

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук
(повна назва)

Кафедра _____ Штучного інтелекту
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти _____ другий (магістерський)

Розпізнавання емоцій з фотографії за допомогою нейронної мережі

(тема)

Виконав:
студент 2 курсу, групи _____ СШМ-20-2
_____ Олійник Д.А.
(прізвище, ініціали)

Спеціальність _____ 122 Комп'ютерні науки

(код і повна назва спеціальності)

Тип програми _____ освітньо-наукова
(освітньо-професійна або освітньо-наукова)

Освітня програма _____ Системи штучного інтелекту

(повна назва спеціалізації)

Керівник _____ доц. Магдаліна І.В.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____
(підпис)

_____ В.О. Філатов
(прізвище, ініціали)

2022 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)
Кафедра Штучного інтелекту
(повна назва)
Рівень вищої освіти другий (магістерський)
Спеціальність 122 Комп'ютерні науки
(код і повна назва)
Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)
Освітня програма Системи штучного інтелекту (СШІ)
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

«_____» _____ 20__ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Олійнику Дмитрові Андрійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Розпізнавання емоцій з фотографії за допомогою нейронної мережі

затверджена наказом університету від 24 березня 2022 р. № 414Ст

2. Термін подання студентом роботи до екзаменаційної комісії _____ 20__ р.

3. Вихідні дані до роботи науково-технічні публікації, дані статей щодо архітектури і розробки глибоких нейронних мереж, результати експериментальних досліджень, дані щодо алгоритмів навчання нейронних мереж

4. Перелік питань, що потрібно опрацювати в роботі _____

1 Аналіз предметної області

2 Комбінований метод розпізнавання емоцій

3 Реалізація програмного додатку для розпізнавання емоцій

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри) Рисунок 1.1 – Отримання знань в глибинному навчанні; Рисунок 1.2 – Базова структура CNN; Рисунок 1.3 – Структура мережі AlexNet; Рисунок 1.4 – Основні підходи, які застосовуються при побудові off-line систем ідентифікації особи; Рисунок 1.5 – Інтерфейс Face Reader; Рисунок 1.6 – Інтерфейс системи EmoDetect; Рисунок 1.7 – Інтерфейс системи Face Security; Рисунок 1.8 – Інтерфейс додатку Emotion Recognition; Рисунок 1.9 – Інтерфейс додатку eMotion Software; Рисунок 1.10 – Робота з маскою MMER_FEASy; Рисунок 2.1 – Кодування опорних точок обличчя; Рисунок 2.2 – Комбінація лицьових рухів; Рисунок 2.3 – Ключові відстані; Рисунок 3.1 – Зображення з розпізнаною емоцією здивування; Рисунок 3.2 – Зображення з розпізнаним нейтральним виразом обличчя

6. Консультанти розділів роботи (п.6 включається до завдання за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Отримання завдання	28.03.22	Виконано
2	Аналіз предметної області і постановка задачі	30.03.22-10.04.22	Виконано
3	Аналіз методів і алгоритмів щодо вирішення задачі	10.04.22-25.04.22	Виконано
4	Розробка вимог до програмної системи	25.04.22-26.04.22	Виконано
5	Програмна реалізація	26.04.22-29.04.22	Виконано
6	Тестування та аналіз результатів	29.04.22-30.04.22	Виконано
7	Оформлення пояснювальної записки	30.04.22-03.05.22	Виконано
8	Попередній захист	04.05.22	Виконано
9	Захист перед ЕК		Виконано

Дата видачі завдання 28 березня 20 22 р.

Студент _____
(підпис)

Керівник роботи _____
(підпис) _____
(посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 61 с., 2 табл., 14 рис., 2 дод., 15 джерел.

ВИЯВЛЕННЯ ОБЛИЧ, КОМП'ЮТЕРНЕ БАЧЕННЯ, КОНВОЛЮЦІЙНА НЕЙРОННА МЕРЕЖА, РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ, СИСТЕМА РОЗПІЗНАВАННЯ ОБЛИЧ

Об'єкт дослідження – розпізнавання облич з допомогою комп'ютерного бачення.

Предмет дослідження – методи побудови конволюційних нейронних мереж для розпізнавання емоцій по фотографіям.

Мета роботи – розробка програмного модулю розпізнавання емоцій по фотографіям.

Методи дослідження – методи розпізнавання облич та емоцій.

В результаті проведених досліджень вирішено задачу вибору найбільш ефективного методу розпізнавання емоцій на фотографії шляхом їх вивчення та порівняння між собою.

На основі результатів виконаних досліджень розроблено програмний додаток для розпізнавання емоцій по фотографії з допомогою обраного методу.

В якості обраної мови програмування обрано Python через ряд переваг даної мови для поставленої задачі над іншими мовами програмування. Пропонована розробка є корисною для вирішення задачі розпізнавання емоцій на фото, що може бути використано в багатьох сферах як повсякденного життя, так і професійної діяльності.

ABSTRACT

Explanatory note: 61 p., 2 tabl., 14 pic., 2 add., 15 sources.

COMPUTER VISION, CONVOLUTIONAL NEURAL NETWORK,
FACE RECOGNITION SYSTEM, FACE DETECTION, IMAGE
RECOGNITION

Research object – face recognition with computer vision.

Subject of study – methods of constructing convolutional neural networks for recognizing emotions in photographs.

The goal of the work – development of a software module for recognizing emotions from photographs.

Research methods – methods of recognizing faces and emotions.

As a result of the research, the problem of choosing the most effective method of recognizing emotions in photography by studying and comparing them was solved.

Based on the results of the research, a software application was developed to recognize emotions in photography using the selected method.

Python was chosen as the chosen programming language due to a number of advantages of this language for the task over other programming languages. The proposed development is useful for solving the problem of recognizing emotions in the photo, which can be used in many areas of both daily life and professional activities.

ЗМІСТ

Вступ.....	7
1 Аналіз предметної області.....	11
1.1 Аналіз сучасного стану в області	11
1.2 Задачі комп'ютерного бачення та області його застосування.....	14
1.3 Переваги та недоліки систем комп'ютерного бачення	20
1.4 Огляд сучасних систем розпізнавання емоцій людини.....	22
1.5 Методи, які використовуються при розпізнаванні обличч людей.....	26
1.6 Огляд існуючих систем розпізнавання емоцій	32
1.6.1 FaceReader Noldus Information Technology	32
1.6.2 EmoDetect.....	34
1.6.3 Face Security	35
1.6.4 Microsoft Oxford Project Emotion Recognition.....	37
1.6.5 eMotion Software	38
1.6.6 MMER_FEASy.....	39
2 Комбінований метод розпізнавання емоцій	41
2.1 Метод на основі локалізації ключових точок на обличчі	41
2.2 Метод на основі інформації про текстуру	44
2.3 Метод на основі геометричного аналізу рухових одиниць обличчя ...	45
3 Реалізація програмного додатку для розпізнавання емоцій	49
Висновки	51
Перелік джерел посилання	52
Додаток А Вихідний код програми	54
Додаток Б Відомість кваліфікаційної роботи.....	61

ВСТУП

Систем комп'ютерного зору (СКЗ) – молода область інформаційних технологій, пов'язана з розпізнаванням образів та зображень. Історія СКЗ бере початок із кінця 50-х років ХХ століття. Проте сучасний світ вже неможливо уявити без технологій розпізнавання зображень. Використання систем для розпізнавання образів є одним із основоположних моментів промислової стратегії розвитку для більшості підприємств світу. Про це говорить експоненційне зростання ринку споживання програмних рішень, що використовують технології розпізнавання зображень, що спостерігається останніми роками. Світовий ринок промислового впровадження систем з елементами використання комп'ютерного зору, що становив наприкінці минулого і на початку нинішнього століття близько 8 млрд. доларів, має тенденцію збільшуватися втричі кожні кілька років. Зростання компаній, що працюють у сфері СКЗ, становлять від 15 до 30% щорічно. Але, наявні комерційні розробки покривають менше 10% усіх теоретичних та потенційних можливостей систем комп'ютерного зору.

Можливості області, у яких можуть застосовуватися СКЗ, тривалий час отримували розвиток, переважно, як машинний аналог людських очей. Основною функцією та призначенням систем комп'ютерного зору є розпізнавання зовнішнього вигляду та розташування у просторі спостережуваних об'єктів, а також, перетворення оцифрування зображень, для подальшої обробки на персональному комп'ютері та (або) передачі. Найбільш значущі напрями використання систем комп'ютерного зору на даний момент є:

- зір роботів;
- пристрої для дистанційного керування;
- засоби обробки та розпізнавання зорових образів та символічної інформації;
- системи для моніторингу та забезпечення безпеки;

- системи медичної діагностики та біометричного контролю.

Також, СКЗ мають досить серйозні переваги перед людським зором. Системи комп'ютерного зору, обладнані сенсорними специфічними пристроями та датчиками, можуть побачити те, що недоступне простому неозброєному оку. Збільшення та розширення спектрального діапазону сприйняття СКЗ за межі сприйняття людським оком розширює функціонал та можливості систем комп'ютерного зору значно. Прикладом пристроїв перевершують за характеристиками людський зір є системи для забезпечення безпеки, телевізори та прилади нічного бачення. Важливою особливістю таких систем є те, що вони забезпечують перетворення невидимого уявлення об'єктів у доступний сприйняття людини діапазон, цим збільшуючи можливості людини у оперативному прийнятті рішень. Реалізація можливостей систем комп'ютерного зору, перерахованих вище, стабільно забезпечується сучасними пристроями (фотоприймачі, відеокамерами). Вони забезпечують високу просторову вибірку освітленості спостережуваних об'єктів. Спектрально-селективні оптичні детектори та джерела світла відкривають новий якісний рівень можливостей систем комп'ютерного зору, освоєння яких перебуває ще на стадії свого зародження. Новий рівень пов'язаний з використанням систем комп'ютерного зору у вирішенні задач для автоматизації моніторингу за спостереженням у контрольованих об'єктах специфічних властивостей, які, наприклад, обумовлені їх хімічним складом.

Сучасний рівень розвитку програмно-апаратних засобів та розробка нових принципів комп'ютерного зору та аналізу зображень розкривають перспективу створення принципово нового та унікального покоління систем комп'ютерного зору. Наявні дослідження вказують, що пристрої з використанням алгоритмів розпізнавання зображень перспективні для багатьох областей практичних розробок програмних продуктів: моніторингу якості та безпеки об'єктів, промислового та природного походження, контролю змін в екосистемах, експрес-діагностики,

забруднення об'єктів невластивими їм, зокрема і невідомими, сполуками. Також, сучасний рівень розвитку технологій дозволяє використовувати системи машинного зору у розпізнаванні осіб, емоцій на фотографіях та у відеорядах, що дозволяє відкрити нові горизонти у використанні систем безпеки. Тема є маловивченою, оскільки недостатньо багато підприємств та наукових лабораторій займаються питаннями розвитку комп'ютерного зору (висновок зроблено з огляду ринку систем комп'ютерного зору).

