проекту	24.11.2020	
9	Захист дипломного проекту	04.12.2020	

Студент

_____ (підпис)

Костров І.А.

(прізвище та ініціали)

Керівник роботи

_____ (підпис)

Кіріленко О.Г.

(прізвище та ініціали)

РЕФЕРАТ

Пояснювальна записка до дипломного проекту містить 88 стор., 31 рис., 54 джерел.

Об'єктом дослідження – процес розпізнавання образів на фотозображенні засобами штучного інтелекту.

Предмет дослідження – моделі та методи штучного інтелекту для генерації структурної моделі обличчя.

Метою роботи є підвищення ефективності розпізнавання особи певної людини по фотозображенню шляхом генерації структурної моделі знайдених обличч засобами штучного інтелекту за рахунок розроблення програмного додатку для детекції обличч на зображенні.

Для досягнення поставленої мети необхідно розв'язати такі задачі: провести критичний аналіз існуючих методів, моделей і програмно-апаратних системи для розпізнавання образів на фотозображенні; розробити метод та модель нейронної мережі для детекції та ідентифікації обличч на фотозображенні; розробити програмний додаток для детекції, розпізнавання та генерації структурної моделі обличч по фотозображенню.

Наукова новизна. Удосконалено метод пошуку обличч людини та ключових точок на ньому на двовимірному зображенні який на відміну від існуючих використовує згорткові нейронні мережи для побудови структурної моделі обличчя, що дає можливість підвищити точність ідентифікації людини по фотозображенню.

Практична значимість отриманих результатів. В результаті роботи був розроблений програмний додаток для детекції та побудови структурної моделі обличчя, а на основі цих даних - ідентифікації особи по фотозображеннях, що наявні в базі даних.

СТРУКТУРНА МОДЕЛЬ ОБЛИЧЧЯ, ЗГОРТКОВІ НЕЙРОННІ МЕРЕЖИ,
РОЗПІЗНАВАННЯ ОБРАЗІВ, ШТУЧНИЙ ІНТЕЛЕКТ

ABSTRACT

Master's thesis contains 88 pp., 31 fig., 54 sources.

The object of study - the process of pattern recognition in photography by means of artificial intelligence.

The subject of research - models and methods of artificial intelligence to generate a structural model of the face.

The aim of the work is to increase the efficiency of specific person face recognition from a photographic image through generating a structural model of found faces with means of artificial intelligence by developing a software application for detecting faces in image.

To achieve this goal it is necessary to solve the following tasks: to conduct a critical analysis of existing methods, models, software and hardware systems for image recognition in the photo image; to develop a method and model of a neural network for detection and identification of faces in a photographic image; develop a software application for detection, recognition and generation of a structural model of faces by photographic image.

Scientific novelty. The method of searching for human faces and key points on a two-dimensional image has been improved, which, unlike existing ones, uses convolutional neural networks to build a structural model of the face, which makes it possible to increase the accuracy of human identification by photographic image.

The practical significance of the results obtained. As a result of the work, a software application was developed for the detection and construction of a structural model of the face, and on the basis of these data - the identification of a person by the photographs available in the database.

**STRUCTURAL FACE MODEL, CONVOLVED NEURAL NETWORKS,
PATTERN RECOGNITION, ARTIFICIAL INTELLIGENCE**

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ	8
ВСТУП.....	9
1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ	12
1.1 Постановка мети й завдань дослідження.....	12
1.2 Засоби автоматизованої детекції обличчя	12
1.1.1 Основні труднощі при розпізнаванні обличчя на фотозображенні	14
1.3 Висновки до розділу 1	20
2 ДОСЛІДЖЕННЯ МЕТОДІВ РОЗПІЗНАВАННЯ ОБРАЗІВ.....	21
2.1 Нейронні мережі.....	21
2.1.1 Навчання штучних нейронних мереж з вчителем	21
2.1.2 Згорткові нейронні мережі	36
2.1.3 Функції активації в навчанні нейронної мережі	41
2.2 Аналіз методів розпізнавання обличчя	45
2.2.1 Структура системи розпізнавання обличчя.....	45
2.2.2 Метод головних компонент	46
2.2.3 РСА з використанням штучних нейронних мереж.....	50
2.2.4 Метод Deep Dense Face Detector	51
2.2.5 Метод Radial Basis Function Neural Networks.....	51
2.2.6 Згорткова нейронна мережа «Cascade».....	52
2.2.7 Метод Bilinear CNN	52
2.2.8 Комбінація методів Back Propagation Network та Radial Basis Function Network для розпізнавання осіб.....	53
2.2.9 Метод Retinal Connected Neural Network (RCNN)	54
2.2.10 Метод Rotation Invariant Neural Network (RINN)	54
2.3 Висновки до розділу 2	55
3 РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ДЕТЕКЦІЇ ТА ГЕНЕРАЦІЇ СТРУКТУРНОЇ МОДЕЛІ ОБЛИЧЧЯ	57
3.1 Опис роботи програмного додатку	57
3.2 Детекція обличчя на зображенні	58
3.3 Генерація структурної моделі	73

3.4 Співставлення обраного зображення у базі даних.....	75
3.5 Висновки до розділу 3	80
ВИСНОВКИ.....	82
ПЕРЕЛІК ПОСИЛАНЬ	83

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

AFR (англ. - Automated face recognition) – автоматизоване розпізнавання облич.

ANN (англ. Artificial Neural Network) – штучна нейронна мережа.

BPН (англ. - Back Propagation Network) – мережа оберненого розповсюдження похибки.

CNN (англ. Convolutional Neural Network) – згорткова нейронна мережа.

DAG (англ. - directed acyclic graph) – орієнтований ациклічний граф.

FLE (англ. - face landmark estimation) – визначення розмітки обличчя.

HOG (англ. - Histogram of Oriented Gradients) – гістограма орієнтованих градієнтів.

NMS (англ. - non-maximum suppression) – алгоритм подавлення не максимумів.

PCA – метод головних компонент.

RBF (англ. - Radial Basis Function Network) – радіально базисна мережа.

ReLU (англ. - Rectified Linear Unit) – випрямлений лінійний елемент.

RPN (англ. - Region proposal network) – мережа пропозицій областей.

Tanh – гіперболічний тангенс.

ВСТУП

Сучасні детектори облич можуть легко знайти та обробити фронтальні зображення. Нещодавні дослідження у цій області фокусуються на менш контрольованому процесі детекції, де число можливих факторів, таких, як зміна пози, вираження емоцій і незвичайне освітлення може призвести до великих візуальних варіацій у вигляді обличчя, що значно погіршує якість роботи детектора облич. Найбільш важкою задачею комп'ютерного зору залишається проблема вирішення неоднозначності, що виникає при перенесенні трьовимірних об'єктів реального простору на плоскі зображення. Дуже багато аспектів залежить від важкопрогнозованих фактів.

Складність у детекції облич в основному диктується двома аспектами: велика кількість візуальних змін облич людини на різноманітному фоні; розмір простору пошуку можливих позицій облич та їх розмірів є занадто великим. Перший потребує вирішувати проблему класифікації, другий - проблеми с потребами у швидкодії.

Механізм детекції облич на зображеннях знаходить використання у різноманітних додатках, а задача - одна з основоположних у комп'ютерному зорі. Область використання систем розпізнавання осіб, вже не обмежується верифікацією особи і спостереженням. Все більше додатків використовують розпізнавання осіб як перший крок до інтерпретації дій людини, його намірів і поведінки. Багато дій та особливості поведінки людини піддаються інтерпретації тільки в тому випадку, якщо ідентифікують його особу і людей, які її оточують. Прикладами можуть слугувати розпізнавання постійного покупця магазину, спостереження за поведінкою пацієнтів, інтерфейси командного управління на військових і промислових об'єктах.

З часів фундаментальної роботи Віюли та Джонса [1], прискорений каскад з простими ознаками залишається найбільш популярним та ефективним підходом у розробці програм для практичної детекції облич. Проста природа

ознак дозволяє швидко оцінювати і вчасно відкидати невірні результати пошуку. Тим часом, прискорений каскад створює групу простих ознак для досягнення точної класифікації облич з іншими об'єктами. Оригінальний детектор Віолі-Джонса використовує ознаки Хаара, які швидко вираховуються, та достатні для опису фронтальних зображень облич. Тим не менш, через простоту ознак Хаара, алгоритм відносно слабкий у неконтрольованому середовищі.

Згодом, у 2005 році, з'явився більш універсальний і зручний метод, що базується на HOG-дескрипторах [3]. Метод демонструє гарні результати і активно використовується для детекції облич. Зокрема алгоритми пошуку облич у сучасних камерах допомагають краще визначити область фокусу.

В останні роки спостерігається сплеск уваги до нейронних мереж. Посилений інтерес виник у 2012 році, коли Алекс Крижевський за допомогою згорткових нейронних мереж переміг у конкурсі ImageNet [3], понизивши рекорд помилок у класифікації з 26% до 15%, що тоді стало проривом. Сьогодні «глибоке навчання» стоїть в основі багатьох систем великих компаній: Facebook використовує нейронні мережі для алгоритмів автоматичного виставлення тегів, Google — для пошуку серед фотографій користувача, Amazon — для генерації рекомендацій товарів, Pinterest — для персоналізації домашньої сторінки користувача, а Instagram — для пошукової інфраструктури.

Фокус на генерації структурної моделі пояснюється великою кількістю можливостей використання цієї інформації. Ця проблема є більш загальною та важливою в контексті обробки зображень облич людини. Інформація може бути використана як для розпізнавання облич, емоцій людини та інших ознак обличчя, так і для обробки зображення.

Отже, актуальною науково-прикладною задачею є створення підходу для розпізнавання людини шляхом пошуку облич людини та ключових точок на ньому на двовимірному зображенні.

Об'єктом дослідження – процес розпізнавання образів на фотозображенні засобами штучного інтелекту.

Предмет дослідження – моделі та методи штучного інтелекту для генерації структурної моделі обличчя.

Метою роботи є підвищення ефективності розпізнавання особи певної людини по фотозображенню шляхом генерації структурної моделі знайдених облич засобами штучного інтелекту за рахунок розроблення програмного додатку для детекції облич на зображенні.

Для досягнення поставленої мети необхідно розв'язати такі задачі:

- провести критичний аналіз існуючих методів, моделей і програмно-апаратних систем для розпізнавання образів на фотозображенні;
- розробити метод та модель нейронної мережі для детекції та ідентифікації облич на фотозображенні;
- розробити програмний додаток для детекції, розпізнавання та генерації структурної моделі облич по фотозображенню.

Методи досліджень. У роботі було використано методи штучного інтелекту та методи розпізнавання образів, методи моделювання для розроблення моделі нейронної мережі.

Наукова новизна. Удосконалено метод пошуку облич людини та ключових точок на ньому на двовимірному зображенні який на відміну від існуючих використовує згорткові нейронні мережи для побудови структурної моделі обличчя, що дає можливість підвищити точність ідентифікації людини по фотозображенню.

Практична значимість отриманих результатів. В результаті роботи був розроблений програмний додаток для детекції та побудови структурної моделі обличчя, а на основі цих даних - ідентифікації особи по фотозображеннях, що наявні в базі даних.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

1.1 Постановка мети й завдань дослідження

Актуальною науково-прикладною задачею є створення підходу для розпізнавання людини шляхом пошуку облич людини та ключових точок на ньому на двовимірному зображенні.

Метою роботи є підвищення ефективності розпізнавання особи певної людини по фотозображенню шляхом генерації структурної моделі знайдених облич засобами штучного інтелекту за рахунок розроблення програмного додатку для детекції облич на зображенні.

Для досягнення поставленої мети необхідно розв'язати такі задачі:

- провести критичний аналіз існуючих методів, моделей і програмно-апаратних системи для розпізнавання образів на фотозображенні;
- розробити метод та модель нейронної мережі для детекції та ідентифікації облич на фотозображенні;
- розробити програмний додаток для детекції, розпізнавання та генерації структурної моделі облич по фотозображенню.

1.2 Засоби автоматизованої детекції обличчя

Автоматизоване розпізнавання обличчя отримало велику кількість уваги від дослідників та представників індустрії протягом декількох останніх поколінь як через великий спектр наукових викликів [5], так і через велику кількість способів застосування у комерційних додатках [6], частково у контексті біометрії та питань безпеки. З недавніх пір технології знайшли використання у сфері мультимедіа та соціальних медіа [4].

Розпізнавання облич це область, що намагається дати відповідь на питання: «Кому належить це обличчя?». Для цієї цілі люди мають природні

здібності за допомогою власних перцептивних і когнітивних систем, тоді як машини потребують комплексні системи, що включають в себе передові алгоритми та великі бази даних облич. Вивчення, розробка та проектування таких методів і технологій - це проблеми області автоматизованого розпізнавання облич.

Автоматизоване розпізнавання облич може згодом стати частиною систем комп'ютерної автоматизованої ідентифікації облич та комп'ютерної автоматизованої верифікації облич. З одного боку, автоматизоване ідентифікації обличчя полягає у один-до-багатьох (1:N) пошуку обличчя серед бази даних, що містить зображення облич багатьох різних людей для відповіді на питання «Чи відоме це обличчя?» [7]. З іншого боку, автоматизована верифікація - це пошук один-до-одного для пошуку відповіді на питання «Чи належить це обличчя до ...?» [8].

Окрім того, автоматизоване розпізнавання облич може бути базисом для рішення питання «Хто знаходиться на зображенні?», що допоможе вирішити проблему автоматизованого виставлення міток облич/найменування облич [9].

Загалом процес автоматизованого розпізнавання облич зображений на рисунку 1.1. Спочатку вирішується питання «Чи присутнє обличчя на зображенні?» (детекція облич). Наступне питання - «Де знаходиться обличчя?». Після цього виконується процес автоматизованого розпізнавання [10].



Рисунок 1.1 - Загальний процес автоматизованої детекції обличчя [4]

1.1.1 Основні труднощі при розпізнаванні обличчя на фотозображенні

Рух голови, що може бути описаний у формі егоцентричного кута оберту, або зміна точку огляду камери може призвести до суттєвих змін у вигляді або формі та створити велику кількість варіацій, що проілюстровані на рисунку 1.2, що робить розпізнавання обличчя у позах важкою задачею.

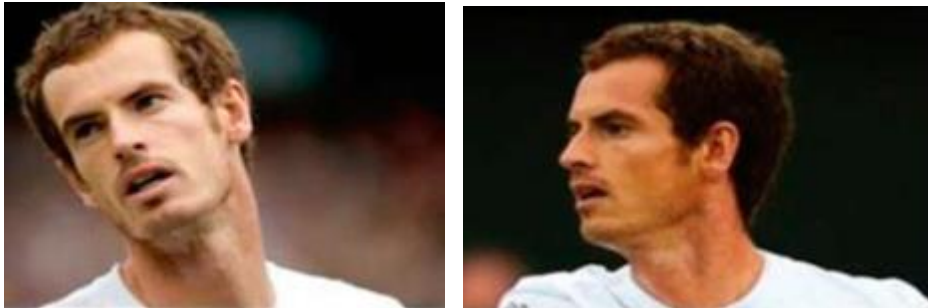


Рисунок 1.2 - Ілюстрація можливих поз голови в залежності від площини нахилу [4]

Оскільки автоматизоване розпізнавання обличчя чутливе до варіацій, корекція пози є важливим елементом і може бути досягнуте використанням різноманітних технік націлених на вирівнювання обличчя [11].

Різнманітність варіантів одного обличчя також може виникати через відсутність певних структурних елементів, або ж через наявність таких компонентів як борода чи вуса, головних уборів, сонцезахисних окулярів, і т.д., або ж через оклюзій лиця елементами фону або переднього плану.



Рисунок 1.3 - Ілюстрація відсутності (a) або наявності (b-d) структурних елементів, а саме бороди та вус(б), кепки(с), сонцезахистних окулярів (d) та часткової оклюзії (e) [4]

Деякі варіації у вигляді обличчя спричинені змінами через вираження емоцій обличчям у варійованих емоційних станах, що зображені на рисунку 1.4.

