

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЧЕРКАСЬКИЙ ДЕРЖАВНИЙ БІЗНЕС-КОЛЕДЖ
кафедра комп'ютерної інженерії та інформаційних технологій

КВАЛІФІКАЦІЙНА РОБОТА

на тему

Дослідження алгоритмів аналізу виразів обличчя на зображенні.

Виконав: студент групи:
№ 1к-20
спеціальності
123 Ком'ютерна інженерія
(шифр і назва спеціальності)

Денис ПРОТАСОВ
(ім'я та прізвище)

Керівник: Наталя ФАЛЬЧЕНКО
(ім'я та прізвище)

ЧЕРКАСЬКИЙ ДЕРЖАВНИЙ БІЗНЕС-КОЛЕДЖ

Кафедра комп'ютерної інженерії та інформаційних технологій

(повна назва випускової кафедри)

Спеціальність 121 Інженерія програмного забезпечення

(шифр і назва спеціальності)

Освітня програма інженерія програмного забезпечення

(назва освітньої програми)

ЗАТВЕРДЖУЮ

Завідувач кафедри
комп'ютерної інженерії та
інформаційних технологій

_____ Хотунов В.І.

(підпис)

« _____ » _____ 20__ р.

ЗАВДАННЯ

НА ВИПУСКНУ РОБОТУ СТУДЕНТУ

_____ Протасову Денису Сергійовичу

(прізвище, ім'я, по батькові студента)

1. Тема випускної роботи : «Дослідження алгоритмів аналізу виразів обличчя на зображенні»

Керівник роботи _____ Фальченко Наталя Григорівна

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

затверджені наказом закладу вищої освіти від “13” жовтня 2023 року №65У

2. Строк подання студентом випускної роботи _____ 31 травня 2024 року

3. Вихідні дані до випускної роботи наукова література, інтернет, власні знання

4. Зміст випускної роботи (перелік питань, які потрібно розробити) _____

1. Виконати огляд існуючих підходів та методів розв'язання задачі розпізнавання виразів обличчя. 2. Проаналізувати і порівняти алгоритми розпізнавання виразів обличчя людини на зображенні. 3. Дослідити нейромережевий алгоритм аналізу виразів обличчя людини на зображенні.

5. Дата видачі завдання 10 жовтня 2023 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Терміни виконання етапів	Примітка про виконання з підписами наукового керівника і студента
1	Вступ	16.10.2023	
2	РОЗДІЛ 1 ОГЛЯД ІСНУЮЧИХ ПІДХОДІВ ТА МЕТОДІВ РОЗВ'ЯЗАННЯ ЗАДАЧІ РОЗПІЗНАВАННЯ ВИРАЗІВ ОБЛИЧЧЯ	6.11.2023	
3	РОЗДІЛ 2 АНАЛІЗ ТА ПОРІВНЯННЯ АЛГОРИТМІВ РОЗПІЗНАВАННЯ ВИРАЗІВ ОБЛИЧЧЯ ЛЮДИНИ НА ЗОБРАЖЕННІ	22.01.2024	
4	РОЗДІЛ 3 ДОСЛІДЖЕННЯ НЕЙРОМЕРЕЖЕВОВОГО АЛГОРИТМУ АНАЛІЗУ ВИРАЗІВ ОБЛИЧЧЯ ЛЮДИНИ НА ЗОБРАЖЕННІ	13.05.2024	
5	Висновки	15.04.2024	
6	Оформлення випускної роботи (чистовий варіант)	20.05.2024	
7	Здача випускної роботи на кафедрі для рецензування	31.05.2024	
8	Перевірка випускної роботи на наявність ознак плагіату	03.06.2024	
9	Подання випускної роботи на затвердження завідувачу кафедри	07.06.2024	

Студент

_____ (підпис)

Протасов Д. С.

(прізвище та ініціали)

Керівник роботи

_____ (підпис)

Фальченко Н. Г.

(прізвище та ініціали)

АНОТАЦІЯ

Завдання автоматичного розпізнавання емоцій за допомогою виразів обличчя є актуальним у наші дні, оскільки має широку область використання у різних сферах людської діяльності, наприклад, в сучасних автомобілях є системи розпізнавання сонливості або втомлюваності людини, при виявленні яких водій отримує попереджувальний сигнал, що попереджує аварійну ситуацію. У кваліфікаційній роботі досліджується алгоритм з використанням згорткової нейронної мережі, що призначена для роботи із зображеннями та вмیه самостійно визначати необхідні ознаки зображень для їх класифікації.

В першому розділі було проаналізовано базові методи розв'язання задачі розпізнавання виразів обличчя людини. Розглянуто сучасні програмні продукти, які широко використовуються для розпізнання виразів обличчя людей, що мають структуру інтегрованого програмного забезпечення або додатка із великим набором функцій для роботи із машинним зором і зображеннями.

У розділі 2 було зроблено аналіз та порівняння алгоритмів розпізнавання виразів обличчя людини на зображенні. Встановлено, що більшість алгоритмів розпізнавання виразів обличчя людини засновані на спільній роботі двох методів: методу пошуку ключових точок обличчя на зображенні (Facial Landmark Detection, Point Distribution Model, 2D-and-3D-face-alignment тощо) та методу класифікації даних (нейронні мережі, дерева прийняття рішень, класифікатор SVM).

У розділі 3 досліджено нейромережевий алгоритм, який вирішує задачу розпізнавання виразів обличчя на зображенні, що має просту архітектуру нейронної мережі. Проведено тестування шляхом навчання нейронної мережі на наборі даних «Facial Expression Recognition Challenge Dataset».

ANNOTATION

The task of automatic recognition of emotions using facial expressions is relevant these days, as it has a wide range of applications in various areas of human activity, for example, modern cars have systems for recognizing drowsiness or fatigue of a person, upon detection of which the driver receives a warning signal that warns of an emergency situation . The qualification paper examines an algorithm using a convolutional neural network, which is designed to work with images and is able to independently determine the necessary features of images for their classification.

In the first chapter, the basic methods of solving the problem of recognizing human facial expressions were analyzed. Modern software products, which are widely used for recognizing human facial expressions, have the structure of integrated software or an application with a large set of functions for working with machine vision and images.

In chapter 2, an analysis and comparison of algorithms for recognizing human facial expressions on the image was made. It has been established that most algorithms for recognizing human facial expressions are based on the joint work of two methods: the method of finding key points of the face in the image (Facial Landmark Detection, Point Distribution Model, 2D-and-3D-face-alignment, etc.) and the method of data classification (neural networks , decision trees, SVM classifier).

Chapter 3 examines a neural network algorithm that solves the problem of facial expression recognition in an image with a simple neural network architecture. Testing was carried out by training a neural network on the "Facial Expression Recognition Challenge Dataset".

ЗМІСТ

ВСТУП	3
РОЗДІЛ 1. ОГЛЯД ІСНУЮЧИХ ПІДХОДІВ ТА МЕТОДІВ РОЗВ’ЯЗАННЯ ЗАДАЧІ РОЗПІЗНАВАННЯ ВИРАЗІВ ОБЛИЧЧЯ.....	6
1.1 Огляд існуючих підходів до розв’язання задачі	6
1.2 Огляд і аналіз методів розв’язання задачі розпізнавання виразів обличчя	10
1.3 Огляд існуючого програмного забезпечення, що використовує алгоритми розпізнавання виразів обличчя	14
Висновки до розділу 1	18
РОЗДІЛ 2. АНАЛІЗ ТА ПОРІВНЯННЯ АЛГОРИТМІВ РОЗПІЗНАВАННЯ ВИРАЗІВ ОБЛИЧЧЯ ЛЮДИНИ НА ЗОБРАЖЕННІ.....	19
2.1 Існуючі алгоритми розпізнавання виразів обличчя людини	19
2.2 Порівняння алгоритмів	23
Висновки до розділу 2	31
РОЗДІЛ 3. ДОСЛІДЖЕННЯ НЕЙРОМЕРЕЖЕВОВОГО АЛГОРИТМУ АНАЛІЗУ ВИРАЗІВ ОБЛИЧЧЯ ЛЮДИНИ НА ЗОБРАЖЕННІ.....	32
3.1 Нейромережевий алгоритм аналізу виразів обличчя людини.....	32
3.2 Dataset для навчання нейронної мережі	38
3.3 Тестування та оптимізація структури нейромережі	42
Висновки до розділу 3	45
ВИСНОВКИ.....	47
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	49

ВСТУП

Емоціями називають суб'єктивні стани людини та тварин, які виникають у відповідь на вплив зовнішніх або внутрішніх подразників і що виявляються у формі безпосередніх переживань [1]. Емоції є невід'ємною частиною життя людей, що супроводжуються: мімікою, жестами, інтонацією, позою, але особливе місце займає серед них – вираз обличчя людини. Завдання, пов'язані з розпізнаванням виразів обличчя людини, завжди були предметом досліджень психологів та вчених. Прикладом може бути опублікована робота Чарльза Дарвіна «Про вирази емоцій у людини і тварин». [2]. Він першим висунув гіпотезу про універсальність мімічних проявів емоцій, причому не тільки різних культур, а і різних видів. Ця теорія була перевірена та продовжена психологом Полом Екманом [3]. У ході спостереження за мімічними виразами мешканців одного із племен у Новій Гвінеї, які жодного разу не мали довготривалого контакту з представниками західної або східної культури, вчений зауважив, що деякі емоції виражалися таким же чином, що і в усьому світі [3]. Пол Екман встановив, що такі емоції як: гнів, здивування, смуток, радість, страх і огида є універсальними та можуть бути зрозумілі людиною, незалежно від її культури.

Завдання автоматичного розпізнавання емоцій за допомогою виразів обличчя є актуальним у наші дні, оскільки має широку область використання у різних сферах людської діяльності, наприклад, в сучасних автомобілях є системи розпізнавання сонливості або втомлюваності людини, при виявленні яких водій отримує попереджувальний сигнал, що попереджує аварійну ситуацію; у правоохоронній діяльності за допомогою таких автоматичних систем, слідчі можуть визначати достовірність показань підозрюваного при допиті. Крім того, системи можуть допомогти виявити людей, демонструють емоційні реакції, притаманні психічно хворим особам, порушниками правопорядку тощо.

В даний час дослідники в галузі штучного інтелекту вивчають можливості застосування технологій комп'ютерного зору для розпізнавання емоцій людини

на основі зображення її обличчя [4-6]. Щорічно проводяться аналізи нових розроблених алгоритмів розпізнавання виразів обличчя, метою яких є отримання максимальної точності розпізнавання у заданому наборі даних. Аналіз результатів змагань AffectNet, AFEW-VA, EmoReact, EmotioNet challenge, 9 th Emotion Recognition in the Wild Challenge, OMG-Emotion challenge, дозволяє виділити наступні особливості реалізації алгоритмів, у яких використовується зв'язок двох методів: пошуку ключових точок обличчя на зображенні; класифікації, яка визначає клас емоцій на основі взаємного розташування ключових точок обличчя. Аналіз [7] вказує на те, що поєднання цих двох методів впливає на точність реалізації алгоритму розпізнавання емоцій.

У моїй роботі досліджується алгоритм з використанням згорткової нейронної мережі, що призначена для роботи із зображеннями та вміє самостійно визначати необхідні ознаки зображень для їх класифікації. Використання згорткової нейронної мережі з малою кількістю шарів, що пропонується у моєму дослідженні збільшує точність результатів у порівнянні з поєднанням «методу пошуку ключових точок та методу класифікації».

Метою роботи є дослідження алгоритму розпізнавання виразів обличчя на зображенні, заснованого на нейромережевих технологіях.

Для досягнення мети необхідно вирішити наступні *задачі*:

- проаналізувати існуючі підходи та методи розв'язання задачі розпізнавання виразів обличчя;
- проаналізувати та порівняти алгоритми розпізнавання виразів обличчя людини на зображенні;
- дослідити нейромережевий алгоритм розпізнавання виразів обличчя на зображенні з вибором структури нейронної мережі.

Предмет дослідження – методи машинного навчання для розпізнавання виразів обличчя людини на зображенні.

Об'єкт дослідження – алгоритми розпізнавання виразів обличчя людини на зображенні.

Методи дослідження – методи, принцип яких базується на геометричному методі розпізнавання; методи статистичної обробки інформації; метод згорткової нейронної мережі; методи засновані на нейронних мережах.

РОЗДІЛ 1. ОГЛЯД ІСНУЮЧИХ ПІДХОДІВ ТА МЕТОДІВ РОЗВ'ЯЗАННЯ ЗАДАЧІ РОЗПІЗНАВАННЯ ВИРАЗІВ ОБЛИЧЧЯ

1.1 Огляд існуючих підходів до розв'язання задачі

Більшість методів, що застосовуються для вирішення задачі класифікації виразів обличчя, складаються з наступних кроків [8]: збір даних та їх попередня обробка, вилучення та вибір ознак, класифікація та прийняття рішення (рис. 1.1).

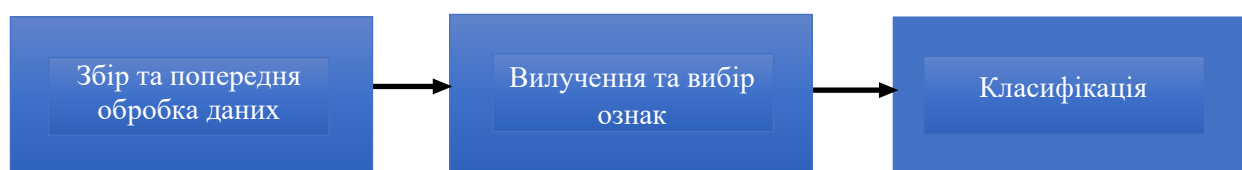


Рисунок 1.1 – Структурна схема функціонування систем виразів обличчя

На відміну від використовуваного в даній роботі підходу на основі згорткової нейронної мережі, усередині якої відбувається вибір ознак та подальше визначення класу зображення, у ряді підходів виділення особливостей зображення і класифікація являють собою окремі, не пов'язані між собою завдання, кожне з яких може бути вирішене одним із можливих способів. У разі використання певних сучасних методів машинного навчання, наприклад глибоке навчання [9-11], на вхід алгоритму подається не набір ознак, а безпосередньо саме зображення і згодом виконується визначення його класу.

Розглянемо існуючі методи отримання ознак, що використовуються у підходах першого типу. Дані підходи можна розділити на групи, залежно від використовуваного простору ознак. Набори ознак, що обчислюються за зображеннями обличчя, можуть включати текстурні або геометричні ознаки [8].

Геометричними ознаками можуть бути відстані, кути тощо. Вони характеризують особливості розташування, орієнтації об'єктів зображення. 2D або 3D моделі зазвичай використовуються для пошуку ключових точок обличчя.