Використання систем для розпізнавання образів є одним із основоположних моментів промислової стратегії розвитку для більшості підприємств світу. Про це говорить експоненційне зростання ринку споживання програмних рішень, що використовують технології розпізнавання зображень, що спостерігається останніми роками. Світовий ринок промислового впровадження систем з елементами використання комп'ютерного зору, що становив наприкінці минулого і на початку нинішнього століття близько 8 млрд. доларів, має тенденцію збільшуватися втричі кожні кілька років. Зростання компаній, що працюють у галузі СМЗ, становлять від 15 до 30% щорічно. Проте, наявні комерційні розробки покривають не більше 10% всіх теоретичних і потенційних можливостей систем комп'ютерного зору. Тому варто розвивати та далі програми, які б використовували алгоритми розпізнавання предметів, які б використовувалися на підприємствах різної спрямованості: медичної, безпеки.

Об'єктом дослідження даної роботи є область знань комп'ютерний зір. Предметом дослідження обрано частину, яка досліджує можливість комп'ютерного зору у його можливостях розпізнавати емоції людей при допомоги аналізу фотографій.

Незважаючи на постійне зростання ринку систем комп'ютерного зору, опрацьованість даної тематики залишається слабкою, цьому є низка проблем, які, на думку деяких експертів, не буде вирішено найближчим часом. Це чинить тиск на розробку та вивчення подібних систем. А така

область як розпізнавання емоцій є ще менш опрацьованою, ніж, наприклад, просто розпізнавання осіб. Це пов'язано здебільшого з несерйозністю ставлення до цієї ідеї і, звичайно ж, відсутність єдиних стандартів розробок, відсутність єдиної бази знань з тематики виявлення та розпізнавання осіб. Ще позначається відсутність зручних і зрозумілих інтерфейсів і дорожнечаявних програмних продуктів при відносно невеликому функціоналі.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Аналіз сучасного стану в області

Як самостійна дисципліна комп'ютерне бачення сформувалася в середині попереднього століття. Даний напрямок виник в момент дослідження штучного інтелекту і в період, коли були суперечки про можливість або неможливість створення «мислячої» машини та штучного інтелекту. З робіт з розпізнавання образів, виділилася окрема дисципліна – комп'ютерний зір саме в той момент, коли була усвідомлена його математична специфіка, що зображення (фотографія) є двовимірною проекцією тривимірного світу, тобто «вироджене» перетворення. Отже, для правильної та повної інтерпретації зображення необхідно мати додаткову інформацію.

Після порівняння комп'ютерного та людського зору, виділили кілька способів отримання вторинної інформації про предмети: стали розглядати активний огляд, використання стереопари, апріорних моделей побаченого, змістовних, семантичних, а також геометричних. Розгляд був в основному «на папері» оскільки відповідної технічної бази ще винайшли, а математичні моделі, що виходили, були занадто складні.

Використання фотозображень для вимірювань почали використовувати задовго до поширення комп'ютерної техніки. Цей підхід пропонувалося використовувати, і він був використаний для аерофотозйомки, а пізніше – в основному для космічної навігації та інших завдань. Але для вирішення інших практичних завдань використовувалися та інші способи отримання додаткової інформації: у різних видах штучна організація середовища, наприклад, спеціальне підсвічування (щільне підсвічування), використання різного роду маяків та маятників (маркування деталей на конвеєрі або триангуляційні вежі при аерофотозйомці), а також стандартне розташування предметів.

Перші завдання та результати були отримані у 70-х роках минулого століття. Завдання були визначені на початку 70-х років, а перші результати у другій половині 70-х. Перший етап робіт з комп'ютерного зору став найскладнішим. Виведення зображення для перегляд, якое сформованого в пам'яті ЕОМ, було неможливим до винаходу системи Графор.

Вона дозволяла виводити на папір за допомогою графобудівника контурний малюнок або графік, вони дуже відрізнялися від зображень, які ми маємо на екранах своїх персональних комп'ютерів. Дещо краще була ситуація і для ЕОМ, які могли виводити зображення на дисплей (М-6000), але і на цих машинах виводилися лише контурні чи точкові зображення. Обробка зображень і моделювання тривимірних об'єктів, предметів та сцен здійснювалася за допомогою самописного та нестандартного матабезпечення, тому комп'ютерне моделювання та технічне макетування на той момент були дуже спрощеними, хоч і вимагали зусиль.

Якісний ривок у галузі комп'ютерного зору стався у 1990-х – 2000-х роках. Це стало можливим завдяки появі суперкомп'ютерів. Але в цьому періоді нічого нового у сфері перевороту усталеної дисципліни не було, стрибок був обумовлений постійним інтересом до комп'ютерного зору і, в результаті, спробами перенесення методів з інших дисциплін (наприклад: математична статистика, штучний інтелект, фотограмметрія та інші), що вирости, в комп'ютерний зір.

Варто зазначити, що першим результатом, який став видимим після такого прориву, з'явився прогрес у сфері розробки способів опису зображень.

Використання згортки зображень за допомогою вейвлетів та фільтрів Габора – дозволяли описувати зображення у компактній формі, зменшуючи їх розмірність. Це дозволило відкрити цілий напрямок: пошук баз даних зображень. Більше того, оскільки тепер опис цікавих областей зображення став удосконаленим (наприклад, дескриптор та його аналоги), стало можливим реалізовувати щільну тривимірну конструкцію. І її точність

приблизно дорівнює точності моделей, які були отримані при активному лазерному скануванні.

Результати, отримані на початку 21-го століття у фотограмметрії, вже дозволяли займатися побудовою тривимірних моделей у медицині майже в режимі реального часу. Також тривимірні моделі широко застосовувалися і використовуються для розробки та побудови комп'ютерних моделей міст. З програмних продуктів існує програма PhotoSynth. Вона обробляє безліч фотографій пам'яток і намагається створити їх тривимірні моделі. Математичні моделі цих об'єктів виходять не дуже точними, але тривимірні об'єкти дозволяють передати набагато більше інформації, ніж окремі фотографії та дозволяють створювати ефект присутності. Це свідчить про те, що синтетичний світ у цьому контексті ближче, ніж здавалося б.

Також довго не займалися обробкою відео, тому що відчувалася нестача потужності комп'ютерів. Але на початку 2000-х це питання стало активно розвиватися в рамках галузі комп'ютерного зору. Системи відеоспостереження та біометрії тепер стали використовуватися комп'ютери, а також повсюдно стало впроваджуватися програмне забезпечення з розпізнавання осіб. На сьогоднішній день навіть ведуться роботи з аналізу розпізнавання поведінки та дій з відеоінформації.

Злиття таких дисциплін як комп'ютерний зір та комп'ютерна графіка дозволило створити «доповнену реальність» (augmented reality), що складається з відчуттів реального світу, зокрема зображень, та доданих до них уявних об'єктів допоміжно-інформативної властивості.

Наприклад, моделлю доповненої реальності є тривимірна модель міста, яка була згенерована за зображеннями та фотографіями, до яких додані графічні елементи: віртуальні дерева, мости та персонажі. Отже, доповнена реальність дозволяє вносити деякі штучні елементи під час моделювання реалістичних об'єктів.

Доповнена реальність відрізняється від віртуальної тим, що не створює нового світу. Вона може бути використана як для розваги, так і в

науці: там, де потрібен синтез інформації різного походження, її уніфікація, подання компактному вигляді.

На даний момент найбільші пошукові системи (Яндекс, Google) надали можливість навігації містами світу, але на їх сайтах можна знайти не тільки карти міст, але й «прогулятися» кварталами, побачити їх у такому вигляді, як побачив би мешканець цих міст, який знаходиться безпосередньо в тому кварталі, за яким ведеться спостереження. Цілком можливо, що за допомогою доповненої реальності незабаром стане можливим отримувати інформацію про будівлі, у т.ч. візуалізувати їх, використовуючи знімки фасадів, а також створювати вигляд із вікна будь-якого поверху.

На сьогоднішній момент теорія комп'ютерного зору як самостійного розділу кібернетики склалася повністю, зараз вона спирається на велику практичну та наукову базу знань. Щорічно з цієї тематики проводяться конференції та симпозіуми, випускається безліч різноманітних програмних та програмно-апаратних комплексів, видаються безліч книг та пишуться наукові статті з цієї тематики. Є кілька наукових організацій, які висвітлюють, підтримують та беруть участь у дослідженнях у галузях сучасних інформаційних технологій, у тому числі системах комп'ютерного зору. Зокрема, до них належать: SPIE (Міжнародне співтовариство з оптичної інженерії), IEEE Computer Society (Institute of Electrical and Electronics Engineers), та ряд інших організацій.

1.2 Задачі комп'ютерного бачення та області його застосування

Головними завданнями, що ставить перед собою машинний зір, є:

- виявлення об'єктів;
- їхнє розпізнавання;
- відстеження об'єктів на статичному зображенні та/або увідеопотоку

при певних заданих умовах.

Стандартне завдання в машинному та комп'ютерному зорі та обробці зображень – це виявлення характерних об'єктів, будь-яких характеристик або дій на відеоданих. Зір людини здатний впоратися з нею без особливого навантаження та за практично будь-яких умов. При цьому варто відзначити, що відсоток розпізнавання об'єктів у нього буде досить високий. Проте в області комп'ютерного зору розпізнавання випадкових об'єктів у випадкових ситуаціях залишається невирішеною проблемою, з якою ще належить розібратися.

Безумовно, вже в наявності різні методи, які дозволяють знаходити об'єкти і справляються з цим успішно. Але вони працюють тільки для конкретних предметів: прості геометричні об'єкти, різного виду символи, людські обличчя та автомобілі. При цьому їм потрібні певні умови для якісного розпізнавання. Як правило, це стосується освітлення, фону та розташування об'єкта щодо детектора. Подібні проблеми широко висвітлено у літературі.

Термінологія, що використовується:

- розпізнавання – це виявлення попередньо заданих об'єктів або класів, як правило, зі своїм дво- або тривимірним положенням на зображенні.

Також використовується, коли програма заздалегідь навчена характеристикам об'єктів;

- ідентифікація – це розпізнавання будь-якого одиничного об'єкта та співвідношення його з певним класом;

- виявлення – це перевірка відеоданих відповідно до будь-яких умов, обговорених заздалегідь. Воно іноді використовується для первинного аналізу, оскільки здатне працювати швидко, ґрунтуючись на простих обчисленнях. Після знаходження потрібної ділянки у відеоматеріалі або зображенні, вони передаються для аналізу найсерйозніших методів, які використовують великі ресурси і відтворюють правильні інтерпретації.