Ефективне автоматичне розпізнавання різних виразів обличчя важливе як для еволюції емоційних станів, так і для автоматизованого розпізнавання обличчя. Загалом, людські вирази обличчя складаються з макро-виразів, що можуть виражати злість, страх, сум, радість, здивування та інші швидкі вирази обличчя, тобто мікро-вирази [4]. Усі ці вирази створюють рух обличчя. Так динаміка обличчя може бути вирахована за допомогою методів optical flow [12].



Рисунок 1.4 - Ілюстрація змін обличчя при вираженні емоцій: злість (a), відраза (b), сум (c) та щастя (d) [4]

Інша причина зміни вигляду обличчя полягає у старінні людського обличчя, що може мати вплив на процес автоматичного розпізнавання обличчя, якщо час поміж кожним зображенням є суттєвим [13].



Рисунок 1.5 - Ілюстрація процесу людського старіння, де одна і та ж сама людина була сфотографована у різному віці [4]

Для вирішення проблеми старіння обличчя у автоматизованому розпізнаванні обличчя, методи мають обробляти патерни старіння обличчя. Окрім цього, з часом змінюються не лише форма та лінії обличчя, а також і інші аспекти, як зачіска, наприклад [14].

Великі зміни у освітленні можуть суттєво знизити якість роботи систем автоматизованого розпізнавання облич. Зокрема, за низького рівня освітлення переднього або заднього плану, детекція та розпізнавання облич відбуваються значно гірше, оскільки тіні на обличчі можуть зробити певні патерни обличчя, які не будуть відрізнятися. З іншого боку, занадто яскраве освітлення може призвести до переэкспонування фотографії і також зробити певні особливості [4].

Швидкі автоматизовані детекція та розпізнавання облич через занадто варійоване освітлення потребують використання таких технік обробки зображення як нормалізація освітлення через виведення гістограми, або методи машинного навчання, що працюють з дійсною загальною інтенсивністю зображення [15, 16].



Рисунок 1.6- Ілюстрація варіантів змін освітлення камери, що призводять до переэкспонування (a), глибоких тіней (b) або часткового затемнення (c)

Інші звичайні фактори, що впливають на якість автоматизованого розпізнавання облич відносяться до якості та роздільної здатності зображення обличчя або до налаштувань можливостей цифрового спорядження для створення зображень. Для цієї цілі ISO/IEC 19794-5 стандарт [17] був

розроблений для специфікації сцени та фотографічних потреб у ролі формату зображення обличчя для автоматизованої системи розпізнавання облич, особливо в контексті біометрії. Однак, реальні ситуації отримання зображень людини включають використання різноманітного заліза для фотографування, що можуть бути ширококутними, мати специфічний нахил-поворот-масштаб, чутливі специфічні сенсори, фотометричне стерео. Камери можуть працювати у спектрі видимого світла, або у інфрачервоному діапазоні, що спричинює складнощі для систем автоматизованого розпізнавання облич. Обличчя отримані в умовах реального світу створюють нові виклики для систем AFR [18].



Рисунок 1.7 - Ілюстрація варіацій масштабу зображень та якості [4]

Наприклад, на рис. 1.7 в деяких ситуаціях обличчя може бути сфотографоване на відстані, що призведе до меншого розміру обличчя у порівнянні зображенням у великому масштабі. Також, деякі камери можуть мати малу роздільну здатність, що призведе до низької якості фотографії обличчя, за якої AFR дуже важко провести [19, 20]. Щоби розібратися з цими обмеженнями, були запропоновані алгоритми реконструкції зображень з низькою роздільною здатністю [21, 22].

Кожна AFR технологія потребує доступний, якісний і реалістичний набір даних облич для виконання 1:N та 1:1 пошуку облич. Якості як повнота(містить різноманітні пози, освітлення, вирази облич), точність (патерни для різного віку) і характеристики(різноманітні формати файлів, рівень кольорів/сірого, роздільна здатність, умови) є ключовими для процесу AFR. Також, при роботі з даними

облич, людські права та привітність мають бути враховані оскільки AFR системи мають бути сумісними з Data Protection Act 2010 [23].

Для дослідницьких цілей, декілька наборів даних були розроблені і є публічно доступними. Слід звернути увагу на наступні:

- ORL являє собою набір даних на 400 зображень з 40 різних об'єктів, PGM форматі і з розділенням 92 x 112, 8-бітного рівня сірого. Чоловічі і жіночі обличчя сфотографовані на однорідному темному фоні, при різних умовах освітлення. Суб'єкти нахиляють голову вгору-вправо, фронтально, з варіаціями в лицьових виразів, деталей особи і обертають в межах $\pm 20\%$ [24].

- The Face Recognition Technology (FERET) - база даних, що складається з 14,126 зображень облич 1199 людей, для кожного 5-11 зображень у градаціях сірого. Кожний набір містить зображення, що зняти під різним кутом з різними виразами обличчя, з використанням декількох камер та різних умов освітлення [25].

- BioID Face database містить 1521 фронтальних зображень облич 23 людей. Розмір зображень 384 x 286 пікселів у pgm форматі, зроблених у реальних умовах, тобто, присутнє різноманіття розмірів, освітлення та інших умов фотографії [26].

- Yale face database містить 165 gif зображень 15 індивідів. На кожну особу приходить 11 зображень, по одному на кожен вираз обличчя і конфігурацію світла ліворуч/праворуч/зверху, з та без окулярів [27].

- Caltech 10,000 web faces складається з 10,524 зображень облич людини у різних наборах (портрети, групові фото і т.д.) від сервісу Google Image. Координати очей, носу та центру роту для кожного фронтального зображення наведені для використання перевірки роботи алгоритмів розпізнавання облич, або для детекції облич перед AFR [28].

Деякі набори даних містять 2D та 3D інформацію про обличчя, такі як Face Recognition Grand Challenge (FRGC) датасет [29], що містить 50,000 контрольованих та не контрольованих зображень від 4003 суб'єктів.

Інші датасети містять декілька модальностей, як наприклад XM2VTSDB мульти-модальний датасет, що є розширенням M2VTS датасету. Це великий, мультимодальний набір зроблений за допомогою високоякісного відео. Він містить 4 записи, кожен з головою, що розмовляє, та головою, що обертається, 295 суб'єктів, збирався він протягом 4 місяців. Ця база включає в себе високоякісні фотографії, 32 kHz 16-bit звукові файли, відео послідовності та 3D моделі [4, 30].

Інша мультимодальна база даних - Surveillance Cameras Face (SCFace) датасет. Він містить записи 4160 статичних людських облич 130 суб'єктів, у видимому та інфрачервоному спектрі, у кімнатних умовах, знятих на набір камер з п'яти відеокамер, що записують різноманітні якості міміки в умовах реального світу [31].

Останні розробки датасетів фокусуються на зборі даних в реальних умовах, тобто невимушених умовах. Наприклад, Face Detection Data Set and Benchmark (FDDDB) це набір з 2845 зображень, кольорових та у градаціях сірого, з 5171 обличчями у реальних умовах, що включають в себе оклюзії, варіації поз, низьку роздільну здатність та погане фокусування на обличчях [32].

Labelled Faces in the Wild (LFW) датасет - це популярний набір даних для вивчення різнопланових облич. Він містить записи 13,233 зображень облич на передньому плані; Інші обличчя асимілюються до фону. Він містить 5749 окремих облич різних людей, що можуть мати одне або декілька фотографій на кожного, окрім того присутні варіації у присутніх позах, освітленні, виразах облич, фоні, расі, етнічній приналежності, віку, полу, одязі, типах зачісок, якості камер, типі освітлення, фокусі і так далі. Зображення 250 x 250 пікселів та у форматі jpeg, більшість - кольорові, лише декілька у відтінках сірого [33].

Деякі інші датасети зібрані для спеціальних цілей. Наприклад, Spontaneous Micro-expression датасет (SMIC) використовується для розпізнавання мікро-емоцій, коли Acted Facial Expression in the Wild (AFEW) датасет, що містить напівавтоматично зібрані зображення зіграних емоцій у фільмах, присвячений

для розпізнавання макро-емоцій у умовах близьких до реальних. З іншого боку, FG- NET Ageing database (FG-NET) може використовуватися для розпізнавання віку, розробки інваріантних до віку AFR та вивченню зміни при віці [34, 35].

1.3 Висновки до розділу 1

В першому розділі обґрунтовано актуальність теми досліджень, адже задача автоматичного розпізнавання обличчя на сьогоднішній день є досить актуальною, як через велику кількість наукових досліджень даної задачі так і через великий потенціал використання даної технології у комерційних проектах.

Однак дана технологія має деякі складності, які не дають можливості застосовувати дану технологію на 100%: варіації пози людини (нахил або поворот голови; наявність/відсутність певних елементів/оклюзій(перекрить); зміни емоцій на обличчі; старіння обличчя; варійовані умови освітлення; роздільна здатність зображення.

Зроблено постановку мети й задач дослідження.

2 ДОСЛІДЖЕННЯ МЕТОДІВ РОЗПІЗНАВАННЯ ОБРАЗІВ

2.1 Нейронні мережі

2.1.1 Навчання штучних нейронних мереж з вчителем

Один з основних підходів, найбільш широко використовувався в області розпізнавання зображень, являє собою застосування класичних моделей-класифікаторів, що навчаються з учителем. Для навчання таких моделей використовуються маркована вибірка даних, що складається з масиву зображень і відповідного їм масиву міток, що визначають категорію, до якої відноситься зображення. В процесі навчання масив даних розділяється на дві нерівні частини - навчальну вибірку і тестову вибірку, потім за допомогою специфічного для конкретного алгоритму правила навчання параметри моделі налаштовуються з використанням навчальної вибірки таким чином, щоб отримавши в якості вхідних даних зображення, модель на виході виробляла б мітку відповідного класу. Цей підхід представлений безліччю моделей, серед яких найбільш широко використовуваними є регресивна модель, штучна нейронна мережа (багатошаровий перцептрон), метод опорних векторів, а також дерева прийняття рішень і моделі-ансамблі, що представляють собою поєднання деяких перерахованих моделей [36]

Спроби відтворити здатність навчатися і виправляти помилки призвели до створення штучних нейронних мереж. Штучні нейронні мережі являють собою сімейство моделей, побудованих по принципу організації і функціонування біологічних нейронних мереж — мереж нервових клітин живого організму. Поняття штучної нейронної мережі було запропоновано ще у 1943 році У. Маккалоком і У. Піттсом в статті [37]. Зокрема, ними була запропонована модель штучного нейрона. Щоб відобразити суть біологічних нейронних систем, штучний нейрон будується наступним чином. Він отримує вхідні сигнали

(вихідні дані або вихідні сигнали інших нейронів нейронної мережі) через кілька вхідних каналів. Кожен вхідний сигнал проходить через з'єднання, що має певне політичне значення. З кожним нейроном пов'язане значне порогове значення. Обчислюється зважена сума входів, з неї віднімається граничне значення і в результаті виходить величина активації нейрона. Сигнал активації перетворюється за допомогою функції активації і в результаті виходить вихідний сигнал нейрона.

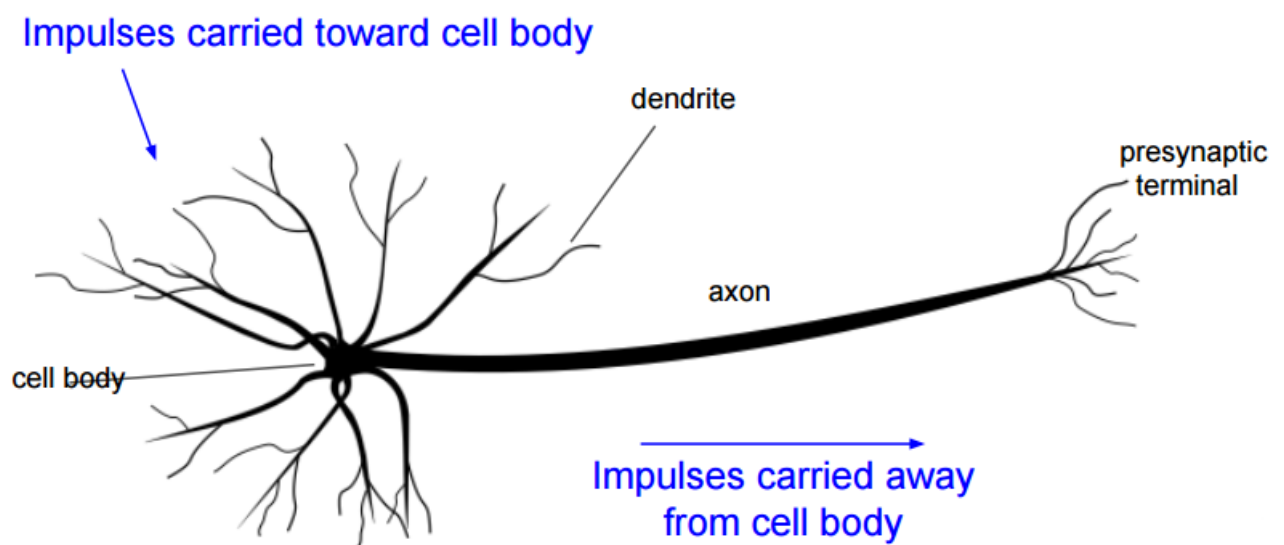


Рисунок 2.1 - Модель нейрону живого організму [53]

Багатошарові перцептрони, які навчаються методом зворотного поширення помилки, широко використовуються для розпізнавання різних категорій зображень, таких як рукописні цифри [16], почерк [39], людські обличчя [40] і дані зорових сенсорів робототехнічних систем [41]. Модель багатошарового перцептрона являє собою сукупність штучних нейронів - обчислювальної одиниці моделі - об'єднаних в рівні (шари), задані в ієрархічному порядку [36].

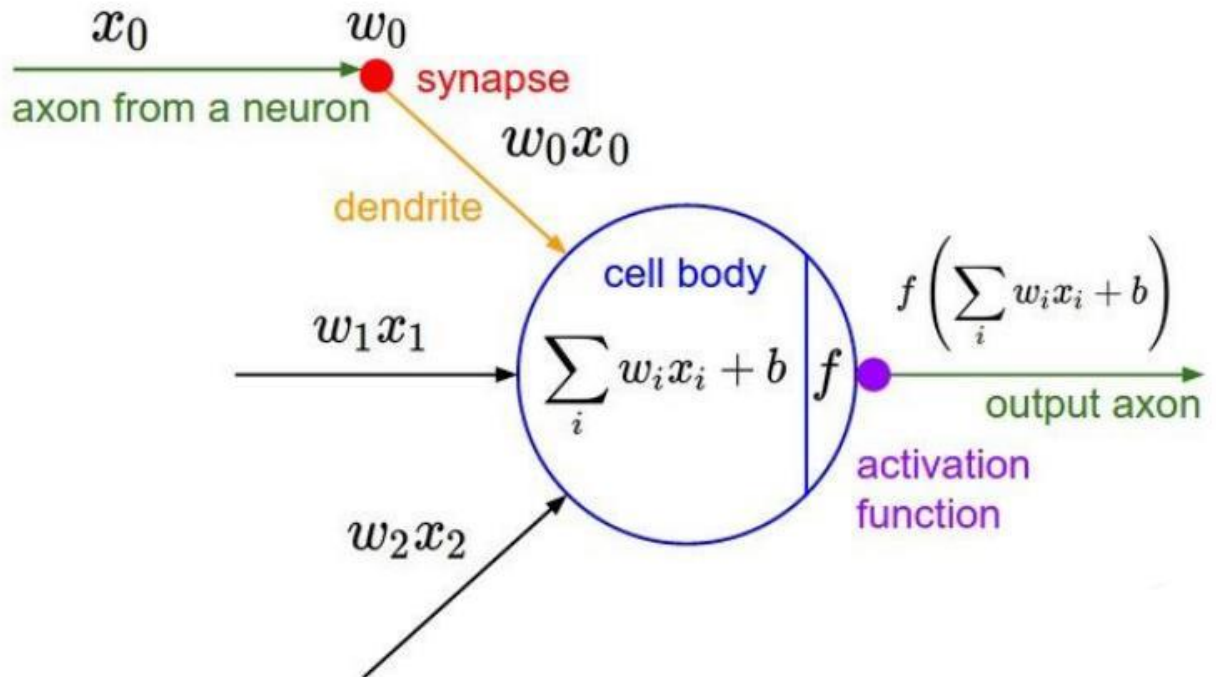


Рисунок.2.2 - Модель штучного нейрону, x_i - вхідний сигнал, w_i - вага вхідного сигналу, $f(\cdot)$ - функція активації [53]

Поведінка нейрона будується наступним чином: нехай є $(m + 1)$ входів, значення яких дорівнюють $x_0, x_1, x_2, \dots, x_m$, а значення їх ваг рівні, w_0, w_1, \dots, w_m , при цьому перший вхідний елемент, як правило, являє собою фіксоване значення зміщення $x_0 = 1$. Тоді вихідне значення нейрона є значення функції активації від зваженої суми його вхідних значень:

$$y = \varphi \sum_{i=0}^m w_i x_i .$$

При об'єднанні штучних нейронів в мережу вхідні значення нейрона шару l являють собою вихідні значення нейронів попереднього шару $(l - 1)$. При цьому нейрони першого (вхідного) шару отримують в якості вхідних значення безпосередньо дані, що підлягають розпізнаванню, які в разі розпізнавання зображення представляють собою значення інтенсивності складових його

пікселів (точкових елементів). Вихідний шар мережі може варіюватися в залежності від завдання, але класична архітектура має на увазі формування його числом нейронів, рівній кількості класів розпізнавання, при цьому вихідне значення кожного нейрона нормується по інтервалу $\{0,1\}$, і являє собою ймовірність приналежності вхідного зображення до відповідного класу. Як відзначають дослідники, такі багатошарові нейронні мережі здатні інкапсулювати будь-яку математичну функцію за допомогою довільного набору нейронів [36].

