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)

Дослідження згорткових нейронних мереж в задачі розпізнавання людських емоцій
(тема)

Виконав:
студент 2 курсу, групи СШІм-19-2
Захарченко Д.О.
(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 проф. Терзіян В.Я.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри _____ В.О. Філатов
(підпис) (прізвище, ініціали)

2021 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 Захарченку Дмитру Олеговичу
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження згорткових нейронних мереж в задачі розпізнавання людських емоцій

затверджена наказом університету від 29 03 20 21 р. № 390 Ст

2. Термін подання студентом роботи до екзаменаційної комісії 20 05 20 21 р.

3. Вихідні дані до роботи Функція: Розробка компонентів побудови графів знань з неструктурованих текстових джерел. Організація даних: файлова з прямим доступом. Перелік використовуваних програмних засобів: ОС Windows 10, середовище розробки Jupyter Notebook, середовище розробки Google Colab, фреймворк TensorFlow, мова програмування Python.

4. Перелік питань, що потрібно опрацювати в роботі 1)Аналіз предметної галузі. 2) Загальна характеристика згорткових нейронних мереж. 3) Огляд методів розпізнавання людських емоцій. 4) Огляд наборів даних. 5) Постановка цілей і задачі дослідження. 6) Розробка та опис програмного додатку. 7) Проведення аналізу виконаної роботи та формування висновків.

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри) Рисунок 1 – Основні емоції; Рисунок 2 – Приклад архітектури з паралельними каналами; Рисунок 3 – Будова нейрона; Рисунок 4 – Слабкозв'язні нейромережі; Рисунок 5 – Повнозв'язані нейромережі; Рисунок 6 – Приклад нейромережі; Рисунок 7 – Основні емоції CK+; Рисунок 8 – Основні емоції MMI; Рисунок 9 – Основні емоції JAFFE; Рисунок 10 – Архітектура VGG19; Рисунок 11 – Архітектура VGG16; Рисунок 12 – Зв'язок швидкого доступу; Рисунок 13 – Функія активації ReLu; Рисунок 14 – Архітектура ResNet34; Рисунок 15 – Блок Inception; Рисунок 16 – Початковий блок зі зменшеними розмірами; Рисунок 17 – Покращений початковий блок; Рисунок 18 – Модифікована архітектура VGG19; Рисунок 19 – Блок у пропонованій внутрішній архітектурі; Рисунок 20 – Архітектура заснована на власному сприйнятті; Рисунок 21 – Модифікована архітектура VGGFace; Рисунок 22 – Приклад неправильної класифікації; Рисунок 23 – Приклад об'єкта з бази даних MMI ; Рисунок 24 – Приклад неправильної класифікації MMI; Рисунок 25 – Приклади с бази даних CK+.

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 на кваліфікаційну роботу	29.03.2021	Виконано
2	Аналіз завдання та об'єкту дослідження	01.04.2021	Виконано
3	Пошук та аналіз тематичної літератури	03.04.2021	Виконано
4	Вибір програмних та технічних засобів	04.04.2021	Виконано
5	Проектування та розробка програмного засобу	08.04.2021	Виконано
6	Аналіз результатів роботи програмного засобу	13.04.2021	Виконано
7	Підготовка пояснювальної записки	18.04.2021	Виконано
8	Проходження нормконтролю та рецензування	23.04.2021	Виконано
9	Перевірка на плагіат	27.04.2021	Виконано
10	Перевірка печатної версії роботи	12.05.2021	Виконано
11	Попередній захист кваліфікаційної роботи	17.05.2021	Виконано
12	Захист кваліфікаційної роботи перед ЕК	20.05.2021	

Дата видачі завдання 29 березня 20 21 р.

Студент _____
(підпис)

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

РЕФЕРАТ

Пояснювальна записка: 68 с., 13 табл., 25 рис., 2 дод., 23 джерела.

ЛЮДСЬКІ ЕМОЦІЇ, МАШИННЕ НАВЧАННЯ, СГОРТКОВІ НЕЙРОННІ МЕРЕЖІ, ЦИФРОВІ ЗОБРАЖЕННЯ, JUPYTERNOTEBOOK, PYTHON, TENSORFLOW.

Об’єкт дослідження – основні підходи розпізнавання людських емоцій.

Предмет дослідження – розпізнавання емоцій, за допомогою згорткових нейронних мереж.

Мета роботи – дослідження та використання методів розпізнавання людських емоцій.

Методи роботи – методи розпізнавання емоцій, бібліотека TensorFlow, методи валідації, аналіз вільних баз даних, отримані результати досліджень із застосуванням інтегрованого середовища розробки Jupyter Notebook та середовища розробки Google Colab, мови програмування Python.

Результати кваліфікаційної роботи – в результаті проведених досліджень вирішено задачу порівняння та аналізу різноманітних методів розпізнавання людських емоцій. На основі проведених досліджень було отримано точність розпізнавання в різних базах даних. А також розроблена власна модель

Область застосування – дана розробка може бути корисною для задач розпізнавання людських емоцій, також застосовувати в медицині.

РЕФЕРАТ

Пояснительная записка: 68 с., 13 табл., 25 рис., 2 доп., 23 источника.

МАШИННОЕ ОБУЧЕНИЕ, СВЕРТОЧНЫЕНЕЙРОННЫЕ СЕТИ, ЦИФРОВЫЕ ИЗОБРАЖЕНИЯ, ЧЕЛОВЕЧЕСКИЕ ЭМОЦИИ, JUPYTER NOTEBOOK, PYTHON, TENSORFLOW.

Объект исследования – основные подходы распознавания человеческих эмоций.

Предмет исследования – распознавание эмоций, с помощью згорткових нейронных сетей.

Цель работы – исследование и использование методов распознавания человеческих эмоций.

Методы работы – методы распознавания эмоций, библиотека TensorFlow, методы валидации, в частности, анализ свободных баз данных, полученные результаты исследований с применением интегрированной среды разработки Jupyter Notebook и среды разработки Google Colab, языки программирования Python.

Результаты квалификационной работы – в результате проведенных исследований решена задача сравнения и анализа различных методов распознавания человеческих эмоций. На основе проведенных исследований были получены точность распознавания в различных базах данных. А также разработана собственная модель

Область применения – данная разработка может быть полезной для задач распознавания человеческих эмоций, также применять в медицине.

ABSTRACT

Explanatory note: 68 p., 13 tabl., 25 fig., 2 ann., 23 sources.

CONVERTING NEURAL NETWORKS, DIGITAL IMAGES, HUMAN EMOTIONS, JUPYTER NOTEBOOK, MACHINE LEARNING, PYTHON, TENSORFLOW

The object of research is the main approaches to recognizing human emotions.

The subject of research is the recognition of emotions with the help of convolutional neural networks.

The purpose of the work is to study and use methods of recognizing human emotions.

Methods of work – emotion recognition methods, TensorFlow library, validation methods, in particular, analysis of free databases, research results using integrated development environment Jupyter Notebook and development environment Google Colab, Python programming languages.

Results of qualification work – as a result of the conducted researches the problem of comparison and analysis of various methods of recognition of human emotions is solved. Based on the research, the accuracy of recognition in different databases was obtained. And also developed its own model

Scope – this development can be useful for the recognition of human emotions, as well as used in medicine.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

CK+ – Cohn-Kanade – розширена база даних;

GPU – Graphics Processing Unit – графічний процесор;

ILSVRC – ImageNet Large-Scale VIsual Recognition Challenge – виклик великомасштабного Visual Recognition;

ImageNet – база зображень призначених для обробки методів розпізнавання;

JAFFE – японська база даних по виразу обличчя жінок;

MMI – база даних;

ReLU – rectified linear unit – функція активації.

ВСТУП

Вираз обличчя – один з найпотужніших природних і універсальних сигналів для спілкування між людьми. Розпізнавання емоцій – це процес виявлення людських емоцій [14]. Людина сильно відрізняється по точності розпізнавання емоцій у інших людей, і для сприяння такому розпізнаванню були розроблені технології. Ми, люди, не в змозі розпізнати емоцію, а тільки емоційний вираз, яке людина проявляє в цей момент (наприклад, людина може бути сумним, але зовні демонструвати вираз радості). У кваліфікаційній роботі будемо прирівнювати два терміни (розпізнавання емоцій і розпізнавання емоційного вираження).

Розпізнавання емоцій використовується в суспільстві в декількох областях. Affectiva [15] розробила програмне забезпечення, яке відстежує виразу обличчя користувача через веб-камеру і на основі цих виразів виражає його стан. Більше від програмного забезпечення включають в себе допомогу дітям-аутистам, допомога сліпим в читанні людських виразів і полегшення взаємодії робота і людини.

Існує кілька областей, в яких розпізнаються емоції [14]: розпізнавання емоцій по тексту, мови, звуку і мальовничим матеріалами, таким як відео і зображення. В кваліфікаційній роботі ми зосередимося на розпізнаванні людських емоцій з образотворчого матеріалу. Ми будемо розрізняти сім людських емоцій. Це радість, печаль, здивування, огида, гнів, страх і нейтральний вираз, що не показує жодної з перерахованих вище емоцій.

Метою кваліфікаційної роботи є вивчення області розпізнавання людських емоцій в цифрових зображеннях людини з використанням згорткових нейронних мереж. Ми хочемо з'ясувати, які методи існують в цій галузі, в тому числі розуміння того, як вони працюють, процедури валідації, які бази даних оцінки використовуються в цій області і т.д. У рамках кваліфікаційної роботи ми протестуємо обрані сучасні нейромережеві архітектури з проблеми розпізнавання емоцій. Грунтуючись на цих

результатах, ми модернізуємо існуючу архітектуру нейромережі або створимо свою власну архітектуру, з метою зробити її продуктивність порівнянної з сучасними архітектурами. У другому розділі ми розглядаємо область сприйняття людських емоцій і даємо огляд відповідних робіт, їх підходів і отриманих результатів. У розділі 3 ми описуємо набори даних і використовувані технології, а також розглядаємо деякі сучасні архітектури згоркових нейромереж. У розділі 3 ми описуємо запропоновані методи вирішення проблеми виявлення людських емоцій. У розділі 4 ми розглядаємо результати, отримані за допомогою методів, описаних в розділі.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧ ДОСЛІДЖЕННЯ

1.1 Виявлення людських емоцій

Вже на початку 20 століття Екман і Фрізен визначили 6 основних емоцій, які відображали певні вирази подібним чином, незалежно від раси, культури людини [1]. Це були емоції гніву, відрази, радості, страху, печалі і подиву, які можна побачити на рисунку 1.1. Пізніші дослідження показують, що основні емоції не універсальні, а специфічні для різних культур.

Рисунок 1.1 – Основні емоції

Вираз обличчя – один з найпотужніших, природних і універсальних сигналів для спілкування. Було проведено багато досліджень по сприйняттю людських емоцій у зв'язку з їх практичною значущістю в комп'ютерно-людському спілкуванні.

Системи виявлення людських емоцій можна розділити на дві групи по вибору ознак: статичні системи і динамічні системи [1]. Статичні системи кодують функції, засновані на поточному зображенні, в той час як динамічні системи кодують функції, засновані на декількох послідовних зображеннях. Якщо спочатку переважна більшість баз даних створювалося в лабораторіях в контрольованих умовах, то в даний час все більшої популярності набувають бази даних, що представляють реальні сценарії і реальне середовище.

У більшості традиційних систем використовуються функції, які добуваються вручну, але великий стрибок в якості апаратного забезпечення і швидкості привів до впровадження таких підходів, як з глибоким вивченням, яке виявилося дуже точним у визначенні людських емоцій. Хоча підходи, засновані на глибокому вивченні, дають дуже хороші результати, в цьому підході є деякі недоліки. Глибокі нейронні мережі потребують дуже великій кількості даних в навчальному наборі, щоб уникнути підгонки класифікатора під тренувальні дані. Міжсуб'єктних відмінності, такі як вік, стать, культура людини і різне освітлення, повороти і пози, також викликають проблеми.

1.2 Огляд супутніх робіт

Існує великий обсяг роботи і досліджень на цю тему. Ми зосередилися на тих роботах і дослідженнях, які використовують згорткові нейронні мережі для вирішення цієї проблеми.