Можна виділити кілька особливих завдань, які концентруються на розпізнаванні конкретно текстів:

- пошук зображень за змістом: процес виявлення всіх зображень з бази, в яких закладено певні характеристики, загальні для групи та важливі для знаходження;

- оцінка положення: ідентифікація положення конкретного об'єкта, його напрямки або орієнтації щодо об'єктива камери;

- оптичне розпізнавання знаків: ідентифікація потрібних символів, друкованих та рукописних, на зображеннях з метою їхнього трансферу в текст, який можливо буде редагувати, індексувати та в цілому зручніше працювати (наприклад, ASCII);

- відновлення сцени. Коли є два або більше зображення сцени, а також відео. Його завдання – відтворити тривимірну модель сцени. Для початку моделлю може бути набір точок тривимірного простору. Більше складні методи відтворюють повну тривимірну модель.

- відновлення 3D форми за 2D зображеннями: таке можливе при використанні реконструкції та стерео реконструкції поля нормалей та карти глибини, з орієнтацією на забарвлення напівтонового зображення; також реконструкції карти глибини за текстурою; визначення форми переміщення.

- відновлення зображень. Як правило, тут вирішується проблема наявності шуму (шум записуючого пристрою, розмитість тощо). Найлегшим способом вирішення цього завдання є використання фільтрів, наприклад, нижніх та середніх частот. Цього достатньо для базового покращення зображення. Для більш серйозної роботи використовується аналіз даних на наявність ліній або кордонів, а потім на основі зібраних даних проводиться більш глибоке виправлення зображення;

- виділення на зображеннях структур певного виду, сегментація зображень;

- аналіз оптичного потоку. Він пов'язаний із ідентифікацією переміщення пікселів між двома зображеннями. Також він здатний

вирішити завдання, пов'язані з оцінкою руху в цілому, де послідовність зображень обробляється таким чином, що стає можливим створити оцінку швидкості кожної точки зображення або 3D сцени. Яскравими прикладами даного аналізу є: слідування камери за переміщеннями тварин, людей або машин, а також визначення тривимірного руху камери. Конволюційна нейронна мережа (CNN) [1], що являється одним із видів глибинного навчання, являється одним з підходів до класифікації та отримання знань з навчальних вибірок, як показано на рисунку 1.1.

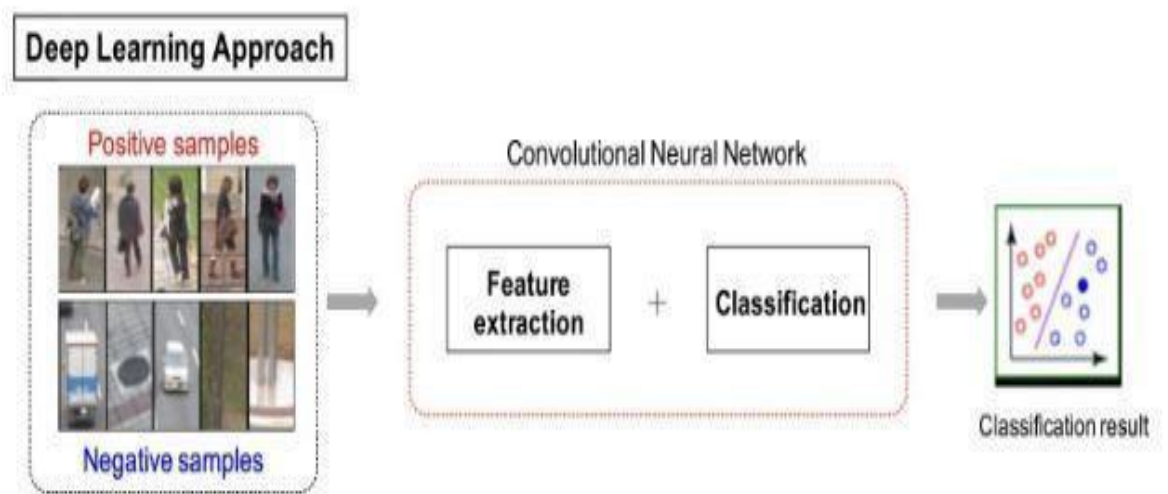


Рисунок 1.1 – Отримання знань в глибинному навчанні

Даний рисунок описує підхід в глибинному навчанні: зосередження на виявленні об'єктів та розумінні сцени.

Як показано на рисунку 1.2, CNN обчислює карту характеристик, що відповідає ядру шляхом згортання ядра на вхідному зображенні.

Кarti характеристик, що відповідають типам ядра, можуть бути порашовані, оскільки існує декілька ядер. Далі розмір карти характеристик збільшується шляхом поєднання декількох карт. Процес поєднання об'єктів виконується неодноразово для отримання однієї карти характеристик.

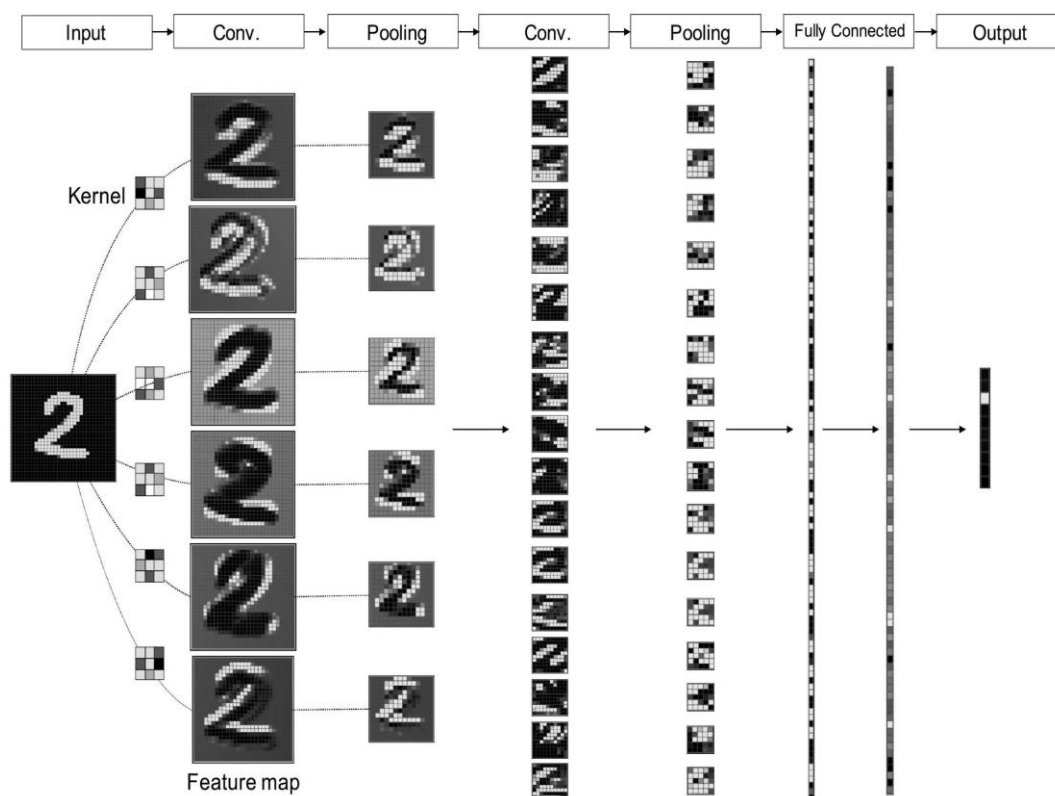


Рисунок 1.2 – Базова структура CNN

Отримана карта характеристик застосовується до повністю поєднаних слоїв та ймовірність кожного класу є виводом. В цьому випадку вхідний та вихідний рівні мають структуру мережі що має частини зображення та велику кількість класів.

Навчання CNN досягається оновленням параметрів мережі методом зворотного розповсюдження. Параметри CNN стосуються ядра конволюційного слою та всіх об'єднаних слоїв. Процес методу зворотного розповсюдження показаний на рис. 1.2.3. Спочатку, навчальні дані вводяться до мережі використовуючи нинішні параметри для отримання прогнозів. Із прогнозів та навчального слою обчислюється похибка і кількість оновлень кожного параметру виходить з похибки. Мережа оновлюється від вихідного слою до вхідного(зворотне поширення). Навчання CNN залежить від повторення даних процесів для отримання якісних параметрів, що допоможуть розпізнати зображення коректно.

На рисунку 1.3 ми бачимо кілька прикладів візуалізації ядер на першому конволюційному рівні «AlexNet», який призначений для завдання класифікації тисячі об'єктних класів.