Active Shape Models (ASM). Даний метод [12] полягає у використанні статистичних моделей форм об'єктів, які можуть бути деформовані для підгонки під досліджуваній об'єкт на новому зображенні.

На обличчях зі знімків навчальної вибірки розмічуються розташування антропометричних точок, причому порядок нумерації точок на зображеннях однаковий. Локалізація точок ASM-моделі на новому зображенні відбувається за рахунок розв'язання оптимізаційної задачі. Для кожної точки визначається найкраще збіжне з шаблоном положення за допомогою градієнта текстури зображення біля цієї точки. В результаті роботи цього методу відбувається локалізація ключових точок обличчя на зображенні (рис. 1.2). Цей метод можна використовувати спільно з системою кодування лицьових рухів Пола Екмана [13].

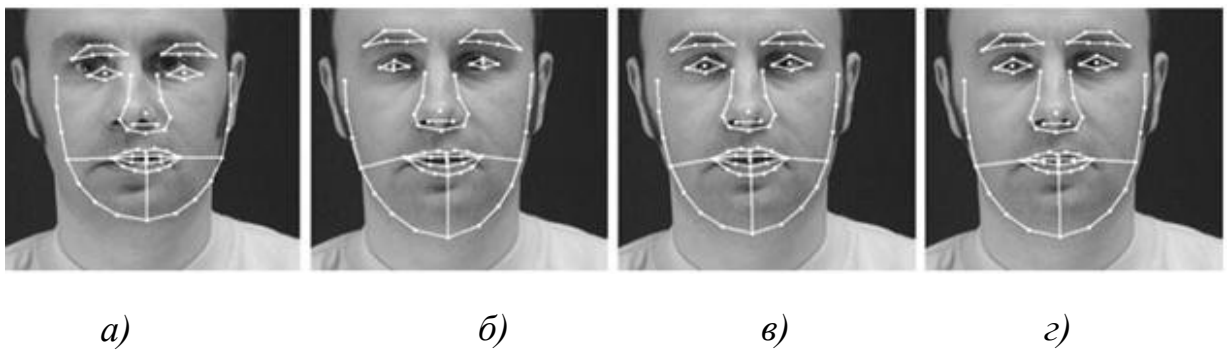


Рисунок 1.2 – Процес локалізації ASM на зображенні: (а) – початкова позиція; (б) – через 5 ітерацій; (в) – через 10 ітерацій; (г) – підсумкова модель

Текстурні ознаки відображають однорідність зображення. Текстурними особливостями зображень обличчя є переважно зміни образу текстури, такі як зморшки на шкірі та опуклості.

Local Binary Pattern (LBP). У процесі роботи даного методу [14] кожен піксель вихідного зображення подається у вигляді бінарного числа, яке залежить від інтенсивностей сусідніх пікселів (рис 1.3).

Вектор ознак створюється наступним чином:

1. Зображення, що досліджується поділяється на комірки.

2. Кожен піксель по черзі порівнюється із одним з його 8-ми сусідами.
3. Якщо значення центрального пікселя більше значення сусіднього, то в цю комірку записується 0, іншому випадку – 1.
4. Сукупність значень всіх сусідніх комірок формує восьмизначне двійкове число, яке визначає модуль пікселя.
5. Для кожної комірки обчислюють гістограму частоти кожного числа, що зустрічається. Гістограма може розглядатися як 256-мірний вектор ознак.
6. Гістограми всіх комірок об'єднуються, що в результаті дає вектор ознак всього зображення.

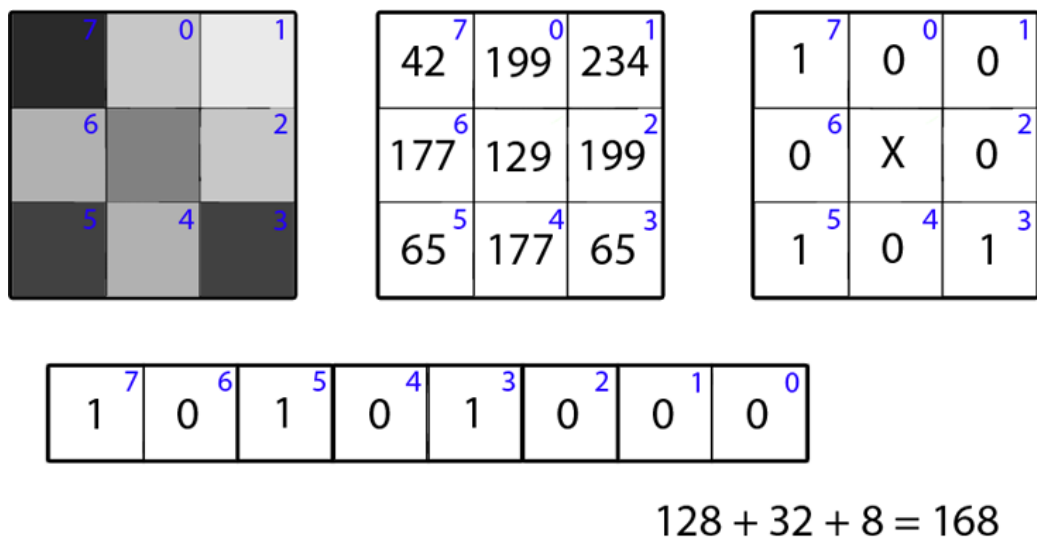


Рисунок 1.3 – Ілюстрація принципу роботи LBP

Вейвлети (фільтри) Габора. Текстурні особливості зображення можуть бути отримані за допомогою вейвлетів Габора [15]. Вейвлет-перетворення є згорткою зображення з функцією ψ -вейвлетом (рис. 1.4). У разі використання двовимірного фільтра Габора з орієнтацією μ та масштабом ν дана функція має вигляд:

$$\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} e^{-\frac{\|k_{\mu,\nu}\|^2 \|z\|^2}{\sigma^2}} \left[e^{ik_{\mu,\nu}z} - e^{-\sigma^2/2} \right], \quad (1.1)$$

де $z(x, y)$ – координати точки;

$\|\cdot\|$ – норма вектора;

$$k_{\mu, \nu} = k_{\nu} \cdot e^{i\varphi_{\mu}}, \quad k_{\nu} = k_{\text{макс}} / f^{\nu}, \quad \varphi = \pi\mu/8,$$

$k_{\text{макс}}$ – максимальна частота;

f – просторовий множник ядрами частотної області.



Рисунок 1.4 – Виходи операції згортки для зображення

Класифікація. На останньому етапі роботи розглянутих систем розпізнавання виразів обличчя вилучені ознаки подаються на вхід класифікатору. Для вирішення задачі класифікації можуть використовуватися такі підходи як: метод опорних векторів, нейронні мережі тощо.

1.2 Огляд і аналіз методів розв'язання задачі розпізнавання виразів обличчя

Після того, як з камер зображення передається у вигляді цифрових даних на електронно-обчислювальну машину – ця інформація обробляється за допомогою спеціального алгоритму, який визначає розміщення області обличчя за його основні риси (рот, очі, ніс, брови тощо). Методів розв'язання задачі розпізнавання виразів обличчя існує багато, а більшість з них є комбінацією декількох методів, внаслідок чого вони можуть бути громіздкими та мати велику обчислювальну складність. Щоб зменшити кількість обчислень та прискорити процес розпізнавання виразів обличчя, багато авторів застосовують методи, які скорочують кількість аналізованих фрагментів. Розглянемо найбільш актуальні методи.

Метод Віола-Джонса (Viola-Jones object detection). Метод був запропонований Майклом Джонсом та Паулом Віолою у 2001 році, який є найпершим методом, що продемонстрував чудові результати при обробці зображень у реальному часі. На сьогоднішній день цей метод має багато реалізацій, особливо у структурі бібліотеки комп'ютерного зору Open CV (функція CV HAAR DETECT OBJECTS) [16].

Переваги методу:

- за рахунок використання каскадного класифікатора – висока швидкість роботи;
- висока точність виявлення повернутих обличчя на кут до 35 градусів (у випадку більшого кута, ефективність цього методу суттєво знижується);

Недоліки методу:

- довготривалий час навчання. Алгоритму потрібно проаналізувати значну кількість тестових зображень;
- побудова неоптимального набору базових алгоритмів;
- бустинг може спричинити побудову громіздких композицій, які складаються із сотень алгоритмів.

Elastic Graph Matching (метод гнучкого порівняння на графах). Цей метод є одним із методів 2D моделювання, основний принцип якого полягає в співставленні графів, що описують вирази обличчя (контури обличчя репрезентується у вигляді сітки з індивідуальним розташуванням ребер і вершин (див. рис. 1.5)) [17].

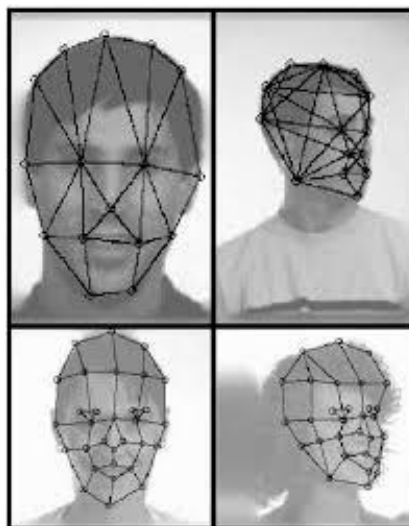


Рисунок 1.5 – Структурна сітка, утворена характерними точками

Процедура розпізнавання виразів обличчя відбувається таким чином: еталонний граф, що характеризує основний параметр розпізнання обличчя, залишається без зміни, у той час як інші спотворюються через структуру обличчя, яка прив'язана до основних антропометричних точок: відстань між вухами, очима, ширина губ, лінія носа тощо. Чим більше таких точок застосовується, тим точніше буде система розпізнавання, що суттєво збільшить час на обробку одного об'єкта [17].

Недоліки методу:

- швидкодія аналітичної системи є обернено пропорційною до розмірів баз даних;
- непроста процедура вводу нових шаблонів у базу даних;
- складний алгоритм розпізнання, що спричиняє необхідність використання значних обчислювальних потужностей.

Приховані Марківські моделі (ПММ). Метод базується на статистичному порівнянні об'єкта з існуючою базою шаблонів. Ці моделі використовують статистичні властивості сигналів та враховують їх просторові характеристики. Складовими частинами моделі є: матриця перехідних ймовірностей, множина спостережних станів, початкова ймовірність станів, множина прихованих станів. Кожному елементу відповідає своя ПММ. Під час розпізнавання виразів обличчя людини, перевіряються всі згенеровані ПММ та знаходиться найкраща з ймовірностей, яка визначає послідовність спостережень для об'єкта згенерованого відповідною моделлю [18].

Недоліки:

- низька швидкість спрацювання;
- система може оптимізувати лише час обробки даних та відклику на власну модель і не має можливості мінімізувати час перебору інших моделей;
- невисока здатність розпізнавати оптимальний і не оптимальний алгоритм навчання.

Principal Component Analysis (PCA – метод головних компонент). Основною метою PCA є зменшення кількості простору ознак без вагової втрати інформації і так, щоб він якнайкраще описував «типові» риси виразів обличчя. У задачі розпізнавання виразів обличчя його використовують головним чином для того, щоб представити міміку у вигляді малорозмірного вектору, який порівнюється із еталонними векторами з бази даних[19].

Набір власних векторів, який отримується з навчальної вибірки, застосовується для кодування інших зображень, які можна отримати шляхом порівняння комбінації власних векторів. При використанні певної кількості власних векторів можна отримати необхідну апроксимацію вхідного зображення, яке згодом можна зберегти у базі даних, у вигляді вектора коефіцієнтів, який є одночасно ключем пошуку в БД [19].

PCA має ефективність використання у додатках розпізнавання. Однак у випадку, коли на зображенні обличчя є вагомі зміни у освітленості або міміці обличчя, ефективність методу вагомо зменшується. Це пов'язано з тим, що метод

головних компонент обирає підпростір з метою максимальної апроксимації вхідного набору даних [19]. Алгоритм на основі PCA представлений на рис. 1.6.

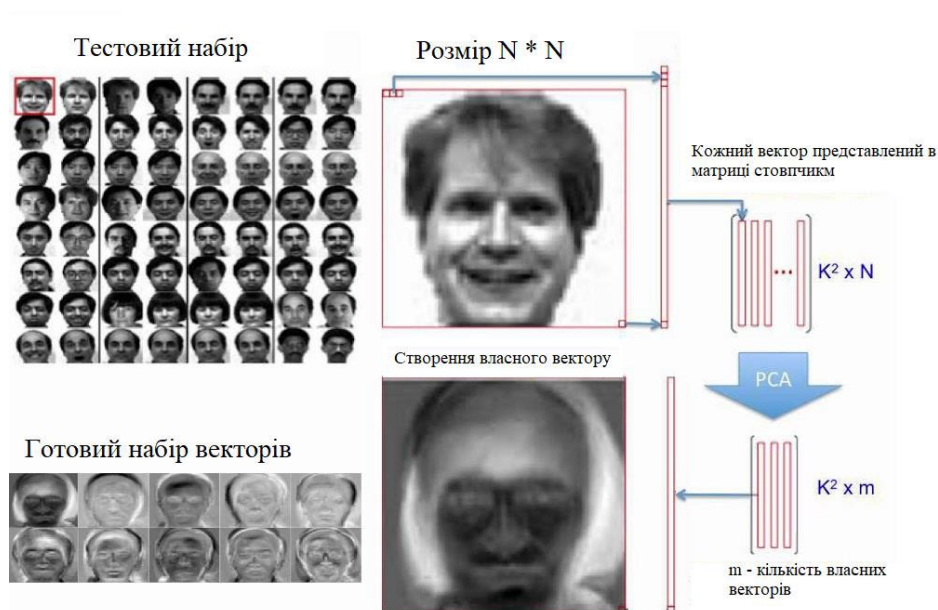


Рисунок 1.6 – Схема роботи алгоритму на основі PCA

Support Vector Machines (SVM – метод опорних векторів) – це набір алгоритмів навчання із учителем, які використовуються у задачах класифікації та регресійного аналізу [20]. Суть методу опорних векторів полягає у знаходженні гіперплощини в ознаковому просторі, що відокремлює класи мімічних зображень. Слід відмітити, що з двох можливих класів гіперплощин, потрібно обрати необхідну гіперплощину, відстань до якої від кожного класу є максимальною.

Недоліки:

– точність роботи методу поступається багатьом методам.

Переваги методу:

– висока швидкість роботи в порівнянні з нейронними мережами;

– висока стійкість до перенавчання;

– можливість зменшення чутливості до завад за рахунок зниження точності.