У першій суміжній роботі [17] автори використовують архітектуру, засновану на моделі Зачаття . Їх архітектура спочатку містила два класичних рівня згортки з двома проміжними рівнями максимальної агрегації. Далі йдуть три блоки, реалізовані за аналогією з блоками Почала, і, нарешті, два повністю з'єднаних шару для сортування. Для остаточного розрахунку результатів використовувався суб'єктно незалежний підхід. Для перевірки їх підходу було використано кілька баз даних; ми зосередилися на результатах з баз даних CK + і MMI. З впровадженої архітектурою, вони досягли 93,2% продуктивності класифікації по базі даних CK +, але по базі даних CK + вони вирізали тільки останнє зображення з відео, що показує терміни, так що у них було тільки 309 зображень.

В [18] автор використовує архітектуру AlexNet для вилучення функцій із зображень. Архітектура AlexNet була розроблена для завдання великомасштабного візуального розпізнавання (ILSVR) ImageNet в 2010

році і складається з 8 шарів [19]. На початку знаходяться п'ять згорткових шарів, за якими слідують три повністю з'єднаних шари. Таким чином, в даній роботі для вилучення особливостей з вхідних зображень використовуються згорткові шари, ці особливості в кінцевому підсумку класифікуються в особливу категорію за допомогою багатокласового класифікатора підтримують векторних машин (SVM). Згорткові шари в цій частині світу були піддані новому навчанню. В результаті роботи була отримана класифікація 94,4% по набору даних CK +, де з кожного відео в наборі даних було вилучено останнє зображення, що показує термін на піку. Додаткова попередня обробка включала перетворення зображень в чорно-біле колірне простір і витяг осіб із зображень за допомогою детектора осіб Viola-Jones.

В останній з розглянутих нами робіт [20] автори використовують підхід, при якому архітектура містить два паралельних каналу згорткових шарів, які, в кінцевому рахунку, з'єднуються між собою, утворюючи повністю з'єднані шари. Приклад такої архітектури показаний на рисунку 2.2. Першим каналом є класичний підхід з згорткові шарами і максимальними шарами агрегування. Другий канал використовується як згортковий автокодувальник, який спочатку кодує зображення (тобто риси витягуються згорткові шарами), а потім декодує зображення на основі витягнутих рисунків [21]. Для тестування цієї моделі використовувалася база даних JAFFE, в якій реалізовані два підходи. У першому підході, для кожного семестру, автори випадковим чином обирали по одному зображенню від кожного випробуваного для тестового набору. В результаті був створений навчальний комплект, що складається з 143 зображень, і тестовий комплект, що складається з 70 зображень. У другому підході використовувалася 10-кратна перехресна перевірка, при якій весь набір даних випадковим чином ділився на 10 груп. При першому підході автори досягли середнього класифікаційного успіху 95,8%. В рамках другого підходу вони домоглися середнього успіху в класифікації на 94,1%.

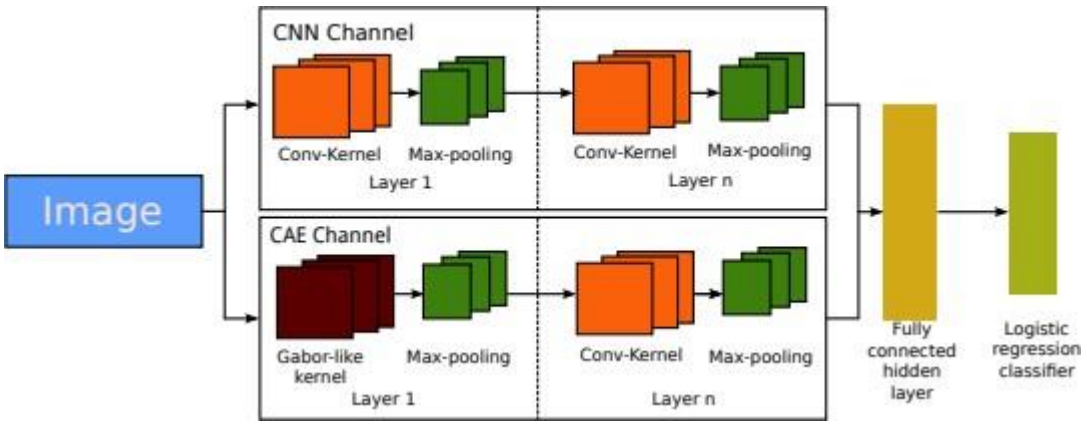

Рисунок 1.2 – Приклад архітектури з двома паралельними каналами

1.3 Основні складові нейронних мереж

Штучні нейронні мережі побудовані по принципу біологічної нейронної мережі, котрі представляють собою мережі нервових клітин, котрі виконують певні фізіологічні функції. Складовою частиною нейронної мережі як очевидно є нейрон рисунок 1.3.

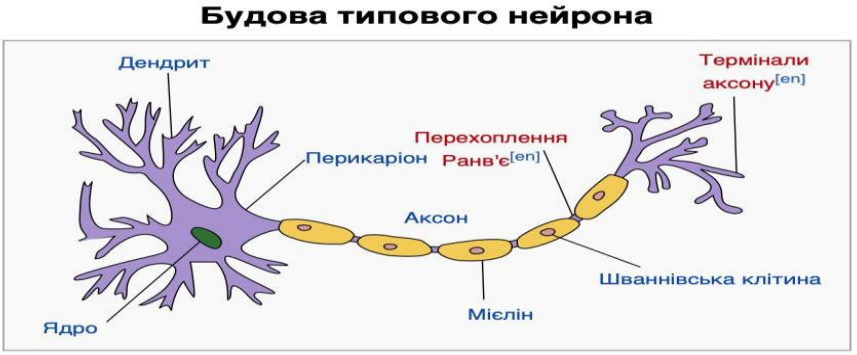

Рисунок 1.3 – Будова нейрона

Синапс – це зв’язок у мережі по котрому вихідний сигнал з одного нейрону потрапляє на вхід інших нейронів. Зв’язки що мають додаткову вагу називають збудливими зв’язками, а ті що мають негативну вагу називають гальмівними. У нейронної мережі штучний нейрон – це деяка нелінійна функція, аргументом якої є лінійна комбінація всіх вхідних

сигналів. Така функція називається активаційною. Потім результат активаційної функції відправляється на вихід нейрона. Об'єднання таких нейронів і називають штучною нейронною мережею. На відміну від справжнього нейрона, штучний складається з нелінійного перетворювача та суматора. Тож бачимо, що штучні нейрони є досить простими та однотипними елементами нейронної мережі, котрі певним чином повторюють роботу нейронів нашого мозку.

1.4 Архітектура нейронних мереж

Базовим елементом штучної нейронної мережі є штучний нейрон. Ці елементи пов’язуються зв’язками та утворюють нейронну мережу. Вона в свою чергу дає можливість досить ефективно обробляти інформацію, а також пристосовуватись до змін зовнішнього середовища. Під час роботи штучної нейронної мережі відбувається перетворення вхідного вектору значень(сигналів) на вихідний. Сам процес перетворення визначається характеристикою нейронів, а також структурою, архітектурою та методом тренування нейронної мережі.

Основними характеристиками штучної нейронної мережі є:

– парадигма нейронної мережі – це спосіб використання та навчання нейронної мережі;

– структура штучної нейромережі – це те як пов’язані нейрони між собою та взаємодіють;

– архітектура штучної нейромережі – це тип або типи нейронів у мережі та те як вони між собою взаємодіють на пов’язані. Варто зазначити що на основі однієї архітектури можуть бути реалізовані різні парадигми та навпаки. Виділяють наступні архітектурні рішення нейронних мереж;

– слабкозв’язні – нейрони пов’зані тільки з сусідніми рисунок 1.4;

– повнозв’язні – це коли кожен нейрон подає сигнал на всі інші нейрони, в тому числі собі рисунок 1.5.

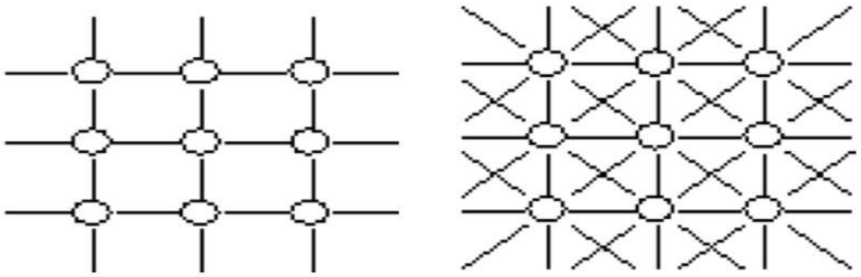

Рисунок 1.4 – Слабкозв'язні нейромережі

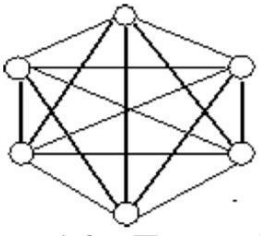

Рисунок 1.5 – Повнозв'язані нейромережі

Розглянемо задачу навчання з учителем. Дано безліч тренувальних прикладів X з мітками Y. Нейронні мережі визначають нелінійну гіпотезу hW, b (x) з параметрами W і b рисунок 1.6.

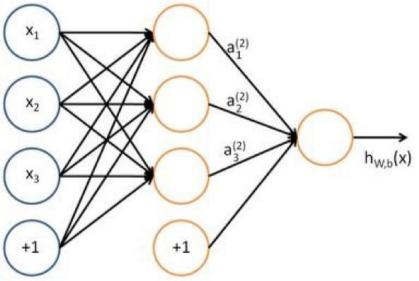

Рисунок 1.6 – Приклад нейромережі

Перший шар на зображенні шар називається вхідним, а останній правий шар – вихідним. Шар посередині є прихованим і називається так

через те, що його значення не спостерігаються в тренувальних прикладах. Таким чином в даній мережі ми спостерігаємо елементи входу, приховані елементи і один вихідний елемент. Мережею прямого поширення називаються штучні нейронні мережі, які використовують вихідні дані одного шару як данні для входу наступного шару.

Мережа прямого поширення сигналу (мережа прямої передачі) – нейронна мережа без зворотних зв'язків. У цій мережі поширення сигналу однонаправлено, тобто немає зворотного зв'язку. Від вхідного шару сигнал обробляється шар за шаром в напрямку виходу. Через відоме число кроків на вихідному шарі з'являється відповідь мережі.

Мережі прямого поширення є добре вивченими і відносно простими в реалізації. Їх недоліком є необхідність великого числа нейронів для виконання складних завдань.

Ланцюги Маркова – свого роду попередники машин Больцмана і мереж Хопфілда, про які сказано нижче. У ланцюгах Маркова ми задаємо ймовірності переходу з поточного стану в сусідні. Крім того, це ланцюга не мають пам'яті: подальший стан залежить тільки від поточного і не залежить від всіх минулих станів. Хоча ланцюг Маркова не можна назвати нейронною мережею, вона близька до них і формує теоретичну основу для машин Больцмана і мереж Хопфілда.

Мережа Хопфілда, є одним з видів мереж асоціативної пам'яті. Це одношарова нейронна мережа, в якій кожен нейрон пов'язаний з усіма іншими, має по одному входу і виходу. Жорстка функція активації генерує два значення: -1 (загальмований) і +1 (збуджений). У моделі використовується принцип зберігання інформації як динамічно стійких атракторів. Енергетична функція зменшується в процесі навчання поки не досягає локального мінімуму , в якому зберігає постійне значення .

Машини Больцмана багато в чому схожі на мережі Хопфілда, але в них деякі нейрони позначені як вхідні, а деякі залишаються прихованими. Вхідні нейрони стають вихідними, коли все нейрони в мережі оновлюють

свої статки. Спочатку вагові коефіцієнти присвоюються випадковим чином, потім відбувається навчання методом зворотного поширення, або останнім часом все частіше за допомогою алгоритму contrastive divergence (коли градієнт обчислюється за допомогою марковського ланцюга). Машини Больцмана – стохастична нейронна мережа, так як в навчанні задіяна ланцюг Маркова.

У 2015 році ResNet здійснила справжню революцію глибини нейромереж. Вона складалася з 152 шарів і знизила відсоток помилок до 3,57% в змаганні класифікації ImageNet. Це зробило її майже в два рази ефективніше GoogleNet. Що ж відбувається з нейромережею, коли ми збільшуємо число шарів? Чи можна, взявши звичайну архітектуру начебто VGG, просто складати все більше і більше шарів один на одного і досягати кращої точності?

Ні, не можна. Швидше за все, більш глибока нейросеть покаже навіть гірші результати як при навчанні, так і при тестуванні. І перенавчання тут ні до чого, оскільки тоді тренувальна помилка була б низькою.