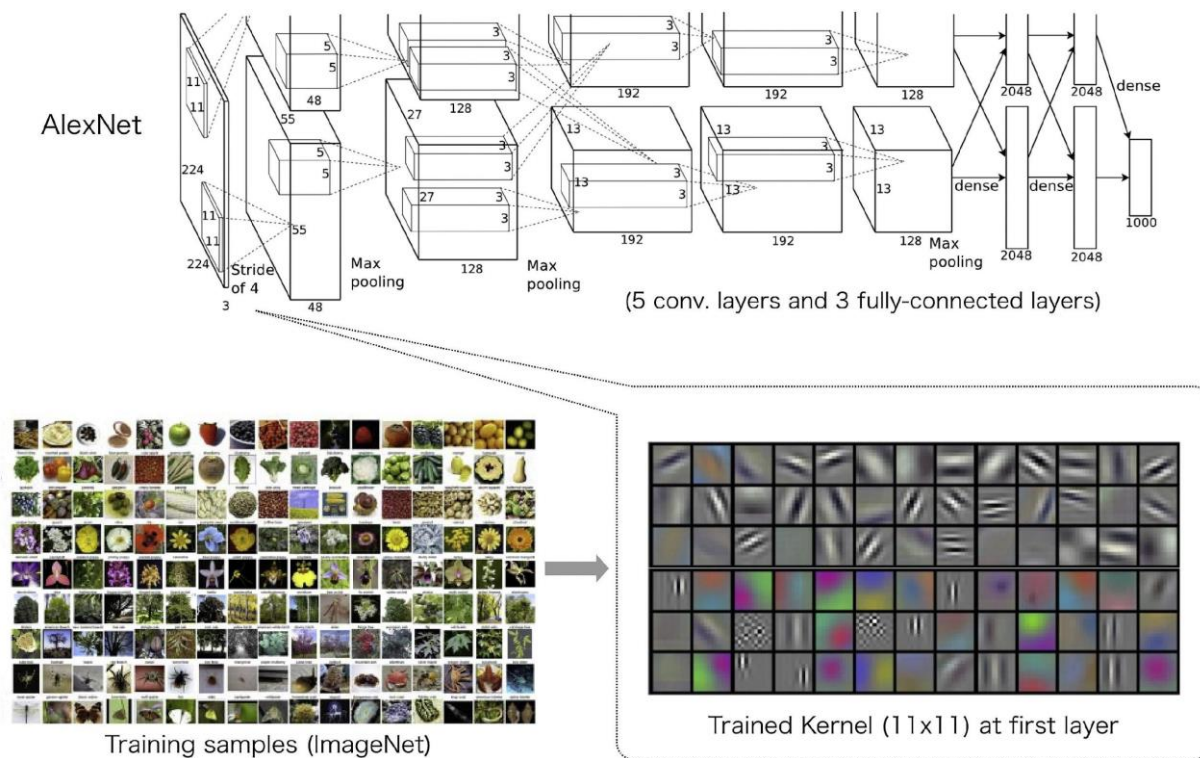


Рисунок 1.3 – Структура мережі AlexNet

AlexNet складається з п'яти конволюційних шарів та трьох повністю з'єднаних шарів, вихідний шар яких має 10 000 частин, що відповідає кількості класів. Ми бачимо, що AlexNet автоматично підібрав різні фільтри, які отримують інформацію про краї, текстуру та колір з компонентами спрямованості. Існує дослідження в якому спостерігали ефективність фільтра CNN, у порівнянні з гістограмою направлених градієнтів, в задачі виявлення людей. Кількість помилок у CNN склала 3%, в той час як у HOG – 8%. Хоча ядро AlexNet не треноване для вправи знаходження людей, все одно точність знаходжень є вищою за HOG, яка є функцією «ручної» роботи.

1.3 Переваги та недоліки систем комп'ютерного бачення

Категорія засобів машинного зору охоплює надзвичайно широку номенклатуру пристроїв від недорогих відеодатчиків, які, втім, стають все більш складними, до систем із широкими функціональними можливостями, мають комп'ютерні модулі формування зображень, швидкісні відеокамери та витончене програмне забезпечення.

П'єрантоніо Бор'єро, керівник напряму з компанії Matrox Imaging (Дорваль, Квебек, Канада), відзначає такі тенденції в галузі високопрофесійних систем: «Повсюдно починають застосовуватися все більш складні, але разом з тим і все зручніші в експлуатації засоби. В якості прикладу можна навести програмне забезпечення, що використовує метод розпізнавання геометричної структури. Добре працюючий у регульованих середовищах ахроматичний аналіз дуже важко реалізується на виробництві. Метод розпізнавання геометричних структур відрізняється куди вищою пристосованістю до використання у несприятливих умовах, а створюване на його основі програмне забезпечення розраховане на користувачів, які не мають наукового ступеня в області обробки зображень. У технічному оснащенні відеокамер починають переважати такі швидкісні послідовні інтерфейси, як IEEE 1394 (Firewire), CameraLink, а також USB 2.0. Використання цих інтерфейсів призводить до зменшення кількості кабелів та спрощує організацію взаємодії камер із системою».

Також, СКЗ мають досить серйозні переваги перед людським зором. Системи комп'ютерного зору, обладнані сенсорними специфічними пристроями та датчиками, можуть побачити те, що недоступне простому неозброєному оку. Збільшення та розширення спектрального діапазону сприйняття СКЗ за межі сприйняття людським оком розширює функціонал та можливості систем комп'ютерного зору значно. Прикладом пристроїв перевершують за характеристиками людський зір є системи для забезпечення безпеки, телевізори та прилади нічного бачення. Важливою

особливістю таких систем є те, що вони забезпечують перетворення невидимого уявлення об'єктів у доступний для сприйняття людини діапазон, тим самим збільшуючи можливості людини у оперативному прийнятті рішень.

Основним недоліком систем комп'ютерного зору, який помітно обмежує зростання ринку цих систем, вважається відсутність єдиних стандартів на програмні інтерфейси та обладнання. Системи різних виробників погано інтегруються, потім постійно скаржаться користувачі, кожна інтеграція таких систем обходиться дуже дорого, оскільки на цьому етапі розвитку, кожна така інтеграція унікальна.

Концепція мультиспектральних систем комп'ютерного зору виглядає перспективною. Вона використовує обробку зображень не тільки в області видимого людиною спектру, але і спектру, який може «засікти» за допомогою високотехнологічних пристроїв (лазерних установок, інфрачервоних камер тощо).

Прикладами таких систем можуть бути телевізори в охоронних системах або, наприклад, під час аналізу якості кондитерської випічки. Але основна перешкода цим системам комп'ютерного зору є висока вартість та складність в експлуатації необхідних датчиків. Складне та тривале калібрування систем з використанням комп'ютерного зору, якщо налаштування відбувається на певну предметну область.

Дуже часто ця процедура вимагає потужних обчислювальних ресурсів та великих витрат часу навчання нейронної мережі. Наявне програмне забезпечення у даній області дуже часто відрізняється незручним інтерфейсом користувача і дуже низькою гнучкістю систем, адже переналаштування системи в режимах реального часу дуже часто є неможливим.

Вимоги споживачів на даний момент до систем, що використовують комп'ютерний зір, такі, щоб аналіз відбувався не поштучно, а оцінюючи

відразу всі об'єкти, які в даний момент часу знаходяться в полі видимості системи.

На даний момент відчувається брак орієнтованих на комп'ютери математичних алгоритмів, які були б орієнтовані на комп'ютери та підтримували б паралельну обробку. Також відкрите питання розпізнавання однакових, але по-різному освітлених об'єктів. Є думка, що з цією проблемою вдасться впоратися у найближчому майбутньому. Відштовхує від придбання таких систем невелика дохідність систем та мала кількість хороших продуктів. У більшості випадків, набагато дешевше найнятиме дешевий персонал для спостереження за виробничим процесом, ніж запровадити систему комп'ютерного зору.

Це є тимчасові проблеми. Вже йдуть перемовини щодо застосування єдиних стандартів, після чого у планах є створити загальнодоступний репозиторій знань та типових помилок у продуктах, там же будуть визначатися програмні інтерфейси, так само планується розробка угоди з мов описання та подання даних.

1.4 Огляд сучасних систем розпізнавання емоцій людини

На даний момент основний напрямок розвитку впроваджуваних у життя людини нових інформаційних технологій – це покращення людино-машинної взаємодії. Варто відзначити, що кількість заходів, присвячених цьому напрямку, велика. Даний напрямок є дуже перспективним, про що говорять щорічні фінансові вливання та щорічне збільшення числа лабораторій, що займаються розробкою систем машинного зору.

Створення роботів, здатних замінити у деяких галузях реальних людей неможливо без подолання бар'єру взаємин «людина – машина». Нездатність машин відчувати і виявляти емоції є одним із факторів, який стримує розвиток цієї галузі.

Ця проблема стає все більш актуальною з розвитком інформаційно телекомунікаційних технологій, які все більше і більше витісняють живе

спілкування. У спілкуванні в мережі люди намагаються висловити емоції за допомогою наборів символів та/або смайликів. Але віртуальне спілкування дедалі більше обмежує наш прояв емоційних реакцій. Ці реакції впливають на когнітивні процеси (у тому числі на процес прийняття рішень).

Розглядаючи системи, де можна визначити передаються чи розпізнаються емоційні реакції вирішено було розбити їх на групи за типами прояву емоційних реакцій:

- за голосом;
- за фізіологічними показниками;
- з міміки;
- за рухами тіла (пантоміміка).

Визначення емоційних реакцій за фізіологічними показниками:

- біологічний зворотний зв'язок (БЗЗ). Використовується при клінічному визначенні емоцій. Ідея та суть методу БЗЗ-терапії полягає в тому, що пацієнт бачить на екрані комп'ютерного монітора або чує в аудіо-формі поточні значення його фізіологічних показників, що визначаються клінічним протоколом: ЕЕГ (електроенцефалограма головного мозку), кардіограма, провідність шкіри, дихання, температура, ЧСС (частота серцевих скорочень), електроміограма, температура, фотоплетизмограма та ін;

- пристрій від NeuroSky містить датчик мозкової активності та надягається на голову – дослідний зразок. Він визначає ступінь розслаблення або занепокоєння, а також концентрації людини, даючи їм оцінку за шкалою від 1 до 100. Це пристрій призначений для підключення до персонального комп'ютеру або ігрових приставок;

- vibraImage – система контролю психоемоційного стану людини. Її головне призначення полягає у виявленні потенційно небезпечних і агресивних людей, за допомогою дистанційного та безконтактного сканування з метою забезпечення безпеки в об'єктах і аеропортах, що охороняються;

- технологія, що використовує спеціальний датчик. Винахід Токійського університету це спроба використовувати «фізичну» передачу емоцій з використанням стільникового телефону.

Визначення емоційних реакцій за мімікою:

- визначення основних емоцій на фотографіях за допомогою програми сервісу FaceReader. Програма може визначити такі емоції: сум, здивування, щастя, агресія, переляк, нейтральне вираження;

- розробка від Політехнічного університету Мадриду та Університету короля Хуана Карлоса – система для розпізнавання та класифікації виразу осіб у режимі реального часу. Вона розподіляє емоції за такими групами: щастя, страх, огида, страх, смуток, подив, щастя;

- система від компанії GFK, для запису міміки людини в автоматичному режимі. Система розроблена для спроби утримати розуміння засобів вираження емоцій людини та середовище їх виникнення;

- auto Smiley – невелика програма для платформи MacOS, вона автоматично генерує «смайлики» в тексті, що набирається у вікні текстового редактора або інтернет-месенджера;

- прилад компанії Media Lab. Складається з мініатюрної камери, прикріпленої до окулярів, яка з'єднана з КПК. Прилад призначений для визначення реакції оточуючих людей на власника устрою. Корисний для аутистів або ораторів;

- truMedia – система відеоспостереження, яка намагається визначити реакцію людини на рекламу. TruMedia стежить, як довго погляд людини спрямований на екран з рекламою та визначає його расову приналежність, підлогу та зразковий вік;

- робот-голова від компанії Machine Perception, може виражати емоції з використанням міміки. Зовнішністю робот схожий на вченого Альберта Ейнштейна і може засмучуватися, сердитись і дивуватись;

- компанія Nokia розробляє систему розпізнавання емоцій та передачу їх за допомогою смартфона. Ця технологія використовує кольорову

різноманітність, кожен використовуваний колір відобразатиме, в якому настрої та емоційному стані буде власник смартфона.