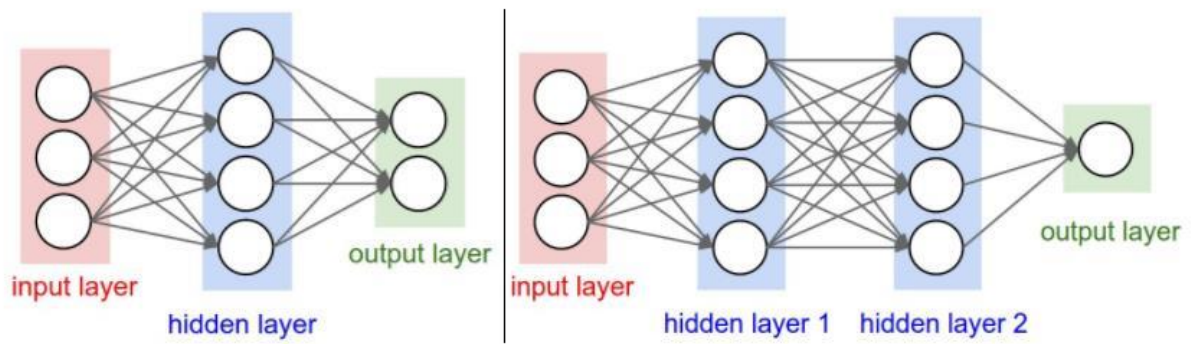


Рисунок 2.3 - Ліворуч: двошарова нейронна мережа, один прихований шар з чотирьох нейронів. Праворуч: 3-шарова мережа з двома прихованими шарами [53]

Навчання мережі методом зворотного поширення полягає в наступному: нехай є деяка невідома функція розпізнавання $g : X - Y$, аргументом котрої є зображення $x_n \in X$, представлені у вигляді вектора довжиною n , а значеннями функції - множина класів (категорій) $y \in Y$. Навчальна вибірка складає собою підмножину значень цієї функції $D = \{(x_o, y_o), (x_1, y_1), \dots, (x_m, y_m)\}$ [36].

Задача навчання моделі розпізнавання полягає у пошуку такої функції $h: X - Y$, яка б апроксимувала функцію g на всій області її визначення, в тому числі в значеннях, що не включені в D , і представляють собою додаток теорії оптимізації [36].

Оскільки сформулювати аналітично правило класифікації зображень за категоріями розпізнавання часто представляється скрутним, здатність навчатися на базі вибірки робить нейронні мережі та споріднені з ними моделі придатними для розпізнавання природних зображень навколишнього світу, що відрізняються нечіткою структурою і безліччю варіацій в межах класу [36].

Нехай $h(x)$ - вихідне значення мережі, отримане шляхом послідовної активації нейронів кожного шару, а $g(x)$ - значення апроксимованої функції для одного й того ж зображення. Далі виконується крок зворотнього поширення, що полягає у вирахуванні часткової похідної для кожного нейрону мережі по відношенню до його вагових коефіцієнтів [36]:

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial o_i} \frac{\partial o_i}{\partial net_j} \frac{\partial net_j}{\partial w_{ij}},$$

$$E = \frac{1}{2} (h(x) - g(x))^2,$$

де E - середньоквадратичне відхилення мережі. Далі на кожному кроці навчання ваги нейронів інкрементуються значеннями часткових похідних в відповідності до методу градієнтного спуску, або інших градієнтних методів навчання. Варіації навчального алгоритму включають в себе використання додаткових параметрів регуляризації з метою захисту від перенавчання, і використання різних оптимізаторів - методу Ньютонна, методу імітації відпалу, L-BFGS і інших [36].

Нейронна мережа - це дуже потужна та швидка техніка класифікації що може бути використана не лише для передбачення відомих даних, а й для невідомої інформації. Вона гарно працює як для лінійно-розділимих даних, так і для лінійно-нерозділних. Нейронні мережі використовуються у багатьох областях [36]. Наведені вище переваги нейромережевої обробки даних визначають сфери їх застосування:

- обробка і аналіз зображень;

- розпізнавання мови незалежно від диктора;
- обробка високошвидкісних цифрових потоків;
- автоматизована система швидкого пошуку інформації;
- класифікація інформації в реальному масштабі часу;
- планування, застосування сил і засобів у великих масштабах;
- вирішення трудомістких задач оптимізації;
- адаптивне управління і передбачення.

В наш час основними напрямками реалізації мереж є:

- програмна реалізація на цифрових ЕОМ традиційної архітектури;
- програмно-апаратна реалізація у вигляді співпроцесорів до ЕОМ загального призначення;
- апаратна реалізація шляхом створення нейрокомп'ютерів на базі нейроплат у вигляді паралельних нейроподібних структур.

Ранні варіанти реалізації нейронних мереж відносяться до перших двох із вказаних напрямів. Перший напрям характеризується універсальністю, дешевизною і низькою швидкістю навчання і функціонування нейронних мереж. Для другого напрямку характерна висока швидкість моделювання функціонування мереж, але при цьому існують серйозні фізичні обмеження числа модельованих елементів і зв'язків між ними, а також можливостей навчання і донавчання. Із розвитком елементної бази ЕОМ став можливим самостійний розвиток третього напрямку, який поклав початок індустрії нейрокомп'ютерів, що подають сукупність апаратних і програмних засобів для реалізації моделей нейронних мереж [36].

2.1.2 Згорткові нейронні мережі

Проблеми, що виникли в процесі використання моделей, які формують цілісні репрезентації, сприяли розвитку нової групи алгоритмів, що використовують локальні ознаки зображень. Необхідність такого підходу була продиктована властивістю стаціонарності природних зображень - об'єкти,

присутні на зображенні, могли вільно переміщатися в межах поля зору, при цьому бажаним результатом розпізнає алгоритму залишалось співвіднесення безлічі таких інваріантних репрезентацій об'єкта до одного класу [36].

Крім іншого, використання локальних ознак при розпізнаванні зображень було підкріплено свідченнями з області нейробіології. У класичній праці Д. Хьюбела і Т. Візела [42], що візуальна кора головного мозку являє собою складний комплекс клітин, кожна з яких чутлива тільки до обмеженого ділянки поля зору. Такі ділянки, інакше звані рецептивних полями, стикуються разом, забезпечуючи перекриття всього поля зору. Відповідні клітини при цьому виконують роль локальних фільтрів вхідних даних, реагуючи на присутність у власному рецептивної поле деяких примітивних структур, таких як краю і кордони. Було виявлено також існування так званих «складних клітин», мають більш широкі рецептивні поля, і демонстрували інваріантність по відношенню до точного розташування об'єкта в полі зору [36].

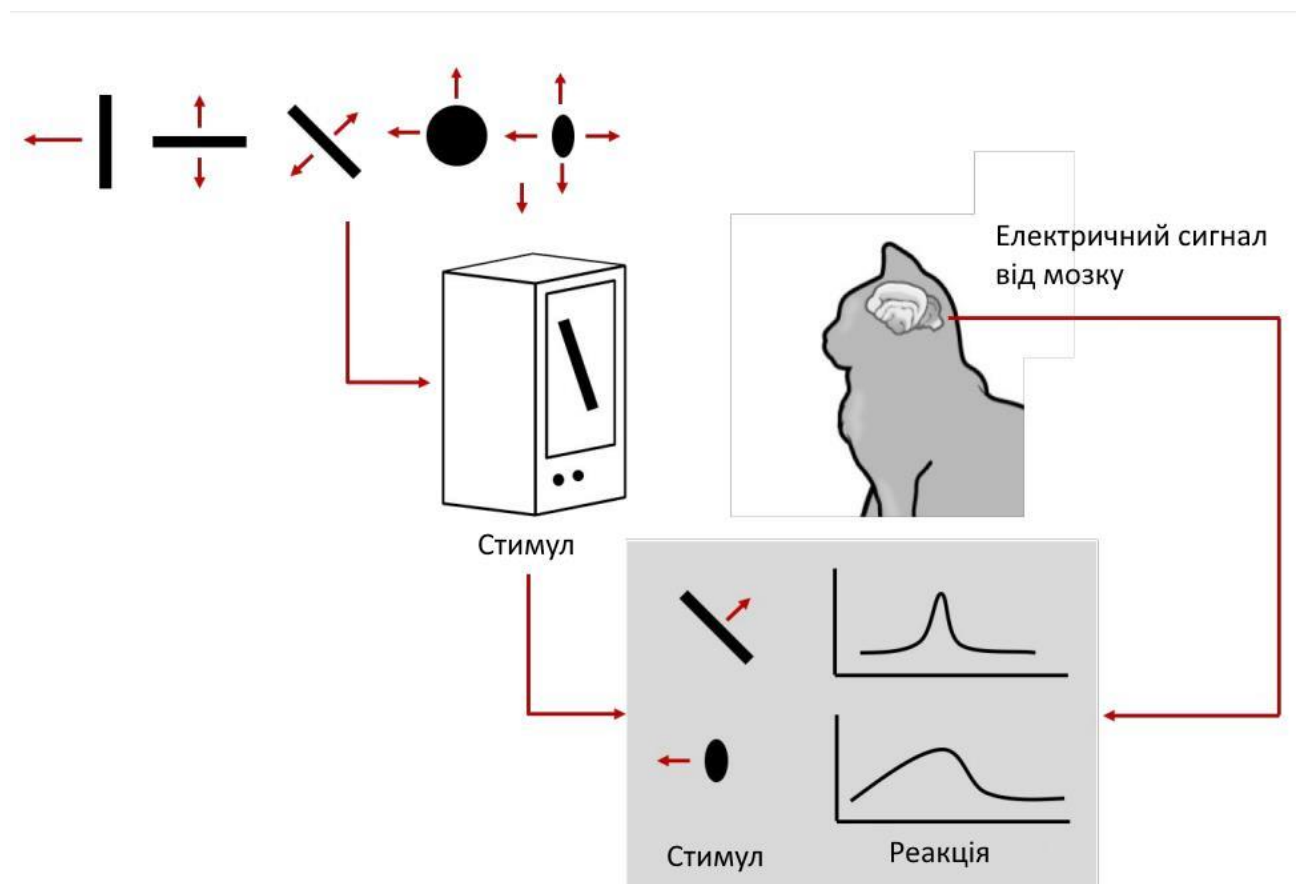


Рисунок 2.4 - Експеримент Хьюбела і Візела

З урахуванням того, що візуальна кора головного мозку являє собою найбільш потужну і гнучку зорову систему з існуючих на даний момент, поява моделей [43,44,45], емулюючих її поведінку, виглядало природним кроком. Однією з найбільш успішних моделей, що вважається визнаним лідером [46] в області розпізнавання зображень, є згорткова нейронна мережа [36]. Згорткові мережі являють собою варіацію архітектури багатошарового перцептрона, і включають в себе згорткові шари, шари підвибірки (субдискретизація), і повнозв'язні шари. Архітектура згорткової мережі використовує переваги двовимірної структури вхідних даних - зображень за допомогою методу локальної зв'язності, обмежуючи кількість зв'язків між нейронами прихованого конвуляційного шару і вхідними даними. Конкретно, кожен нейрон прихованого шару пов'язаний тільки з обмеженою локальною ділянкою зображення [36].

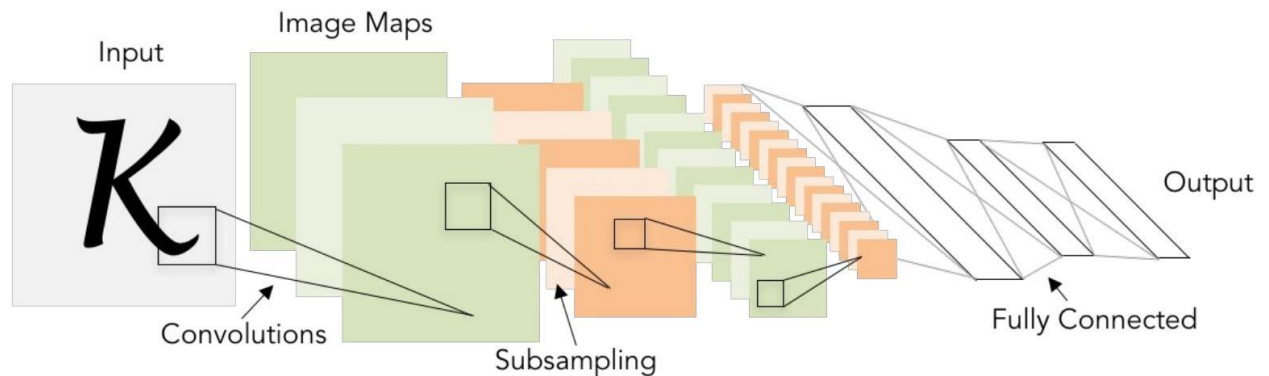


Рисунок 2.5- Архітектура згорткової нейронної мережі [53]

Крім цього, нейронна мережа використовує загальні, або розподілені вагові коефіцієнти, накладаючи штучне обмеження на алгоритм навчання зворотним поширенням помилки, так, щоб кожен нейрон прихованого шару мав набір ваг, спільний з іншими нейронами цього шару. При прямому поширенні така мережа здійснює математичну операцію згортки вхідного зображення

набором фільтрів, які подаються вагами нейронів прихованого шару. Проміжними результатами мережі є так звані «карти ознак» - двовимірні матриці, що представляють собою результат згортки окремим фільтром [36]. Шар субдискретизації виконує операцію групування мап ознак, розглядаючи регіони $p * p$ і агрегуючи значення, отримані в результаті згортки. Основне призначення субдискретизації – знизити варіативність даних, забезпечуючи стійкість к трансляціям локальної ознаки в рамках даного регіону. Таким чином, в умовах, коли одна і та сама ознака опиняється здвигнутою на деяке значення $(\Delta x, \Delta y)$ в межах p , відповідний нейрон, що інкапсулює локальну субдискретизовану ознаку зображення, буде все ще активним. Таким чином мережа забезпечує інваріантність до просторових спотворень. В якості агрегуючої функції шару субдискретизації зазвичай розглядається функція вирахування середнього чи максимального значення [36].

Таким чином, значення за рахунок використання декількох поперемінних шарів згортки і субдискретизації згорткова нейронна мережа дозволяє отримувати уявлення, незалежні від конкретного розташування локального ознаки в зображенні, і однаково чиним реагувати на об'єкти, що цікавлять (наприклад, людські обличчя), присутні на будь-якій ділянці фотографії [36].