Рекурентні мережі – це глибокі мережі, в яких присутні зворотні зв'язки. Це означає, що є присутнім хоча б один шар, сигнали з якого надходять на нього ж, або на один з попередніх шарів. Нейрони беруть участь в обробці інформації багаторазово, що дозволяє використовувати динамічні властивості мережі. Такі мережі дозволяють скоротити число нейронів. На основі рекурентних мереж розроблені різні моделі асоціативної пам'яті. Особливо ці мережі стали в нагоді в області розпізнавання мови.

Мережа Хемінга (Класифікатор по мінімуму відстані Хеммінга) – інше приклад нейронної мережі асоціативної пам'яті. Принцип роботи заснований на обчисленні відстані Хеммінга від вхідного вектора до всіх векторів-зразків, відомих мережі. При надходженні вхідного образу, мережа вибирає зразок з найменшим до нього відстанню Хеммінга і відповідний йому вихід активізується.

Глибокі мережі довіри – це мережі, що представляють собою каскад Обмежених Машин Больцману. Стандартна Машина Больцмана складається з повнозв'язних «видимих» і «прихованих» нейронів, які беруть бінарні значення, певні векторами. Обмежені машини відрізняються тим, що нейрони одного класу не пов'язані між собою. Ці мережі цікаві тим, що можуть грати роль генеруючих моделей. Іншими словами, мережа навчена розпізнавати, наприклад, рукописний текст в теорії може бути використана для генерації зображень, які виглядають як рукописний текст.

Згорткові нейронні мережі і глибокі згорткові нейронні мережі кардинально відрізняються від інших мереж. Вони використовуються в основному для обробки зображень, іноді для аудіо та інших видів вхідних даних. Типовим способом їх застосування є класифікація зображень. Такі мережі зазвичай використовують «сканер», що не обробляє всі дані за один раз.

Навчання є головною потребою штучного інтелекту (ШІ), який люди можуть забезпечити на сучасному етапі ШІ еволюція. У людському світі ця потреба вирішується освітою, яка є важливим фактором, що сприяє розвитку людини, що передбачає розширення можливостей (також творчих) та свободи. Школи та університети служать центрами для нарощування інтелектуального потенціалу, перевіреного обміну інформацією і обмін. Так само сучасна система ШІ повинна бути добре навченою[23]. Ми стверджуємо, що (глибоке) навчання для машини – це динамічний, еволюційний процес, дуже схожий на традиційну вищу освіту, однак, з новими викликами та особливостями.

1.5 Постановка задач дослідження

Метою даної кваліфікаційної роботи є дослідження та використання методів та алгоритмів розпізнавання людських емоцій.

З урахуванням сформульованої мети в даній кваліфікаційній роботі

необхідно вирішити низку таких задач:

– розглянути існуючі рішення поставленої задачі і порівняти їх переваги і недоліки;

– провести порівняльний аналіз основних алгоритмів які вирішують поставлену задачу і проаналізувати їхні переваги і недоліки, вибрати найбільш відповідний ;

– розглянути існуючі мови програмування, вибрати той, який найбільш підходить для реалізації програми;

– вибір платформи і технології створення, на основі аналізу існуючих інструментів і засобів розробки;

– розглянути основні напрямки роботи згорткових нейронних мереж;

– здійснити вибір та формування наборів даних для експериментальних досліджень;

– здійснити вибір програмних засобів для перевірки працездатності побудованого програмного додатку;

– здійснити дослідження можливостей розширення функціоналу програмного додатку.

– провести тестування та порівняльний аналіз розробленого методу розпізнавання емоцій з методами які вже існують.

Далі в подальших розділах будуть детально розглянуті бази даних з різними наборами даних, також будуть розглянуті самі популярні та актуальні методи розпізнавання людських емоцій, практичне застосування отриманих результатів досліджень щоб зрівняти ефективність методів та використання особистого методу також його зрівняння з існуючими методами .

2 ВИКОРИСТАНІ БАЗИ ДАНИХ І ТЕХНОЛОГІЇ

2.1 Бази даних

У практичній частині магістерської дисертації ми використовуємо згорткові нейронні мережі, тому нам потрібні дані для навчання, валідації та тестування. Оскільки побудова набору даних адекватного розміру і якості займає багато часу і є дорогим, у своїй роботі ми використовували вільно доступні дані з бази даних зображень емоцій людини. Ми вирішили використовувати три добре зарекомендували себе бази даних, а саме CK +, MMI і JAFFE.

База даних CK +, або розширена база даних Cohn-Kanade, була випущена в 2010 році для вивчення і автоматичного визначення людських емоцій і виразів [5]. CK + – це оновлення бази даних Кун-Канади з 2000 р База даних складається з 123 предметів з 593 відеоматеріалами про людські емоції людей віком від 18 до 50 років різного полу та походження в кожному відео показана зміна виразу обличчя від нейтрального до того яке буде відображене на малюнку чи камері, в той час як базова база даних КК складається з 97 предметів і 486 відеоматеріалів про людські емоції. Кожне відео з колекції складається з серії зображених зображень, що представляють собою перехід від нейтральної емоції на початку запису до кінцевої емоції в кінці запису. Сім останніх емоцій (за винятком основних нейтральних емоцій) в цьому збірнику: гнів, презирство, відраза, страх, радість, смуток і здивування. Приклад об'єкта CK + показаний на рисунку 2.1.

База даних CK+ широко вважається найбільш широко використовуваної доступною базою даних класифікації виразів обличчя з лабораторним контролем і в більшості методів класифікації виразів обличчя.

Рисунок 2.1– Основні емоції з бази даних CK+

База даних MMI була створена в 2002 році як ресурс створення і оцінки алгоритмів в області виявлення людських виразів [6]. Спочатку розроблений для збору великої кількості одиниць дії, пізніше він був доповнений даними, що дозволяють розрізняти шість основних емоцій. Розширені дані містили спонтанні реакції суб'єктів, які були отримані шляхом показу суб'єктам різних коротких відеокліпів, які потім привели суб'єктів до вираження певної емоції. Загальна база даних складається з більш ніж 2900 послідовностей або відеозаписів і 740 зображень 25 суб'єктів, які демонструють емоційні вирази в контрольованих умовах. На рисунку 2.2 показаний приклад суб'єкта з шістьма основними виразами (виключаючи нейтральне) з бази даних MMI.

Звернення до бази даних MMI здійснюється за допомогою web-інтерфейса, який складається з двох основних областей: області для завдання умов пошуку і вікно результатів пошуку. Також фільтр містить в собі 12 параметрів які можливо задавати при потребі.

Рисунок 2.2– Основні емоції з бази даних MMI

Колекція JAFFE (Японська жіноча експресія особи) складається з 213 зображень, знятих з використанням 10 японських жіночих моделей [7]. На зображеннях 7 емоційних виразів. Це радість, печаль, здивування, гнів, відраза, страх і нейтральне вираз. Приклад суб'єкта з цього набору даних показаний на рисунку 2.3.

Рисунок 2.3 – Основні емоції з бази даних JAFFE

2.2 Tensorflow

Tensorflow [2] був розроблений в 2015 році компанією Google як наступник існуючої системи DIstBelief. Tensorflow – це інтерфейс для впровадження і розгортання великомасштабних моделей машинного навчання. Система дозволяє виконувати обчислення на різних мобільних платформах, різних типах обладнання і на менших або великих розподілених архітектурах. Наявність тільки однієї системи, яка охоплює такий великий набір платформ, дозволяє дуже легко використовувати машинне навчання в реальному світі. Система може бути використана для реалізації великої кількості алгоритмів, в тому числі для навчання і тестування глибоких нейронних мереж. Вона вже успішно застосовується в області комп'ютерного зору, робототехніки, обробки природної мови та інших галузях інформатики. Ми використовували Tensorflow при розробці наших архітектур.

Бібліотека версія 2.1.0. Так як ця бібліотека також дозволяє завантажуватися на GPU з Nvidia, нам довелося встановити додаткові драйвери для підтримки цієї функції. Базовий драйвер для GPE був випущений у версії 446.14. Для підтримки Tensorflow були встановлені додаткові програмні пакети, CUDA Toolkit версії 11.3 і cuDNN Software Development Kit (SDK) версії 8.2

Переваги фреймворка TensorFlow:

– підтримує безліч мов програмування, що уможливлює використання даного інструментарію, навіть не маючи великого досвіду в інших середовищах програмування;

– велике кількостей посібників для вивчення та тренувальних матеріалів, TensorFlow, володіє написаними туторіали, його легко зрозуміти і використовувати на практиці;

– згорткові нейронні мережі використовуються для розпізнавання зображень, рекомендаційних систем, а також обробки природної мови

(NLP). CNN складаються з набору різних верств, що перетворюють вхідні дані в оцінки для залежної змінної з заздалегідь відомими класами. В результаті аналізу, легкість в побудові моделей використовуючи TensorFlow, тимчасова верстка CNN в Torch виділяють даний фреймворк серед інших;

– рекурентні нейронні мережі (RNN) використовуються для розпізнавання мови, прогнозування часових рядів, захоплення зображень і вирішення інших завдань, в яких потрібна обробка послідовної інформації. Tensorflow має деякий набір матеріалів по RNN, а TFLearn і Keras включають в себе велику кількість прикладів RNN, що використовують TensorFlow;

– має легкий у використанні, модульний призначений для користувача інтерфейс, створюючи інтуїтивно зрозуміле середовище для розробки;

– показує найкращий результати при тестуванні продуктивності згортальних нейронних мереж;

Використовуючи бібліотеку глибокого навчання Keras, яка:

– дозволяє легко і швидко створювати прототипи (завдяки зручності, модульності і розширюваності) ;

– підтримує як згорткові мережі, так і повторювані мережі, а також комбінації цих двох;

– легко працює на процесорі і графічний процесор.

Переваги бібліотеки Keras:

– зручність для користувача. Keras слід найкращим методам зниження когнітивної навантаження: він пропонує послідовні і прості API, він мінімізує кількість дій користувача, необхідних для випадків загального використання, і забезпечує чітку і ефективну зворотний зв'язок з помилкою користувача;

– модульність. Під модельністю розуміється послідовність або графік автономних повністю настроюються модулів, які можуть бути підключені разом з мінімальними обмеженнями. Зокрема, нейронні шари, функції

витрат, оптимізатори, схеми ініціалізації, функції активації, схеми регуляризації – це автономні модулі, які ви можете комбінувати для створення нових моделей;

– легка розтяжність, нові модулі просто додавати (як нові класи та функції), а існуючі модулі надають безліч прикладів. Щоб мати можливість легко створювати нові модулі, ви можете повністю виразити свою виразність, що робить Keras відповідним для передових досліджень;

– робота з Python. Немає окремих файлів конфігурації моделей в декларативному форматі. Моделі описані в коді Python, який компактний, легше налагоджується і забезпечує простоту розширюваності.

2.3 Сучасні згорткові нейронні мережі

Згорткові нейронні мережі – це тип глибоких нейронних мереж, які найчастіше використовуються для класифікації зображень, аналізу медичних зображень, розпізнавання об'єктів в зображеннях і / або відео і т.д [1]. Згорткові нейронні мережі були натхненні біологічними процесами, що відбуваються в нейронах всередині мозку, де певні нейрони стимулюються над обмеженими областями поля зору. Ці окремі нейрони перекривають один одного, охоплюючи все поле зору. Сучасні згорткові нейронні мережі, як правило, складаються з трьох шарів [1]. Це згортковий шар, пул і повністю з'єднаний шар.

Згорткові нейронні мережі мають перевагу використання вхідних даних без великої попередньої обробки в порівнянні з іншими алгоритмами класифікації зображень. Мережа будує свої власні фільтри, а не ми вручну визначаємо їх на підставі попередніх знань. Це безперечно є великою перевагою при розробці рішення [3].

Назва архітектура мережі отримала через наявність операції згортки, суть якої в тому, що кожен фрагмент зображення множиться на матрицю згортки поелементно, а результат підсумовується і записується в аналогічну

позицію вихідного зображення.

Згорткові нейронні мережі. Звучить як дивне поєднання біології та математики з домішкою інформатики, але як би воно не звучало, ці мережі – одні з найвпливовіших інновацій в області комп'ютерного зору. Вперше нейронні мережі привернули загальну увагу в 2012 році, коли Алекс Крижевський завдяки їм виграв конкурс ImageNet, знизивши рекорд помилок класифікації з 26% до 15%, що тоді стало проривом. Сьогодні глибинне навчання лежить в основі послуг багатьох компаній: Facebook використовує нейронні мережі для алгоритмів автоматичного створення тегів, Google – для пошуку серед фотографій користувача, Amazon – для генерації рекомендацій товарів, Pinterest – для персоналізації домашньої сторінки користувача, а Instagram – для пошукової інфраструктури.