Визначення емоційних реакцій з рухів тіла (пантоміміка):

- проект компанії Microsoft – Natal. програмний комплекс забезпечує повне тривимірне розпізнавання голосу, міміки обличчя та рухів тіла. Natal здатний розпізнавати емоції по обличчю та голосу. Але цей програмний комплекс не забезпечує аналізу рухів;

- технологія Sony Computer Entertainment America. Направлена на визначення таких емоцій, як туга, сміх, радість, гнів, печаль. Розпізнає емоції за голосом та рухами. Немає можливості тривимірного моделювання;

- система визначення емоційного стану з міміки та жестикуляції від дослідників Natice Gunes, Massimo Piccardi та Maja Pantic з групи комп'ютерного зору (Computer Vision Research) у Сіднеї. Рухи ніг, нахил голови та торса не враховуються;

- система визначення емоційного стану по жестикуляції та міміці від дослідників C. Shan, S. Gong, та P. W. McOwan (Факультету Інформатики (Department of Computer Science) Лондонського інституту). Аналізується лише зміна положення тіла, статичні положення та пози не враховуються.

Визначення емоційних реакцій за голосом:

- програма Emotive Alert від Affective Computing Research Group розпізнає емоції за голосом. Встановлюється безпосередньо на автовідповідач власника та індексує вхідні повідомлення. Класифікує кожне повідомлення з тим або іншим емоційним забарвленням;

- програмне забезпечення для автомобілів від Affective Media. Розпізнає емоційний стан водія під час руху. Технологія дозволяє постійно стежити за емоційним станом водія, аналізуючи його інтонацію, тембр голосу під час запитів до системи навігації автомобіля;

- технологія детектування «ступінь кохання та дружелюбності» в голосі людини від мобільного оператора KTF. Система голосового аналізу

може з'ясувати, чи бреше людина чи говорить правду, чи зацікавлений у розмові чи тема бесіди здається нудною та нецікавою;

- технологія компанії Sound Intelligence. Застосовується лондонською поліцією для оснащення камер спостереження розумними голосовими сенсорами, які зможуть уловлювати агресивну інтонацію у розмові. Якщо пристрій помітить, що хтось із перехожих свариться, воно сповістить про це чергового. Той, у свою чергу, може наблизити зображення за допомогою зуму та з'ясувати, у чому там справа;

- технологія розпізнавання емоцій компанії Federal Express. Використовується, щоб з'ясувати, наскільки хороше чи погане враження склалося у клієнтів. В основі технології лежить верифікація голосу, аналізується гучність звуку та його висота, щоб знаходити у записаних дзвінках слова, що звучать як, наприклад, «вау».

1.5 Методи, які використовуються при розпізнаванні облич людей

Найбільша кількість публікацій, що належать до теми аналізу зображень людей за їхніми фотографіями, присвячені знаходженню оптимальних методів розв'язання задачі першого класу. Існує близько десятка принципово різних підходів до вирішення цієї проблеми (див. рис. 1.4). Основними критеріями при оцінці наведених нижче методів є обчислювальна вартість алгоритмів, побудованих на їх основі, та достовірність розпізнавання.

Наступним, найбільш опрацьованим після методу аналізу антропометричних характеристик особи можна назвати метод власних векторів (іноді його називають методом основних компонентів облич). Він є прикладом того, як математичні методи (метод аналізу основних компонентів), що успішно застосовувалися в інших областях, виявилися ефективно адаптованими до розпізнавання людей за їхніми портретами.



Рисунок 1.4 – Основні підходи, які застосовуються при побудові off-line систем ідентифікації особи

Будь-яке цифрове зображення може бути представлене у вигляді вектора у просторі ознак. Якщо зображення описується $w \times h$ пікселями, розмірність найпростішого векторного простору, до якого даний вектор належить, дорівнюватиме добутку w на h і, відповідно, базис подібного векторного простору буде складатися з $w \times h$ векторів. Проте у зв'язку з тим,

що всі людські особи схожі між собою (овальна форма з носом, ротом, очима і т.д.), всі вектори, що описують зображення обличчя, будуть розміщуватись у вузько обмеженій області даного векторного простору. Тому при вирішенні задачі розпізнавання людей по портрету опис та зберігання всього векторного простору не є раціональним.

Таким чином, постає питання побудови простору меншої розмірності, в якому зображення людських обличчя описуються більш компактно. Одним з варіантів є простір, базовими векторами якого служать основні компоненти всіх зображень осіб, що містяться в ньому. Розмірність такого простору заздалегідь визначити неможливо, але вона набагато менша за розмірність векторного простору всіх зображень. Зі сказаного вище випливає, що головною метою методу аналізу принципів компонентів є значне зменшення розмірності простору ознак таким чином, щоб воно якнайкраще описувало типові образи, що належать безлічі портретів. У разі застосування даного методу для ідентифікації осіб такими іноді будуть навчальні зображення.

Іншими словами, за допомогою аналізу основних компонентів вдається виявити всілякі мінливості в навчальному наборі зображень осіб та описати цю мінливість за допомогою кількох змінних. Ці змінні є $w \times h$ – розмірні вектори, які називаються власними. Якщо перетворити подібні вектори на зображення, то одержувані картинки будуть відображати основні компоненти представленої навчальної множини (також звані власними особами). Таким чином, за рахунок зниження розмірності простору базисних векторів, в якому знаходяться зображення, добиваються хороших показників як швидкості, так і точності розпізнавання.

Імовірнісний підхід (Probability estimation). Подібно до попереднього методу в ймовірнісних моделях також використовується навчальний набір. У цьому формуються два класи з усіх варіантів уявлення об'єктів: внутрішньооб'єктної та зовнішньої мінливості. Відбираються ознаки, за якими всі портрети поділяються на два класи:

- портрет даної людини;
- усі інші портрети.

Функції щільності ймовірності кожного класу оцінюються за допомогою згаданої вище навчальної множини і згодом використовуються для обчислення міри схожості, яка таким чином ґрунтується на отриманих досвідченим шляхом ймовірностей. Крім того, для отримання більш точних результатів іноді використовується імовірнісна модель деякого фізичного процесу, за допомогою якої і формується остаточний захід схожості двох зображень.

У методі зіставлення з еталоном процес розпізнавання розбивається на частини, відповідні окремим рисам особи. Кожна фотографія, що надходить на вхід системи, що розпізнає, повинна являти собою фронтальне зображення особи людини з певною для конкретної бази даних кількістю масок, що представляють основні для ідентифікації регіони особи (наприклад, очі, ніс, рот та нижня частина особи). Крім того, розташування даних масок мають бути однаково нормалізовані (наприклад, щодо очей) для всіх зображень у базі даних.

Під час процесу розпізнавання, коли частини вхідного зображення по черзі порівнюються з частинами зображення, що зберігається в базі, використовується вектор, що відображає результат порівняння в балах (один бал за кожен збігається рису особи) та обчислюваний шляхом нормалізованої взаємної кореляції (втім, методи порівняння можуть бути різними). Після чого вхідне зображення класифікується відповідно до максимально набраних балів. Є також деякі різновиди даного підходу, наприклад, з тими, що змінюються в процесі порівняння зразками.

Принципи функціонування систем, побудованих на нейронних мережах (іноді їх називають авто асоціативною пам'яттю), полягають у тому, щоб у відповідь на деяку вхідну сукупність даних, звану ключом, видати на вихід збережену в мережі і найближчу до вхідної за значеннями сукупність даних такої ж розмірності. У разі розпізнавання осіб ключем

служить обличчя людини. Нейроном називається осередок мережі, яка є найпростішим елементом пам'яті.

Порядок роботи нейронних мереж наступний: 1-й крок – зображення оцифровується та кодується у вигляді вектора; 2-й крок – кожна координата вектора розташовується в окремому осередку, пов'язаному з усіма іншими осередками (навчання або налаштування системи відбувається шляхом зміни ваг зв'язків між осередками); 3-й крок – зображення осіб фільтруються через нейромережу, при цьому вхідне зображення трансформується в найближче запам'ятоване, яке та подається на вихід. На даний момент цей підхід є одним із найбільш популярних. Однак оцінити обчислювальну складність алгоритмів поза обчислювальними машинами з паралельною архітектурою важко. Проте повідомляється про досить ефективне використання нейронних мереж в області аналізу зображень осіб за трьома напрямками: класифікація людей за статтю, безпосередньо розпізнавання та визначення емоційних виразів осіб.

Застосування аналізу оптичних потоків з метою ідентифікації осіб визнано досить ефективним, але дорогим з обчислювальної точки зору і в практичних додатках не використовується. Суть методу така: порівнювані зображення $A(i)$ і $B(i)$ перетворюються на багат шарові усічені піраміди шляхом багаторазового згортання чотирьох сусідніх пікселів в один із середнім арифметичним значенням яскравості. Після завершення цього процесу на відповідних шарах двох різних пірамід проводять пошук відповідних між собою найкращим чином груп пікселів. До кожного блоку зображення $A(i)$ визначається вектор зсуву.

Цей вектор уточнює зсув між центрами блоку $A(i)$ і найбільш близьким до нього блоком $B(i)$. Аналогічно будуються вектори для зображення $B(i)$. Аналізуючи системи векторів, що вийшли, можна зробити висновок про рівень схожості порівнюваних зображень.

Одним із найефективніших методів детектування об'єктів на фото та відео зображеннях, у реальному часі є метод Віоли-Джонса[1]. У даного детектора дуже мала ймовірність помилкового детектування об'єктів.

Алгоритм відмінно виявляє риси обличчя, якщо навіть нахил об'єкта спостерігається під невеликим кутом, орієнтовно до 40 гр. Точність детектування, з використанням даного алгоритму, може досягати шкали понад 90%. Це дуже чудовий результат. Якщо кут нахилу більше 30 гр., то можливість детектування осіб падає. Ця особливість цього методу, зазвичай у стандартній ситуації не дозволяє знаходити особу, якщо вона повернута під великим кутом, це значною мірою може затруднити або унеможливити користування цього методу в сучасних модернізованих системах, якщо врахувати їх великі потреби [3] .

Наступним методом є порівняння шаблонів. Принцип даної методики полягає у відмітці ділянок особи на картинці, та у порівнянні даних ділянок для обох різних малюнків. І будь-який збіг на ділянках підвищує ступінь подібності малюнків. Для звіряння ділянок використовуються найпростіші алгоритми, наприклад, попиксельні звіряння [2].

Слабкість цієї методики полягає в тому, що вона потребує чималих засобів, для запам'ятовування ділянок, а також для порівняння. Також застосовується найпростіший алгоритм звіряння, малюнки мають бути зняті в спеціально встановлених умовах, таких як недопущення особливих змін ракурсу, емоційних виразів, освітлення та ін. Безпомилковість виявлення із застосуванням даної методики показує результат 80%, і це є непоганим результатом.