Нейромережеві методи. Вони є найпоширенішими методами, які складаються з десятка різних алгоритмів. Особливістю таких мереж є їхнє

навчання на наборі готових прикладів заздалегідь занесених до бази даних. У період навчання нейронна мережа автоматично обирає ключові ознаки та створює взаємозв'язок. Після цього, навчена нейронна мережа застосовує отриманий досвід для розпізнання раніше невідомого об'єкта. Нейромеревеві методи показали одні з найкращих результатів у галузі розпізнавання виразів обличчя людини, але вважаються найскладнішими для реалізації [5].

Переваги даного методу:

- висока точність розпізнавання при правильному налаштуванні параметрів мережі.

Недоліки:

- складно формалізувати мережеву архітектуру (шарів, кількість нейронів, характер зв'язків);
- складна процедура внесення змін (внесення будь-якої зміни потребує перенавчання мережі);
- висока обчислювальна складність.

1.3 Огляд існуючого програмного забезпечення, що використовує алгоритми розпізнавання виразів обличчя

FaceReader розроблена нідерландською компанією «Noldus Information Technology» [21, 24]. Програма здатна класифікувати вирази обличчя за 6 базовими емоціями, а також визначати нейтральний вираз обличчя. Крім цього, *FaceReader* здатна визначати стать та вік людей.

Програма використовує геометричні ознаки зображення для класифікації, зокрема, використовується метод *Active Appearance Model*, що створює модель обличчя людини за ключовими точками з врахуванням деталей поверхні (рис. 1.7). Розв'язання задачі класифікації здійснюється нейронною мережею.

Переваги:

- система здатна аналізувати як статичні кадри так і відеофайли;

– для розпізнавання виразів нахил обличчя у площині може бути будь-яким, якщо система зможе його виявити;

– у програмі чудова візуалізація: присутні діаграми, гістограми, відсотки емоцій, що виражаються.

Недоліки:

– система не натренована для розпізнавання виразів обличчя дітей віком до п'яти років;

– точність розпізнавання знижується, якщо користувач в окулярах;

– програма не детектує профіль обличчя.



Рисунок 1.7 – Інтерфейс програми FaceReader

Project Oxford проєкт розроблений компанією «Microsoft» [22]. Даний проєкт є набором готових REST API, що реалізують алгоритми комп'ютерного зору. Крім сервісу розпізнавання осіб, мови та тексту, даний проєкт включає

також сервіс із розпізнавання виразів обличчя. Програма здатна розрізняти такі типи емоцій: щастя, страх, смуток, огиду, гнів, зневагу, подив, і навіть нейтральний вираз обличчя. Розробники можуть використовувати ці сервіси всередині своїх додатків.

Переваги:

- можливість використання всередині різних програм;
- демонстрація відсотків виражених емоцій.

Недоліки:

- програма не може розпізнати обличчя з нахилом понад 40°;
- розпізнавання виразів обличчя проводиться лише на статичних зображеннях.

FaceVACS-VideoScan – просте у використанні програмне забезпечення розпізнавання виразів обличчя, яке пропонує компанія «Cognitec Systems» [23].

Система *FaceVACS-VideoScan* побудована на основі декількох системних компонентів: сервера відеосканування, що призначений для координації усіх компонентів системи та реалізує основні біометричні операції; відеосervera, який управляє відеопотоками; обчислювального вузла, що застосовується для розподілу обчислювального процесу; інтерфейсу користувача; диспетчера обслуговуючого мобільні пристрої та інформаційні сигнали отримувача повідомлення про події; операційну базу даних та комплекту інтеграторів [23].

Технологія *FaceVACS* використовує алгоритм розпізнавання виразів обличчя *B10T9*, який стійкий до змін виразів, поворотів обличчя (на $\pm 17^\circ$) часткового його перекриття, використання окулярів та зміни освітлення [23].

Система *eMotion Software* [24] відома через тим, що її винахідники розпізнали емоційний вираз обличчя на картині «Мона Ліза». Результат аналізу вказали на те, що вона була на 85 % щасливою, 9% показали огиду, 6% страх і всього на 2% Мона Ліза сердилася.

Першим користувачем *eMotion Software* була компанія Unilever, яка використала систему розпізнавання обличчя в автоматі з продажу морозива –

Unilever Share Happy. Покупці посміхалися автомату, і він видавав за посмішки безкоштовне морозиво [24]!

Якщо людина проявляє емоції, гнів, здивування, смуток, радість, страх, огиду або корчить гримасу, тисячі дрібних м'язів обличчя знаходяться в роботі. Система розпізнавання емоцій, або Emotion-recognition system (ERS) створює 3D-модель зображення обличчя з ідентифікацією 12 ключових областей, таких як кутки очей та роту тощо (див. рис. 1.8) [24].

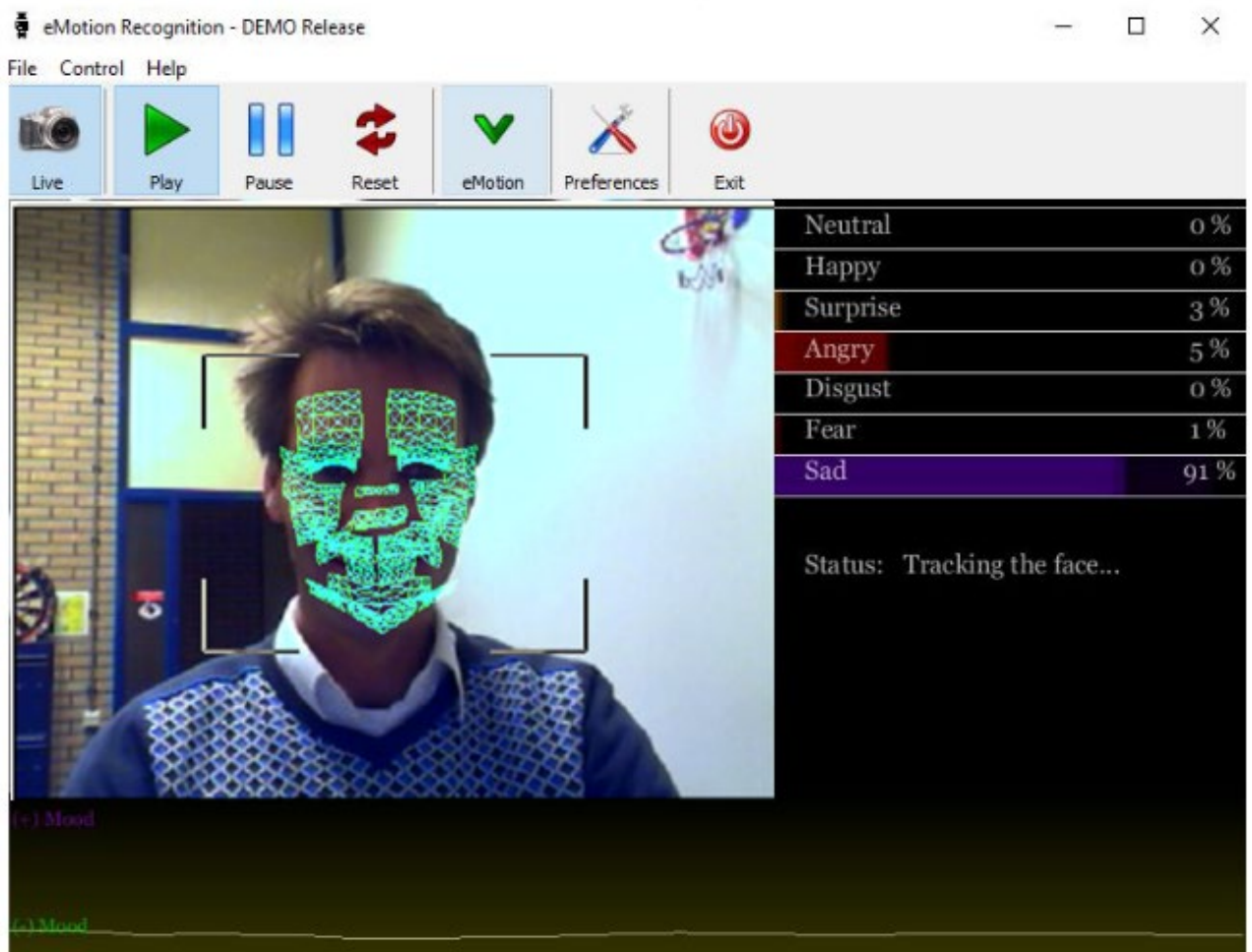


Рисунок 1.8 – Приклад роботи системи eMotion Software

Висновки до розділу 1

Задача автоматичного розпізнавання емоцій є актуальною задачею сьогодення, оскільки має широке застосування в різних сферах людської діяльності.

У першому розділі проаналізовано основні підходи та методи розв'язання задачі розпізнавання виразів обличчя; розглянуті приклади програмного забезпечення, що використовують алгоритми та методи розпізнавання виразів обличчя

Всі розглянуті методи мають свої переваги у певних задачах. Наприклад, нейромережеві методи показують кращі результати у задачах класифікації зображень, тоді як SVM може бути ефективнішим у задачах класифікації тексту.

Нейромережеві методи, такі як згорткові нейронні мережі, мають більшу точність у задачах ідентифікації та розпізнавання виразів обличчя людини порівняно з іншими розглянутими у роботі методами, такими як PCA, SVM тощо.

Таким чином, у першому розділі було розглянуто сучасні програмні продукти, які широко використовуються для розпізнавання виразів обличчя людей. Вони мають структуру інтегрованого програмного забезпечення або додатка із великим набором функцій для роботи із машинним зором і зображеннями.

РОЗДІЛ 2. АНАЛІЗ ТА ПОРІВНЯННЯ АЛГОРИТМІВ РОЗПІЗНАВАННЯ ВИРАЗІВ ОБЛИЧЧЯ ЛЮДИНИ НА ЗОБРАЖЕННІ

2.1 Існуючі алгоритми розпізнавання виразів обличчя людини

Алгоритми машинного навчання є однією з ключових технологій штучного інтелекту, що забезпечує здатність вчитися вирішувати поставлені завдання користувачем. Спочатку алгоритми машинного навчання використовували для аналізу даних, які можна подати у вигляді простих таблиць з невеликою кількістю стовпців та рядків. Наприклад, у вигляді таких таблиць можна подати дані працівників підприємств або здобувачів освіти, які навчаються в коледжі. У цьому випадку аналіз таких даних за допомогою алгоритмів машинного навчання дозволяє автоматизовано визначити закономірності в наявних параметрах та будувати на їх основі різні прогнозні моделі. За допомогою даних працівників підприємства, можна побудувати модель, яка визначає кваліфікаційні характеристики та досвід роботи тощо, за результатами яких можна підвищувати кваліфікацію, встановлювати надбавки та преміювання, а для здобувачів освіти, на основі попередніх оцінок з дисциплін можна прогнозувати ймовірність успішного завершення навчання [25].

З часом алгоритми машинного навчання також використовуються при аналізі текстової інформації, що сформувало напрям у штучному інтелекті під назвою «Обробка природної мови (Natural Language Processing, NLP)». А застосування машинного навчання у завданнях аналізу даних відео потоку та окремих кадрів призвело до виникнення напряму у штучному інтелекті під назвою «Комп'ютерний зір (Computer Vision, CV)» [24].

Комп'ютерний зір – це науковий напрямок зі створення апаратно-програмних комплексів, здатних проводити виявлення, локалізацію та розпізнавання об'єктів. Технології комп'ютерного зору застосовуються на вирішення завдань різних напрямків діяльності людини, таких як [24]:

- пошук інформації на основі зображень;

- автоматизація охоронних функцій на різних об'єктах;
- верифікація клієнтів банків із зображення;
- встановлення зв'язків людей з фотографій соціальних мереж;
- автоматичне покращення якості фотографій;
- забезпечення безпеки на масових заходах.

Однією із сучасних завдань комп'ютерного зору є розпізнавання виразів обличчя людини з урахуванням аналізу зображення. Технологія розпізнавання виразів обличчя може застосовуватись для автоматизованої оцінки реакції людини на зовнішні подразники. Сфери застосування даної технології багато, наприклад, в освіті, за мімікою людини можна контролювати рівень залучення здобувачів освіти до освітнього процесу. У різних соціальних мережах розпізнавання виразів обличчя може бути використане для персоналізованих рекомендацій контенту (за емоціями користувача можна визначати, який контент йому подобається, а який ні). Також розпізнавання виразів обличчя може застосовуватися як додатковий інструмент під час тестування на поліграфі.

Для підтримки інтересу до проблеми розпізнавання виразів обличчя на зображенні щорічно проводяться змагання з розробки та тестування алгоритмів.

У таблиці 2.1 представлені сучасні відомі змагання з розробки та тестування алгоритмів розпізнавання емоцій на зображеннях.

Як видно з таблиці 2.1, весь спектр емоцій можна поділити на різну кількість класів. Загальний підхід щодо уявлення та класифікації емоцій людини описав у своїй роботі Дж. Рассел [24, 26]. У ній він виділив два базові параметри будь-якої емоції – «Інтенсивність/arousal» та «Знак/valance». На колі, центр якого розташований на перетині осей, а рівнина і значення розташовані назви емоцій. Графічне уявлення спектра емоцій за моделлю Дж. Рассела представлено на рисунку 2.1. Прикладом позитивної емоції (велике позитивне значення параметра valance) з низькою інтенсивністю (близьке до нуля значення arousal) є «задоволеність».