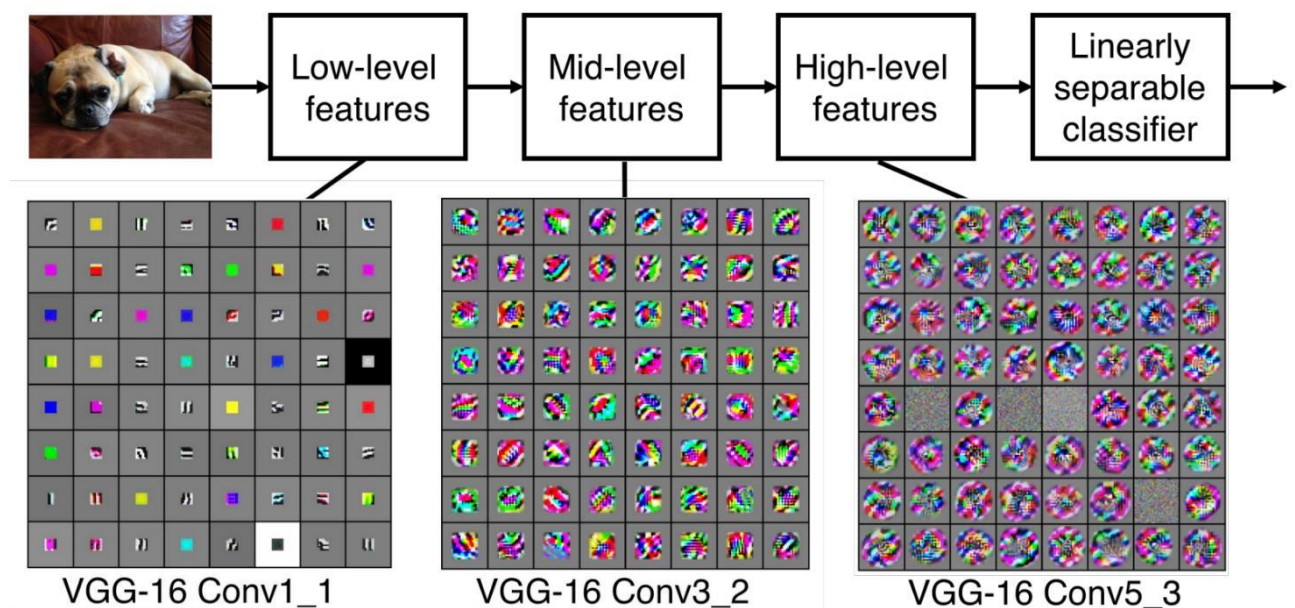


Рисунок 2.6 - Візуалізація активацій фільтрів згорткової нейронної мережі[53]

Згорткові мережі - один з найбільш успішних існуючих на сьогоднішній день підходів до розпізнавання зображень. Моделям, які реалізують відповідну архітектуру, належать перші місця в змаганнях алгоритмів розпізнавання, таких як ImageNet: згорткові мережі розпізнають рукописні цифри вибірки MNIST з 0.23% помилок [47], людські обличчя з 2.4% помилок [48] (використовувалася вибірка розміром в 5600 фотографій облич більш ніж 10 осіб); результат змагання ImageNet для переможця 2014 року - згорткової мережі GoogLeNet компанії Google становить 0.06656% помилкових розпізнавань [49], що являє собою найкращий результат на поточний момент і можна порівняти з помилкою, що допускається людиною на вибірці ImageNet. Серед недоліків виділяють труднощі при обробці маленьких об'єктів, і нездатність справлятися з спотвореннями, такими як розмиття або сильний шум (такі спотворення присутні в навколишньому світі, наприклад, при погляді через товсте скло). При цьому згорткові мережі порівняно легко справляються з проблемами високоточного розпізнавання, які викликають труднощі у людей - наприклад, розпізнавання окремих моделей машин або порід собак, і інші завдання, що вимагають виділення вузькоспеціфічних ознак [36].

Операція згортки забезпечує інваріантність по відношенню до трансляції локальних ознак (здвигу по осях x і y). Архітектура згорткових мереж не передбачає стійкості до інших афінних перетворень, таким як обертання, дзеркальне відображення і масштабування. Для вирішення цієї проблеми, як правило, використовуються евристичні методи (вирівнювання зображення по лінії горизонту, використання просторових пірамід і різних відображених копій оригіналу) [50]. Перетворення, які не є афінними по відношенню до плоского зображення, такі як обертання об'єкта в тривимірній площині, аналогічним чином не можуть бути оброблені за допомогою операції згортки. Для об'єктів, представлених в різних орієнтаціях, моделі доводиться навчатися різним,

непересічним набором ознак [51]. Згорткові нейронні мережі та інші підходи, які використовують локальні інваріантні ознаки, в процесі розпізнавання періодично вимушено навчаються ознакам, максимізуючим інваріантність в рамках представленої вибірки, але не забезпечують дискримінативного вигляду об'єкта в цілому [52].

Альтернативні підходи до виділення локальних ознак включають в себе методи класичного комп'ютерного зору, які не використовують навчальні моделі. Ці методи здійснюють пошук на зображенні характерних ділянок, що відповідають алгоритмічно явно заданим умовам.

2.1.3 Функції активації в навчанні нейронної мережі

Кожна функція активації додає нелінійність в навчанні нейронної мережі. Вона бере окреме число, що є результатом роботи попереднього шару і виконує фіксовану математичну операцію з ним таблиця 2.1.

Таблиця 2.1 - Класифікація функцій активацій [54]

Назва	Формула	Область значень
Порогова	$\Psi = F(\psi) = \begin{cases} 0, & \psi < \theta, \\ 1, & \psi \geq \theta. \end{cases}$	$(0, 1)$
Знакова	$\Psi = F(\psi) = \begin{cases} 1, & \psi > 0, \\ -1, & \psi \leq 0. \end{cases}$	$(-1, 1)$
Сигмоїдна	$\Psi = F(\psi) = \frac{1}{1 + e^{-\psi}}$	$(0, 1)$
ReLU	$\Psi = F(\psi) = \frac{1}{1 + e^{-\psi}}$	$(0, \infty)$
Лінійна	$F(\psi) = \psi$	$(-\infty, \infty)$
Радіальна базисна	$\Psi = F(\psi) = e^{-\psi^2}$	$(0, 1)$

Закінчення таблиці 2.1

Назва	Формула	Область значень
Напівлінійна	$\Psi = F(\psi) = \begin{cases} 0, \psi \leq 0, \\ \psi, 0 < \psi < 1, \\ 1, \psi \geq 1. \end{cases}$	$(0,1)$
Лінійна з насиченням	$\Psi = F(\psi) = \begin{cases} -1, \psi \leq -1, \\ \psi, -1 < \psi < 1, \\ 1, \psi \geq 1. \end{cases}$	$(-1,1)$
Tanh	$\Psi = F(\psi) = \frac{e^{\psi} - e^{-\psi}}{e^{\psi} + e^{-\psi}}$	$(-1,1)$
Трикутна	$\Psi = F(\psi) = \begin{cases} 1 - \psi , \psi \leq 1, \\ 0, \psi > 1. \end{cases}$	$(0,1)$

Найбільш часто у практиці зустрічаються наступні:

Сигмоїдна. Суть її роботи полягає в тому, що вона бере значення і «втискає» його в межі від 0 до 1. По суті, великі негативні числа стають близькими до 0, а великі позитивні ближче до 1.

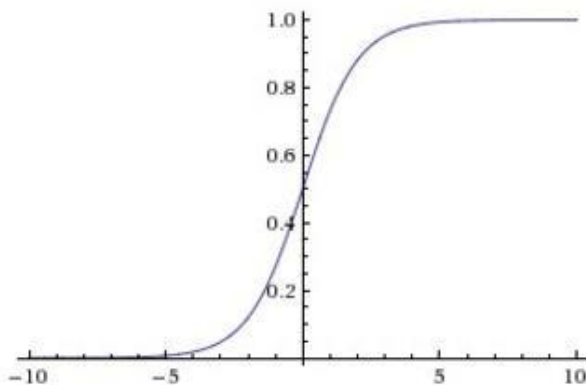


Рисунок 2.7 - Графік сигмоїдної функції [53]

Сигмоїдна функція історично часто використовується через подібність до рівня активації нейрона: від зовсім неактивного (0), до повністю активованого (1). На практиці, сигмоїдна нелінійність не так давно вийшла з моди і досить рідко використовується. Вона має два суттєві недоліки:

– сигмоїда подавляє та вбиває градієнти. Зовсім непотрібна особливість сигмоїдного нейрону полягає в тому, що коли активізація близька до обох «кінців» біля 1 або 0, градієнт у цих областях близький до 0. Під час зворотного розповсюдження похибки локальний градієнт буде помножений на вихідний. Таким чином, якщо локальний градієнт дуже малий, то він зробить і глобальний дуже малим, і розповсюдження помилки також буде малим, рекурсивно поширюючи цю проблему. Окрім того, слід слідкувати за ініціалізацією вагових коефіцієнтів.

– вихід сигмоїди не централізований відносно нуля. Це є проблемою, оскільки нейрони у глибших шарах нейронної мережі отриматимуть не також не центровані відносно нуля значення. Це матиме вплив на динаміку під час градієнтного спуску, тому що дані на вході нейрону будуть завжди позитивними, тоді градієнт вагових коефіцієнтів буде або цілком позитивним, або цілком негативним. Це призведе не зиг-загообразної динаміки оновлень градієнту. Хоча ця проблема може бути вирішеною використанням батчів, під час яких градієнти сумуються, а лише потім йде оновлення вагових коефіцієнтів, все одно це призводить до незгручностей [53].

Tanh. Tanh нелінійність «стискає» дійсне число у проміжок (-1,1). Подібно до сигмоїдної функції її активації демонструють насичення, але вона є центрованою відносно 0. На практиці завжди обирають Tanh у порівнянні з сигмоїдною функцією. Окрім того, Tanh фактично видозмінена сигмоїда.

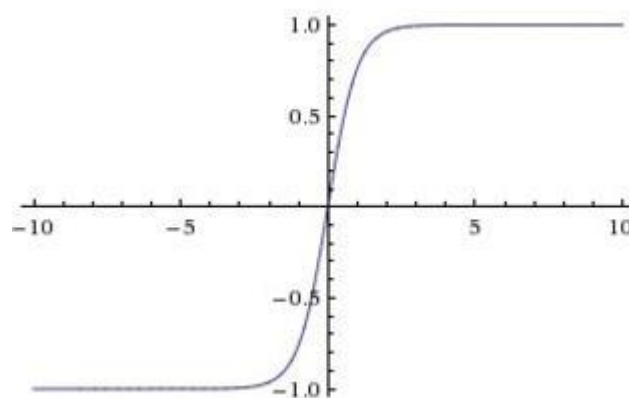


Рисунок 2.8 - Графік Tanh [53]

ReLU. Ця функція активації стала дуже популярною в останні роки. Фактично вона робить рівними нулю всі значення менше нуля. Вона має декілька переваг та недоліків:

- вона значно пришвидшує швидкість стохастичного градієнтного спуску у порівнянні з сигмоїдною або Tanh функціями [53];

- у порівнянні з сигмоїдною або Tanh функціями, що містять складні операції(експонента і т.д.), ReLU легко імплементується та швидко вираховується [53];

- нажаль, ReLU об'єкти можуть бути хрупкими і вбивати навчання. Наприклад, великий градієнт, що проходить через ReLU нейрон може спричинити оновлення вагових коефіцієнтів у такий спосіб, що нейрон не активується за жодного набору даних. Якщо це виникає, тоді градієнт, що проходить через цей нейрон завжди буде рівний нулю. Через цей ефект ви можете з'ясувати, що 40% вашої мережі не активізується, якщо ваша інтенсивність навчання буде занадто високою. За правильних налаштувань навчання ця проблема зустрічатиметься рідше [53].

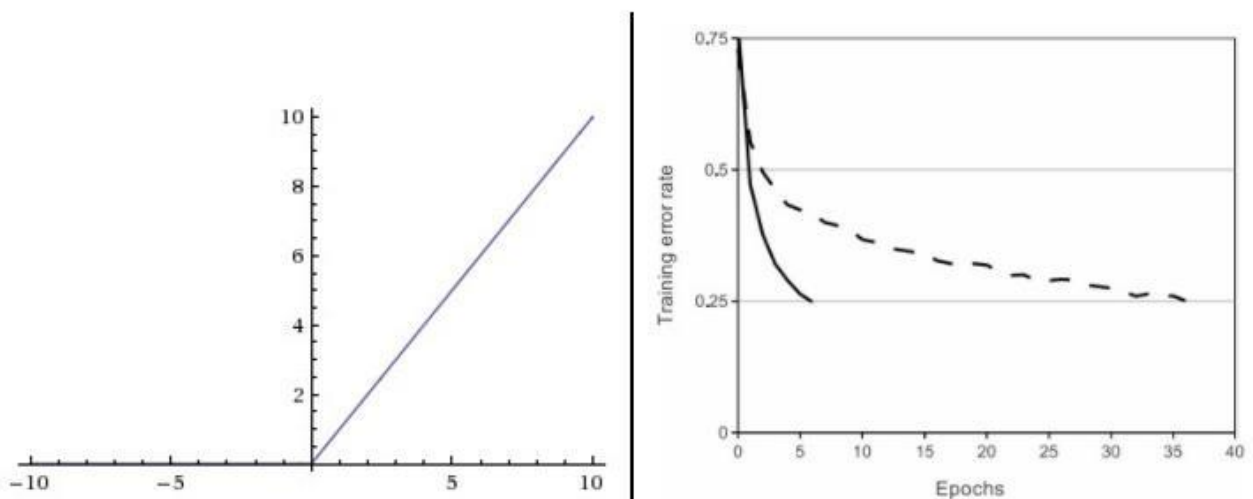


Рисунок 2.9 - Ліворуч: графік ReLU. Праворуч: графік що демонструє швидкість навчання у порівнянні з Tanh [53]

Leaky ReLU. Спроба вирішити проблему вмираючого ReLU. Замість присвоювання 0 від'ємним значенням, Leaky ReLU матиме дуже малий коефіцієнт, на який значення буде домножуватися. Формула:

$$f(x) = 1(x < 0)(\alpha x) + 1(x \geq 0)(x).$$

Деякі дослідники кажуть, що досягли успішних результатів з використанням цієї функції, але результати не завжди вдалі [53].

2.2 Аналіз методів розпізнавання обличчя

2.2.1 Структура системи розпізнавання обличчя

Будь-який з розглянутих методів розпізнавання обличчя базується на архітектурі, в основі якої наступні чотири етапи.

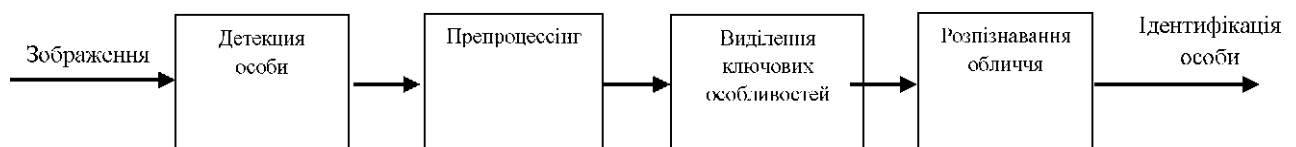


Рисунок 2.10 - Архітектура алгоритмів розпізнавання обличчя [55]

Першим кроком є детекція особи з вхідного зображення. Суть етапу полягає в підтвердженні чи має вхідне зображення обличчя і локалізація цієї області. Після цього ця область подається на препроцесінг.

Під час препроцесінгу зображення позбавляється небажаного шуму, розмиття, різних умов освітленості, тіней за допомогою різних технік. Необхідність його застосування визначається підходом, використовуваним для виділення ключових особливостей.

На третьому етапі відбувається виділення ключових особливостей за допомогою обраного для цього підходу. Результатом цього найчастіше є вектор фіксованого дозволу або набір певних точок з відповідним розташуванням.