Визначення топології мережі орієнтується на вирішувану задачу, дані з наукових статей і власний експериментальний досвід. Можна виділити наступні етапи що впливають на вибір топології :

– визначити вирішувану задачу нейромережею(класифікація, прогнозування, модифікація);

– визначити обмеження у вирішуваній задачі(швидкість, точність відповіді);

– визначити вхідні(тип: зображення, звук, розмір ,формат і вихідні дані(кількість класів).

2.3.1 Архітектура VGG19 і VGG16

Однією з найбільш сучасних згорткових нейронних мереж є мережа VGG19 [4], яка виявилася успішною для класифікації об'єктів в ImageNet Large-Scale VIsual Recognition Challenge (ILSVRC) [8].

Мережа VGG19 є глибокою нейронну мережу з 19 шарами і є хорошим прикладом архітектури, що ілюструє нашарування. Входом в нейронах зображення є кольорове зображення розміром 224 x 224 пікселів.

Нейронна мережа складається з 5 зготкових блоків з проміжним злиттям і, нарешті, три повністю з'єднаних шари, які використовуються для класифікації.

Архітектура VGG19 виглядає наступним чином:

– перший блок згортки складається з двох шарів згортки, що містять 64 фільтра;

– другий блок згорток містить два шари згорток зі 128 фільтрами;

– третій блок згортки має чотири шари згортки з 256 фільтрами;

– Четвертий і п'ятий шари згортки мають по чотири шари згортки з 512 фільтрами;

– всі шари мають розмір фільтра 3 x 3 і включають функцію активації ReLU;

– між блоками згортки, шари агрегування засновані на максимальному значенні для зменшуючи розмір даних;

– далі йдуть два повністю пов'язаних блоку 4096 нейронів з функцією активації ReLU і кінцевий повністю пов'язаний шар тисячі нейронів з функцією активації Softmax, яка призначена для класифікації даних, отриманих з вхідного зображення. Графічне представлення архітектури VGG19 представлено на рисунку 2.4.

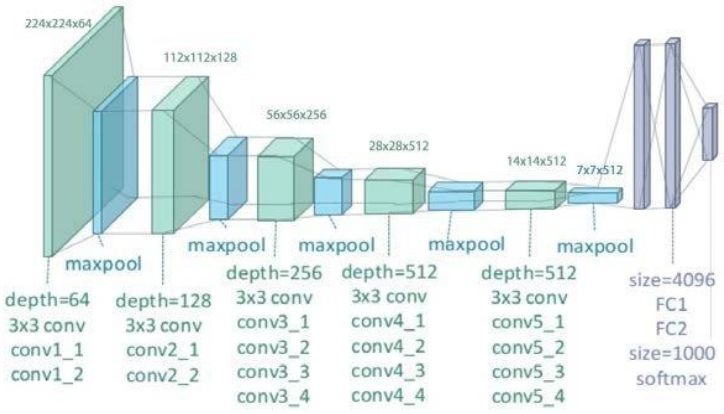

Рисунок 2.4 -Архітектура VGG19

Автори [4], які реалізували VGG19, також створили більш дрібну версію цієї архітектури, яку вони назвали VGG16. Її мета була та ж, що і у VGG19, вона використовувалася для класифікації об'єктів в конкурсі ILSVRC. Менша версія була створена для того, щоб вони могли порівняти, як глибина коливань згорткової нейронної мережі впливає на її класифікаційні характеристики, або знайти об'єкти на зображенні. Грунтуючись на результатах, вони виявили, що чим глибше мережі дають найкращі результати.

Архітектура VGG16 складається з наступних будівельних блоків:

– перший блок згортки складається з 2 шарів згортки з 64 фільтрами, розміром 3 x 3;

– другий блок згортки складається з 2 шарів згортки з 128 фільтрами, розміром 3 x 3;

– третій блок згортки складається з 3 шарів згортки з 256 фільтрами розміром 3 x 3;

– четвертий і п'ятий блоки містять по 3 шари згортки з 512 фільтрами розміром 3 x 3;

– між блоками згортки існують шари агрегування, засновані на максимальному значенні для зменшення розмірності даних. Функція активації ReLU використовується на всіх згорткових шарах;

– за згорткові шарами слідують два повністю пов'язаних шару з 4096 нейронами на шар. Нарешті, є додатковий повністю підключений шар з функцією активації Softmax для цілей класифікації.

Всі приховані шари забезпечені ReLU. Відзначимо також, що мережі (за винятком однієї) не містять шару нормалізації (Local Response Normalisation), так як нормалізація не покращує результату на датасет ILSVRC, а веде до збільшення споживання пам'яті та часу виконання коду.

Завдяки цьому дана архітектура незважаючи на деякі недоліки є гарним будівельним блоком для навчання , та її легко реалізувати за допомогою технології TensorFlow.

Графічне представлення архітектури VGG16 можна побачити на рисунку 2.5.

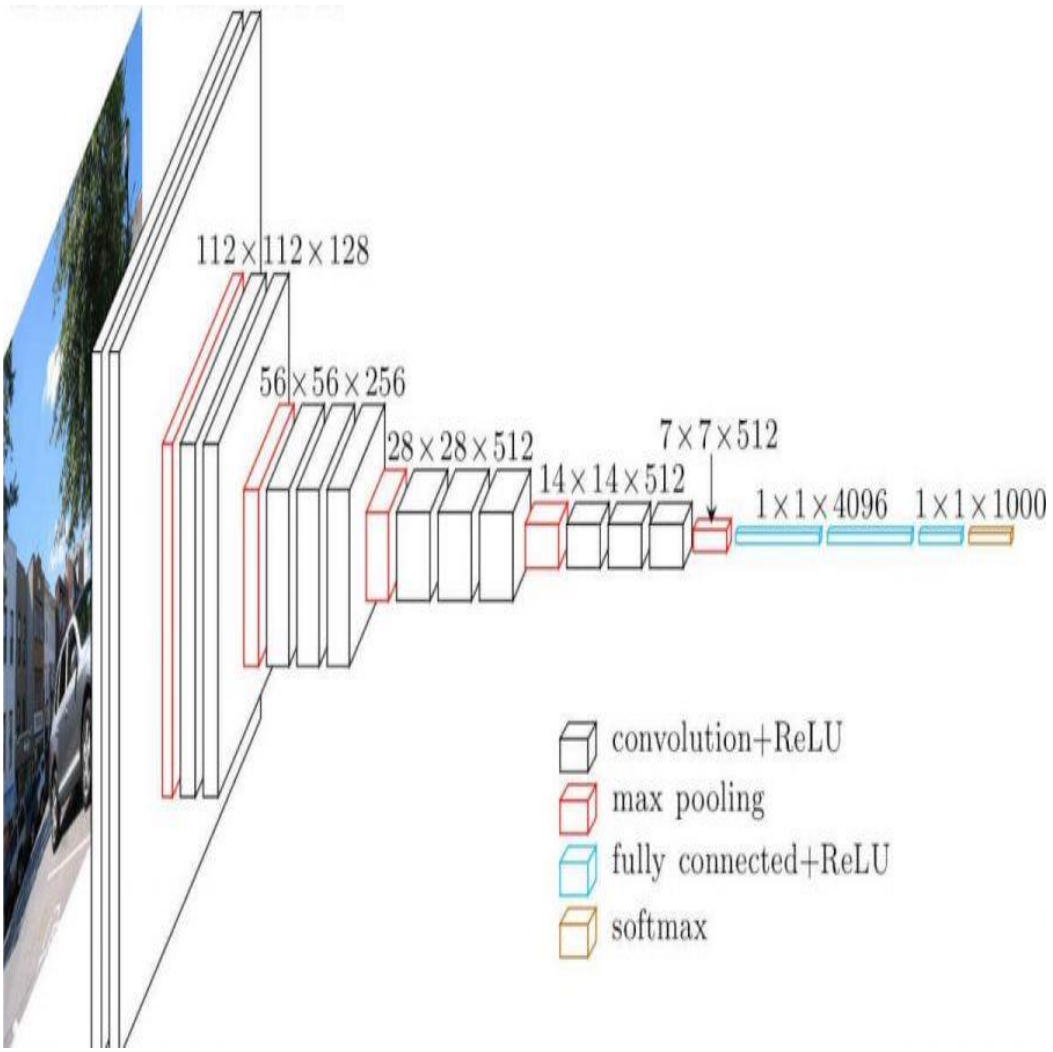

Рисунок 2.5 – Архітектура VGG16

Архітектури VGG16 і VGG19 виявили, що глибина сітки впливає на продуктивність моделі. Однак у міру поглиблення нейронних мереж може виникнути так званий феномен градієнтної деградації [11]. Деградація не є феноменом переналаштування моделі, але вона призводить до більшої помилки в продуктивності моделі. Тому автори ввели в архітектуру залишковий блок [11], який виглядає наступним чином: вхід в блок додається до виходу блоку, тобто вихід блоку являє собою суму проміжних

шарів в блоці і входу. Вони називають цей стрибок проміжних шарів ідентифікаційною картою. Графічне представлення такого блоку можна побачити на рисунку 2.6.

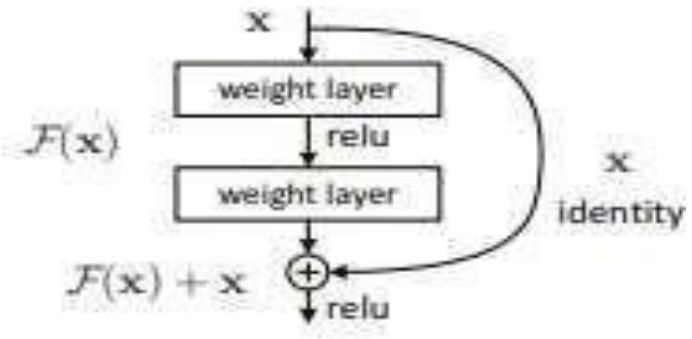

Рисунок 2.6 – Зв’язок швидкого доступу

Одним з етапів розробки нейронної мережі є вибір функції активації нейронів. Вид функції активації багато в чому визначає функціональні можливості нейронної мережі і метод навчання цієї мережі. Класичний алгоритм зворотного поширення помилки добре працює на двошарових і тришарових нейронних мережах, але при подальшому збільшенні глибини починає зазнавати проблеми. Одна з причин – так зване загасання градієнтів. У міру поширення помилки від вихідного шару до вхідного на кожному шарі відбувається домножение поточного результату на похідну функції активації. Похідна у традиційної сигмоидной функції активації менше одиниці на усій області визначення, тому після декількох шарів помилка стане близькою до нуля. Якщо ж, навпаки, функція активації має необмежену похідну(як, наприклад, гіперболічний тангенс), то може статися вибухове збільшення помилки у міру поширення, що приведе до нестійкості процедури навчання.

Відомо, що нейронні мережі здатні наблизити скільки завгодно складну функцію, якщо в них досить шарів і функція активації є нелінійними. Функції активації ніби сигмоидной або тангенціальними є

нелінійними, але призводять до проблем із загасанням або збільшенням градієнтів. Проте можна використати і набагато простіший варіант – випрямлену лінійну функцію активації(rectified linear unit, ReLU), виражається формулою зображеною на рисунку 2.7.

$$f\ (s) = \max\ (0, s)$$

Рисунок 2.7 – Функція активації ReLu

Переваги використання ReLu її похідна дорівнює або одиниці, або нулю, і тому не може статися розростання або загасання градієнтів, оскільки помноживши одиницю на дельту помилки ми отримаємо дельту помилки, якщо ж ми б використали іншу функцію, наприклад, гіперболічний тангенс, то дельта помилки могла, або зменшитися, або зрости, або залишитися такою ж, тобто, похідна гіперболічного тангенса повертає число з різним знаком і величиною, що можна сильно вплинути на загасання або розростання градієнта:

– більше того, використання цієї функції призводить до проріджування вагів;

– обчислення сигмоиды і гіперболічного тангенса вимагає виконання ресурсоємних операцій, таких як піднесення до степеня, тоді як ReLU може бути реалізований за допомогою простого порогового перетворення матриці активацій в нулі;

– відсікає непотрібні деталі в каналі при негативному виході.