Таблиця 2.1 – Змагання з розробки алгоритмів для розпізнавання емоцій на зображеннях

Назва змагання (рік)	Кількість класів емоцій	Тип аналізованих даних
AffectNet (2017)	7 класів	Зображення (фото)
EmoReact (2018)	17 класів	Відео та аудіо
9th Emotion Recognition in the Wild Challenge (2023)	10 класів	Відео та аудіо
OMG-Emotion challenge (2018)	7 класів	Відео та аудіо
EmotioNet challenge (2020)	18 класів	Зображення (фото)
AFEW-VA Database for Valence and Arousal Estimation In-The-Wild (2023)	2 класи	Відео

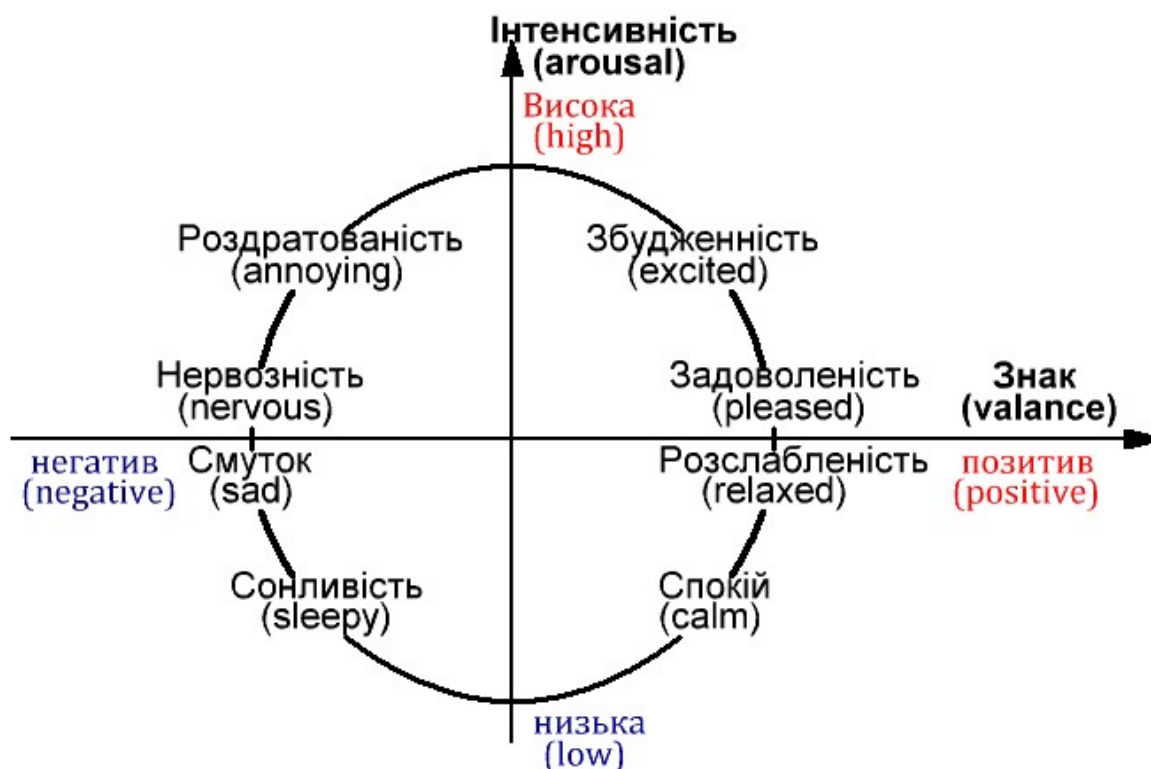


Рисунок 2.1 – Спектр виразів обличчя людини за моделлю Дж. Рассела

Навчальна вибірка даних необхідна для налаштування алгоритмів розпізнавання емоцій. Для формування навчальної вибірки даних беруться статичні зображення, аудіо та відео дані, де вказуються різні емоції, і розмічаються вручну. У ході ручної розмітки даних спеціаліст повинен визначити, яка емоція показана на об'єкті з вибірки даних, а також записати відповідну мітку класу емоції до системи обліку [26].

Розмітка даних для формування навчальної вибірки даних є трудомістким та тривалим процесом. Одні й самі зображення можна розмітити по-різному залежно від вибраної кількості категорій емоцій. Наприклад, можна при розмітці об'єктів навчальної вибірки використовувати лише 4 категорії емоцій, розділивши всі емоційні відтінки групи так, як це представлено на рисунку 2.2, а.

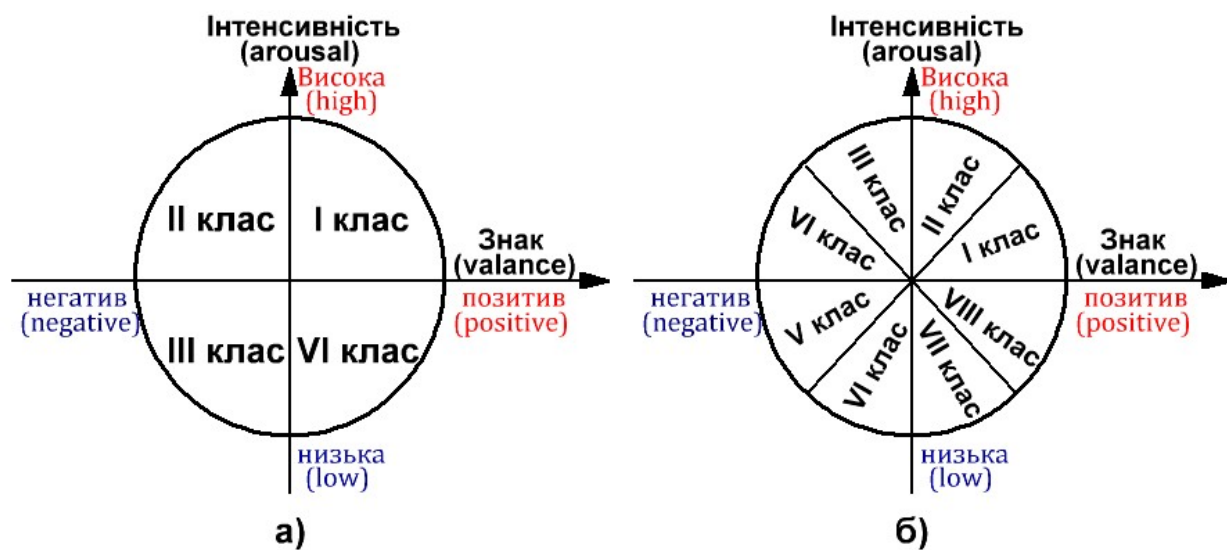


Рисунок 2.2 – Розпізнавання виразів обличчя як завдання класифікації за різної кількості класів у вихідних даних: а – при 4 класах, б – при 8 класах

У деяких випадках може знадобитися точніша класифікація емоцій. І тут потрібно збільшити кількість класів, наприклад, до 8 (див. рис. 2, б). Це збільшить трудомісткість розмітки навчальної вибірки, але, зрештою, алгоритм розпізнавання виразів обличчя, який навчається на цих даних, зможе відрізнити більше емоційних відтінків.

У всіх змаганнях алгоритмів розпізнавання виразів обличчя, які представлені у таблиці 2.1, застосовувалися унікальні навчальні вибірки, що розмічені з використанням різної кількості класів (від 2 до 17).

Варто також відзначити, що розпізнавання виразів обличчя людини може бути представлене у вигляді регресійного аналізу. У цьому випадку, на основі аналізу зображення виразів обличчя необхідно визначити числові значення компонентів *valence* і *arousal*. За отриманими результатами значень можна точно визначити відтінок емоції, скориставшись моделлю Дж. Рассела (див. рис. 2.3) [26].

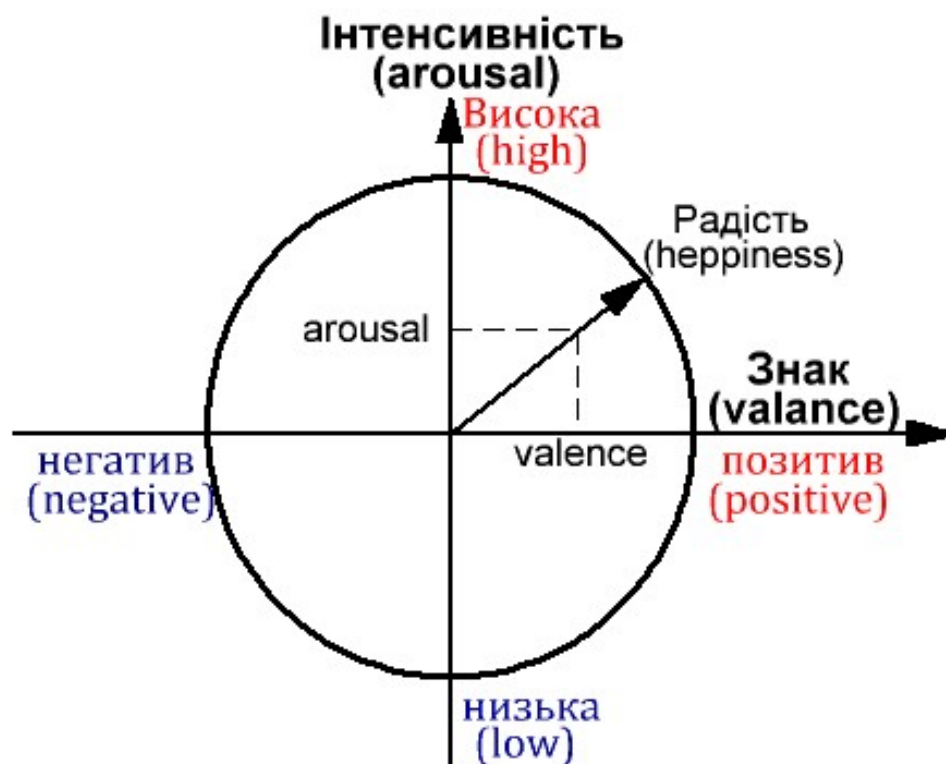


Рисунок 2.3 – Розпізнавання емоцій як задача регресійного аналізу

2.2 Порівняння алгоритмів

Аналіз найкращих рішень, представлених на змаганнях OMG-Emotion challenge, EmoReact, 9th Emotion Recognition in the Wild Challenge, AFEW-VA Database for Valence and Arousal, EmotioNet challenge та AffectNet з розробки

алгоритмів розпізнавання виразів обличчя вказує на такі особливості їх реалізації.

Робота більшої частини представлених алгоритмів складається з виконання 3 етапів: завантаження зображення, пошук на зображенні ключових точок обличчя та аналіз взаємного розташування ключових точок людини для визначення класу емоцій [24].

На першому етапі завантажене зображення перетворюється на тривимірний числовий масив. Розмір масиву залежить від розмірів зображення (висоти та ширини), а також від колірної моделі зображення. У наборах даних, що використовуються на змаганнях, всі зображення вже приведені до єдиного розміру та формату.

На другому етапі проводиться аналіз числового масиву даних з метою виявлення ключових точок людини на зображенні. Під ключовими точками розуміють контур обличчя (у деяких випадках - контур вилиць), контур брів, контур носа, контур меж очей, контур меж губ.

Параметрами ключових точок людини є: координати в просторі (3 числових значення), типи точок залежно від елемента людини з яким вони пов'язані (1 числове значення), номер точки з послідовності (1 числове значення). Приклад опису ключових точок для контуру лівої брови (масив *left_eyebrow*) представлений у вигляді формули (2.1)

$$left_eyebrow = \begin{bmatrix} 122 & 201 & 196 & 2 & 1 \\ 130 & 204 & 198 & 2 & 2 \\ 134 & 206 & 198 & 2 & 3 \\ 138 & 204 & 200 & 2 & 4 \\ 141 & 202 & 201 & 2 & 5 \end{bmatrix}. \quad (2.1)$$

На третьому етапі проводиться класифікація отриманих числових значень з використанням одного з відомих методів. Основні етапи класифікації емоцій з

використанням технології визначення ключових точок обличчя представлено на рис. 2.4.

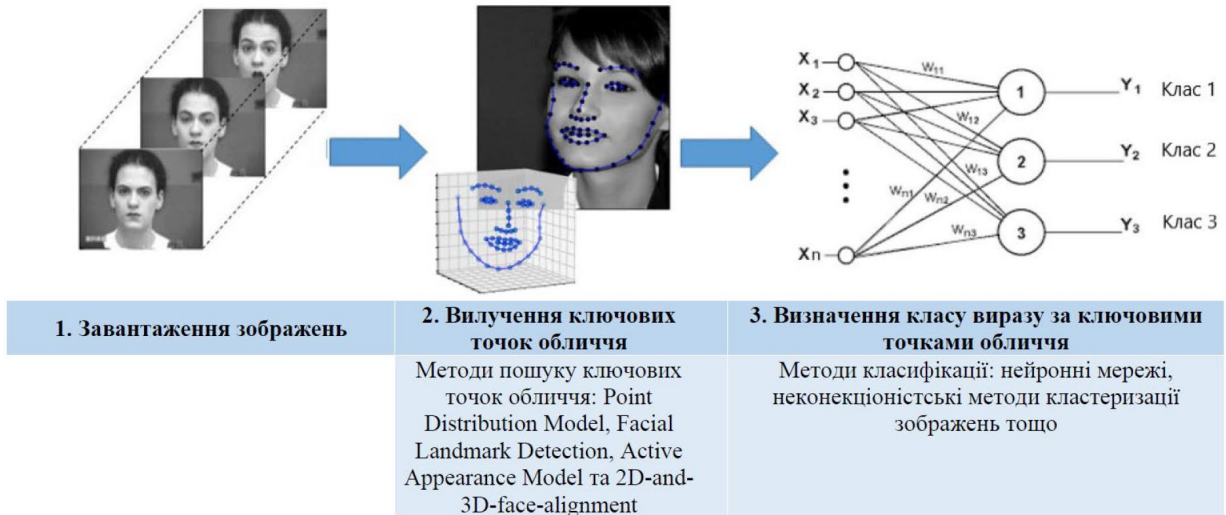


Рисунок 2.4 – Основні етапи класифікації виразів обличчя з використанням технології визначення ключових точок

Основною відмінністю алгоритмів розпізнавання виразів обличчя одного від іншого полягає у використанні різних поєднань методів пошуку ключових точок обличчя та методу класифікації даних. При такому підході завдання розробки алгоритму розпізнавання виразів обличчя людини зводиться до підбору оптимального поєднання методів пошуку ключових точок і класифікації під навчальну вибірку розмічених зображень. Так, наприклад, якщо в навчальній вибірці даних містяться зображення людей, що знаходяться під великим кутом до площини зображення, і частина обличчя перекрита волоссям, рукою або знаходиться в тіні, то необхідно використовувати методи пошуку ключових точок, які вміють самостійно моделювати приховані області обличчя (див. рис. 2.5) [24].

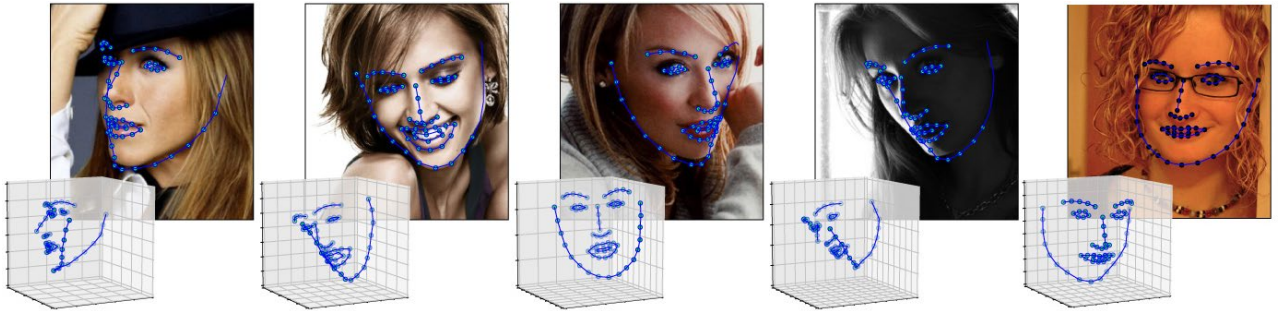


Рисунок 2.5 – Визначення ключових точок при великих кутах розташування обличчя до площини зображення, а також при частковому перекритті обличчя

Найбільш відомими алгоритмами пошуку ключових точок обличчя на зображенні є:

– Active Appearance Model – це статистична модель, яка має властивість зміни форми та зовнішнього вигляду класу завантажених зображень. Вона є генеративною моделлю, метою якої є підбір та відновлення параметричного опису виразу за допомогою оптимізації [27]. Приклад результату роботи алгоритму з Active Appearance Model представлено на рис. 2.6.