На четвертому етапі відбувається аналіз відображень кожної особи. Тут відбувається ідентифікація особи на основі наявної бази даних[55].

Порівняння середньої точності на різних наборах даних наведених нижче методів і інших популярних підходів [55].

Таблиця 2.2 - Порівняння середньої точності популярних підходів ідентифікації обличчя [55]

PCA with ANN Face recognition system	95.45%
DDFD	91.79%
RBNNs	97.56%
CNN	85.1%
B-CNN	95.3%
BPN+RBF	98.88%
RCNN	90.3%
RINN	90.6%
MRC & MLP Neural Network	91.6%
Gabor Wavelet Faces with ANN	93%
WNN	89.22%

2.2.2 Метод головних компонент

Метод головних компонент (PCA) - математичний метод, який перетворює ряд ймовірно-корельованих змінних в меншу кількість некоррелірованих змінних, які називаються основними компонентами. PCA є дуже відомим підходом, який використовується для розрахунку набору функцій для розпізнавання особи. Це може відноситися до будь-якої особи:

1 Особа може бути представлено мінімальним набором характеристик[56].

2 Може бути побудовано з використання невеликої колекції власних фотографій. Зображення особи проектується на деяку кількість шаблонів осіб, зване власними обличчями, потім обчислюється різниця між цими обличчями, яка буде вважатися характерною рисою для зображення. Коли набір власних обличчя обчислений, зображення особи може бути реконструйовано зваженою комбінацією власних обличчя. Ваги утворюють вектор ознак для подання обличчя і розпізнавання. Коли подається нове тестове зображення, ваги вираховані завдяки проекції зображення на вектор власних осіб. Класифікація між зображеннями виходить з порівняння відстаней між векторами ваг тестового зображення і навчальної вибірки (вхідних даних). Є можливість реконструювати вихідне зображення з власних обличчя так, що вхідне зображення повинно точно відповідати оригінальному зображенню, використовуючи всі власні особи, витягнуті з оригінальних зображень. PCA є математичним методом, який заснований на перетворенні змінних, за допомогою якого можна перетворити набір корельованих змінних в набір некоррелірованих змінних. Змінні, які некоррелюють один з одним - основні компоненти. Кількість основних компонентів означає, що некоррелірованих має бути менше або дорівнює кількості вихідних змінних. У такого роду трансформації перші головні компоненти будуть мати найвищий пріоритет, який покаже максимальне відхилення. Це допоможе обчислити точність даних. Якщо набір даних розподілено нормально, тільки тоді головні компоненти можуть бути незалежними від інших змінних [56].

Алгоритм PCA

Нехай є такий набір осіб для навчання: $\Gamma_1 \Gamma_2 \Gamma_3 \dots \Gamma_M$. Середнє обличчя набору може бути визначено[56]:

$$\psi = \frac{1}{M} \sum_{n=1}^M \Gamma_n.$$

Кожне обличчя відрізняється від середнього вектором ознак [56]:

$$\Phi_i = \Gamma_i - \psi$$

Матриця коваріацій будується завдяки вектору [56]:

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T, \quad \Phi_n^T = A A^T,$$

де матриця $A = (\Phi_1 \ \Phi_2 \ \Phi_3 \ \dots \ \Phi_M)$. До множин великих векторів застосовують PCA $U_1 \ U_2 \ \dots \ U_M$. Отримуємо вектор ваг Ω зображення обличчя.

Обличчя проецирується на простір облич [56]:

$$W_k = U_k^T (\Gamma - \psi).$$

Для $k = \overline{1, M'}$, де $M' \leq M$ номер власного обличчя, яке використовується для ідентифікації. Ваги формують вектор $\Omega = (w_1, w_2, w_3 \dots w_M)$ [56].

Навчальний процес багатосарової нейромережі зі зворотним поширенням помилки вимагає набори вхідних і бажаних вихідних даних. Вихідний вектор "о" порівнюється з бажаним вихідним. У разі виявлення різниці між цими векторами, ваги підлаштовуються для досягнення мінімізації різниці. Спочатку випадкові ваги підлаштовуються до нейромережі. Ці ваги оновлюються під час кожної ітерації з метою мінімізувати середньоквадратичну помилку між вихідним і бажаним вихідним векторами.

Вхідні дані для прихованого шару надаються таким чином [56]:

$$net_m = \sum_{Z=1}^n \sum x_z w_{mz}.$$

Значення вихідного вектора прихованого шару після проходження функції активації надаються таким чином [56]:

$$h_m = \frac{1}{1 + \exp(-net_m)}$$

і значення вихідного вектора вихідного шару надаються таким чином [56]:

$$o_k = \frac{1}{1 + \exp(-net_k)}$$

Для поновлення ваг необхідно обчислити помилку таким чином [56]:

$$E = \frac{1}{2} \sum_{i=1}^k (o_i - t_i)^2$$

Це становить реальний вихід і бажаний на i -тому нейрону вихідного шару. Якщо помилка менше заданого порогу, навчання припиняється, інакше ваги повинні бути оновлені. Для ваг між прихованим і вихідним шарами зміна ваг відбувається так [56]:

$$\Delta w_{ij} = a \delta_i h_j$$

Де швидкість навчання в межах $[0.01, 1.0]$, вихід з нейрона j в прихованому шарі, може бути отримана таким чином [56]:

$$\delta_i = (t_i - o_i) o_i (1 - o_i)$$

Аналогічно, зміна ваг між прихованим і вихідним шарами відбувається таким чином [56]:

$$\Delta w_{ij} = \beta \delta_{ni} x_j$$

де β - коефіцієнт швидкості навчання в межах [0. 01, 1. 0], вихід з нейрона j у вхідному шарі, може бути отримана таким чином [56]:

$$\delta_{ni} = x_i(l - x_i) \sum_{j=1}^k \delta_j w_{ij},$$

вихід на нейроні i у вхідному шарі, і сумарно становить зважену суму всіх значень, що відповідають нейронам вихідного шару, отриманих в обчисленні. Після обчислень вага змінюється на всіх шарах, вона може бути оновлена таким чином [56]:

$$w_{ij}(new) = w_{ij}(old) + \Delta w_{ij}.$$

Цей процес повторюється поки похибка не буде мінімізована[56].

2.2.3 PCA з використанням штучних нейронних мереж

PCA це метод зменшення розмірності заснований на перетворенні Карунена-Лоева. У задачі розпізнавання осіб його застосовує головним чином для представлення зображення особи за допомогою вектора малої розмірності (головні компоненти), який порівнюється з еталонними векторами, закладеними у базу даних [57].

Головною метою методу головних компонент є значне зменшення розмірності простору ознак таким чином, щоб вони якомога краще описували "типові" риси, які притаманні великій кількості осіб. Використовуючи цей метод можна виявити різні особливості в навчальній вибірці зображень облич і описати ці особливості в базисі декількох ортогональних векторів, які називаються власними [55].

В даному підході для кожної особи відбувається навчання нейронної мережі на основі власного обличчя, яка в подальшому буде визначати схожість осіб. Метод має високу точність, але недоліком є те, що для будь-якої людини потрібна своя нейронна мережа [55].

2.2.4 Метод Deep Dense Face Detector

Даний метод базується на глибокій згортковій нейронній мережі і використовується для детекції осіб. Основними ідеями є висока здатність мереж в класифікації і виділення особливостей для побудови єдиного класифікатора для детекції осіб і мінімізація обчислювальної складності за рахунок спрощення архітектури детектора. Автор методу використовував архітектуру AlexNet і навчив її для завдання детекції саме облич. Обличчя одного і того ж класу (людини) в різних умовах були навчальною вибіркою. Їхні обличчя були модифіковані за допомогою різних методів аугментації даних для збільшення розмірів самої вибірки [58].

Детектор аналізує оцінки DDFD детектора і шукає кореляції між цими оцінками і розподілом позитивних прикладів в навчальній вибірці. Метод показав себе добре в багатьох випадках і умовах, був стійкий до різних типів освітлення, хоча відсоток точності розпізнавання значно впав при сильному повороті особи. Хоча автор вказав можливі шляхи вирішення даної проблеми за допомогою різних підходів до роботи з вхідними даними. Недоліком є необхідність повного перенавчання мережі для додавання нових осіб [55].

2.2.5 Метод Radial Basis Function Neural Networks

Даний підхід відмінно себе зарекомендував і має одні з найвищих показників в точності. Як частина препроцесінга, використовується гібридний підхід на базі Моделі Активної Форми (Active Shape Models) і PCA (Principal Component Analysis). За допомогою ASM відбувається виділення контуру та

форм особи для створення персонального профілю, а після використовується PCA алгоритм, щоб зменшити розмірність відображень осіб [55]. У розпізнавальній частині використовуються Radial Basis Function Neural Networks (RBF NN) для ідентифікації унікального шаблону асоційованого з кожною людиною. Така архітектура складається з трьох модулів займаються фазою стану, фазою виведення і фазою укладення, яка працює за допомогою правил нечіткої логіки в заміні стандартних if-then умов[59].

2.2.6 Згорткова нейронна мережа «Cascade»

Naohang Liy, Zhe Linz, Xiaohui Shenz, Jonathan Brandtz, Gang Hua [60] запропонували згорткову нейронну мережу «Cascade» для розпізнання осіб. Цей метод має потужні здатності розділення ознак, зберігаючи при цьому високу продуктивність. Запропонований CNN каскад швидко відкидає фонові області на швидких етапах з низьким дозволом, і ретельно оцінює невелику кількість кандидатів в останній стадії. Для підвищення ефективності, а також зменшення кількості кандидатів на більш пізніх етапах, автор ввів етап калібрування CNN. У результаті генерування калібрувальної сітки отримується точніша локалізація обличчя, використовуючи більші вікна сканування по меншій кількості параметрів [55].

2.2.7 Метод Bilinear CNN

Aruni RoyChowdhury Tsung-Yu Lin Subhranshu Maji Erik Learned-Miller [61] запропонували метод Bilinear CNN для ідентифікації особи, який показав різке зростання продуктивності на деяких дрібнозернистих зразках. Модель ліквідує розрив між текстурами і деталями на основі алгоритмів CNN (зазвичай використовується дві CNN).

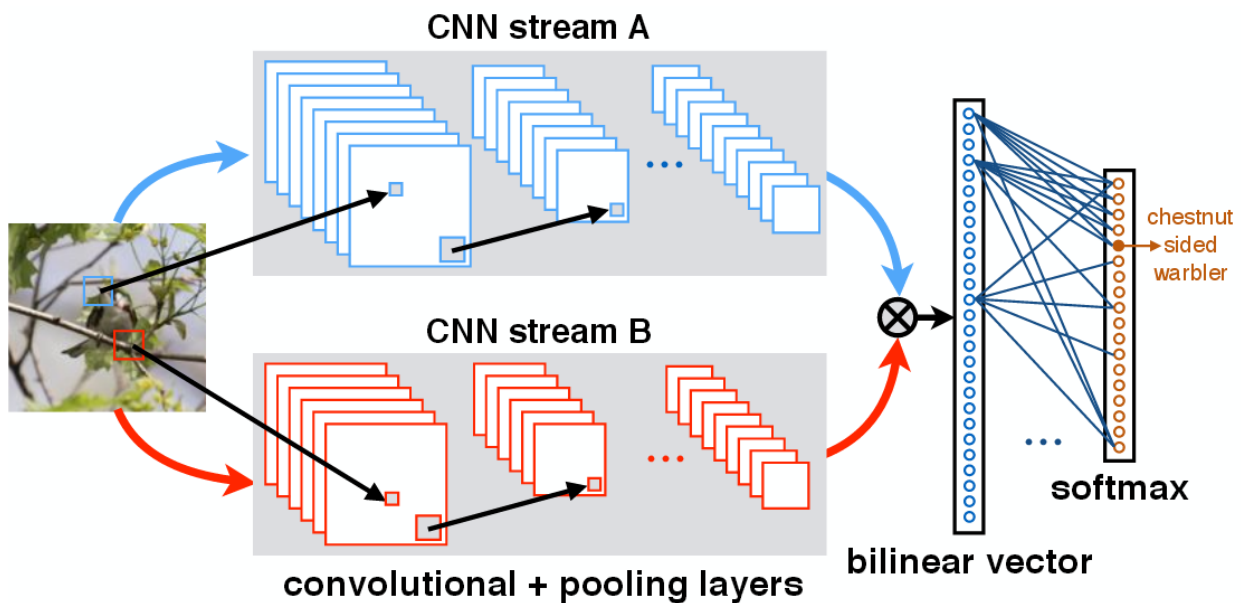


Рисунок 2.11 - Структура Bilinear CNNs [57]

Архітектура являє собою орієнтований ациклічний граф (DAG), обидва мережі можуть бути навчені одночасно за допомогою градієнтного back-propagating. Замість того, щоб навчати CNN для розпізнавання осіб з нуля, що потребує пошук оптимальної архітектури нейронної мережі та масивні бази даних, B-CNN може використовувати заздалегідь підготовлені мережі і адаптувати їх до задачі розпізнавання особи.

2.2.8 Комбінація методів Back Propagation Network та Radial Basis Function Network для розпізнавання осіб

M.Nandini, P.Bhargavi, G.Raja Sekhar [62] запропонували комбінацію Back Propagation Network та Radial Basis Function Network для розв'язку задачі розпізнавання осіб. Цей метод пропонує новий підхід для розпізнавання людських облич. Розпізнавання здійснюється шляхом порівняння характеристик нового обличчя з уже відомими. Навчання нейронної мережі (NN) здійснюється за допомогою Back Propagation і Radial Basis Function (RBF) мереж. Radial Basis

Function нейронні мережі є вельми привабливими для інженерних завдань. Вони мають дуже компактну топологію, універсальні наближення, дуже велику швидкість навчання. Нейронна мережа Back Propagation має трьохшарову архітектуру (вхідний шар, прихований шар і вихідний шар). RBF використовується як ідентифікатор у системі розпізнавання особи і входи в цій мережі є результати, отримані з BPN. Точність розпізнавання за допомогою такої моделі дуже висока. Що дає можливість розширити і використовувати цей метод для рухомих зображень і зображень з різних фоном [55].

2.2.9 Метод Retinal Connected Neural Network (RCNN)

Henry A. Rowley, студент [63] запропонував, RCNN, що досліджує невеликі вікна зображення і визначає, чи містить кожне вікно людину. Перший компонент системи є фільтр, який приймає в якості вхідних даних пікселі області 20×20 зображення, і генерує вихідний сигнал в межах від 1 до -1, що означає наявність або відсутність особи, відповідно. Для виявлення особи в будь-якому місці на вході, фільтр застосовується в кожній точці в зображенні. Для виявлення особи, розмір яких перевищує розмір вікна, вхідне зображення багаторазово зменшується. Цей фільтр повинен мати деяку інваріантність в положенні і масштабі. Кількість інваріантності визначає число ваг і положень, в яких він повинен бути застосований [55].

Основним недоліком існуючої системи є те, що вона виявляє тільки в вертикальному положенні особи, дивлячись на камеру. Окремі версії систем можуть бути навчені для кожної орієнтації голови, і результати можуть бути об'єднані за допомогою арбітражних методів [55].

2.2.10 Метод Rotation Invariant Neural Network (RINN)

Henry A. Rowley, Shumeet Baluja, Takeo Kanade [64] запропонував обертальний інваріантний метод навчання нейронної мережі для системи

розпізнавання осіб, яка використовується для ідентифікації об'єктів будь-якого ступеня обертання в площині зображення. На відміну від своїх аналогів, які розпізнають об'єкти тільки у вертикальному положенні, розпізнає обличчя в будь-якого ступеня обертання в площині зображення [55].

2.3 Висновки до розділу 2

Нейронні мережі зарекомендували себе як потужний механізм для класифікації зображень, але протягом останнього часу вони активно використовуються і в інших областях комп'ютерного зору. Їм належать перші місця в змаганнях алгоритмів розпізнавання, таких як ImageNet, а якість розпізнавання близька, а іноді і перевищує людську.

Окрім того, згорткові мережі відносно легко вирішують проблеми високоточного розпізнавання, що визивають проблеми у людей, наприклад, визначення породи собаки чи марки автомобіля, та інші задачі, що потребують виділення вузькоспецифічних ознак.