З недоліків можна відмітити, що ReLU не завжди досить надійна і в процесі навчання може виходити з ладу. Наприклад, великий градієнт, що проходить через ReLU, може привести до такого оновлення вагів, що цей нейрон ніколи більше не активується. Якщо це станеться, то, починаючи з цього моменту, градієнт, що проходить через цей нейрон, завжди дорівнюватиме нулю. Відповідно, цей нейрон буде безповоротно виведений

з ладу. Наприклад, при занадто великій швидкості навчання(learning rate), може виявитися, що до 40% ReLU «мертві»(тобто, ніколи не активуються). Ця проблема вирішується за допомогою вибору належної швидкості навчання.

На основі цього блоку були розроблені архітектури ResNet. Існує кілька архітектур ResNet, в залежності від їх глибини. У нас є ResNet з глибиною 18 рівнів, глибинами 34, 44, 56, 101 і 152 рівня. Архітектури були протестовані на базі даних ImageNet з змагання ILSVRC. Архітектури також були протестовані на проблемі CIFAR-10 [13], де були проведені додаткові тести, щоб побачити, як глибина нейронної мережі впливає на продуктивність класифікації.

Архітектура ResNet глибиною 34 має 0.46 M параметрів і складається з обох (графічне уявлення архітектури з такою глибиною показано на рисунку 2.8):

– по-перше, згортковий шар з 64 фільтрами 7 x 7;

– далі йдуть перші три залишкових блоки з згорткові шарами, які мають 64 фільтра розміром 3 x 3;

– далі йдуть чотири залишкових блоки, які збільшують кількість фільтрів на згорткових шарах з 64 до 128. Розміри 3 x 3;

– за яким слідують шість залишкових блоків, в яких кількість фільтрів знову збільшується тільки до 256;

– останній набір залишкових блоків складається з 3 блоків, знову ж просто збільшивши кількість фільтрів до 512;

– ми закінчуємо залишкові блоки з середнім пулів шаром і повністю пов'язаним рівнем від 1000 нейронами для класифікації (в базі даних ImageNet є 1000 класів).

ResNet дозволяє відносно легко оптимізувати: «прості» мережі показують велику помилку навчання, коли глибина збільшується.

ResNet дозволяє відносно легко збільшити точність завдяки збільшенню глибини, чого з іншими мережами домогтися складніше.

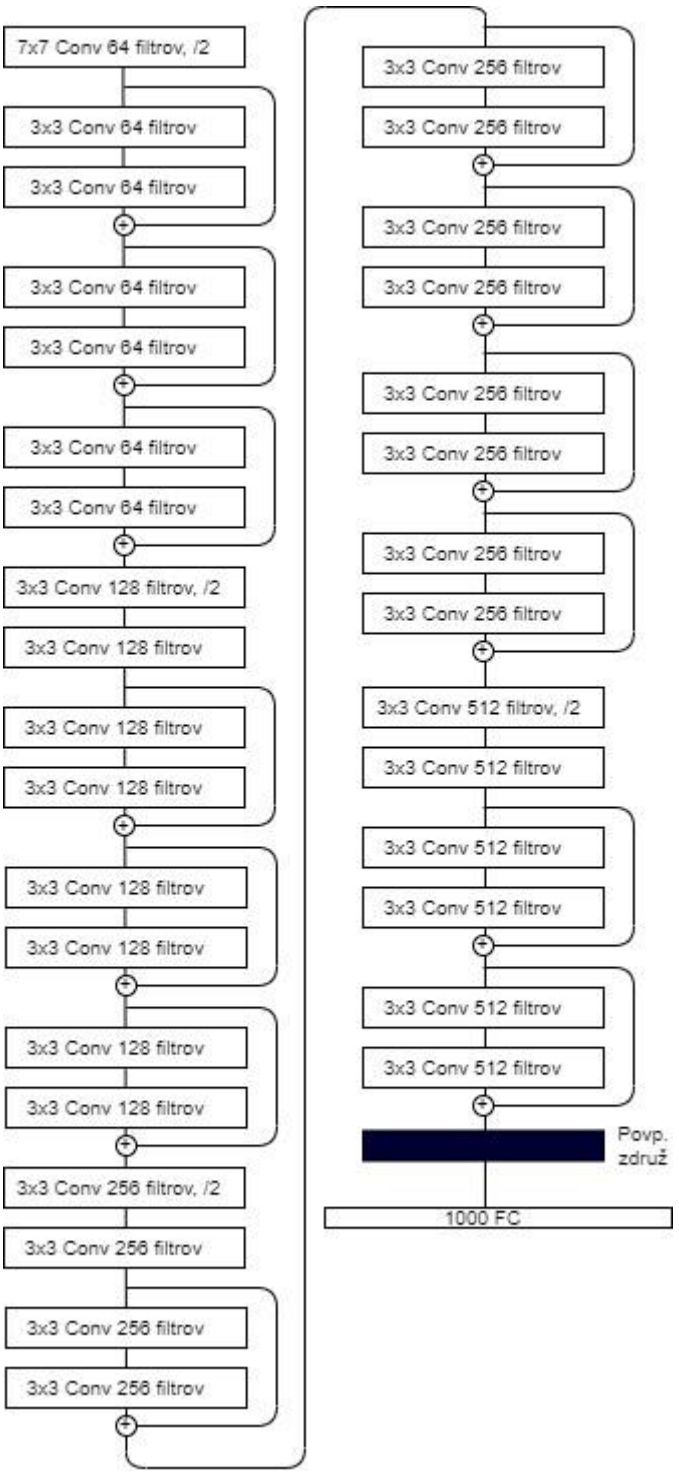

Рисунок 2.8 – Архітектура ResNet34

Ідея першої версії моделі Inception виникла тоді, коли автори [10] досліджували способи підвищення ефективності об'єктної класифікації за допомогою згорткових нейронних мереж. Найпростіший спосіб поліпшення просто «збільшити» ємність нейромережі тобто зробити її глибше (додати

більше рівнів), збільшити кількість параметрів на шар (наприклад, збільшити кількість фільтрів в згорткових шарах, збільшити кількість нейронів в повністю з'єднаних шарах і т.д.). Однак таке поліпшення може привести до переналаштуванню моделі через обмежений розміру навчального набору (тобто обмеженої кількості мічених даних).

Іншим недоліком є підвищена обчислювальна складність і, як наслідок, збільшення споживаної обчислювальної потужності (більш тривалий час навчання, більш висока споживана потужність і т.д.). Тому автори [10] впровадили так звані «вхідні блоки». За допомогою цих блоків Зачаття було вирішено кілька завдань. Одне із завдань полягала в тому, що шари, що містяться в цьому блоці, можна було використовувати для пошуку на зображенні ознак з кількох вимірах.

У простій реалізації цього блоку, блок складається з чотирьох згорткових шарів, що йдуть паралельно. Один шар містить фільтри 1 x 1, другий шар містить фільтри 3 x 3, третій шар містить фільтри 5 x 5 і останній шар містить фільтри 3 x 3 зі зменшенням розмірів за рахунок максимального об'єднання в пули. Графічне представлення наївною версії блоку «Початок» можна побачити на Рисунку 2.9

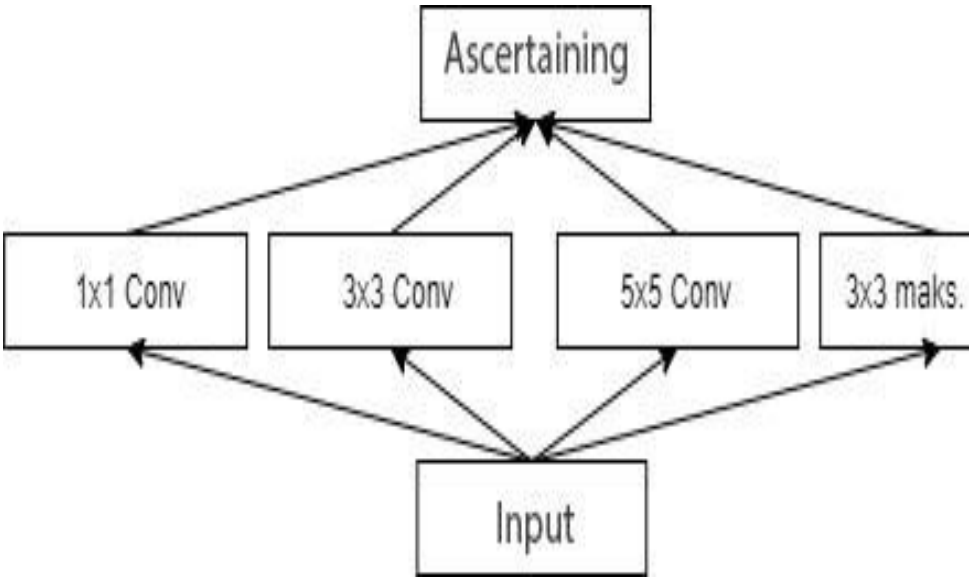

Рисунок 2.9 – Блок Inception

Це принесло додатковий вимір в нейронні мережі, а саме ширину. Одним з недоліків цього наївного блоку є те, що при розмірі 5 х 5 складно обчислити згорткові операції. Тому в наївною версії введено розмірне зменшення вхідних даних за допомогою конверсійних фільтрів розміром 1 х 1. Графічне представлення блоку «Початок» з доданим розмірним зменшенням можна побачити на рисунку 2.10.

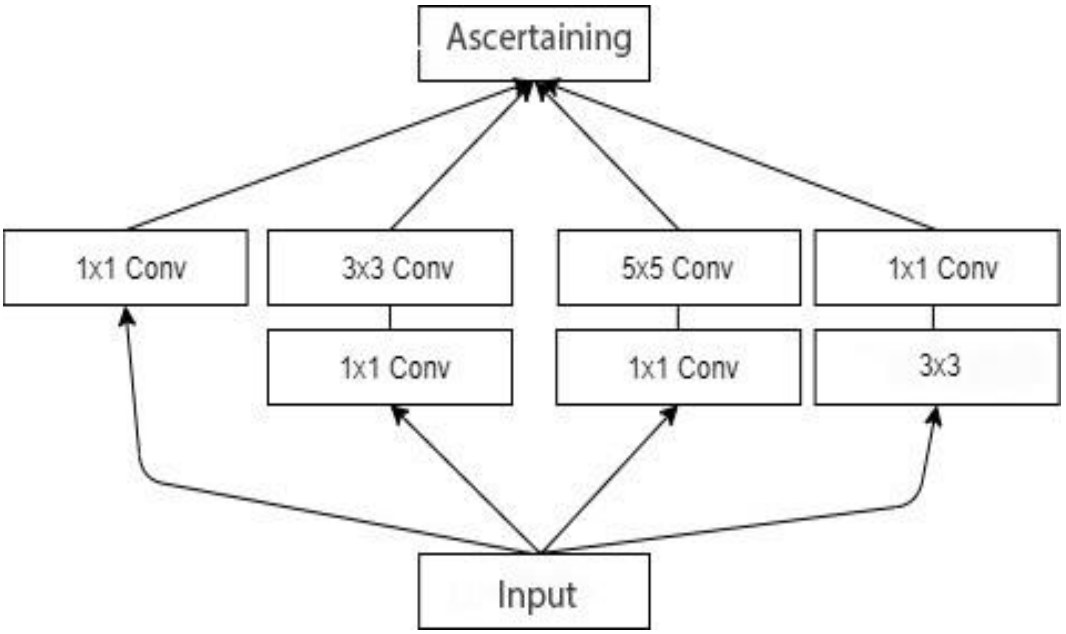

Рисунок 2.10 – Початковий блок зі зменшеними розмірами

Остаточна архітектура Inception має глибину 27 шарів. У лівій колонці показана початкова частина архітектури, а в правій колонці – фінальна частина.

Удосконалення вже були реалізовані для оригінальної версії Inception, а саме в [16] автори протестували методи зменшення кількості параметрів в загальній архітектурі. Придумане ними рішення полягало в заміні конвекції, що мають розмір фільтра 5 х 5, на дві конвекції, мають розмір фільтра 3 х 3.

Графічне представлення модифікованого блоку представлено на рису нку 2.11. Первісна реалізація архітектури Inception має 5 М параметрів, в той час як удосконалення оригінальної версії Inception має 23 М параметра.