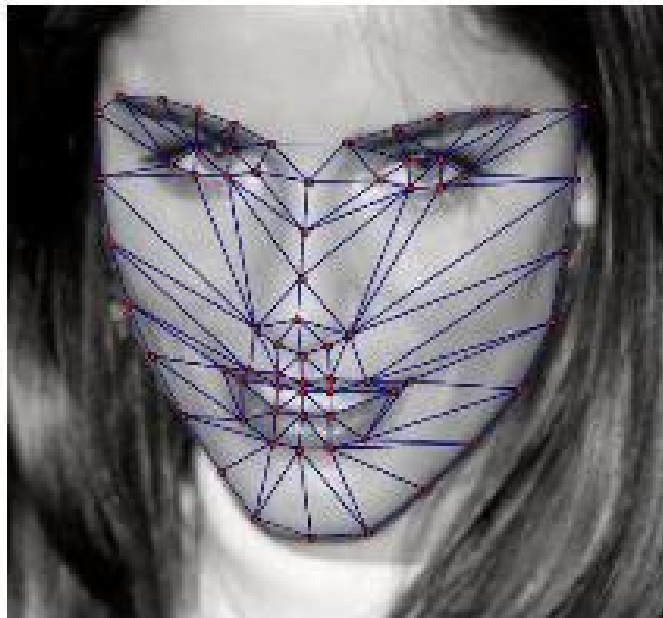


Рисунок 2.6 – Приклад результату роботи алгоритму з Active Appearance Model

– Facial Landmark Detection – це алгоритм комп'ютерного зору, в якому за допомогою моделі необхідно визначити ключові точки, що становлять області або орієнтири на обличчі людини - очі, ніс, губи тощо [28]. Виявлення орієнтирів людини є базовим алгоритмом, який можна використовувати для виконання інших завдань комп'ютерного зору, включаючи оцінку положення голови, визначення напрямку погляду, виявлення мімічних жестів та заміну осіб (див. рис. 2.7).

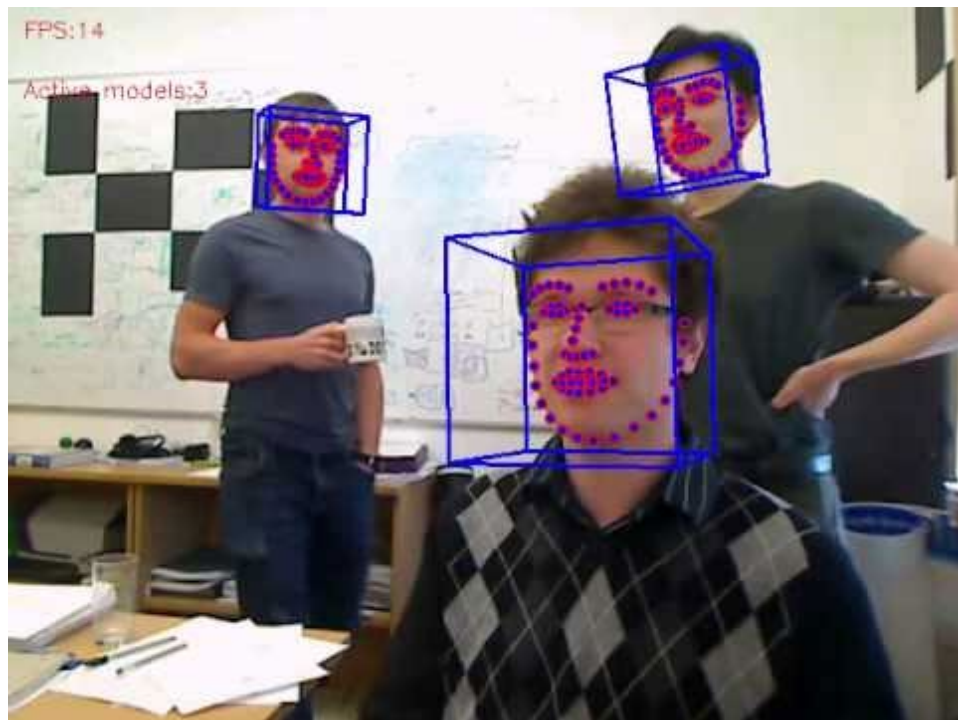


Рисунок 2.7 – Приклад візуалізації алгоритму Facial Landmark Detection

– Point Distribution Model – модель розподілу точок, концепції якої було розроблено Cootes, Taylor et al., що є стандартом у комп'ютерному зорі для статистичного дослідження форми та для сегментації зображень, де апріорні значення форми дійсно допомагають інтерпретувати шумові та малоконтрастні пікселі/вокселі. Останній пункт призводить до активних моделей форми (ASM) та активних моделей зовнішнього вигляду (AAM) [27]. Приклад результату роботи алгоритму з Point Distribution Model представлено на рис. 2.8.



Рисунок 2.8 – Приклад результату роботи алгоритму з Point Distribution Model

– 2D-and-3D-face-alignment – даний алгоритм використовується для реконструкції дво- або тривимірної моделі обличчя людини (див. рис. 2.9) [28].

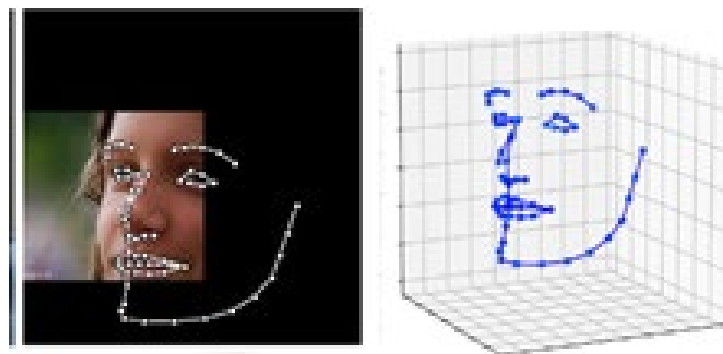


Рисунок 2.9 – Приклад результату роботи алгоритму 2D-and-3D-face-alignment

Слід відмітити, що розглянуті алгоритми відрізняються один від одного математичним апаратом, кількістю ключових точок та наявністю різних обмежень, пов'язаних із положенням обличчя на аналізованому зображенні.

Існують також інші методи пошуку ключових точок обличчя, але через їхню малу популярність у моїй роботі я їх не розглядаю.

Для того, щоб зв'язати дані про положення ключових точок обличчя з виразами його, що виявляються, необхідно навчити класифікатор. Як

класифікатори при вирішенні даної задачі зазвичай використовуються: нейронні мережі прямого поширення (neural networks), дерева прийняття рішень (decision trees) або класифікатор SVM (support vector machine) [29].

При навчанні нейронної мережі прямого поширення проводиться налаштування вагових коефіцієнтів зв'язків, що з'єднують нейрони різних шарів. У процесі навчання нейронної мережі ці коефіцієнти підбираються таким чином, щоб нейронна мережа максимально точно працювала на вибірці навчальних даних. В якості алгоритму навчання найчастіше використовують метод зворотного поширення помилки [29]. Структура нейронної мережі складається з:

- вхідного шару X , розмір якого відповідає сумарній кількості n параметрів усіх ключових точок обличчя;
- одного або кількох прихованих шарів, які реалізують обчислення нейронної мережі;
- вихідного шару Y , у якому кількість нейронів дорівнює кількості класів емоцій у навчальній вибірці даних [29].

Приклад нейронної мережі прямого поширення, здатної розпізнавати 3 класи емоцій на основі даних ключових точок обличчя, представлено на рис. 2.10.

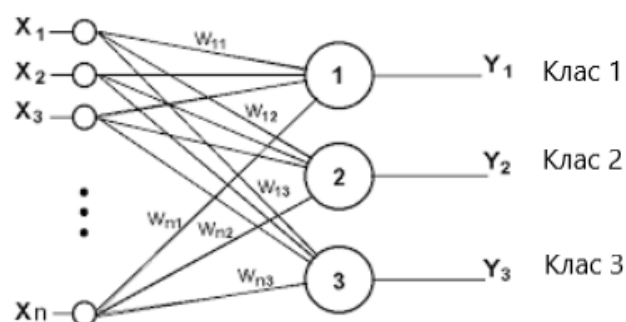


Рисунок 2.10 – Приклад нейронної мережі для розпізнавання 3 класів виразів обличчя за даними ключових точок

Дерево прийняття рішень також можна використовувати як класифікатор даних. Дерево рішень – це граф, у вузлах якого розміщуються умови переходу по ребрах до наступних вузлів дерева. Робота такого класифікатора починається з визначення послідовності переходів від кореневого вузла одного з листів. У листі міститься мітка класу, яка повертається користувачеві як результат класифікації даних. Для побудови дерев ухвалення рішень застосовуються різні алгоритми, серед яких найбільш популярні: CART та C5.0 [24]. Таким чином, на основі переміщення по вузлах дерева рішень за даними ключових точок обличчя здійснюється визначення класу емоцій (див. рис. 2.11).

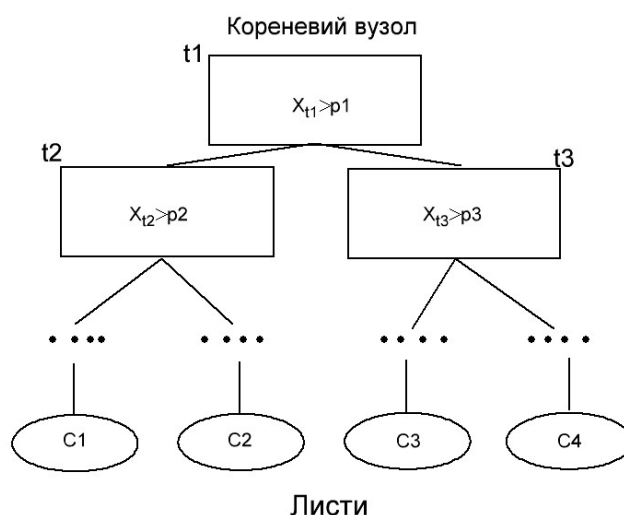


Рисунок 2.11 – Фрагмент дерева прийняття рішень: $t1, t2 \dots t3$ – позначення вузлів; $X_{t1}, X_{t2} \dots X_{t3}$ - параметри ключових точок обличчя; $C1, C2, C3, C4$ – позначення класів виразів обличчя

Робота класифікатора SVM заснована на побудові гіперплощини у просторі параметрів ключових точок обличчя, здатної відокремити різні класи виразів.

Таким чином, більшість алгоритмів для розпізнавання виразів обличчя людини ґрунтується на ідеї необхідності визначення ключових точок обличчя на зображенні. Такий підхід використовується з таких причин:

- простота ідеї - у реальному світі визначення емоцій людини можливе за рахунок спостереження за мімікою особи;

- технічні обмеження методів класифікації – вони адаптовані під аналіз зображень, оскільки графічна інформація містить велику кількість ознак даних.

Для аналізу графічної інформації вченими та інженерами було розроблено згорткові нейронні мережі [30]. За ідеєю, такий тип нейронних мереж за рахунок шарів згортки здатний самостійно визначати необхідні ознаки зображення, а за допомогою повнозв'язкового шару нейронів – здатний вирішувати завдання класифікації даних. За ідеєю, зв'язку методу пошуку ключових точок обличчя та методу класифікації можна замінити на згорткову нейронну мережу. І це дозволить не лише спростити математичний апарат розпізнавання виразів обличчя людини на зображенні, а й полегшити налаштування класифікатора емоцій за рахунок використання стандартних алгоритмів навчання нейронних мереж.

Висновки до розділу 2

Таким чином, до розділу 2 можна зробити висновки:

- встановлено, що більшість алгоритмів розпізнавання виразів обличчя людини засновані на спільній роботі двох методів: методу пошуку ключових точок обличчя на зображенні (Facial Landmark Detection, Point Distribution Model, 2D-and-3D-face-alignment та Active Appearance Model) та методу класифікації даних (нейронні мережі, дерева прийняття рішень, класифікатор SVM);

- запропоновано зв'язку методу пошуку ключових точок обличчя та методу класифікації, що можна замінити на згорткову нейронну мережу, яка дозволяє полегшити налаштування класифікатора виразів обличчя людини за рахунок використання стандартних алгоритмів навчання нейронних мереж.

РОЗДІЛ 3. ДОСЛІДЖЕННЯ НЕЙРОМЕРЕЖЕВОВОГО АЛГОРИТМУ АНАЛІЗУ ВИРАЗІВ ОБЛИЧЧЯ ЛЮДИНИ НА ЗОБРАЖЕННІ

3.1 Нейромережевий алгоритм аналізу виразів обличчя людини

Розпізнавання класу виразів обличчя людини за зображенням відноситься належить до задач класифікації. Формальний опис задачі класифікації інтерпретується наступним чином, існує завчасно відома і фіксована множина класів, кожен який є окремим типом емоцій. Для розпізнавання виразів обличчя необхідно сформуванати класифікатор, здатний на основі зображення, що подається на його вхід, визначати номер клас емоції, який людина відчуває на зображенні. Як класифікатор у моїй роботі запропоновано використовувати згорткову нейронну мережу.

Згорткові нейронні мережі можуть мати різну архітектуру, причому заздалегідь невідомо, яка архітектура покаже найкращі результати класифікації при наявному наборі зображень. Щоб підібрати архітектуру згорткової нейронної мережі необхідно вибрати стратегію її проектування [30].

У своїй роботі я вибрав наступні підходи для визначення структури нейронної мережі:

- віддавати перевагу використанню структури нейронної мережі з меншою кількістю шарів за умови збереження її працездатності на наявному наборі даних;
- додавати додаткові шари в нейронну мережу лише за неможливістю покращити результати роботи мережі іншими методами;
- для збільшення швидкості навчання нейронної мережі та спрощення її математичного апарату надавати переваги більш простим функцій активації - функція \max для шарів підвибірки, функція ReLU для проміжних шарів мережі [29];
- віддається перевага модифікації навчальної вибірки зображень замість ускладнення структури нейронної мережі.