Серед недоліків можна виділити проблеми при обробці елементів малого розміру, і відсутність можливості справлятися зі спотвореннями, такими як розмиваючий фільтр або шум. Також недоліком є проблема вибору набору даних, який задовольнить проблему, що вирішується та буде досить якісним і великим для навчання нейронної мережі

Був проведений аналіз різних методів у розпізнаванні облич. Окремо можна виділити загальну структуру, що притаманна всім методам. Першим кроком зазвичай є детекція особи на вхідному зображенні. Суть етапу полягає в підтвердженні чи має вхідне зображення обличчя і локалізація цієї області. Після цього ця область подається на препроцесінг. Під час препроцесінгу зображення позбавляється небажаного шуму, розмиття, різних умов освітленості, тіней за допомогою різних технік. Необхідність його застосування визначається підходом, використовуваним для виділення ключових особливостей. На третьому етапі відбувається виділення ключових особливостей

за допомогою обраного для цього підходу. Результатом цього найчастіше є вектор фіксованого дозволу або набір певних точок з відповідним розташуванням. На четвертому етапі відбувається аналіз відображень кожної особи.

У огляді методи на нейронних мережах суттєво краще демонструють себе в тестах і мають кращу точність, але тим не менш, важливо відмітити, що одні з найкращих методів є модифікаціями класичних підходів з використанням нейронних мереж.

Незважаючи на гарні результати на тестових даних, деякі методи мають властивість погано демонструвати себе на інших наборах, або в реальних умовах. Тобто проблема універсальної системи все ще є відкритою.

3 РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ДЕТЕКЦІЇ ТА ГЕНЕРАЦІЇ СТРУКТУРНОЇ МОДЕЛІ ОБЛИЧЧЯ

3.1 Опис роботи програмного додатку

Програмний додаток призначений для генерації структурної моделі обличчя, яка б використовувала структурні нейронні мережі у своїй роботі, для підвищення ефективності роботи додатку в умовах реального світу у порівнянні з додатками, що демонструють гарну роботу лише в штучних умовах.

Додаток складається з трьох модулів:

- модуль який відповідає за детекцією облич.
- модуль, що відповідає за генерацію структурної моделі обличчя. основні функції: визначення пози обличчя, якості освітлення та проведення нормалізації. після цього він видає унікальні особливості, які можна використовувати для відрізнення однієї людини від іншої.
- модуль, що відповідає за пошук та співставлення обраного зображення обличчя з базою даних зображень.

На рисунку 3.1 подано алгоритм роботи додатку для детекції та генерації структурної моделі обличчя. Після отримання зображення для аналізу, в дію вступає модуль, що відповідає за пошук облич на зображенні. В результаті роботи цього модуля знаємо кількість та розташування облич на зображенні. Аналіз обличчя та генерація структурної моделі є наступним етапом. На основі даних отриманих від цього модуля нейронна мережа генерує обличчя, які вже потім використовуються для генерації. У ролі модуля для генерації пропозиції використовується система для розпізнавання облич, але можливі і інші варіанти використання, такі як розпізнавання емоцій, віку і т.д.

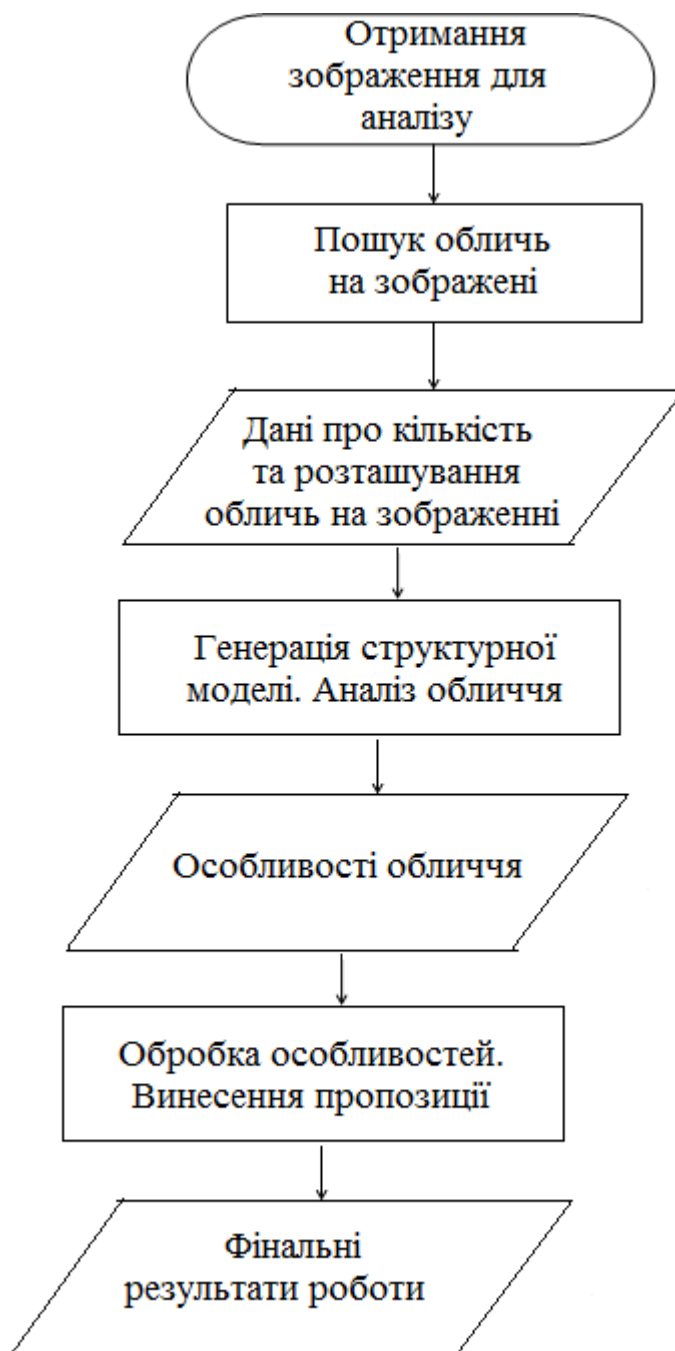


Рисунок 3.1 - Алгоритм роботи додатку для детекції та генерації структурної моделі обличчя

3.2 Детекція обличч на зображенні

Для детекції обличч у програмному додатку необхідно знати розташування обличч на зображенні. Можлива відсутність обличч на зображенні робить виклик наступних модулів марним, оскільки, не знатимемо зон, на яких запускати їх.

Для вирішення задачі детекції було обрано архітектуру Faster R-CNN. У 2015 році вона перемогла у змаганні COCO. Останні модифікації Faster R-CNN демонструють вражаючі результати в задачах класифікації. Тому було вирішено використати її для задачі детекції облич. В основу лягли преднавчані шари від VGG16, що навчена для класифікації об'єктів. Основною перевагою і особливістю Faster R-CNN є вдале поєднання швидкості та перенесення задачі класифікації на високоточну задачу детекції.

Ключовими аспектами Faster R-CNN є використання преднавчаної мережі для класифікації, на базі якої будується RPN, що складається з згортальних шарів 3×3 , два шари 1×1 додані для гілок класифікації та регресії. Розмір входу мережі відповідає VGG16 і становить 228×228 . Архітектура мережі наведена на рисунку 3.2.

Для навчання був обраний датасет WIDER. Він містить 32,203 зображення з 393,703 розміченими людьми з великим різноманіттям масштабів, поз, оклюзій. Метрика евалюації аналогічна PASCAL VOC.

Детектор дає точність 95.2% на тестовій вибірці, що можна вважати гарним результатом на фоні інших state-of-the-art підходів. Недоліком слід вважати досить велику кількість false-positive помилок. Ця особливість є недоліком архітектури. Алгоритм NMS дозволяє суттєво зменшити кількість false-positive, але зазвичай він подавляє інколи і вірні детекції. Існують більш точні архітектури, але вони працюють значно довше. Процес тренування цієї архітектури займає приблизно 2 дні.

Досі для детекції обличчя використовують HOG методи, через їх швидкість та простоту.

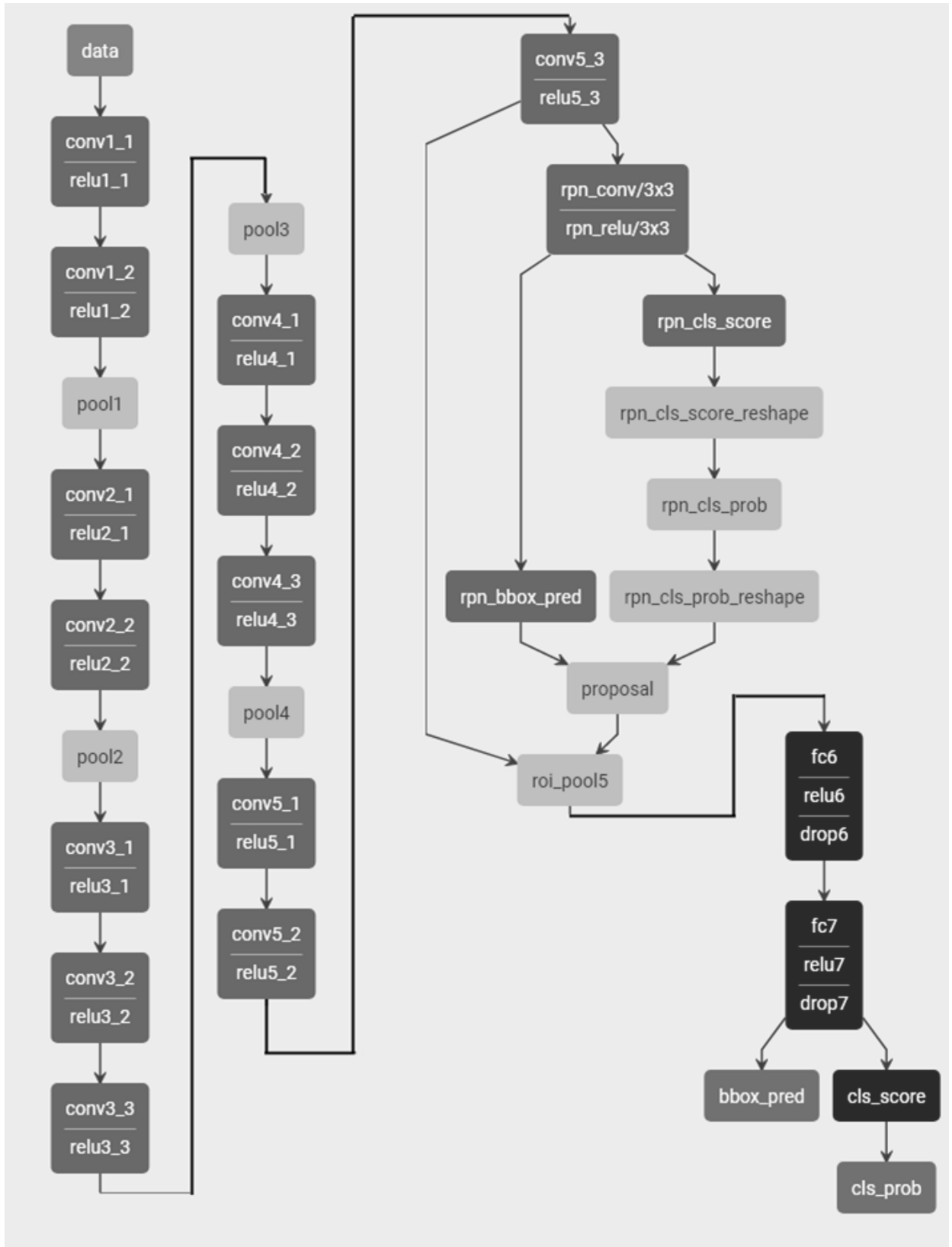


Рисунок 3.2 - Архітектура Faster-R-CNN для детекції облич

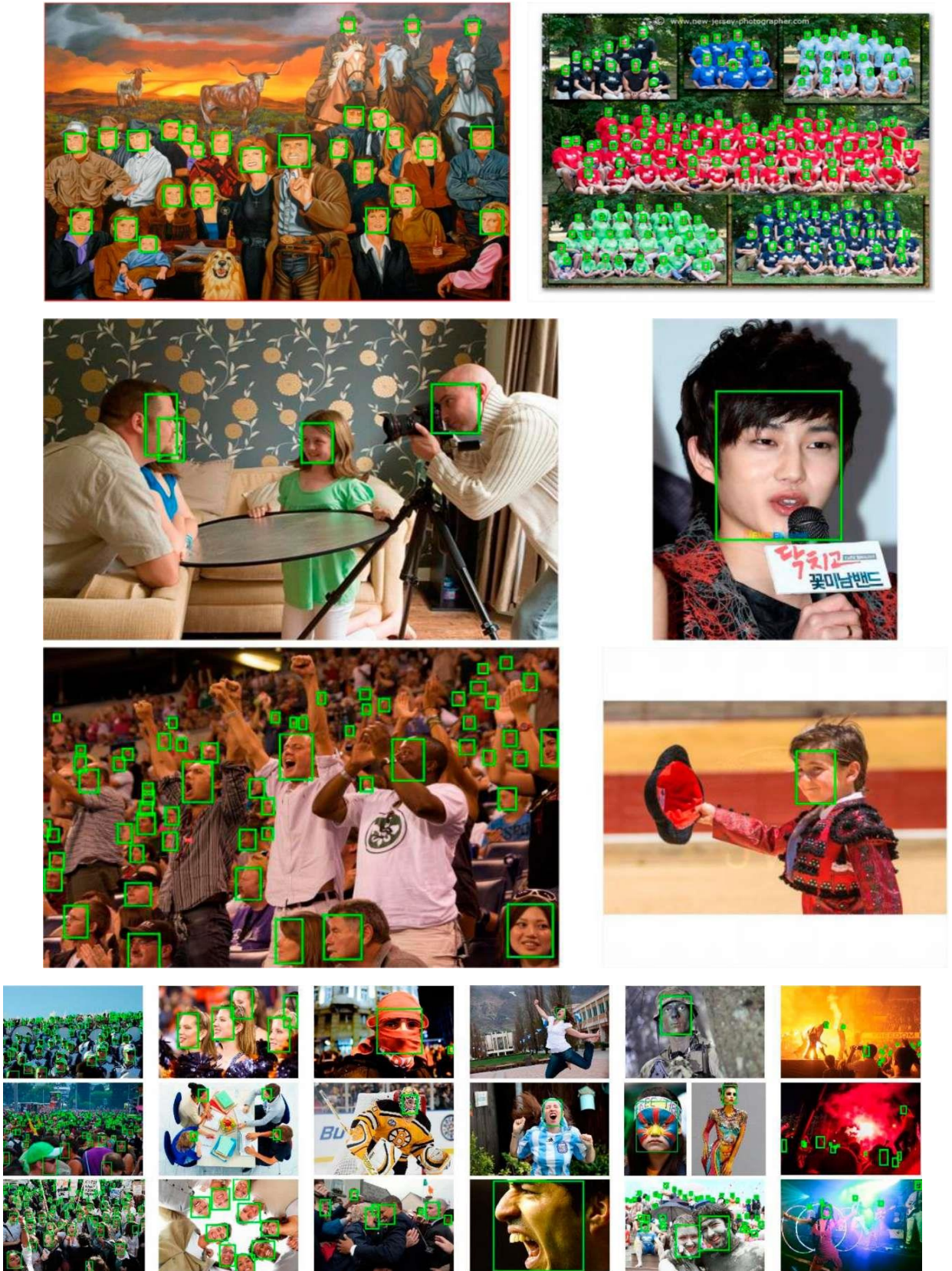


Рисунок 3.3 - Приклади зображень з датасету WIDER

3.3 Генерація структурної моделі

Після того, як ізолювали обличчя на зображенні необхідно визначити чи є оберт або нахил обличчя. Для цього проведемо нормалізацію зображення так, щоб очі та губи завжди були на одному місці. Це спростить задачу порівняння облич.