Також одним з найбільш популярних методів є нейронна мережа Хопфілда. Методика практики системи Хопфілда має відмінну від звичайних способів тренування перцептронів. Різниця в тому, що замість почергового прискорення до необхідного результату з підрахунком помилок і всі ознаки вагової матриці обчислюються за єдиною формулою,

за єдиний цикл, і після цього система вже готова працювати [2]. Ліміти методики:

- образи, що відзначаються, не повинні бути схожі між собою;
- особи не повинні бути зміщені або повернені щодо вхідного вихідного зображення.

Для видалення даних мінусів розглядаються різні зміни типових нейронних мереж Хопфілда. Система Хопфілда з ортогональною модифікацією може відновити образи, які дуже скорельовані, за рахунок перетворення їх вхідної величини до дуальної суми набору. Отже, створюється нейронна мережа, що може записувати певну чисельність векторів, і навіть при відправці на вхід кожного вектора, може з'ясувати, на котрий із записаних він найбільше схожий. Безпомилковість виявлення з використанням цієї методики показує результат вище 90%. У деяких випадках близько 100%, це чудовий результат. Більшість сучасних продуктів виявлення та розпізнавання осіб має завдання зіставлення невідомого зображення осіб з реєстром зображень образів з реєстру даних.

Основні проблеми систем розпізнавання емоцій – зашумлений фон, різниця освітлення, складність зачіски, присутність бороди або окулярів на обличчі. Якщо врахувати всі можливі проблеми та ризики, та створити гібридні методики, то вони покладають на себе великі надії, які використовують плюси та мінуси розглянутих раніше різних відокремлених підходів. І вони могли б показати вищі результати, ніж нинішні методи окремо.

1.6 Огляд існуючих систем розпізнавання емоцій

1.6.1 FaceReader Noldus Information Technology

Ця система може, чітко розпізнавати емоції за виразом обличчя. Система знаходить такі емоції: радість, смуток, подив, злість, страх, огиду, нейтральність. Крім цього, FaceReader (рис. 1.5) може визначити стать, вік,

расу людини. Програмне забезпечення створюється без машинного навчання та додаткових налаштувань. У системі реалізовано стандарти комп'ютерного зору. Зокрема, метод Active Templates, який полягає в накладенні шаблону, що деформується, на фотографію особи.

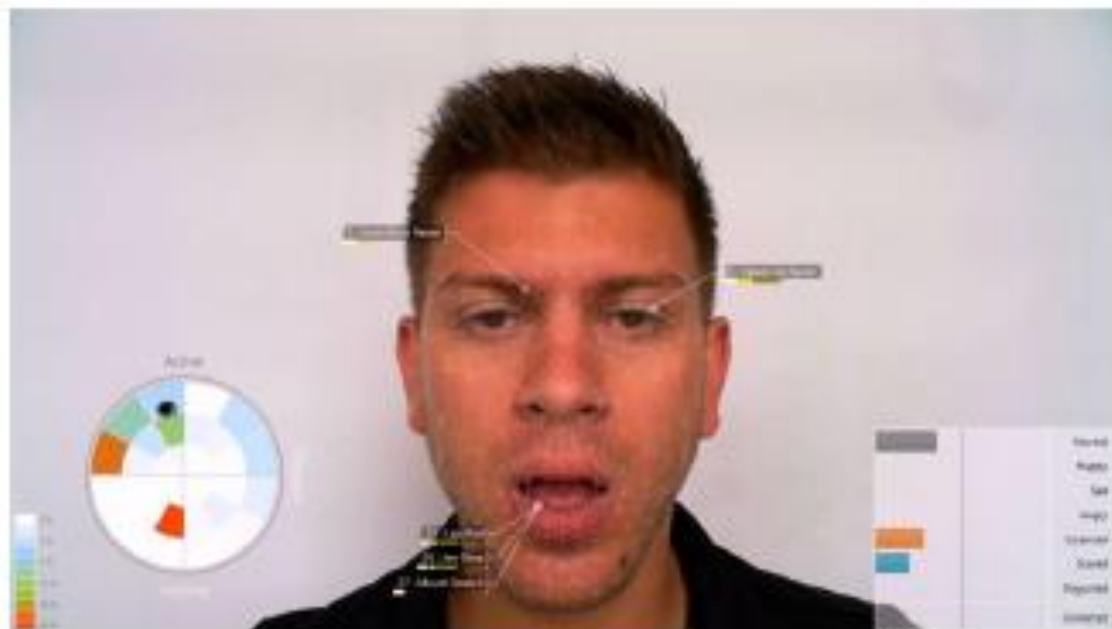


Рисунок 1.5 – Інтерфейс Face Reader

Переваги системи:

- середній відсоток розпізнавання емоцій цією системою дорівнює 89%, також не залежить від повороту та нахилу облич;
- програмне забезпечення підтримує безліч форматів відео - файлів, з вкладеннями MPEG1, MPEG2, DV-AVI, DivX5, DivX4, DivX6. Розпізнавання емоцій відбувається покадрово, або загалом під час перегляду відео – файла;
- система працює не тільки із завантаженим відео, FaceReader так само працює і із статичними і динамічними зображеннями, тобто в реальному часі, які беруться зі зчитування даних з камери пристрою користувача [3];
- система має такі можливості візуалізації: наявність діаграм, гістограм, присутня сітка, що накладається, відсотки виражених емоцій.

Недоліки:

- відсоток визначення віку та емоції дуже сильно падає при розпізнаванні осіб дітей віком менше 5 років;
- виявлення емоції може бути неточним, при носінні людиною окуляри;
- програмне забезпечення не знаходить особи у профіль.

1.6.2 EmoDetect

EmosDetect – це програмне забезпечення, що дозволяє визначати психоемоційний стан людей, на їх зображення (у відео, або фотографії). Даний емоційний класифікатор (рис. 1.6) може визначити шість головних емоцій, таких як радість, сум, подив, страх, злість, огида. Емоції розпізнаються з допомогою нейронних мереж. Це відбувається так, на обличчі людини знаходяться ключові точки, далі класифікатор повинен побудувати цілісну картину на основі положення точок відносно один одного. Компанія Нейроботикс реалізує системи, що досліджують людську фізіологію та фізіологію тварин. Система використовує ці знання, створює технології в галузі сенсорики у робототехніці, біоуправлінні, біомеханіці [3].

Переваги:

- здійснює специфікацію емоцій трьома вільними реєстрами такими як система вирішальних законів, нейронна мережа, класифікатор за зваженою кількістю сигналів;
- може побудувати графік динаміки перетворення емоцій, що перевіряється у часі; також формує реєстр, про результати обробки відеофайлів;
- програма має можливість записувати відеофайли з веб-камери;
- робить аналіз усіх форматів, які підтримуються ОС.

Недоліки:

- програмне забезпечення не визначає особи у профіль.



Рисунок 1.6 – Інтерфейс системи EmoDetect

1.6.3 Face Security

Дана система сканує потік даних і якщо особи знаходяться в кадрі, порівнюється з готовою базою даних, на наявність схожості (збігу). У разі знаходження схожості, програма відправляє повідомлення операторам, все це відбувається в реальному часі. Ця програма має широку поширеність у використанні в системах безпеки країн Європи, для детектування підозрілих осіб, в громадських місцях. Великими компаніями використовується для ідентифікації значних клієнтів [1]. Рядовими фірмами система

використовується для виявлення та розпізнавання осіб своїх співробітників (рис. 1.7), а також для виявлення незнайомих відвідувачів, тут використовуються бази даних із фотографіями своїх працівників. Також ця програма може допомогти у слідстві, ідентифікації особи у місцях злочину за зображеннями на фотографії та відеоспостереження, шляхом зіставлення зображень осіб у реєстрі агенцій [7].



Рисунок 1.7 – Інтерфейс системи Face Security

Переваги:

- система працює у реальному часі. Має можливість стеження за детектованою особою, не тільки на одному потоці даних одночасно;
- при скануванні режиму реального часу виявлені особи порівнюються із зображеннями, заздалегідь занесеними до реєстру;
- реєстрація при нерухомому зображенні;
- ручне або автоматичне керування відеопотоком;

- можливість лупи, тобто масштабування в межах «контрольного реєстру»: кількість, розмір, відеопоток та кількість відомих осіб на зображенні.

Недоліки:

- не має можливості виконання обчислень для окремих осіб, які мають відхилення від звичайного фронтального положення більш ніж на 15 градусів.

- не може розпізнавати обличчя, якщо на нього потрапляють тіні та відблиски.

1.6.4 Microsoft Oxford Project Emotion Recognition

Принцип роботи системи Emotion Recognitions аналогічний перерахованим вище системам: за допомогою машинних алгоритмів система аналізує наявність осіб у вхідному зображенні, після цього визначає по міміці, ймовірні емоції людини (рис. 1.8).



Рисунок 1.8 – Інтерфейс додатку Emotion Recognition

Переваги:

- обчислює відсотки емоцій, що виражаються;
- доступний Emotion API, який можна використовувати в різних програмах.

Недоліки:

- визначає емоцій лише на статичних зображеннях;
- програма не може розпізнавати особи, які розташовані під кутом нахилу, залежно від фронтального положення (40 градусів).

1.6.5 eMotion Software

Система реалізована у Нідерландах, компанією Visual Recognition. eMotion Software – це система, яка стала відомою тим, що розробники системи визначили емоції на шедеврї Леонардо Да Вінчі «Мона Лізи». Кінцевий результат був таким: Мона Ліза відчувала 83% щастя, 9% вона висловлювала огиду, 6% страху, на 2% вона гнівалася. Також додаток популярний тим, що це перший комерційний додаток, і є платним «коробковим» рішенням. У зв'язку із цим рішенням група засновників запустила веб-ресурс GladsOrSad.com. Система Visual Recognition взяла на себе лідерство у відкритті порталу, присвяченого визначенню емоцій в онлайн режимі. Інтерфейс даного додатку додається на рисунку 1.9.