На основі описаних вище принципів прийняття проектних рішень визначимо структуру нейронної мережі.

Перший шар будь-якої нейронної мережі завжди є вхідним [29]. Це означає, що на нього подаються вхідні дані, які потім обробляються нейронною мережею. У моєму випадку вхідними даними будуть цифрові зображення людей, тому кількість нейронів вхідного шару слід обирати в залежності від параметрів зображення:

- колірна модель зображення (RGB, Lab або відтінки сірого).
- кількість пікселів зображення (розмір зображення);

При аналізі деяких зображень з інших предметних областей колір пікселів може відігравати значну роль їх класифікації. Наприклад, при розпізнаванні дорожніх знаків, колір елементів на зображенні відіграє важливу роль, тому що у цьому випадку кількість вхідних нейронів буде розраховуватися як добуток висоти на ширину зображення в пікселях, помножений на кількість компонентів пікселя, що описує його колір. Наприклад, у колірній моделі RGB кількість компонентів, що описують колір одного пікселя дорівнює трьом. Завдання аналізу виразів обличчя, колір не несе у собі значні ознаки, тому для зниження кількості нейронів у вхідному шарі можна аналізувати зображення без врахування кольору (зображення у відтінках сірого). Відмова від кольору є звичайною практикою у системах комп'ютерного зору.

Так як ми будемо аналізувати зображення без врахування світлової складової, то кількість нейронів вхідного шару дорівнюватиме кількості пікселів аналізованого зображення. Кількість вхідних нейронів постійне і не підлягає зміні після налаштування нейронної мережі. Якщо кількість пікселів у аналізованого зображення більше ніж нейронів у вхідному шарі, то вихідне зображення необхідно зменшити (масштабувати) до необхідного розміру, перш ніж подати на вхід нейронної мережі.

Так як навчання нейронної мережі буде проводитися на існуючому наборі даних Facial Expression Recognition Challenge [31], то було підібрано розмір вхідного шару, що дорівнює розміру зображень, що є в ньому – 48×48 пікселів.

Основними шарами згорткової нейронної мережі є: шар згортки та шар підвибірки. Перший забезпечує виділення ознак на числовому наборі даних, а другий забезпечує зниження розмірності ознакового простору [29]. Мінімальна кількість вхідних блоків, що складаються з поєднання цих шарів дорівнює двом. Тому включимо їх до нашої структури.

Обов'язковим елементом надточної нейронної мережі є шар зниження розмірності, який перетворює всі дані попереднього шару у вектор. В кінці згорткової нейронної мережі використовується звичайний шар нейронів, що робить класифікацію даних, виділених із зображення за допомогою попередніх шарів мережі.

Таким чином, попередня структура нейронної мережі складатиметься з наступних шарів [30].

- вхідний шар для отримання зображення у вигляді числового масиву $45 \times 45 \times 1$;
- згортковий шар (convolution) з ядром 5×5 , кроком згортки 2, кількістю площин (каналів) 32 та функцією активації ReLU;
- шар підвиборки (max-pooling) із групою ущільнення 2×2 ;
- згортковий шар (convolution) з ядром 3×3 , кількістю площин (каналів) 64 та функцією активації ReLU;
- шар підвиборки (max-pooling) із групою ущільнення 2×2 ;
- шар зниження розмірності (flatten) даних до одновимірного вектора;
- шар (core layer ReLU) із 500 нейронів прямого поширення із функцією активації ReLU;
- вихідний (core layer SoftMax) шар із функцією активацією SoftMax.

Графічно архітектуру нейронної мережі представлено на рис. 3.1.

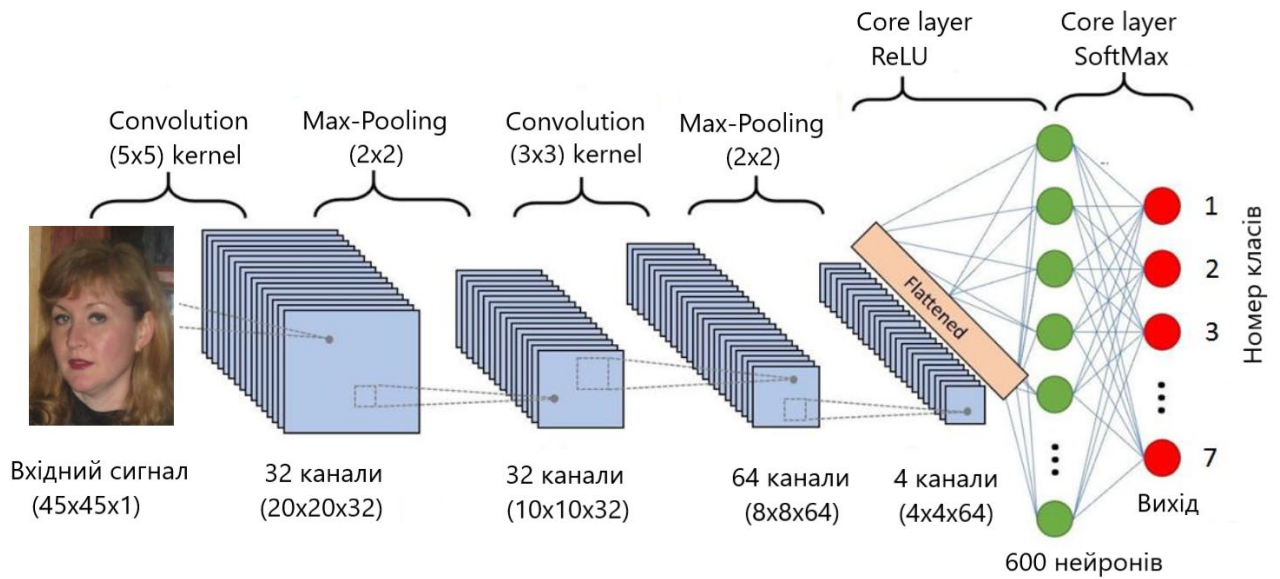


Рисунок 3.1 – Обрана архітектура нейронної мережі

Розглянемо математичний апарат роботи шарів нейронної мережі. Розмір ядра згорткового шару визначається за виразом:

$$\begin{cases} w_3 = w_i - K + 1, \\ h_3 = h_i - K + 1, \end{cases} \quad (3.1)$$

де w_3, h_3 – ширина та висота згорткової площини,

w_i, h_i – ширина та висота матриці сигналів попереднього шару,

K – ширина ядра згортки.

В якості функції активації використовується Rectified linear unit (ReLU):

$$f(a) = \begin{cases} 0, & a < 0, \\ a, & a \geq 0, \end{cases} \quad (3.2)$$

де $f(a)$ – шукане значення елемента,

a – зважена сума сигналів попереднього шару.

Ця функція є монотонною, з монотонною похідною і не апроксимує тотожну функцію поблизу початку координат.

Перевагами функції ReLU у порівнянні з іншими варіантами функції активації (гіперболічний тангенс, логістична функція) є:

- функція є найпростішою з математичної точки зору;
- функція має не складну похідну, що дає економію обчислень під час навчання та роботи нейронної мережі.

Вираз функціонування нейрона згорткового шару:

$$y_k^{(i,j)} = b_k + \sum_{s=1}^K \sum_{t=1}^K w_{k,s,t} \cdot x^{((i-1)+s)(j+t)}, \quad (3.3)$$

де $y_k^{(i,j)}$ – нейрон k -ї площини (k -ого каналу) згорткового шару,

b_k – нейронне зміщення k -ї площини,

K – розмір рецептивної області нейрона (розмір ядра згортки),

$w_{k,s,t}$ – матриця синаптичних коефіцієнтів,

x – виходи нейронів попереднього шару.

Схема роботи згорткового шару для кольорового зображення з трьома каналами (RGB) представлена на рис.3.2.

Значення компонентів ядра визначаються під час навчання мережі. Початкові ваги спочатку мають випадкові значення і коригуються в процесі навчання. Перемноження та підсумовування повторюються для кожної локації, через яку проходить ядро.

Вираз функціонування шару підвиборки (підвибіркового шару) з групою ущільнення 2×2 має вигляд:

$$y_k^{(i,j)} = b_k + w_k \cdot \max \left(x^{((2i,2j),(2i,2j))}, x^{((2i,2j),(2i,2j)+1)}, x^{((2i,2j)+1,(2i,2j))}, x^{((2i,2j)+1,(2i,2j)+1)} \right), \quad (3.4)$$

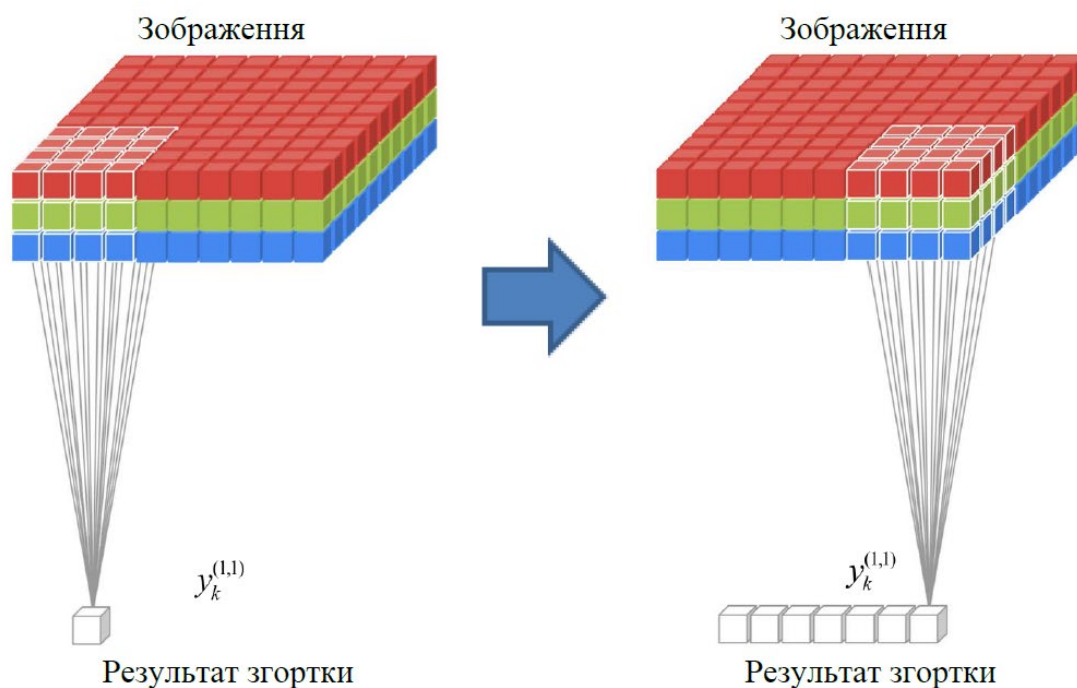


Рисунок 3.2 – Схема роботи згорткового шару k -ї площини

Приклад виконання операції підвиборки із використанням функції $\max()$ з групою ущільнення 2×2 представлено на рис. 3.3.

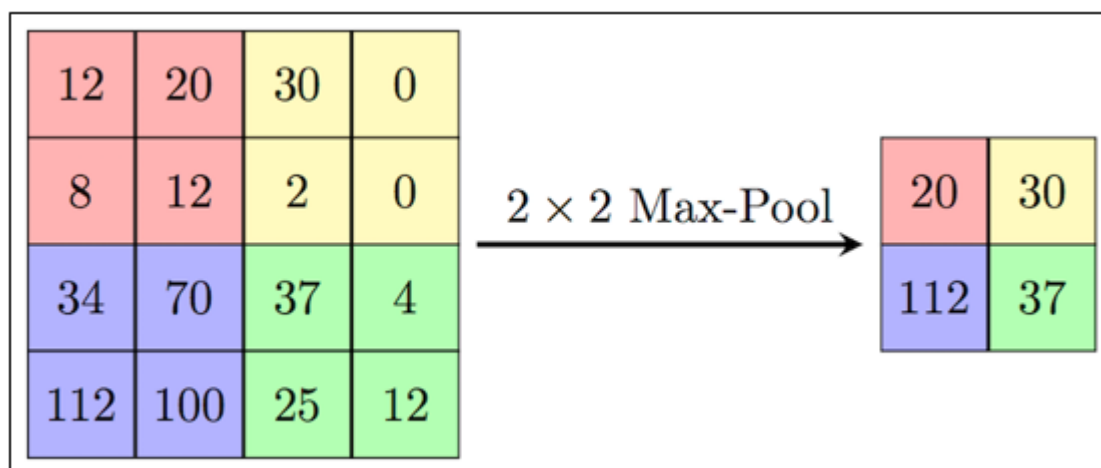


Рисунок 3.3 – Приклад виконання субдискретизації з використанням функції $\max()$ та групою ущільнення 2×2 (групи виділені кольором)

Аналогічно, на основі виразів (3.1-3.4) функціонують другий згортковий шар та другий шар підвибірки.

Шар зниження розмірності (flatten) виконує єдину дію – перетворює дані попереднього шару на одномірний вектор.

Після шару зниження розмірності фактично створюється стандартна нейронна мережа прямого поширення з одним прихованим шаром, що складається з 500 нейронів. Вихідний шар складається з 43 нейронів (43 класу зображень) із функцією активації SoftMax. Таким чином, на виході кожного нейрона вихідного шару буде число, що виражає ступінь впевненості мережі в кожному з класів для зображення, що розглядається.

3.2 Dataset для навчання нейронної мережі

Навчання нейронної мережі повинно здійснюватися з урахуванням розміченої вибірки даних. Чим складніша нейронна мережа, тим більшою повинна бути навчальна вибірка. Для обраної в п. 3.1 архітектури нейронної мережі наближена кількість навчальних прикладів, що визначається на основі кількості вагових коефіцієнтів, повинна становити приблизно від 20000 до 100000 зображень.

Для навчання та тестування нейронної мережі було прийнято рішення використовувати вже підготовлений набір даних Challenges in Representation Learning: Facial Expression Recognition Challenge [31], що наявний у відкритому доступі на веб-сторінці змагання ICML Workshop on Representation Learning (див. рис. 3.4).

Цей набір був створений у 2013 році з метою проведення змагань на платформі «Kaggle» вченими, які працюють у галузі машинного навчання, - П'єром Люком Карр'є та Аароном Курвіллом [31]. В результаті змагання найкращими стали три команди, які використовували згорткові нейронні мережі (CNN – convolutional neural networks) та методи перетворення зображень.

Переможець, Ічуань Тан, зміг досягти точності 71,162%, використавши як функцію втрат методом опорних векторів.