Для цього використаємо алгоритм FLE [53]. Отже визначаємо 68 специфічних точок, що характеризують обличчя. Схема цих точок зображена на рисунку 3.4.

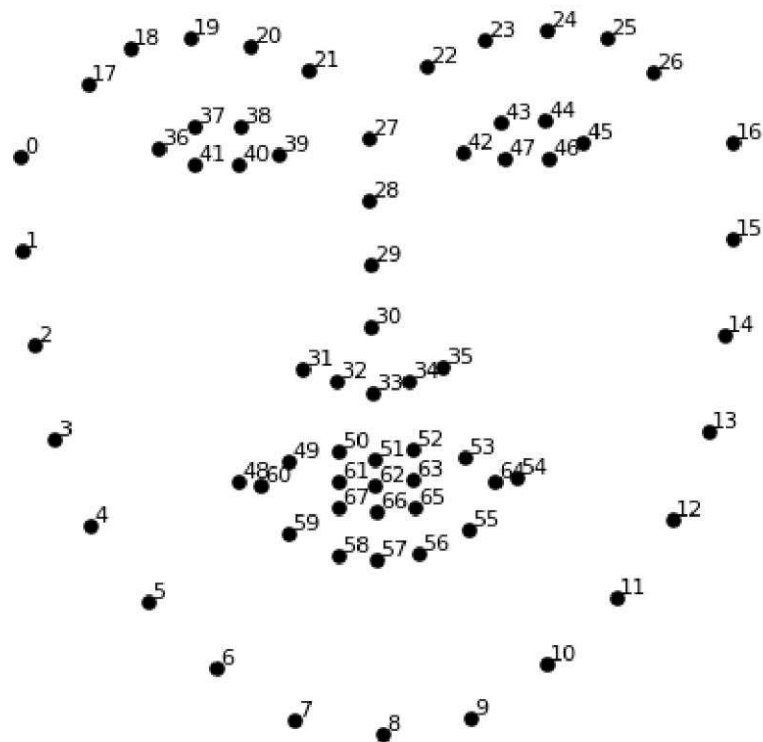


Рисунок 3.4 - Точки, що характеризують обличчя згідно з алгоритмом FLE

Після того, як отримано позицію очей і рота можна «вирівняти» його, тобто відцентрувати у приблизно На рис. 3.5 наведені приклади накладення розмітки на обличчя.

Наступний крок полягає у аналізі отриманої розмітки. Для використання її для процесу розпізнавання людини необхідно визначити які саме особливості слід порівнювати, тобто спосіб для визначення декількох базових метрик з кожного обличчя. Метрики, такі як колір очей, не мають жодного змісту для комп'ютера, тому для цих цілей навчимо нейронну мережу, яка б генерувала 128 вимірів.

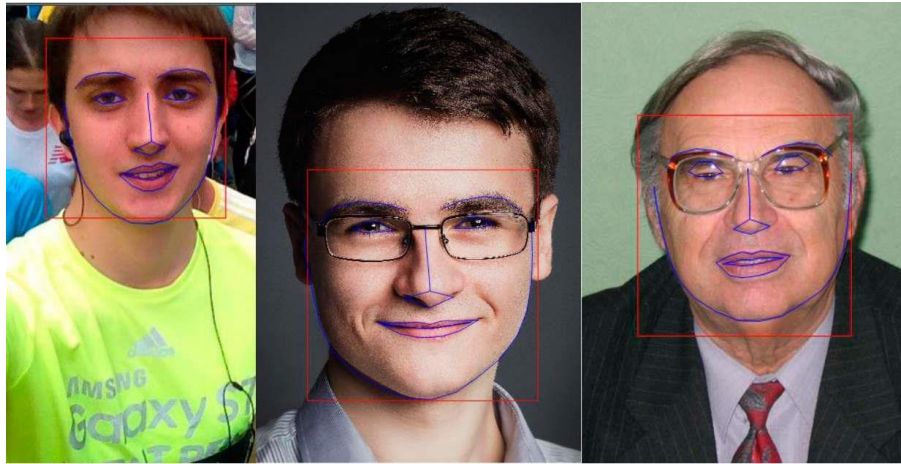


Рисунок 3.5 - Приклади роботи алгоритмів детекції та роботи FLE

Процес навчання полягає в порівнянні трьох зображень: два різних зображення однієї людини, та одне зображення абсолютно іншої.

Результат роботи мережі виділення особливостей показано на рис. 3.6.

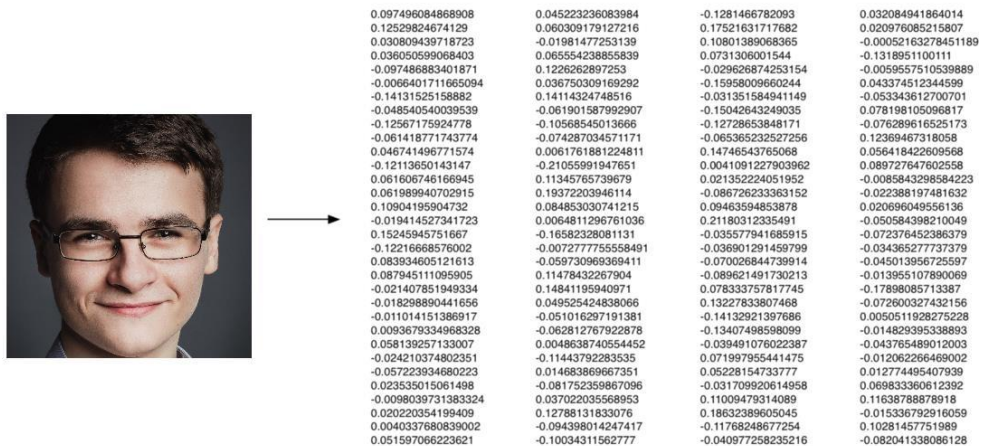


Рисунок 3.6 - Результат роботи мережі виділення особливостей

Архітектура мережі: nn.SpatialConvolutionMM: 1, 64, 48, 48
 nn.SpatialBatchNormalization: 1, 64, 48, 48
 nn.ReLU: 1, 64, 48, 48
 nn.SpatialMaxPooling: 1, 64, 24, 24
 nn.SpatialCrossMapLRN: 1, 64, 24, 24
 nn.SpatialConvolutionMM: 1, 64, 24, 24
 nn.SpatialBatchNormalization: 1, 64, 24, 24
 nn.ReLU: 1, 64, 24, 24
 nn.SpatialConvolutionMM: 1, 192, 24, 24
 nn.SpatialBatchNormalization: 1, 192, 24, 24
 nn.ReLU: 1, 192, 24, 24
 nn.SpatialCrossMapLRN: 1, 192, 24, 24
 nn.SpatialMaxPooling: 1, 192, 12, 12
 nn.Inception: 1, 256, 12, 12
 nn.Inception: 1, 320, 12, 12
 nn.Inception: 1, 640, 6, 6
 nn.Inception: 1, 640, 6, 6
 nn.Inception: 1, 1024, 3, 3
 nn.Inception: 1, 736, 3, 3
 nn.Inception: 1, 736, 3, 3
 nn.SpatialAveragePooling: 1, 736, 1, 1
 nn.View: 736 nn.Linear: 128
 nn.Normalize: 128

3.4 Співставлення обраного зображення у базі даних

Останній крок – співставлення обраного зображення. Для цього використаємо SVM класифікатор. В якості даних, на яких будемо навчати класифікатор виберемо фото відомих ведучих (рис. 3.7 – рис.3.9).



Рисунок 3.7 - Фото ведучого Джеремі Кларксона

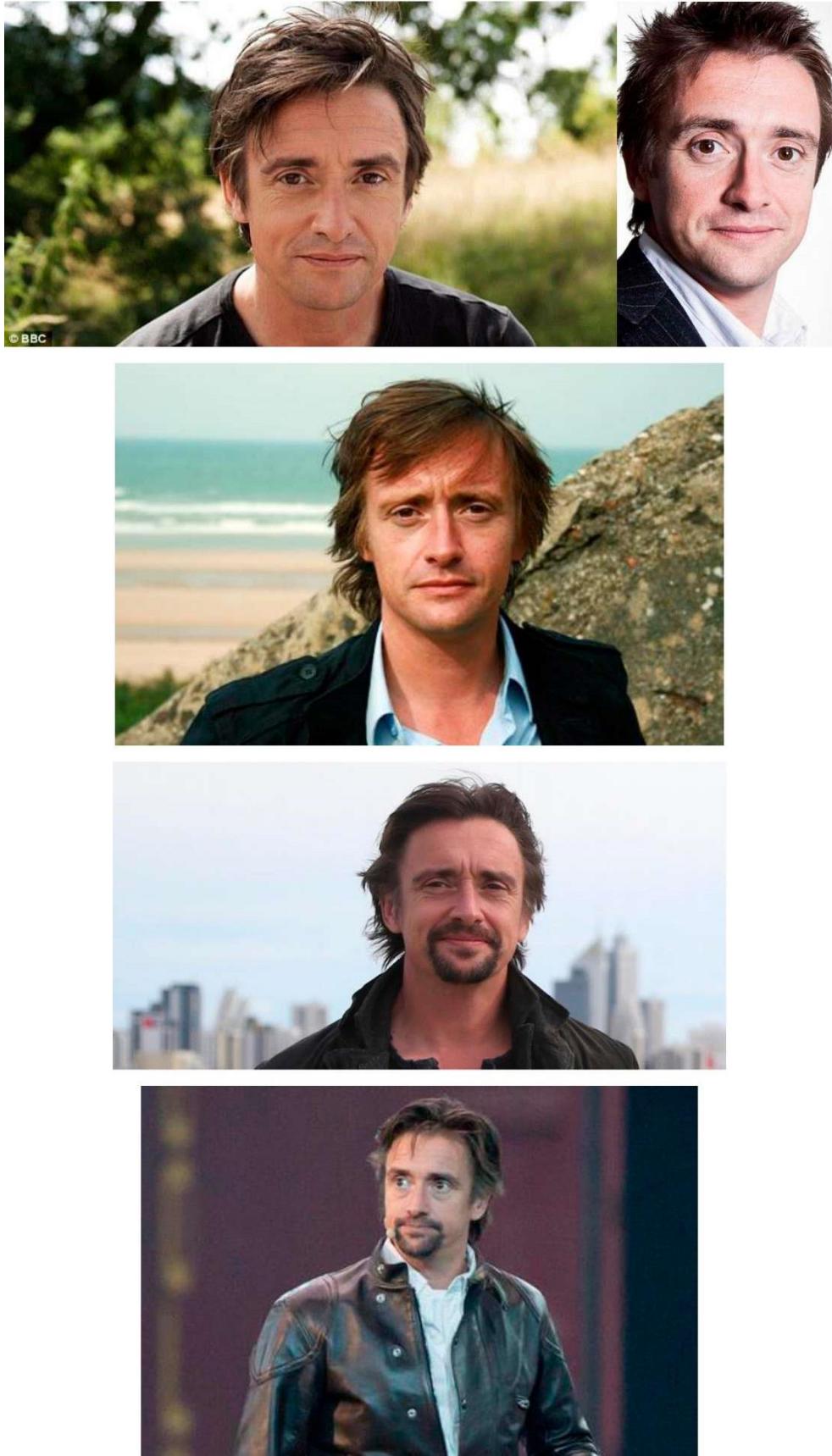


Рисунок 3.8 - Навчальна вибірка фото ведучого Річарда Хаммонда

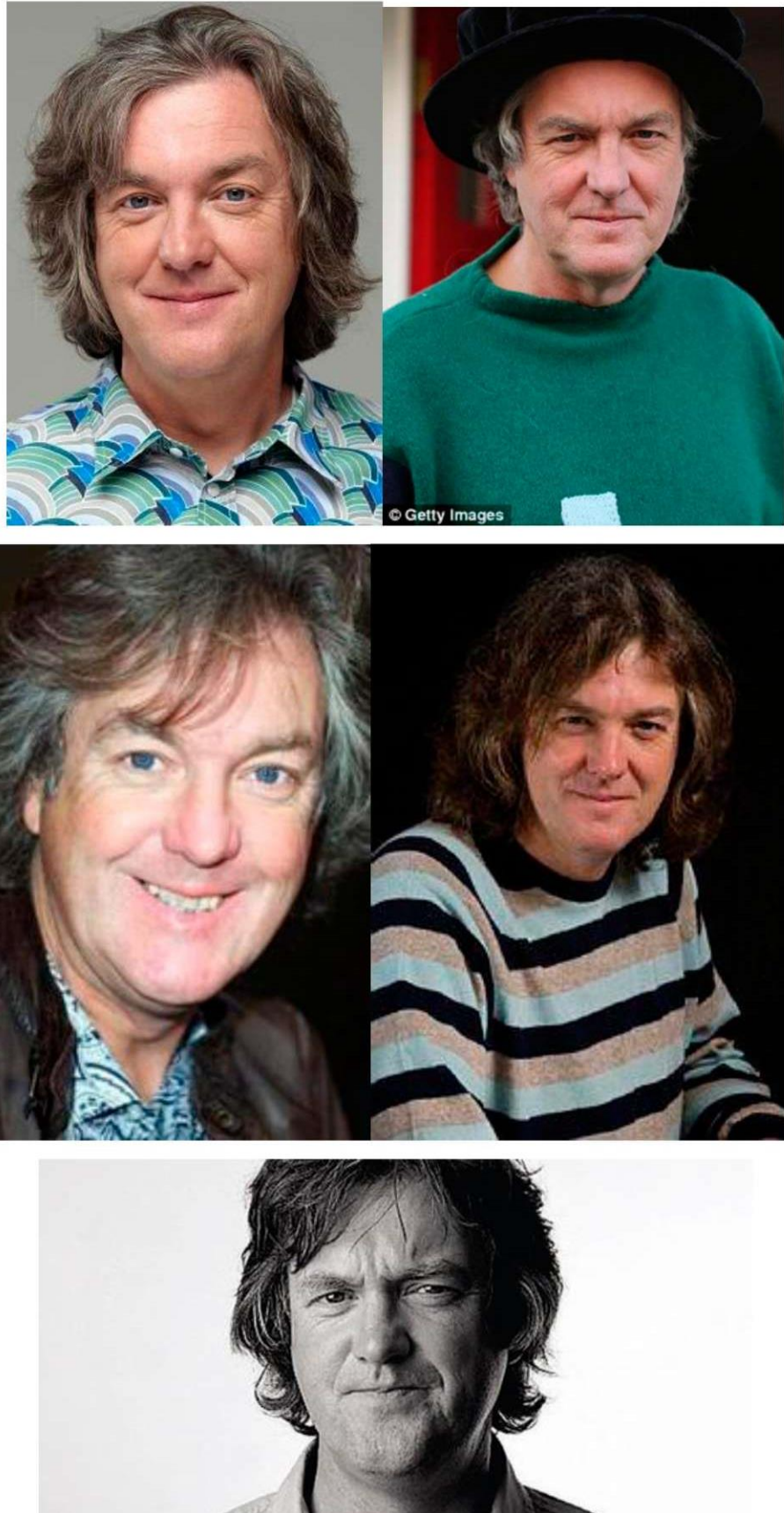


Рисунок 3.9 – Фото ведущего Джеймса Мея

Спочатку запускаємо детектор облич. Потім проводимо нормування облич за допомогою створеної структурної моделі перед поданням на мережу, що буде виявляти ознаки кожного обличчя.