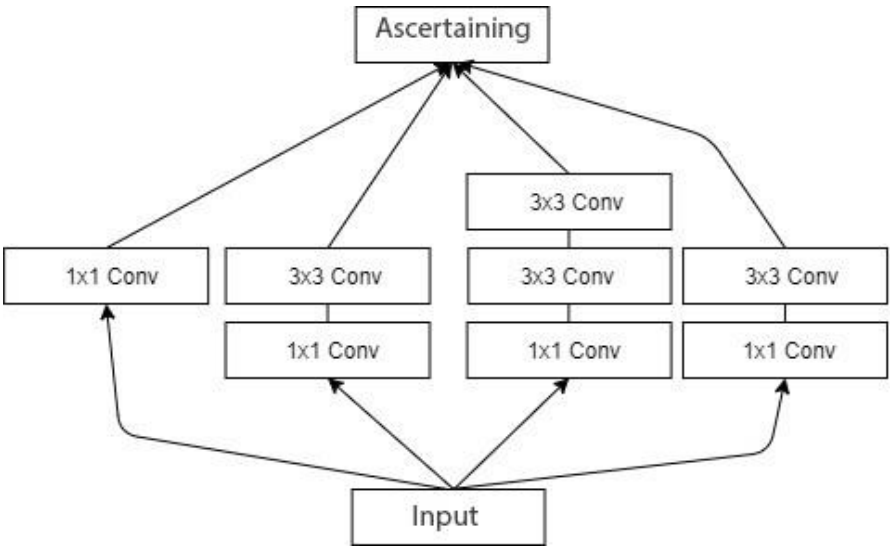

Рисунок 2.11–Покращений початковий блок

В цьому розділі були розглянуті дві архітектури VGG16 та VGG19, в підсумок можна сказати що архітектури мають недоліки серед них повільна швидкість навчання , сама архітектура займає багато місця як для такої архітектури , але вона дозволяє виконувати всі забаганки і досить простим способом за допомогою інших технологій.

3 ПРОПОНОВАНІ МЕТОДИ РОЗПІЗНАВАННЯ ЛЮДСЬКИХ ЕМОЦІЙ

3.1 VGG19 з переносним навчанням

Перший запропонований метод, VGG19 transfer, для розпізнавання людських емоцій – використання архітектури VGG19, описаної в підрозділі.

ImageNet– це база даних, яка містить понад 14 мільйонів зображень, які були помічені вручну [8]. Зображення містять об'єкти різних категорій. Колекція нараховує понад 20 000 категорій об'єктів. Архітектура була попередньо вивчена на усіченої версії цієї колекції, яка включає тільки 1000 категорій.

Для вирішення нашої проблеми ми впровадили метод трансфертного навчання [9]. Трансферне навчання – метод, широко використовуваний в машинному навчанні. Перевага такого підходу полягає в тому, що знання, які ми отримали по одній проблемі, можуть бути застосовані по відношенню до іншої проблеми. У нашому випадку це використання архітектури, яка вже була вивчена по одній проблемі і адаптована до нашої проблеми за допомогою ітеративного процесу навчання. Застосовуючи VGG19 з трансферним навчанням до нашої проблеми, ми вивчаємо тільки повністю зв'язані шари архітектури в процесі навчання, де на останньому рівні ми змінюємо кількість нейронів з 1000 до 7, як ми сприймаємо 7 людських емоцій. Ми не пристосовували згорткові шари за допомогою навчання, так як вони використовувалися тільки для вилучення функцій з вхідних зображень. Вхідні зображення архітектури мали розмір 224 x 224 пікселів.

3.2 VGG19 з модифікацією ResNet

Запропоноване покращення засноване на ідеї архітектури ResNet і залишкового блоку, де вхідні дані «пропускають» певні рівні, а потім

додаються до вихідних даних з пропущених рівнів. Це призводить до зменшення деградації, яка відбувається під час сходження ваг в глибоких мережах. Додавання цих значень додає нових значень і не збільшує обчислювальну складність архітектури. Крім того, кількість нейронів в повністю пов'язаних рівнях було скорочено з 4096 до 2048 нейронів на рівень, що призвело до значно меншої кількості параметрів для вивчення. Остаточна модель показана на рисунку 3.1.

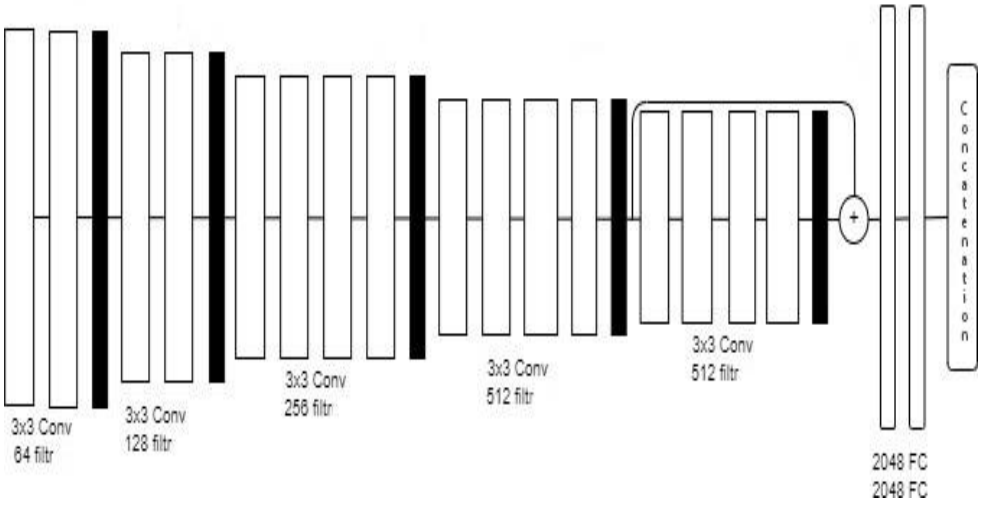

Рисунок 3.1 – Модифікована архітектура VGG19 на основі архітектури ResNet

3.3 VGGFace з трансферним навчанням

Наступним пропонованим методом, VGGFace_transfer, є використання моделі VGGFace [12]. Модель VGGFace заснована на архітектурі VGG16. Вона менше, ніж VGG19, так як має всього 16 рівнів. Розмір вхідного зображення 224 x 224 пікселів, як в архітектурі VGG19. Архітектура VGGFace вирішує проблему розпізнавання людей в зображеннях або відео. Модель була отримана по базі даних 2622 відомих людей з близько 1000 зображеннями на людину. Знову ж таки, для вирішення нашої проблеми був використаний метод трансфертного

навчання. Знову ж, тренувалися тільки повністю з’єднані шари, де ми коректували кількість нейронів в останньому повністю з’єднаному шарі. В останньому повністю пов’язаному шарі кількість нейронів було дорівнює кількості категорій емоцій, тобто 7. Знову ж, ми використовували згорткові шари тільки для вилучення функцій з вхідних зображень, і ми не змінювали їх в процесі навчання, як і для методу VGG19 transfer , що позитивно відобразилося на результаті.

3.4 Власна архітектура на основі концепції

Четвертий запропонований метод, заснований на особистому Inception, передбачає створення власної архітектури, натхненної моделлю Inception. Особливістю моделі Inception, яку ми використовували при створенні власної архітектури, є пошук особливостей різного розміру в рівнях згортки [10]. Наївна версія Блоку Зачаття містить згорткові шари розміром 1 x 1, 3 x 3, 5 x 5 і максимальний шар розміром 3 x 3. Вхідні дані проходять паралельно через ці шари і остаточно об’єднуються в єдину структуру.

Сама архітектура була структурована таким чином. Входом в сітку було зображення 64 x 64. За вхідним рівнем слід блок згортки з двома шарами 64 фільтрів кожен розміром 5 x 5. Наступний блок – блок, натхненний моделлю Зачаття. Блок складається з трьох пар шарів згортки, за якими слідують шар агрегування максимального значення і шар кластеризації або конкатенації. Після цього відбувається нормалізація всіх значень за допомогою процесу пакетної нормалізації.

Шари згортки мають розміри фільтрів 3 x 3, 2 x 2 і 1 x 1 з 64 фільтрами на шар. Блок, побудований з наявного блоку «Початок» і використаний у пропонованій нами внутрішній архітектурі, показаний на рисунку 3.2.

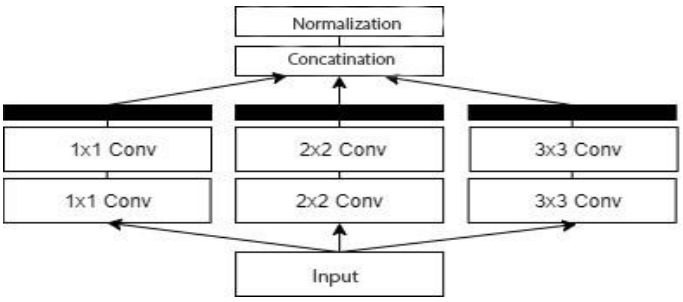

Рисунок 3.2 – Блок у пропонованій внутрішній архітектурі

За першим блоком ідуть ще два ідентичних блоку, але зі збільшенням кількості фільтрів до 128 Функція активації ReLU використовується для всіх згорткових шарів. Три повністю з'єднаних шару йдуть в кінці пропонованої архітектури. Перші два шари містять 4096 нейронів, а останній шар містить кількість нейронів, що дорівнює нашого завдання (тобто 7). Загальна архітектура показана на рисунку 3.3.

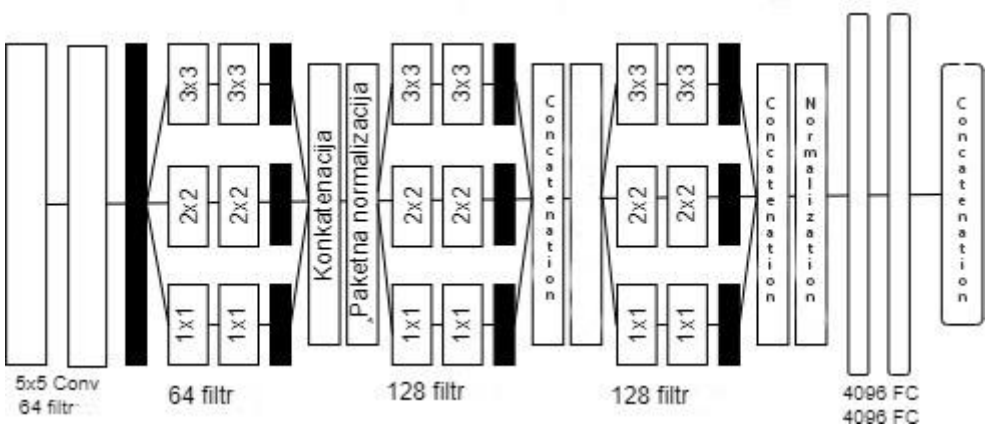

Рисунок 3.3 – Архітектура заснована на власному сприйнятті

Для пропонованої нами архітектури ми протестували два підходи до навчання. Перший підхід полягав у попередньому вивченні бази даних СІФАР-100 і використанні трансфертного навчання. База даних CIFAR-100 містить 100 категорій і 600 зображень розміром 32 x 32 пікселя для кожної категорії. Другий підхід до навчання полягав у тому, щоб вчитися з нуля,

використовуючи дані виключно з наших баз даних CK +, MMI і JAFFE.

3.5 VGGFace модифікований блоком з власної архітектури

У п'ятому запропонованому методі, VGGFace_modified, ми використовували попередньо вивчену модель VGGFace (описану в підрозділі 4.3) в поєднанні з нашої власної архітектурою. Остаточна архітектура складалася з визначених рівнів, за якими послідував блок власної архітектури, що містить 128 фільтрів. Вхідні функції, витягнуті з моделі VGGFace на рівень нашого блоку, мають розмір 3 x 3, а додавши наш власний блок, ми додали додаткові фільтри розміром 3 x 3, 2 x 2 і 1 x 1 з 128 фільтрами на шар. Остаточна кількість додаткових конвекційних фільтрів збільшилася на 768, при цьому не тільки зосереджуючись на великих областях, а й з огляду на особливості менших областей 1 x 1 і 2 x 2. Повна архітектура показана на рисунку 3.4.

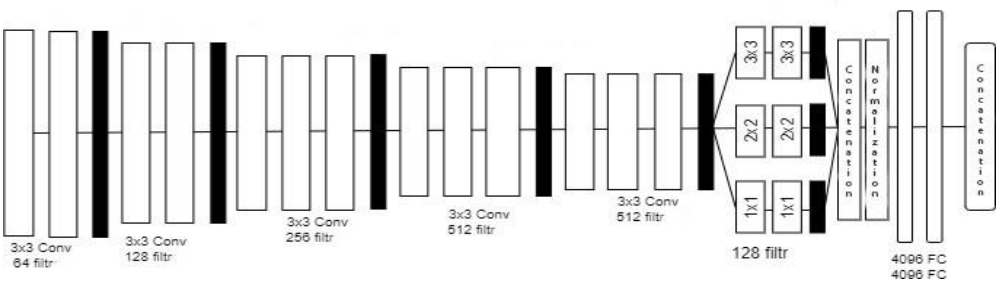

Рисунок 3.4– Модифікована архітектура VGGFace

Нарешті, після додавання власного блоку, є два більш повнозв’язних шари з 4096 нейронами і остаточний класифікаційний шар.