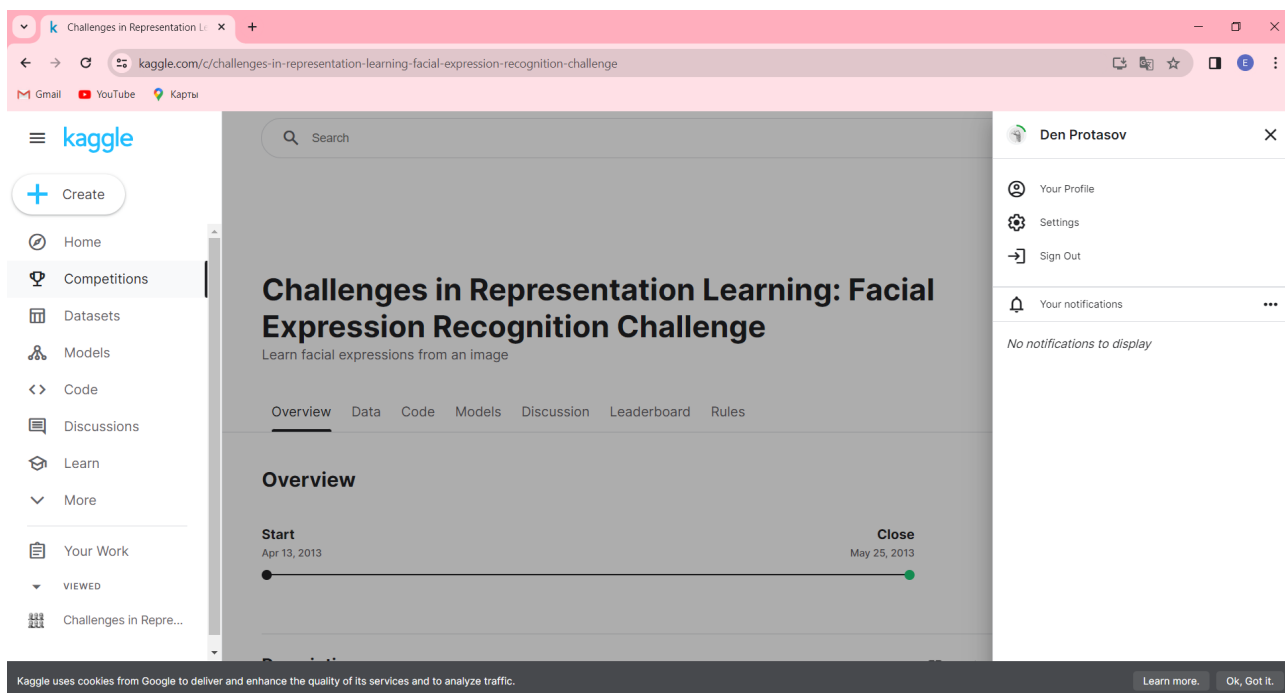


Рисунок 3.4 – Веб-сторінка Challenges in Representation Learning: Facial Expression Recognition Challenge

Архів з навчальним та тестовим наборами даних (Facial Expression Recognition Challenge Dataset) займає близько 777,8 МБ. Вміст архіву представлено на рис. 3.5. Він складається з файлів:

1. train.csv, що містить два стовпці: «пікселі» та «емоції». Стовпець «Пікселі» містить рядок, укладений у лапки для кожного зображення. Стовпець «емоція» містить чисельний код від 0 до 6 включно для емоції, яка наявна на зображенні. У цьому файлі зберігається навчальна вибірка з прикладами даних у вигляді 28709 зображень.

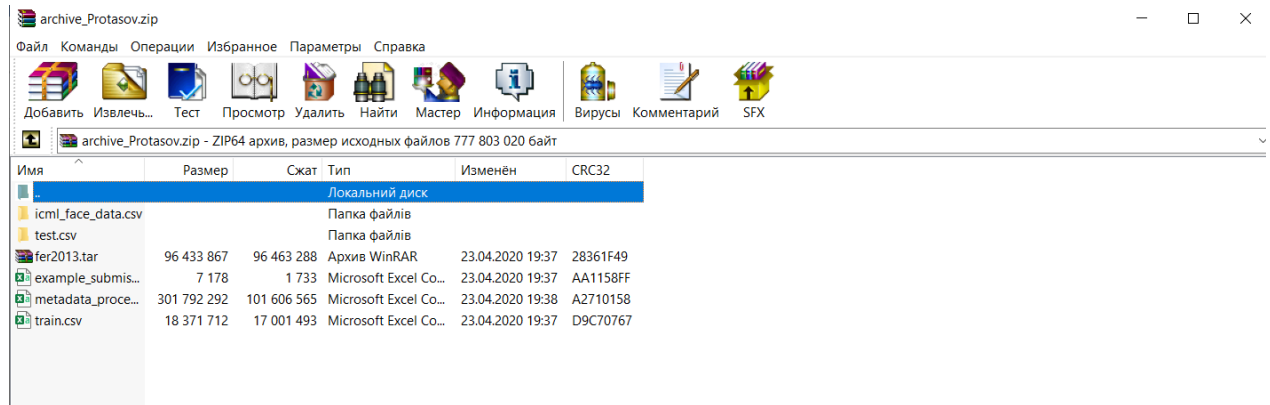


Рисунок 3.5 – Структура архіву з навчальними та тестовими даними

Вміст файлу організовано таким чином. Спочатку у вигляді числа заданих класів емоцій, представлених на зображенні, а потім після коми в лапках відображаються значення пікселів самого зображення, який представлений у вигляді вектора. Так як зображення мають розмір 48×48 пікселів у відтінках сірого, то довжина вектора – 2304 значення (див. рис. 3.6).

2. test.csv – це файл у якому зберігаються дані для тестування навчального класифікатора емоцій. У ньому містяться дані 7178 зображень, організовані таким же чином, як у файлі train.csv.

3. Папка icml_face_data.csv є результатом об'єднання файлів train.csv і test.csv. Вона необхідна у тому випадку, якщо виникне потреба самостійно розділити дані на навчальну та тестову вибірку.

4. example_submission.csv містить приклад того, як повинен працювати класифікатор емоцій на зображенні.

5. Файл містить fer2013.tar.gz ті ж набори даних із зображеннями обличчя, але представлені в іншому форматі.

	A	B	C	D	E	F	G	H	I
1	emotion,pixels								
2	0,"70 80 82 72 58 58 60 63 54 58 60 48 89 115 121 119 115 110 98 91 84 84 90 99 110 126 143 153 158								
3	0,"151 150 147 155 148 133 111 140 170 174 182 154 153 164 173 178 185 185 189 187 186 193 194 1								
4	2,"231 212 156 164 174 138 161 173 182 200 106 38 39 74 138 161 164 179 190 201 210 216 220 224 2								
5	4,"24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 19 43 52 13 26 40 59 65 12 20 63 99 98 98 111 75 62 4								
6	6,"4 0 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50 58 84 115 127 137 142 151 156 155 149 153 152 157 160 162								
7	2,"55 55 55 55 55 54 60 68 54 85 151 163 170 179 181 185 188 188 191 196 189 194 198 197 195 194 1								
8	4,"20 17 19 21 25 38 42 42 46 54 56 62 63 66 82 108 118 130 139 134 132 126 113 97 126 148 157 161								
9	3,"77 78 79 79 78 75 60 55 47 48 58 73 77 79 57 50 37 44 56 70 80 82 87 91 86 80 73 66 54 57 68 69 68								
10	3,"85 84 90 121 101 102 133 153 153 169 177 189 195 199 205 207 209 216 221 225 221 220 218 222 2								
11	2,"255 254 255 254 254 179 122 107 95 124 149 150 169 178 179 179 181 181 184 190 191 191 193 19								
12	0,"30 24 21 23 25 25 49 67 84 103 120 125 130 139 140 139 148 171 178 175 176 174 180 180 178 178								
13	6,"39 75 78 58 58 45 49 48 103 156 81 45 41 38 49 56 60 49 32 31 28 52 83 81 78 75 62 31 18 19 19 20 :								
14	6,"219 213 206 202 209 217 216 215 219 218 223 230 227 227 233 235 234 236 237 238 234 226 219 2								
15	6,"148 144 130 129 119 122 129 131 139 153 140 128 139 144 146 143 132 133 134 130 140 142 150 1								

Рисунок 3.6 – Вміст файлу train.csv

У наборі даних «Facial Expression Recognition Challenge Dataset» усі зображення обличчя розмічені та поділені на 7 класів емоцій. Кожен клас емоцій (emotion) позначений числом від 0 до 6 (див. рис. 3.7):

- мітка 0 – клас «Angry» (злість);
- мітка 1 – клас «Disgust» (огида);
- мітка 2 – клас «Fear» (страх);
- мітка 3 – клас «Happy» (щастя);
- мітка 4 – клас «Sad» (сум);
- мітка 5 – клас «Surprise» (здивування);
- мітка 6 – клас «Neutral» (спокій).

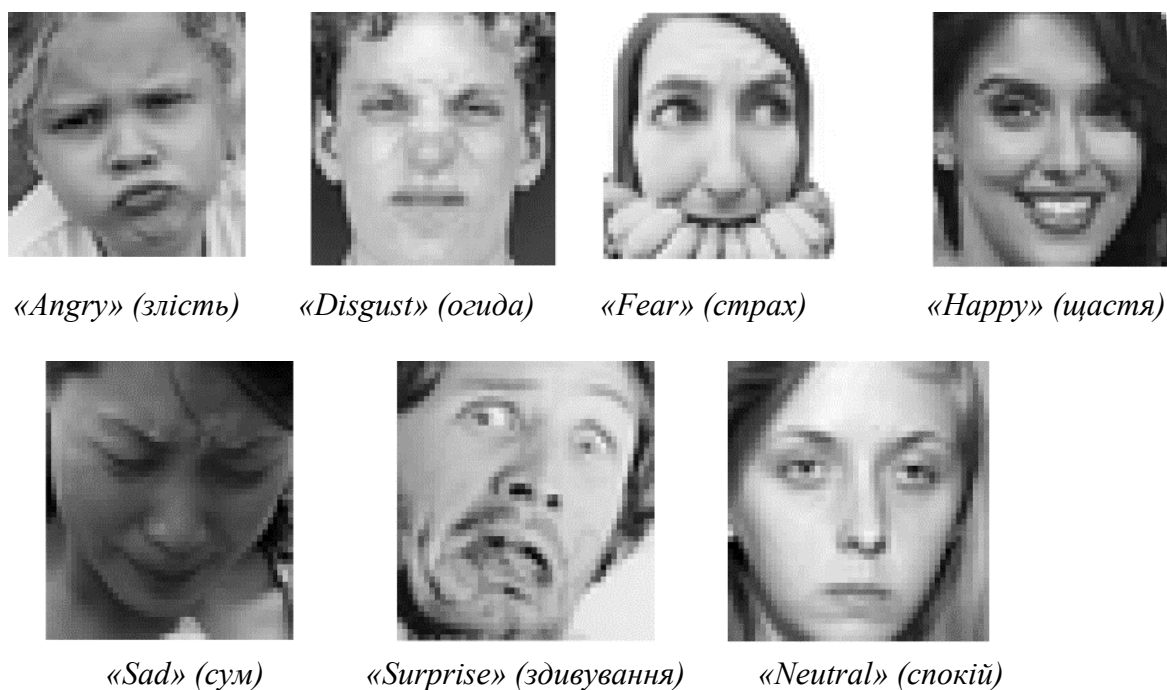


Рисунок 3.7 – Приклади зображень навчальної вибірки для кожного класу

3.3 Тестування та оптимізація структури нейромережі

Проектування нейронної мережі є ітераційним процесом, це означає, що може знадобитися кілька ітерацій до того, як буде знайдено оптимальну під наявні дані архітектуру мережі.

Було проведено кілька циклів навчання нейронної мережі з архітектурою, представленою на рис. 3.2, на наборі даних «Facial Expression Recognition Challenge Dataset». За результатами оцінки роботи навченої мережі на тестовому наборі даних було встановлено, що середня точність розпізнавання виразів обличчя на зображеннях людей становить 65 %. Ці результати є незадовільними, тому було проведено аналіз причин низької точності роботи мережі.

У ході такого аналізу вивчалися результати роботи на матриці помилок, що показує, на зображення яких класів найчастіше виникає некоректна робота мережі. У ході тестування з'ясувалося, що нейронна мережа найчастіше помиляється на зображеннях з класу Disgust (огида). Підрахувавши програмним способом розподіл зображень у навчальній вибірці за класами, була побудована

стовпчаста діаграма, представлено на рис. 3.8. З діаграми видно, що клас «Disgust» є малочисленним, тобто прикладів у навчальній вибірці недостатньо, щоб нейронна мережа навчилася коректно працювати із зображеннями цього класу.

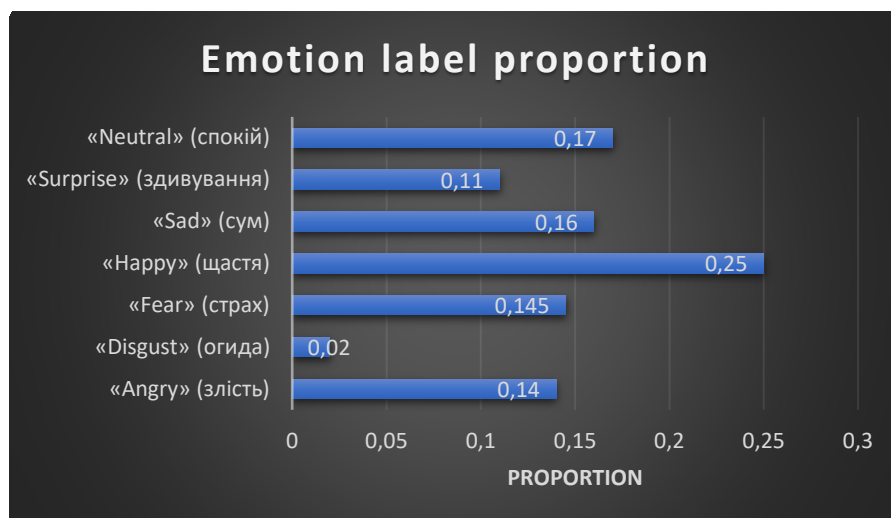


Рисунок 3.8 – Діаграма розподілу зображень у навчальній вибірці за класами

Для вирішення цієї проблеми необхідно збільшити чисельність прикладів у даному класі. Для вирішення цього завдання було прийнято рішення використовувати процедурну генерацію нових зображень на основі наявних.

Для цього до існуючих зображень застосовуються послідовно операції обертання по осі та усунення кутів зображення. З використанням такого підходу створювалися нові приклади зображень, аж до збільшення частки класу Disgust до рівня Surprise.

Приклади згенерованих зображень представлено на рис. 3.9.