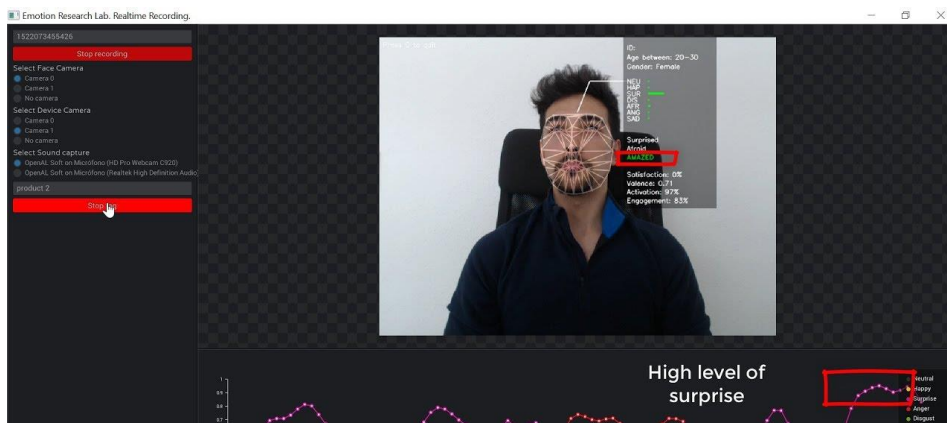


Рисунок 1.9 – Інтерфейс додатку eMotion Software

У цій програмі ведеться відстеження і ведеться ідентифікація тих самих емоцій, злість, страх, смуток, подив, огида, щастя. ПЗ не дуже вимогливий до технічних та системних характеристик комп'ютера. Про деталі розробки алгоритму нічого невідомо, оскільки методика тримається у таємниці [5].

1.6.6 MMER_FEASy

Система була розроблена у Німеччині, компанією MMERSystems. У реалізації цієї системи був використаний метод вставки на особу людини встановленого деформованого шаблону та перевірки на сумісність. ActiveAppearanceModel methodology це методологія, яка може вираховувати необхідні ознаки в онлайн режимі. Програма розпізнає шість основних почуттів, також має можливість визначення особи вік, расу і стать. Додаток ідентифікує особистість людини, якщо раніше фотографія особистості була введена в реєстр системи. Робота з маскою представлена на рисунку 1.10.



Рисунок 1.10 – Робота з маскою MMER_FEASy

Також у додаткових можливостях системи є синхронізація інших програм, на даний момент програма виступає в ролі опції віддаленого помічника, або помічників водія транспорту, або домашнього мультимедійного сервісу та маркетингових досліджень [6].

До недоліків системи можна віднести неповне охоплення завантажених даних, оскільки робота можлива тільки з веб-камерою. Незадовільні результати, тобто замість шаблону, вставляється та ж особа, але з іншою емоцією [8].

2 КОМБІНОВАНИЙ МЕТОД РОЗПІЗНАВАННЯ ЕМОЦІЙ

2.1 Метод на основі локалізації ключових точок на обличчі

Мета створення комбінованого методу у тому, щоб поліпшити точність та швидкість систем розпізнавання емоцій. Методи розпізнавання об'єднуються в чотири великі класи, такі як: холістичні методи, локальні методи, методи, що обчислюють форму, методи, що обчислюють динаміку. У даній роботі йде обробка статичних зображень, з використанням динамічних методів. Були взяті за основу наступні методи: метод на основі локалізації ключових точок на обличчі (локальний метод) [10], Метод на основі інформації про текстуру (Холістичний метод), Геометричний метод рухових одиниць особи (метод, що обчислює форму, локальний метод) У ролі класифікатора емоцій був використаний класифікатор SVM (локальний метод).

Міжнародний стандарт ISO14496 надає детальну інформацію щодо параметрів рухових одиниць обличчя Специфікація цього стандарту визначає 68 анімаційних параметрів особи. Вони є повним базовим набором лицьових рухів і, отже, дозволяють уявити найбільш природні вирази обличчя. У блоці контрольних точок описуються координати ключових точок обличчя, і кожній точці надається свій номер (рис. 2.1).

Також у стандарті описуються вимоги до:

- до фронтального фотографування;
- до параметрів цифрового образу для отримання переднього фотографування;
- до параметрів цифрового образу повного переднього зображення;
- до форматів запису даних для повного фронтального фотографування особи;

- до форматів запису інформації для основного типу зображення обличчя і т. д.

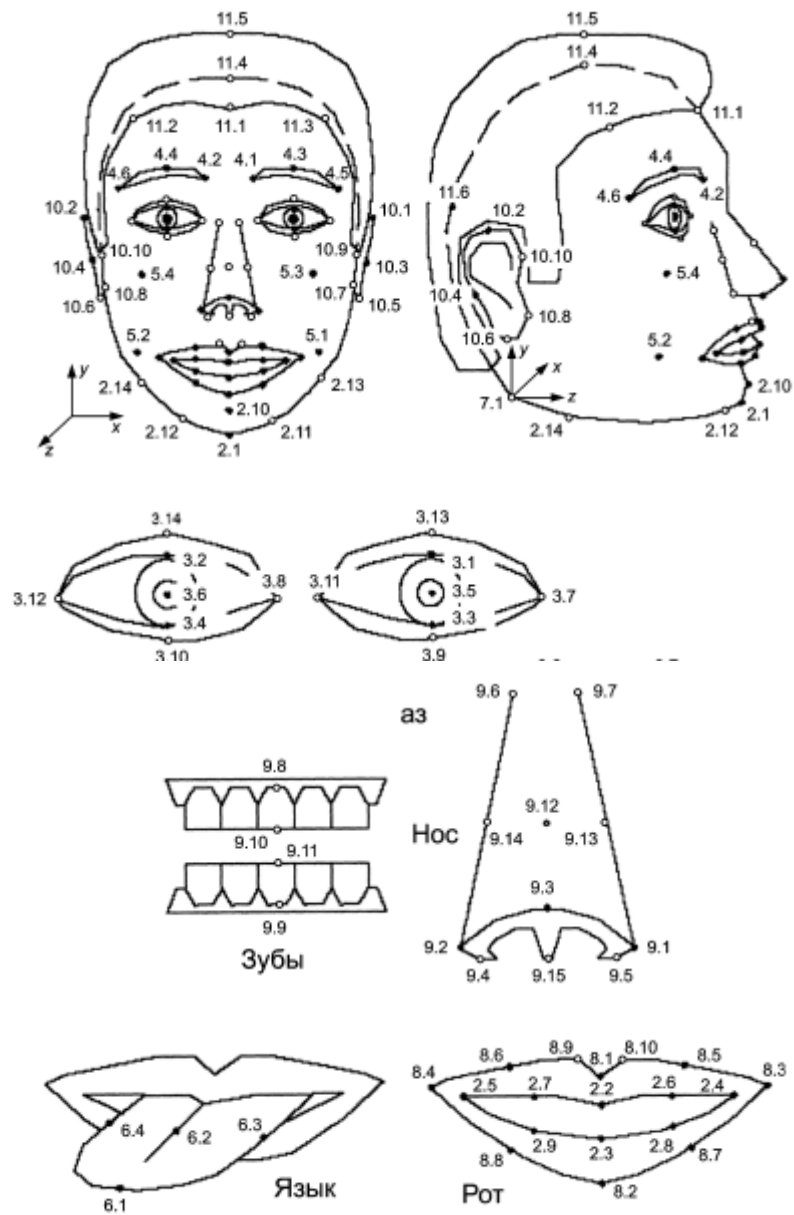


Рисунок 2.1 – Кодування опорних точок обличчя

Вибрані засоби та інструменти для виконання роботи:

- мова програмування Python.

Причини вибору цієї мови:

- кросплатформова мова програмування;

- ця мова підтримує головні парадигми програмування, які використовуватимуться під час виконання завдань;

- великий набір корисних бібліотек.

- бібліотека NumPy.

Ця бібліотека призначена для використання на мові Python, вона додає підтримку великих n-вимірних масивів, а також бібліотеку високорівневих і математичних значень для роботи з даними масивами, вони працюють досить швидко за рахунок застосування вставок мовами: C, і C++.

- бібліотека OpenCV [10].

Бібліотека комп'ютерного зору (з відкритим кодом), застосовується для обробки зображення, охоплює навіть функції машинного навчання. У цій роботі було використано такі функції описаної бібліотеки:

- метод Віоли-Джонса для знаходження осіб [13];

- алгоритми для детектування осіб, таких як LBPH, EigenFaces;

- метод Лукас – Канади для відстеження об'єкта (у нашому випадку – обличчя);

- функції для роботи із зображеннями, їх фільтрування та перетворення.

- бібліотека OpenFace.

Бібліотека з відкритим кодом, яка використовується при розпізнаванні осіб, у якій використовується згортова нейронна мережа, заснована на FacesNet.

- бібліотека Dlib [10].

Бібліотека, що містить у собі методи машинного навчання [14] та різноманітних допоміжних інструментів. Також у цій бібліотеці є готова навчена модель активного зовнішнього вигляду (AAM) [12], у роботі вона використовується для знаходження ключових точок на обличчях людей.

2.2 Метод на основі інформації про текстуру

Метод інформації про текстуру будується на основі зморшок, тіней та інших елементів текстури на обличчі.

Для розрахунку текстурних ознак на сферах особи були використані різні методи та способи фільтрації зображень. Для виявлення дрібних зморшок (таких як область чола) було використано обчислення міри середньоквадратичних розкид S значень пікселів по заданій області. Була обрана область I розміром $m * n$ пікселів (2.1).

$$S = \sqrt{\frac{1}{N-1} \sum_{i=1}^m \sum_{j=1}^n (I_{ij} - \mu)^2}, \quad (2.1)$$

де $\mu = \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^n I_{ij}$.

Для виявлення великих зморшок (наприклад, носогубна складка) було використано двовимірні фільтри Габора [5].

Двовимірні перетворення Габора є гармонійною функцією, згладжена гауссовским вікном. Комплексна гармонійна функція представляється наступним чином (2.2):

$$S(x,y) = \exp(j(2\pi(u_0x + v_0y) + P)), \quad (2.2)$$

де (u_0, v_0) і P – просторові частоти і фаза гармонійної функції відповідно.

Гауссівське вікно виражається таким чином (2.3):

$$\omega_r(x, y) = K \exp\left(-\pi(a^2(x - x_0)_r^2 + b^2(y - y_0)_r^2)\right), \quad (2.3)$$

де (x_0, y_0) – центр вікна;

a, b – параметри, що характеризують форму вікна;

K – амплітуда вікна.

Підсумкова функція Габора виглядає так (2.4):

$$g(x, y) = K \exp\left(-\pi(a^2(x - x_0)_r^2 + b^2(y - y_0)_r^2)\right). \quad (2.4)$$

Значення фільтр та просторові частоти розраховуються в залежності від орієнтації зморшки та його розмірів.

2.3 Метод на основі геометричного аналізу рухових одиниць обличчя

Третій метод який був використаний в даній роботі – геометричний аналіз, він проводиться на основі Системи кодування лицьових рухів, розробленої в 1978 році Полом Екманом і Уоллесом Фризенем. Система описана у 600 сторінках. У цій роботі використовували лицьові рухи, пов'язані з емоційними станами (Таблиця 2.3.1) [15].