4 ПРАКТИЧНЕ ЗАСТОСУВАННЯ ОТРИМАНИХ РЕЗУЛЬТАТІВ ДОСЛІДЖЕНЬ

4.1 Обґрунтування вибору програмного середовища та набору даних

У цьому розділі описано деталі реалізації і розглянуто результати, отримані за допомогою методів розпізнавання людських емоцій, запропонованих в розділі 3.

Всі архітектури були реалізовані з використанням модуля Keras з бібліотеки Tensorflow. Навчання і тестування проводилися на комп'ютерній системі з апаратним і програмним забезпеченням:

– Intel Core i7 4x ядерний процесор;

– NVIDIA gtx 1660 Super з 4 ГБ пам'яті;

– 8 ГБ оперативної пам'яті;

– Операційна система Windows 10;

– Python, версія 3.9.4

– JupyterNotebook

Навчання та валідація засновані на суб'єктно-незалежному валідаційному підході. Це означає, що використовувались образи предметів, які не були використані в процесі навчання, для перевірки або тестування архітектури. Іншими словами, ми не використали жодного зображення людини під час тестування, якщо на етапі навчання використовувався хоча б один образ цієї людини. Ми використовували 5-кратну перехресну перевірку. Було розділено всі випробувані групи на 5 груп, 4 з яких були використані для навчання, а одна група – для тестування вивченої моделі. Кількість суб'єктів і кількість зображень на емоцію на групу для бази даних CK + приведено в таблиці 4.1. Кількість суб'єктів і кількість зображень на емоцію на групу для бази даних MMI приведено в таблиці 4.2.

Таблиця 4.1– Кількість суб'єктів і зображень в розрахунку на емоції по папці для бази даних CK +

Група	Суб’єкти	Нейтральний	Гнів	Відраза	Страх	Радість	Сум	Сюрприз
1	21	110	7	0	5	0	5	8
2	21	140	2	7	2	1	7	1
3	21	118	0	3	4	0	1	9
4	21	134	18	8	5	1	9	0
5	22	116	18	9	9	5	2	1

Таблиця 4.2 – Кількість суб'єктів і зображень в розрахунку на одну емоцію для бази даних MMI

Група	Суб’єкти	Нейтральний	Гнів	Відраза	Страх	Радість	Сум	Сюрприз
1	6	64	12	18	18	15	12	21
2	6	78	18	21	15	21	18	24
3	6	84	30	15	15	18	21	27
4	6	106	27	21	21	30	27	33
5	5	82	9	21	15	42	18	18

Так як бази даних CK + і MMI складаються з відео, ми повинні були

вибрати правильні зображення з відео, щоб отримати емоції на піку або якомога ближче до нього. У колекції CK + ми вирішили вибрати перші два зображення і останні три зображення у відео. Перші два зображення представляють собою нейтральну емоцію, а останні три цільову емоцію суб'єкта. У колекції MMI ми вибрали зображення аналогічним чином.

Таблиця 4.3– Кількість суб'єктів і зображень на емоції в розбивці по папках для бази даних JAFFE

Група	**Суб' єкти**	**Нейтральний**	**Гнів**	**Відраза**	**Страх**	**Радість**	**Сум**	**Сюрприз**
1	2	6	6	7	7	7	6	6
2	2	6	6	5	6	6	7	5
3	2	6	6	6	6	6	6	6
4	2	6	6	5	6	6	6	6
5	2	6	6	6	7	6	6	6

Було використано перші два зображення на відео в якості нейтрального вираження, а середні три зображення – як зображень, що представляють пікові емоції. Колекція JAFFE вже представлена зображеннями в базовій версії, і в цьому додатковому кроці немає необхідності.

Також вирізали особи з відеозображень так, щоб на кінцевому зображенні було видно тільки обличчя, без фону. Для обрізки особи ми використовували машинну навчальну модель, яка використовує гістограму

градієнтів і лінійний класифікатор [22]. Всі зображення були додатково перетворені з колірного простору RGB в чорно-біле колірне простір.

Оскільки нейронні мережі очікують великих обсягів даних, ми ще більше розширили тестові дані. Для цього необхідно дзеркально відобразити кожне зображення і обертати його з кроком в 1 максимум до 20 за і проти годинникової стрілки. Початкове і кінцеве кількість зображень після збагачення даних по кожному емоції показано в таблиці 4.4 для колекції CK +, в таблиці 4.5 для колекції MMI і в таблиці 4.6 для колекції JAFFE.

Таблиця 4.4 – Кількість знімків до і після збагачення даних для бази CK+

Емоції	**Кількість зображень**	**Кількість зображень після збагачення**
Нейтральний	618	50676
Гнів	135	11070
Відраза	177	14514
Страх	75	6150
Радість	207	16974
Смуток	84	6888

Таблиця 4.5 – Кількість знімків до і після збагачення даних для бази MMI

Емоції	**Кількість зображень**	**Кількість зображень після збагачення**
Нейтральний	414	33948
Гнів	96	7872
Відраза	96	7872
Страх	84	6888
Радість	126	10332
Смуток	96	7872
Сюрприз	123	10086
Усього	1035	84870

Таблиця 4.6 – Кількість знімків до і після збагачення даних для бази JAFFE

Емоції	**Кількість зображень**	**Кількість зображень після збагачення**
Нейтральний	30	2460
Гнів	30	2460
Відраза	29	2378
Страх	32	2624
Радість	31	2542
Смуток	31	2542
Сюрприз	30	2460
Усього	213	17467

Кожен із запропонованих методів був протестований на оригінальних зображеннях і на збагачених даних (зображеннях). Ми проводили тести окремо для кожної колекції, а також перевіряли продуктивність методу при об'єднанні даних з усіх трьох колекцій. У підході до тестування, де ми об'єднали дані з усіх колекцій, ми також використовували 5-кратну перехресну перевірку. Суб'єкти в цьому підході були згруповані по папках так само, як і при роздільному тестуванні кожної колекції. Для оцінки продуктивності моделей ми провели 2-3 прогону нашого класифікаційного підходу в кожній папці для кожної колекції. Для обчислення продуктивності моделі ми використовували вбудовану в бібліотеку Tensorflow функцію «evaluate», яка обчислює відсоткове співвідношення продуктивності моделі для тестових даних. По-перше, ми усереднили класифікаційні показники всіх 7 емоцій по всьому прогонів (часткові середні показники) для кожної групи. Остаточне середнє значення було розраховане як середнє з раніше розрахованих часткових середніх показників.

Для більш достовірного порівняння результатів все тренінги проводилися однаково або використовувалися одні і ті ж параметри. Максимальна кількість епох було встановлено на 100 з розміром пакета 32. Для адаптивного навчання ми використовували алгоритм оптимізації Адама і категоричну функцію вартості ентропії. Алгоритм Адама являється популярним алгоритмом в області глубокого навчання, тому що він швидко досягає гарних результатів. Швидкість навчання підтримується для кожної ваги параметра та окремо адаптується по мірі розвитку навчання. Метод вираховує індивідуальні адаптивні швидкості навчання для різних параметрів з оцінок першого та другого моментів градієнту.

Тренувальний комплект був розділений в процесі навчання наступним чином: 80% від загального комплекту використовувалося як навчальний комплекту, а решта 20% – в якості валідаційного комплекту. Ми перестали вчитися, якщо успішність не покращилася протягом 10 епох. Перед початком навчального процесу навчальний комплект був також

перетасувати таким чином, щоб вхідні зображення змішувалися один з одним, щоб запобігти появі послідовних зображень одного і того ж предмета. Щоб гарантувати, що перетасування даних завжди відбувалася однаково, вручну встановлюємо всі можливі насіння для генераторів випадкових чисел. Таким чином, зображення були перетасувати в процесі попередньої обробки. Зображення також перетасовувалися перед виконанням кожної епохи. Середня продуктивність розпізнавання емоцій методом VGG19 transfer для кожного набору даних приведена в таблиці 4.7. Результати наведені окремо для випадку використання вихідних даних і збагачених даних.

Таблиця 4.7 – Середня продуктивність розпізнавання емоцій методом VGG19

Колекція	**Середня точність** **Вихідні дані**	**Збагачені дані**
CK+	78,94 %	83,03 %
MMI	50,94 %	58,85 %
JAFFE	48,78 %	54,68 %
CK+,MMI,JAFFE	68,71 %	69,64 %

Середня продуктивність розпізнавання емоцій за методом VGG19 ResNet для кожного набору даних приведена в таблиці 4.8. Результати знову наводяться окремо для випадку використання вихідних даних і збагачених даних.

Таблиця 4.8 – Середня продуктивність розпізнавання емоцій методом VGG19 ResNet

Колекція	**Середня точність–Вихідні дані (зображення)**	**Збагачені дані (зображення)**
CK+	90,73 %	87,99 %
MMI	58,29 %	45,49 %
JAFFE	54,92 %	25,83 %
CK+,MMI,JAFFE	65,42 %	39,10 %

Середня ефективність розпізнавання емоцій методом VGGFace_transfer для кожного набору даних показана в таблиці 5.9. Результати знову наводяться окремо для випадку використання вихідних даних і збагачених даних.

Таблиця 4.9 – Середня продуктивність розпізнавання емоцій методом VGGFace transfer

Колекція	**Вихідні дані (зображення)**	**Збагачені дані (зображення)**
CK+	89,46 %	91,49 %
MMI	55,97 %	57,24 %
JAFFE	51,35 %	66,53 %
CK+,MMI,JAFFE	75,94 %	75,95 %

Середня продуктивність методу розпізнавання емоцій на основі власних інцепцій в наборі даних CIFAR-100 для кожного набору даних показана в таблиці 4.10.

Таблиця 4.10 – Середня продуктивність розпізнавання емоцій по власному методу «Власне сприйняття »

Колекція	**Вихідні дані (зображення)**	**Збагачені дані (зображення)**
CK+	78,16 %	82,83 %
MMI	47,58 %	59,08 %
JAFFE	39,24 %	48,99 %
CK+, MMI,JAFFE	68,58 %	73,17 %

Середня ефективність розпізнавання емоцій за методом Own inception при відсутності раніше здобутої освіти (модель навчалася тільки за розділами даних CK +, MMI і JAFFE) для кожного набору даних показана в таблиці 4.11.

Таблиця 4.11 – Середня продуктивність розпізнавання емоцій по власному методу «Власне сприйняття» без навчання.

Колекція	**Вихідні дані (зображення)**	**Збагачені дані (зображення)**
CK+	85,95 %	85,56 %
MMI	46,05 %	47,55 %
JAFFE	43,97 %	59,47 %
CK+, MMI, JAFFE	69,50 %	73,75 %

Середня ефективність розпізнавання емоцій за методом VGGFace_modified для кожного набору даних показана в таблиці 5.12. Результати знову наводяться окремо для випадку використання вихідних

даних і збагачених даних.

Таблиця 4.12 – Середня продуктивність розпізнавання методом VGGFace modifield

Колекція	**Вихідні дані**	**Збагачені дані**
CK+	90,88 %	91,65 %
MMI	57,86 %	55,91 %
JAFFE	61,59 %	67,69 %
CK+,MMI,JAFFE	79,05 %	79,86 %

4.2 Огляд результатів

У цьому розділі ми розглядаємо результати і аналізуємо, наскільки добре виконані запропоновані нами методи. Ми розглянемо, як відмінності в продуктивності між базовими і збагаченими навчальними наборами і як наші кращі методи порівнюються з суміжними роботами.

З отриманих результатів видно, що для переважної більшості моделей більший тренажерний набір допомагає поліпшити ефективність класифікації. Підхід VGG19 ResNet був єдиним виходом в цьому сенсі, так як ми навіть бачили значне зниження продуктивності з цієї архітектурою. Причина деградації результатів з збагаченими навчальними множинами може бути пов'язана зі зменшенням кількості нейронів в повністю з'єднаних шарах, і модель могла почати підлаштовуватися. Точність для набору даних CK + погіршилася на 3% для збагаченого набору даних, для набору даних MMI – на 13%, а для набору даних JAFFE точність змінилася найбільше, погіршившись більш ніж на 29%. Для спільного навчального набору точність знизилася майже наполовину, тобто на 25 відсотків.