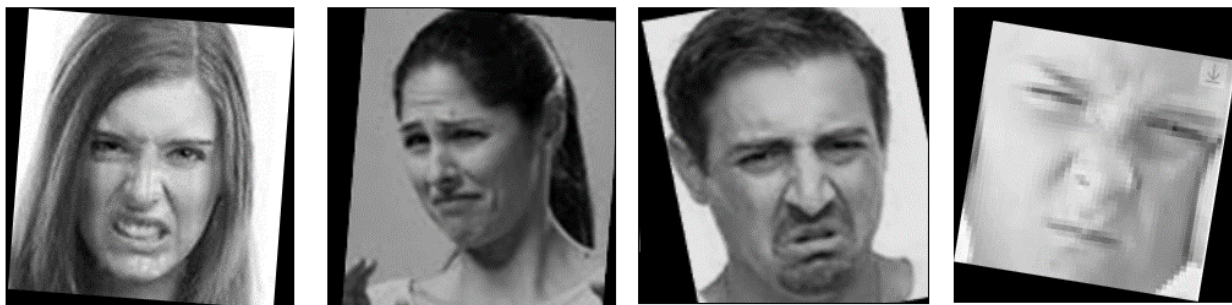


Рисунок 3.9 – Приклади синтезованих зображень на основі наявних у навчальній вибірці зображень для класу емоцій «Disgust»

Така генерація додаткових зображень не може бути повноцінною заміною розширення навчальної вибірки за рахунок додавання унікальних зображень, тому що згенеровані зображення багато в чому мають збіжність з оригінальними зображеннями, тому необхідно вживати заходів для захисту від ефекту перенавчання мережі.

У цьому випадку стандартним рішенням є додавання спеціальних шарів – «dropout» [32]. Дані шари функціонують в такий спосіб. У процесі навчання вони випадково блокують передачу заданої частини сигналів у наступний шар нейронів. Це змушує нейронну мережу кожної ітерації навчання звертати увагу до різних ознак зображень, знижуючи домінування одних ознак з іншими. Це особливо важливо, при вибраному методі збільшення зображень класу «Disgust» в навчальній вибірці даних. Також варто відзначити, що шари «dropout» функціонують тільки в процесі навчання мережі, а при використанні навченої мережі ці шари не впливають на сигнали.

Таким чином, оптимізована структура нейронної мережі включає наступні шари (шари «dropout» працюють тільки в процесі навчання мережі). Шари «dropout» були додані після двох шарів підвиборки і після одного шару «core layer ReLU».

При повторному навчанні нейронної мережі на вибірці даних з доповненим класом «Disgust» та з використанням шарів «dropout» було отримано наступні

результати: Швидкість навчання знизилася в 1,5 рази, але точність розпізнавання виразів обличчя на тестовій вибірці зображень збільшилася до 77,5 % (див. рис. 3.10).

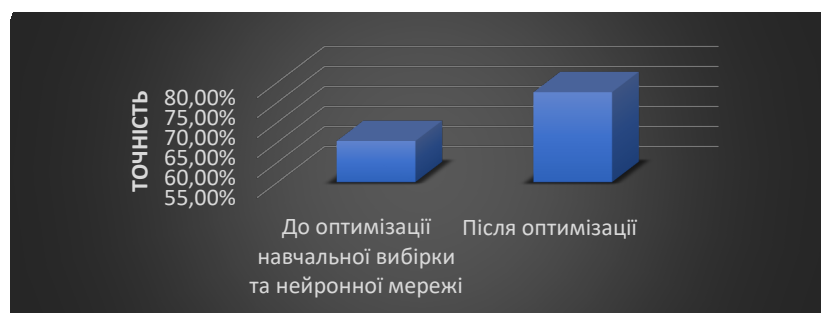


Рисунок 3.10 – Вплив оптимізації на точність роботи мережі на тестовій вибірці даних

Висновки до розділу 3

Досліджено нейромеревий алгоритм, який призначений для вирішення задачі розпізнавання виразів обличчя на зображенні, має просту архітектуру нейронної мережі, що складається з вхідного шару для отримання зображення у вигляді числового масиву $45 \times 45 \times 1$; згортковий шар (Convolution) з ядром 5×5 , кроком згортки 2, кількістю площин (каналів) 32 та функцією активації ReLU; шар підвибірки (Max-pooling) із групою ущільнення 2×2 ; шар «dropout», що блокує сигнали 25% нейронів попереднього шару; згортковий шар (Convolution) з ядром 3×3 , кількістю площин (каналів) 64 та функцією активації ReLU; шар підвибірки (Max-pooling) із групою ущільнення 2×2 ; шар «dropout», що блокує сигнали 25% нейронів попереднього шару; шар зниження розмірності даних (Flatten) до одновимірного вектора; шар (Core layer ReLU) із 500 нейронів прямого поширення з функцією активації ReLU та вихідний (Core layer SoftMax) шар із функцією активацією SoftMax.

Проведено тестування шляхом навчання нейронної мережі на наборі даних «Facial Expression Recognition Challenge Dataset». За результатами оцінки роботи

навченої мережі на тестовому наборі даних було встановлено, що середня точність розпізнавання виразів обличчя на зображеннях людей становить 65 %, що є незадовільним.

Збільшення точності роботи нейронної мережі з 65,3 % до 77,3% вдалося досягти за рахунок додавання шарів «dropout», що функціонують тільки в процесі налаштування мережі та послаблюють ефект перенавчання, а також за рахунок синтезування додаткових зображень у найменший клас (класу емоцій «Disgust») зображень навчальної вибірки. Синтез додаткових зображень здійснювався за рахунок застосування з випадковими параметрами операцій масштабування та обертання до вихідних зображень, що містяться у класі.

ВИСНОВКИ

На основі аналізу базових методів розв'язання задачі розпізнавання встановлено:

- всі розглянуті методи мають свої переваги та можуть показати кращі результати у певних задачах порівняно з іншими. Наприклад, нейромережеві методи можуть показати кращі результати у задачах класифікації зображень, тоді як SVM може бути ефективнішим у задачах класифікації тексту;

- нейромережеві методи, такі як згорткові нейронні мережі, мають більшу точність у задачах ідентифікації та розпізнавання виразів обличчя людини порівняно з іншими розглянутими у роботі методами, такими як PCA, SVM тощо.

- у системах розпізнавання виразів обличчя нейронні мережі можуть бути використані для навчання при великому наборі зображень, а також для розпізнавання нових зображень на основі цього навчання.

Розглянуто сучасні програмні продукти, які широко використовуються для розпізнавання виразів обличчя людей. Вони мають структуру інтегрованого програмного забезпечення або додатка із великим набором функцій для роботи із машинним зором і зображеннями. Слід відмітити, що результати роботи програмних продуктів для розпізнавання виразів обличчя можуть бути не зовсім точними, особливо в умовах мінливості освітлення, кута огляду та інших факторів. Це може призвести до помилкових спрацьовувань та помилок ідентифікації.

Встановлено, що більшість алгоритмів розпізнавання виразів обличчя людини засновані на спільній роботі двох методів: методу пошуку ключових точок обличчя на зображенні (Facial Landmark Detection, Point Distribution Model, 2D-and-3D-face-alignment та Active Appearance Model) та методу класифікації даних (нейронні мережі, дерева прийняття рішень, класифікатор SVM).

Запропоновано зв'язку методу пошуку ключових точок обличчя та методу класифікації, що можна замінити на згорткову нейронну мережу, яка дозволяє

полегшити налаштування класифікатора виразів обличчя людини за рахунок використання стандартних алгоритмів навчання нейронних мереж.

Досліджено нейромережевий алгоритм для вирішення задачі розпізнавання виразів обличчя на зображенні, який має просту архітектуру нейронної мережі, що складається з вхідного шару для отримання зображення у вигляді числового масиву $45 \times 45 \times 1$; згортковий шар (Convolution) з ядром 5×5 , кроком згортки 2, кількістю площин (каналів) 32 та функцією активації ReLU; шар підвибірки (Max-pooling) із групою ущільнення 2×2 ; шар «dropout», що блокує сигнали 25% нейронів попереднього шару; згортковий шар (Convolution) з ядром 3×3 , кількістю площин (каналів) 64 та функцією активації ReLU; шар підвибірки (Max-pooling) із групою ущільнення 2×2 ; шар «dropout», що блокує сигнали 25% нейронів попереднього шару; шар зниження розмірності даних (Flatten) до одновимірного вектора; шар (Core layer ReLU) із 500 нейронів прямого поширення з функцією активації ReLU та вихідний (Core layer SoftMax) шар із функцією активацією SoftMax.

Проведено тестування шляхом навчання нейронної мережі на наборі даних «Facial Expression Recognition Challenge Dataset». За результатами оцінки роботи навченої мережі на тестовому наборі даних було встановлено, що середня точність розпізнавання виразів обличчя на зображеннях людей становить 65 %, що є незадовільним.

Збільшення точності роботи нейронної мережі з 65,3 % до 77,3% вдалося досягти за рахунок додавання шарів «dropout», що функціонують тільки в процесі налаштування мережі та послаблюють ефект перенавчання, а також за рахунок синтезування додаткових зображень у найменший клас (класу емоцій «Disgust») зображень навчальної вибірки. Синтез додаткових зображень здійснювався за рахунок застосування з випадковими параметрами операцій масштабування та обертання до вихідних зображень, що містяться у класі.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Червяк П. І. Медична енциклопедія. Видання третє, доповнене. Київ, Вид. центр «Просвіта», 2012. 1504 с.
2. С. Darwin. The Expression of the Emotions in Man and Animals. With an introduction, afterword, and commentaries by Paul Ekman. 3rd ed. Oxford University Press, 1998, URL: <http://darwin-online.org.uk/content/frameset?pageseq=1&itemID=F1142&viewtype=text>
3. P. Ekman, Universals and cultural differences in facial expressions of emotion. Nebraska, Lincoln University of Nebraska Press, 1971, Vol. 19, Lincoln University of Nebraska Press, URL: <https://www.paulekman.com/wp-content/uploads/2013/07/Universals-And-Cultural-Differences-In-Facial-Expressions-Of.pdf>
4. Md. Abdur Rahim, Md. Najmul Hossain, Tanzillah Wahid and Md. Shafiul Azam. "Face recognition using local binary patterns (LBP)", Global journal of computer science and technology graphics & vision, Vol - 13, Issue 4, Version 1.0, 2013.
5. Khorrami P., Le Paine T., Brady K., Dagli C., Huang T. S. How Deep Neural Networks Can Improve Emotion Recognition on Video Data. 2016 IEEE International Conference on Image Processing (ICIP). Piscataway, IEEE Publ., 2016, pp. 619–623.
6. Ульяновко А.Л., Дорофеев Ю.І. (2020). Огляд методів автоматичного розпізнання емоційного стану людини по зображенню. Вісник Національного технічного університету «ХПІ». Серія: Системний аналіз, управління та інформаційні технології, 1 (3), 85–89. URL: <https://doi.org/10.20998/2079-0023.2020.01.15>
7. Computer imaging: digital image analysis and processing. By S E Umbaugh. pp. 696, 2005 (CRC Press, Taylor & Francis Group, Boca Raton, FL), £38.99 ISBN 0-8493-2919-1.

8. Y. Chen, H. Wu, T. Wang, Y. Wang. A Comparison of Methods of Facial Expression Recognition. Proceedings of the 1st WRC Symposium on Advanced Robotics and Automation 2018 Beijing, 2018. P. 261-262.
9. O. Arriaga, P. Plöger, Matias Valdenegro. Real-time Convolutional Neural Networks for Emotion and Gender Classification. ArXiv, 2017.
10. D. Duncan, G. Shine, C. English. Facial Emotion Recognition in Real Time. 2016.
11. I. Swathi, S.S. Gantayat. Facial Expression Recognition Based On Emotion For Visually Challenged People. ICICSS, 2018.
12. Gil-Jin Jang, Jeong-Sik Park, Ahra Jo, Ji-Hwan Kim. Facial Emotion Recognition Using Active Shape Models and Statistical Pattern Recognizers. Ninth International Conference on Broadband and Wireless Computing, Communication and Applications, 2014.
13. Ekman, P., Friesen, W. V., & Hager, J. C. Facial Action Coding System. Manual and Investigator's Guide, Salt Lake City, UT: Research Nexus, 2002.
14. C. Shan, S. Gong, and P. W. McOwan, Facial expression recognition based on local binary patterns: A comprehensive study, Image and Vision Computing, 2009. P. 803 – 816.
15. N. Mehta, S. Jadhav. Facial Emotion recognition using Log Gabor filter and PCA. International Conference on Computing Communication Control and automation, 2016.
16. Метод Віюли-Джонса: URL: <https://habr.com/ru/articles/133826/>
17. Elastic Bunch Graph Matching: URL: http://www.scholarpedia.org/article/Elastic_Bunch_Graph_Matching
18. Прихована Марковська модель: URL: https://uk.wikipedia.org/wiki/Прихована_марковська_модель
19. Метод ГОЛОВНИХ КОМПОНЕНТ: URL: https://uk.wikipedia.org/wiki/Метод_ГОЛОВНИХ_КОМПОНЕНТ
20. Support Vector Machines: URL: https://en.wikipedia.org/wiki/Support_vector_machine

21. FaceReader: URL: <https://www.noldus.com/human-be-havior-research/products/facereader>
22. Project Oxford: URL: <https://azure.microsoft.com/ru-ru/services/cognitive-services/>
23. Cognitec: URL: <https://www.cognitec.com/facevacs-videoscan.html>
24. Хабб: URL: <https://habr.com>
25. Calonder M. BRIEF: binary robust independent elementary features / M. Calonder, V. Lepetit, C. Strecha, P. Fua // European Conference on Computer Vision, 2010. P. 778-792.
26. Технологія модульного навчання. Дж. Расел (США), П. Юцевічене, М. А. Чошанов: URL: <https://fschool2.jimdo.com/технологія-модульного-навчання/>
27. The Menpo Project: URL: <https://www.menpo.org/menpofit/pdm.html>
28. GITHUB: URL: <https://github.com/TadasBaltrusaitis/OpenFace?tab=readme-ov-file>
29. Лисенко В.П., Решетюк В.М., Штепа В.М., Заєць Н.А. та ін. Системи штучного інтелекту: нечітка логіка, нейронні мережі, нечіткі нейронні мережі, генетичний алгоритм. – К: НУБІП України, 2014. 336 с.
30. Khandelwal R. Convolutional Neural Network (CNN) Simplified. 2018.: URL : <https://medium.com/datadriveninvestor/convolutional-neuralnetwork-cnn-simplified-eca4d4ee52c5>
31. Facial Expression Recognition Challenge: URL: <https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge>
32. Хабб: URL: <https://habr.com/ru/companies/wunderfund/articles/330814/>