Рисунок 3.10 - Приклади «нормованих» облич

```

0.014849368482828,0.10054879635572,0.10990730673075,-0.13388541340828,-0.017203044146299,-0.060
0.088297910988331,-0.004823740106076,0.099746376276016,-0.14157070219517,-0.060170289129019,0.0
0.09831478446722,0.070236377418041,0.068821899592876,-0.06771082431078,0.014927946962416,0.0332
0.094958402216434,0.045980162918568,0.081006750464439,-0.093230009078979,0.016965072602034,-0.0
0.030756305903196,0.12479083240032,0.1537029594183,-0.12454158067703,0.018650038167834,-0.07123
0.066001996397972,0.11692174524069,0.0035676613915712,-0.11703211069107,-0.073685929179192,-0.0
-0.010681649670005,0.015157928690314,-0.018830694258213,-0.016133764758706,-0.075115852057934,-
0.03742715716362,-0.034199804067612,-0.0365827716887,0.015391842462122,-0.089156538248062,0.026
0.014625946991146,-0.033607393503189,0.055713146924973,0.041101723909378,-0.049529023468494,0.1
0.052812650799751,-0.029624357819557,0.078864328563213,-0.018065735697746,-0.098214991390705,0.
0.020143695175648,-0.0006854044040665,-0.026975683867931,-0.0053625041618943,-0.062140122056007
0.011664037592709,-0.02269085496664,-0.019086403772235,0.012104541994631,-0.048881955444813,0.0
-0.047883406281471,0.19454275071621,0.15070277452469,-0.2516727745533,-0.014277837239206,-0.056
-0.1213983297348,0.11870682239532,0.065691709518433,-0.30478924512863,0.098245792090893,-0.0503
-0.050967533141375,-0.012118018232286,0.10637626051903,-0.21983744204044,0.10834065824747,0.006
-0.038127541542053,-0.010182742029428,0.096340551972389,-0.21043434739113,0.11051530390978,0.01
-0.0060945032164454,0.12262935936451,0.058394853025675,-0.24557285010815,0.0080501642078161,-0.

```

Рисунок 3.11 - Вихід нейронної мережі. Кожен рядок містить ознаки одного зображення

З отриманих ознак навчаємо SVM. Тестуємо її роботу на зображенні, якого не було в навчальній вибірці (рис. 3.12, 3.13).

Як видно з рис. 3.13 класифікатор відпрацював правильно. Обраних 17 фотографій для навчання вистачило для класифікації з впевненістю 68 відсотків.



Рисунок 3.12 - Тестове зображення

```
ml@gladosML:~/workspace/facefinder$ ./demos/classifier.py infer ./generated-embeddings/classifier.pkl 1.jpg
/home/ml/.local/lib/python2.7/site-packages/sklearn/lda.py:4: DeprecationWarning: lda.LDA has been moved to discrim
inant_analysis.LinearDiscriminantAnalysis in 0.17 and will be removed in 0.19
  "in 0.17 and will be removed in 0.19", DeprecationWarning)

=== 1.jpg ===
Predict hammond with 0.68 confidence.
ml@gladosML:~/workspace/facefinder$
```

Рисунок 3.13 - Робота на тестовому зображенні

3.5 Висновки до розділу 3

В третьому розділі дипломного проекту було створено програмний додаток для детекції та генерації структурної моделі обличчя використовуючи штучні нейронні мережи. Метою програмного забезпечення було продемонструвати, що дані про структурну модель обличчя можуть бути корисними для використання у практичних цілях, а саме для пошуку на ідентифікації людини по фотозображенню.

Описано алгоритм роботи модуля детектору облич. Він базується на провідній архітектурі Faster-R-CNN. Детектор демонструє відмінні результати на датасеті WIDER, а саме 95.2% на тестовій вибірці.

На основі координат облич працює метод для генерації структурної моделі обличчя, а саме використано алгоритм FLE. Дані про структурну модель обличчя

служать для нормалізації зображення. Після цього процесу використовується нейронна мережа, що навчена відрізнати одну особу на зображенні від іншої за своїми ознаками.

Потім використовується SVM для ідентифікації людини по зображенню, що відрізняє одну від іншої на основі ознак, отриманих від нейронної мережі.

Даний спосіб використання структурної моделі обличчя не є єдиним. Ці дані можуть стати основою для систем розпізнавання активності, стану, настрою і інших ознак. Надалі при розвитку програми для цих цілей можна використати детектори на базі архітектури Stacked Hourglass Network та Convolutional Pose Machine.

ВИСНОВКИ

На сьогодні задачам автоматичного розпізнання обличчя присвячено велику кількість наукових досліджень, також має великий потенціал використання даної технології у комерційних проектах, тому тема дипломного проекту є актуальною.

В дипломному проекті проведено структурування описаної задачі, та розбито на ряд підзадач: детекцію та ідентифікацію. Зроблено обґрунтування використання апарату штучного інтелекту, а саме нейронних мереж для рішення поставлених завдань.

Був проведений аналіз різних методів розпізнаванні образів. У огляді методи на нейронних мережах суттєво краще демонструють себе в тестах і мають кращу точність, але тим не менш, важливо відмітити, що одні з найкращих методів є модифікаціями класичних підходів з використанням нейронних мереж.

Було створено програмний додаток для пошуку та ідентифікації особи. Ціллю платформи були продемонструвати, що дані про структурну модель обличчя можуть бути корисними для використання у практичних цілях. ПП базується на провідній архітектурі Faster-R-CNN. Детектор демонструє відмінні результати на датасеті WIDER, а саме 95.2% на тестовій вибірці.

Розроблено метод для генерації структурної моделі обличчя з використанням нейронної мережі, що навчена відрізнити одну особу від іншої за своїми ознаками. Для ідентифікації особи використовується SVM, що відрізняє одне фотозображення від іншого на основі ознак, отриманих від нейронної мережі.

Розроблені у дипломному проекті засоби можуть стати основою для систем розпізнавання активності, стану, настрою і інших ознак.

ПЕРЕЛІК ПОСИЛАНЬ

- 1 P. A. Viola and M. J. Jones. Rapid object detection using a boosted cascade of simple features. In Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2001.
- 2 Dalal N., Triggs B. Histograms of oriented gradients for human detection //Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. - IEEE, 2005. - Т. 1. - С. 886-893.
- 3 Olszewska J. Automated Face Recognition: Challenges and Solutions//Intech - Open science Open minds , [Електронний ресурс]. - Режим доступу: <https://cdn.intechopen.com/pdfs-wm/52911.pdf>
- 4 Sirovich L., Kirby M. Low-dimensional procedure for the characterization of human. Journal of the Optical Society of America// Optics, Image Science and Vision. 1987. 4(3):519-524.
- 5 Zhao W., Chellappa R., Rosenfeld A., Phillips P.J. Face recognition: A literature survey. ACM Computing Surveys. 2003. 35(4):399-458.
- 6 Iosifidis, A., Gabbouj, M. Scaling-up class-specific kernel discriminant analysis for large-scale face verification. IEEE Transactions on Information Forensics and Security. 2016.11(11):2453-2465
- 7 Almudhahka N., Nixon M., Hare J. Human face identification via comparative soft biometrics. In: Proceedings of the IEEE International Conference on Identity, Security and Behavior Analysis (ISBA).Sendai, JP, 29 Feb-02 Mar 2016. pp. 1-6.
- 8 Berg T.I., Berg A.C., Edwards J., Forsyth D.A. Who's in the Picture?. In: Proceedings of the Neural Information Processing Systems Conference (NIPS). Springer. 2004. pp. 137-144.
- 9 Torres L. Is there any hope for face recognition? In: Proceedings of the IEEE International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS). 2004.

- 10 Nash S., Rhodes M., Olszewska J. I. iFR: Interactively-pose-corrected face recognition. In: Proceedings of the INSTICC International Conference on Bio-Inspired Systems and Signal Processing (BIOSIGNALS). 2016. pp. 106-112.
- 11 Shreve M., Godavarthy S., Goldgof D., Sarkar, S. Macro- and micro-expression spoJing in long videos using spatio-temporal strain. In: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition Workshops(AFGR). 2011. pp. 51-56.
- 12 Liu L., Xiong C., Zhang H., Niu Z., Wang M., Yan S. Deep aging face verification with large gaps. IEEE Transactions on Multimedia. 2016. 18 (1):64-75.
- 13 Julian P., Dehais C., Lauze F., Charvillat V., Bartoli A., Choukroum A. Automatic hair detection in the wild. In: Proceedings of the IEEE International Conference on Pattern Recognition (ICPR). 2010. pp. 46174620.
- 14 Wood R., Olszewska J. I. Lighting-variable AdaBoost based-on system for robust face detection. In: Proceedings of the INSTICC International Conference on Bio-Inspired Systems and Signal (BIOSIGNALS). 2012. pp. 494-497.
- 15 Shan S., Gao W., Cao B., Zhao D. Illumination normalization for robust face recognition against varying lighting conditions. In: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (AFGR). 2003. pp. 157-164.
- 16 ISO/IEC 19794-5:2011. The Face Image Format Standards. In: Information Technology Biometric data interchange formats: Part 5: Face image data. 2nd Ed. 2011.
- 17 Prince S.J.D., Elder J., Hou Y., Sizinstev M., Olevskiy E. Towards face recognition at a distance. In: Proceedings of the IET Conference on Crime and Security. 2006. pp. 570-575.
- 18 Mudunuri S.P., Biswas S. Low resolution face recognition across variations in pose and illumination. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2016. 38(5): 1034-1040.
- 19 Zou W.W.W., Yuen P.C. Very low resolution face recognition problem. IEEE Transaction on Image Processing. 2012. 21(1):327-340.

- 20 Matas J., Hamouz M., Jonsson K., Kittler J., Li Y., Kotroupolous C., Tefas A., Pitas I., Tan T., Yan H., Ameraldi F., Bigun J., Capdevielle N., Gerstner W., Ben-Yacoub S., Abd jaoued Y., Mayoraz E. Comparison of face verification results on the XM2VTS database. In: Proceedings of the IEEE International Conference on Pattern Recognition (ICPR). 2000. pp. 858863.
- 21 Grgic M., Delac K., Grgic S., Klimpak B. SCface - Surveillance cameras face database. Multimedia Tools Applications. 2011. 51:863-879.
- 22 FDDB. A Benchmark for Face Detection in Unconstrained Settings Jain V., Learned - Miller E. Technical Report. UM-CS-2010-009. University of Massachusetts Amherst. USA [Электронный ресурс]. - Режим доступа: <http://www.cs.umass.edu/fddb>.
- 23 Learned-Miller E., Huang G.B., Roy-Chowdhury A., Li H., Hua G. Labeled Faces in the Wild: A Survey. In: Kawulok M., Emre Celebi M., Smolka B., editors. Advances in Face Detection and Facial Image Analysis. Springer. 2016. pp. 189-248.
- 24 Li X., Pefister T., Huang X., Zhao G., Pietikainen M. A spontaneous micro-expression database: Inducement, collection and baseline. In: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (AFGR). 2013. pp. 1-6.
- 25 Panis G., Lanitis A., Tsapatsoulis N., Cootes T.F. Overview of research on facial ageing using the FG-NET ageing database. IET Biometrics. 2016. 5(2):37-46.
- 26 Хуршудов, А.А. Разработка систем распознавания визуальных образов в потоке данных: дис.кан. тех. наук. 05.13.01: защищена 22.01.15: утв. 15.07.5 / Хуршудов Артем Александрович. - Краснодар, 2015. - 130 с.
- 27 Hinton, G.E. Learning multiple layers of features from tiny images / A. Krizhevsky, G. Hinton // Computer Science Department, University of Toronto, Tech. Rep. — 2009. — no. 1.4. — Pp. 7-10.

- 28 Deng, L. The MNIST database of handwritten digit images for machine learning research / L. Deng // *IEEE Signal Processing Magazine*. — 2012. — no. 29.6. — Pp. 141-142.
- 29 Bengio, Y. Representation learning: A review and new perspectives / Y. Bengio, A. Courville, P. Vincent // *Pattern Analysis and Machine Intelligence*. — 2013. — no. 35(8). — Pp. 1798-1828.
- 30 Fergus, R. A sparse object category model for efficient learning and exhaustive recognition / R. Fergus, P. Perona, A. Zisserman // *Computer Vision and Pattern Recognition*. — 2005. — no. 1. — Pp. 380-387.
- 31 Murphy-Chutorian, E. Head pose estimation in computer vision: A survey. / E. Murphy Chutorian, M.M. Trivedi // *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 31.4. — 2009. — Pp. 607-626.
- 32 Hubel, D. H. Receptive fields and functional architecture of monkey striate cortex / D. H. Hubel, T. N. Wiesel // *The Journal of physiology*. — 1968. — no. 195(1). — Pp. 215-243.
- 33 Kreutz-Delgado, K. Dictionary learning algorithms for sparse representation / K. Kreutz-Delgado // *Neural computation*. — 2003. — no. 15.2. — Pp. 349-396.
- 34 Fukushima, K. Neocognitron: A hierarchical neural network capable of visual pattern recognition / K. Fukushima // *Neural networks*. — 1988. — no. 2. — Pp. 119-130.
- 35 Serre, T. Object recognition with features inspired by visual cortex / T. Serre, L. Wolf, T. Poggio // *Computer Vision and Pattern Recognition*. — 2005. — no. 2. — Pp. 994-1000.
- 36 Krizhevsky, A. Imagenet classification with deep convolutional neural networks / A. Krizhevsky, I. Sutskever, G.E. Hinton // *Advances in neural information processing systems*. — 2012. — Pp. 1097-1105.
- 37 Ciresan, D. Multi-column deep neural networks for image classification / D. Ciresan, U. Meier, J. Schmidhuber // *Computer Vision and Pattern Recognition (CVPR)*. — 2012. — Pp. 3642-3649.

38 Matsugu, M. Subject independent facial expression recognition with robust face detection using a convolutional neural network / M. Matsugu // Neural Networks. — 2003. — no. 16(5). — Pp. 555-559.

39 Yahia, S. Human detection based on integral Histograms of Oriented Gradients and SVM / S. Yahia, M. Atri, R. Tourki // Communications, Computing and Control Applications. — 2011. — Pp. 1 - 5.

40 Deng, L. Recent advances in deep learning for speech research at Microsoft /L. Deng // Acoustics, Speech and Signal Processing (ICASSP). — 2013. — Pp. 8604-8608.

41 Konolige, K. Projected texture stereo / K. Konolige // Robotics and Automation (ICRA). — 2010. — Pp. 23-28.

42 Keyzers, D. Comparison and combination of state-of-the-art techniques for handwritten character recognition: topping the mnist benchmark / D. Keyzers // arXiv. — 2007. — no. 0710.2231. — Pp. 21-27.

43 Convolutional Neural Networks for Visual Recognition: [Электронный ресурс]. - Режим доступа: <http://cs231n.github.io>.

44 Выбор функции активации обучения нейронной сети [Электронный ресурс]. - Режим доступа: <https://monographies.ru/ru/book/section?id=2465>.

45 N Jindal, V Kumar, "Enhanced Face Recognition Algorithm using PCA with Artificial Neural Networks", International Journal of Advanced Research in Computer Science and Software Engineering, vol 3, no. 6, (2013), pp. 864-872.

46 A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Image net classification with deep convolutional neural networks", Edited F. Pereira and C.J.C. Burges and L. Bottou and K.Q. Weinberger, publish by Curran Associates, Inc. , (2012), pp. 1097-1105.

47 S-H Yooa, S-K Oha, Witold Pedrycz, "Optimized face recognition algorithm using radial basis function neural networks and its practical applications", International journal on Neural Networks, volume 69, (2015), pp. 111-125.

48 H Liy, Z Linz, X Shenz, J Brandtz, GHua," A Convolutional Neural Network Cascade for Face Detection", IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, (2015), pp. 5325 - 5334.

49 A Roy Chowdhury Tsung-Yu Lin Subhranshu Maji Erik Learned-Miller," Face Identification with Bilinear CNNs", Computer vision and pattern recognition, (2015).

50 M.Nandini, P.Bhargavi, G.Raja Sekhar,"Face Recognition Using Neural Network", International Journal of Scientific and Research Publications, vol. 3, no. 3, (2013), pp. 1-5.

51 H A. Rowley, Student Member," Neural Network-Based Face Detection",IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA , (1996), pp. 203 - 208.

52 H A. Rowley, S Baluja, T Kanade," Rotation Invariant Neural Network-Based Face Detection", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Santa Barbara, CA, (1998), pp. 38 - 44.

53 One Millisecond Face Alignment with an Ensemble of Regression Trees. [Электронный ресурс]. - Режим доступа: <http://www.csc.kth.se/~vahidk/papers/KazemiCVPR14.pdf>.

54 Cmusatyalab/openface [Электронный ресурс]. - Режим доступа: <https://github.com/cmusatyalab/openface>.