Зірочкою позначено інтенсивність кожного руху. Двигуни це головні рухи м'язів, які здійснюються окремими м'язами або декількома. Дескриптори – це рух групових м'язів, наприклад, нижня щелепа випирається. Інтенсивність у системі кодування позначаються латинськими літерами від А до Е і пишуться з додаванням до номера. А – важко помітні, В – непомітні, С – яскраво виражені, D дуже помітні, Е – максимально помітні. По симетричності ДЕ поділяються на двосторонні (bilateral) і односторонні (unilateral-U).

У роботі було використано 68 лицьових рухів, які утворюють 6 базових емоцій.

Класифікація рухових одиниць розраховується за ключовими відстанями (рис. 2.2):

Таблиця 2.1 – Комбінація лицьових рухів [5]

Емоція	Коди
Здивування	1+2+27 5+26 5В+27 2+5В+26 1+2+5+27 1+2+25
Страх	1+2+4+5* 1+2+5*, с/без 25, 26, 27 5*+20* с/без 25, 26, 27 1+2+4+5*+20*+26, 26, или 27 1+2+4+5*+25, 26, или 27 1+2+4+5*+L или R20*+25, 26, или 27
Печаль	6+15* с/без 54+64 1+4+11+15В с/без 54+64 1+4+15* с/без 54+64 1+4+11, 1+4+15В 11+15, 11+17
Радість	6+12* 12С/D
Гнів	3+5*+7+10*+22+26, 25 4+5*+7+10*+23+26, 25 4+5*+7+23+26, 25 4+5*+7+17+25 4+5*+7+23
Відраза	9 8+16+15, 26 10* 10*+16+25, 26

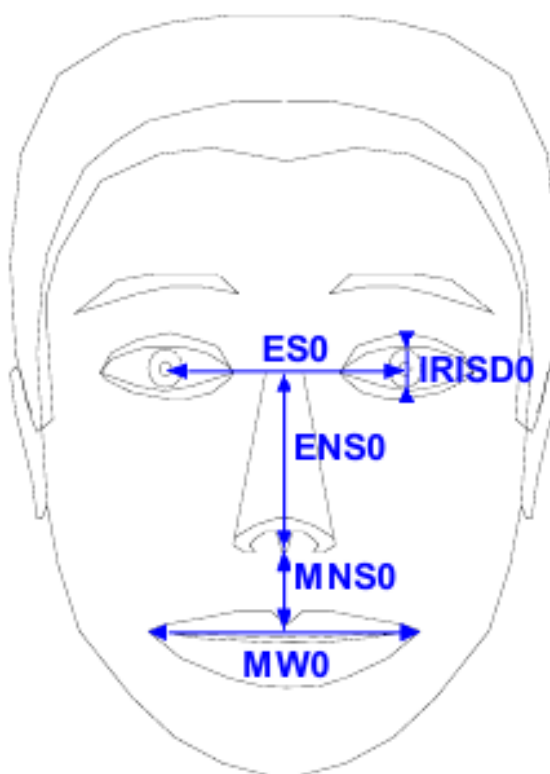


Рисунок 2.2 – Ключові відстані

Основні відстані вказані у таблиці 2.2, де розраховуються відстані та діляться на 1024, щоб одиниця була досить мала, щоб FAP могли представлятися цілими числами.

Таблиця 2.2 – Комбінація лицьових рухів [5]

Опис		FAPU значення
$IRISD0 = 3.1.y - 3.3.y = 3.2.y - 3.4.y$	Діаметр ока (за визначенням дорівнює відстані нижньої точки верхньої повіки і верхньої точки нижньої повіки) у нейтральній емоції	$IRISD = IRISD0 / 1024$
$ES0 = 3.5.x - 3.6.x$	Відстань між зіницями	$ES = ES0 / 1024$
$ENS0 = 3.5.y - 9.15.y$	Довжина носу від відстані зіниць	$ENS = ENS0 / 1024$
$MNS0 = 9.15.y - 2.2.y$	Відстань роту та носу	$MNS = MNS0 / 1024$
$MW0 = 8.3.x - 8.4.x$	Довжина роту (ширина)	$MW = MW0 / 1024$

Нейтральне положення обличчя (коли всі FAR дорівнюють 0) визначається наступним чином:

- система координат права;
- погляд у напрямку осі Z;
- усі м'язи обличчя розслаблені;
- повіки стосуються райдужної оболонки ока;
- губи перебувають у контакті;
- лінія губ горизонтальна та кути губ на однаковій висоті;
- рот закритий.

У комбінованому методі кожен метод може працювати самостійно, і кожен окремо показує непогані результати (85–87%). Але об'єднавши їх, планується отримати метод, який показав би вищі результати. Також планується використовувати комбінацію ще низки інших методів та класифікаторів. У майбутньому також планується робота з динамічними методами, які детектуватимуть об'єкти в реальному часі та визначатимуть емоцій в онлайн режимі.

3 РЕАЛІЗАЦІЯ ПРОГРАМНОГО ДОДАТКУ ДЛЯ РОЗПІЗНАВАННЯ ЕМОЦІЙ

Для реалізації програмного додатку для розпізнавання емоцій була використана бібліотека Tensorflow. Вона являє собою загальнодоступну програмну бібліотеку для машинного навчання, розроблену для вирішення задач побудови та тренування нейронної мережі з метою автоматичного знаходження і класифікації образів, досягаючи якості людського сприйняття.

Для реалізації класифікатору поставленої задачі була обрана бібліотека OpenCV. Вона являє собою відкриту бібліотеку комп'ютерного бачення, яка призначається для аналізу, класифікації та обробки зображень.

Також була використана бібліотека Keras. Вона являє собою бібліотеку для Python, що дозволяє легко та швидко створювати нейронні мережі. Вона сумісна з TensorFlow, Theano, Microsoft Cognitive Toolkit та MXNet. Перші дві платформи найбільш використовуються розробки алгоритмів глибокого навчання, але досить складні в освоєнні. Keras ж, навпаки, є чудовим варіантом для тих, хто тільки починає вивчення нейронних мереж.

В ході реалізації програмного додатку також була використана бібліотека Pillow. Вона була заснована на коді PIL, а потім перетворилася на покращену, сучасну версію. Надає підтримку при відкритті, керуванні та збереженні багатьох форматів зображення. Багато властивостей у ній працює так само, як і в оригінальній PIL.

Був використаний датасет FER2013. Для навчання за даним датасетом було потрібно до чотирьох годин часу.

Програмний код додається в додатку А. Результат роботи можна спостерігати на рисунку 3.1 та рисунку 3.2



Рисунок 3.1 – Зображення з розпізнаною емоцією здивування



Рисунок 3.2 – Зображення з розпізнаним нейтральним виразом
обличчя

ВИСНОВКИ

В роботі показано, як нейронні мережі застосовуються в завданнях розпізнавання зображень, а також які новітні технології розпізнавання зображень з допомогою глибинного навчання впроваджуються. Технологія розпізнавання емоцій людини на зображеннях – проблема пошуку відповідної мап-функції з великої кількості даних.

Розпізнавання є достатньо трудомісткою технологією. Вона потребує потужних графічних процесорів, високопродуктивних відеокарт, великого об'єму пам'яті. Крім того, даній технології потребується більше часу для навчання в порівнянні з традиційним машинним навчанням. Технології глибинного навчання суттєво покращили та спростили алгоритми семантичної сегментації зображень, проложивши путь для більш широкого використання в реальному житті.

Незважаючи на значний прогрес в галузі комп'ютерного бачення, все ще залишається багато відкритих питань. Одним з нагальних питань є застосування таких систем в умовах обмежених ресурсів – безпілотних літальних апаратах, мобільних пристроях, роботизованих та супутникових системах.

Було створено систему розпізнавання емоцій, з урахуванням комбінованого методу. Система була протестована і показала дуже добрий результат розпізнавання – 90,5% точності.

У комбінованому методі кожен метод може бути самостійно, і кожен окремо показує непогані результати (85-87%). Але об'єднавши їх, був отриманий метод, який показав вищий результат.

Надалі планується покращити роботу системи за допомогою збільшення кількості тренувальних зображень, додавання додаткових функцій, таких як: виявлення окулярів, бороди та інших подібних доповнюючих на обличчі. Також планується функція розпізнавання емоцій у реальному часі (через веб-камеру).

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Гудфеллоу Я., Бенджио І., Курвіль А. Глибинне навчання. Москва: ДМК, 2018. 604 с.
2. Tawfiq A., Ahmed J. Object detection and recognition by using enhanced Speeded Up Robust Feature. *International Journal of Computer Science and Network Security*. 2016. Vol. 16, No. 4. P. 66-71.
3. Бабій М.С., Тарановський А.В., Ідентифікація та розпізнавання емоційно-психічного стану людини за зображенням обличчя. *Вісник СумДУ*. 2013. С. 29.
4. Фосайт Д., Понс Ж. Комп'ютерний зір. Сучасний підхід. Видавничий дім «Віл'ямс». 2004. 928 с.
5. AI-facilitated coating corrosion assessment system for productivity enhancement [Електронний ресурс] / IEEE. 2018. Режим доступу до ресурсу: <https://ieeexplore.ieee.org/document/8397787>.
6. Хайкин С. Нейронные сети: полный курс. Киев: Вильямс. 2016. 1104 с.
7. Gabor D. Theory of communication. Part 1: The analysis of information. *Journal of the IEE*. 2005 Vol. 93, No. 27. P. 430–440.
8. Viola P., Jones M. Fast Multi-view Face Detection. Mitsubishi Electric Research Laboratories. 2005. TR2005-097. С. 67.
9. Turk M., Pentland, A. Eigenfaces for recognition / *Journal of Cognitive Neuroscience* 4. 1998. P. 72–88.
10. Batlett M.S., Haget J.C, Ekman P., Sejnowskie T.J. Measuring facial expressions by computer image analysis. Cambridge University Press. public. 2000. P. 254–265.
11. Breazieal P., Washeef A Robots Emotion: A functional perspective. Who Need Emotions: The Brain Meet the Robots. MIT Press, 2003. P. 138–169.

12. Chandran S., Washeef A., Somar M., Debasis M. *Facial Expressions: A Cross Cultural Study / Emotion Recognition: A Pattern Analysis Approach*. Wiley, 2016. P. 89.

13. Coots T. R., Edward G. J. Active appearance model. *British Machines Vision Conference*. 2011. №5(3). P. 480-488.

14. Kanade T., Cohn J. F., Tian Y. Comprehensive database for facial expression analysis. *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*. Grenoble, France, 2001. P. 47–56.

15. Lee Y. O., Shim J. A. Curvature based human face recognition using depth weighted Hausdorff distance. *Proc. ICIP 3 2013*. P. 1430–1433.