Для підходу VGG19 transfer точність підвищилася приблизно на 4% для набору даних CK + при використанні збагаченого навчального набору, приблизно на 8% для MMI і майже на 6% для JAFFE. Суттєвого підвищення

точності при загальному наборі тренувань не відбулося. Ймовірною причиною є значна межсуб'єктних різниця між окремими колекціями. Варто відзначити, що це тільки модель, яка вивчається для розпізнавання об'єктів, а риси, витягнуті з згорткових рівнів, не зовсім пристосовані для розпізнавання осіб.

Для наступного підходу, VGGFace transfer, який націлений на розпізнавання осіб, продуктивність значно вище, ніж для підходу VGG19_transfer, за винятком бази даних MMI, де продуктивність нижче на один процентний пункт. Для бази даних CK + продуктивність підвищилася майже на 9% в порівнянні з VGG19_transfer, а для бази даних JAFFE точність підвищилася майже на 12%. Для спільного навчального набору, що складається з усіх колекцій, середня продуктивність моделі залишилася колишньою між базовою і збагаченої колекціями, з приростом в 6% в порівнянні з VGG19_transfer. Для цієї моделі найбільша різниця в точності між базовим навчальним набором і збагаченими даними спостерігається для набору JAFFE, який має найменшу кількість зображень для початку. Це також свідчить про перевагу перекладу навчання, а саме про те, що такий підхід в значній мірі усуває необхідність у великому комплексі навчання.

Для підходу на основі Own Inception, заснованого на принципі архітектури Inception, де функції шукаються одночасно за кількома розмірами зображень, нейромережеве навчання з випадково Ініціалізувати вагами показало кращу продуктивність, за винятком випадку з базою даних MMI, де модель працювала значно краще при використанні методу трансфертного навчання, а модель була попередньо підготовлена з розпізнавання об'єктів з набору даних CIFAR-100. У випадку з базою даних MMI найкраща продуктивність була знайдена в разі методу нейромережевого навчання. У разі навчання з випадковою ініціалізацією ваг, ми спостерігаємо, що продуктивність моделі з використанням наборів даних CK + і MMI не покращилася значно зі збагаченим навчальним набором, поліпшення було тільки між 1 і 2%. У наборі даних JAFFE,

В останньому підході VGGFace modified, де ми об'єднали модель VGGFace з блоком, який ми реалізували при побудові власної архітектури, ми отримали найкращу середню продуктивність серед всіх запропонованих підходів. Знову ж, найбільша різниця між збагаченим набором і базовим набором знаходиться в наборі JAFFE, а в наборі MMI продуктивність навіть впала в міру збільшення тренувального набору. Точність тут знизилася з 57,8% до 55,9%. Однак, цей підхід все ж забезпечив кращу середню класифікацію в колекціях CK + і JAFFE, а також в тренувальному наборі, що складається з даних з усіх трьох колекцій. Середні показники склали 91,6% для CK +, 67,6% для JAFFE і 79,8% для спільного комплекту. Найкращі результати в наборі даних MMI були досягнуті при використанні підходу «Власне сприйняття»

Ґрунтуючись на результатах, ми бачимо, що всі запропоновані методи показали гірші результати в базі даних MMI. Набір даних MMI відомий в дослідницькому співтоваристві як більш складний, тому що він має більш високу міжособистісну варіативність. Сюжети з різних культур; у деяких з них є такі аксесуари, як окуляри, вуса і так далі. Суб'єкти також відрізняються один від одного інтенсивністю відображаються емоцій. Навіть в суміжних роботах [17], присвячених цьому питанню, продуктивність методів в цій колекції гірше, ніж в порівнянні з результатами в колекціях CK + і JAFFE. У відповідній роботі, описаної в підрозділі 2.2, результати для колекції MMI склали близько 77% класифікаційного успіху.

З іншого боку, розмір бази даних JAFFE є проблемою, так як вона містить тільки 213 зображень. Вона потребує значного надалі збагаченні, щоб бути порівнянної за розмірами з базами даних CK + і MMI. Найкращий результат для бази даних JAFFE був досягнутий при використанні модифікованого підходу VGGFace modified, де середня продуктивність склала 67,69%. У відповідній роботі, описаної в підрозділі 2.2, середній показник ефективності класифікації для колекції JAFFE становить близько

94,9%. Що показує найвищий результат серед всіх пропонованих баз даних але при використанні модифікаційного підходу булу отримано на багато менший відсоток що не дає нам високий показник.

Для бази даних CK + підхід VGGFace modified досяг найбільшої точності 91,65%. Для цього набору даних отримані результати найбільш близькі до результатів відповідної роботи, описаної в підрозділі 3.2, де точність результатів складає близько 93,8%. В підсумок можемо отримати результати , де можна поглянути на точність проведених робіт .

У суміжній роботі підходи, які використовуються для валідації, відрізнялися від наших. У наших підходах використовувалася 5-кратна перехресна перевірка, в той час як в суміжних роботах застосовувалася 10-кратна перехресна перевірка. У суміжних роботах також застосовувалася додаткова попередня обробка зображення, наприклад, підсвічування і нормалізація контрастності, тоді як в наших підходах але ми не використали експозицію і нормалізацію контрастності. Це тому, що ми тестували продуктивність, коли моделі проходять по необроблених зображеннях. Це відбувається тому, що від згорткових нейронних мереж очікують власних фільтрів, які попередньо обробляють вхідні зображення і компенсують помилки, що виникають під час захоплення зображення. Однак, ґрунтуючись на отриманих результатах, це не так, тому що ми бачимо, що в суміжних роботах, де вони додатково препроцесурують зображення, вони мають більш високі класифікаційні характеристики, ніж в наших підходах, де ми використовували необроблені зображення.

На цьому етапі також видно що база даних MMI за допомогою власного методу показує низький результат але використовуючи інші бази даних та власний метод отримали підвищення результатів на кілька відсотків.

Порівняння отриманих результатів з результатами суміжних робіт наведено в таблиці 4.13.

Таблиця 4.13 – Кращі результати отримані з допомогою методів запропонованих в розділі 3.

База даних	**Найвища точність**	**Точність супутніх робіт**
CK+	91,65 % (VGGFace modified)	93,8 %
MMI	59,08 % Власний метод	77,0 %
JAFFE	67,69 % (VGGFace modified)	94,9 %

Ґрунтуючись на всіх отриманих результатах, ми бачимо, що підхід VGGFace_modified, тобто модель, в якій ми розширили підхід VGGFace_transfer блоком з нашої власної архітектури, в середньому показав найкращі результати. Таким чином, ми використовували частина моделі VGGFace для вилучення рис обличчя, серед яких ми потім додатково шукали риси, які з'являються в декількох різних розмірах. Додавання цього додаткового рівня згорток з різними розмірами пошуку збільшило продуктивність на 1-2 відсотки.

У базі даних JAFFE наш кращий метод насилу класифікував емоції страху, смутку і нейтральності. У деяких випадках метод не міг класифікувати ці три емоції і, наприклад, класифікувати цільову емоцію страху як нейтральну, а печаль як гнів. Один з предметів, для якого метод мав труднощі з класифікацією, показаний на рисунку 4.1. Як видно на рисунку, показані виразу обличчя дуже мало відрізняються один від одного для деяких предметів.

Рисунок 4.1– Приклад неправильної класифікації в базі JAFFE

База даних MMI вважається дуже складною для класифікації за визначенням, оскільки вона містить велику кількість об'єктів, які сильно відрізняються один від одного. Обидві статі включені, різні вікові групи, деякі суб'єкти носять окуляри, деякі ні, волосся на обличчі виявлені і т.д. Для одного і того ж суб'єкта, іноді трапляється, терміни сильно варіюються, що дуже ускладнює навчання і класифікацію. Один з таких об'єктів з колекції MMI показаний на рисунку 4.2. Найбільш важко класифікувати тут нейтральний термін, термін відрази, термін страху і термін несподіванки. На рисунку 4.3 ми бачимо приклад літньої жінки-суб'єкта, де в нашому кращому підході нейтральний термін був позначений, відповідно, як термін для позначення відрази і страху.

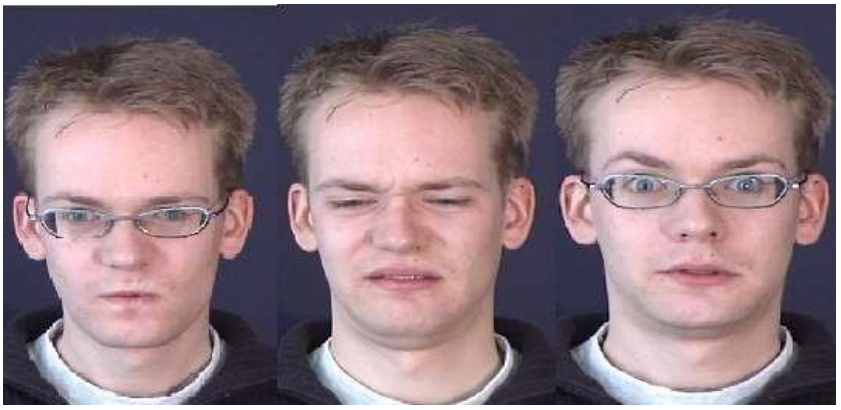

Рисунок 4.2– Приклад об’єкта з бази даних MMI

Рисунок 4.3 – Приклад неправильної класифікації MMI

У колекції CK + наш кращий підхід не мав ніяких серйозних проблем з сортуванням по потрібним цільовим емоціям. Емоції, які викликали найбільше проблем, були сумом. У деяких випадках він класифікував цю емоцію як нейтральний термін, що, ймовірно, пов'язано з дуже нерівномірним розподілом емоцій в колекції CK +, як видно з таблиці 4.4. На рисунку 4.4 показані деякі суб'єкти, які при нашому кращому підході класифікувалися як нейтральний термін, незважаючи на те, що суб'єкти проявляли емоції смутку.

Рисунок 4.4 – Приклади с бази CK+

У цьому розділі було показано практичне застосування та приведені отримані результати роботи методів в різних колекціх записані в таблицях, також були взяті рисунки для прикладу з різних баз даних.

ВИСНОВКИ

В даній кваліфікаційній роботі було показано, класифікацію шести основних людських емоцій, включаючи нейтральний вираз, в цифрових зображеннях з використанням згорткових нейронних мереж. По-перше, було вивчено область розпізнавання людських емоцій, проведено огляд деяких пов'язаних з цим робіт і розглянуті існуючі сучасні згорткові нейронні мережі. Було реалізовано кілька існуючих моделей згорткових нейромереж і розроблено власні моделі, використовуючи бібліотеку Tensorflow і мова програмування Python. Протестовано рішення на вільно доступних базах даних CK +, MMI і JAFFE. Попередньо розширено зображення з баз даних для отримання більшого обсягу даних «зображення були віддзеркалені та повернуті».

Запропоновано п'ять методів розпізнавання людських емоцій. Перший метод був заснований на моделі VGG19, яку було адаптовано до нашої проблеми шляхом трансфертного навчання VGG19_transfer.Другим методом була комбінація архітектури VGG19 і лежить в її основі архітектури ResNet VGG19_ResNet. Додано додаткове з'єднання, яке додало дані з входу на останній згортковий шар до виходу цього згорткового шару. Додавання цього посилання допомагає моделі вирішити так звану проблему зниження ваги. Третій метод використовує модель VGGFace метод VGGFace transfer, який спрямований на розпізнавання осіб в зображеннях. Також в цьому методі ми використовували метод трансфертного навчання, щоб адаптувати модель VGGFace до нашої проблеми. Четвертий метод, Own Inception based, був нашою власною розробленою архітектурою, яка працює за принципом, схожим з архітектурою Inception. Перевагою цієї архітектури є пошук можливостей на зображенні декількох різних розмірів. Для цієї архітектури ми використовували два підходи до навчання. Перший підхід полягав у попередньому навчанні по другому набору даних, де ми потім використовували

Метод трансфертного навчання для повторної адаптації моделі до нашої проблемі. У другому підході ми навчали модель тільки з даними з наборів даних CK +, MMI і JAFFE. У п'ятому методі так званий метод VGGFace modified використовувалась модель VGGFace, до якої додали один блок з нашої власної розробленої моделі після згорткових шарів. Це додало додаткові рівні.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. S. Li., W.Deng. Deep Facial Expression Recognition: A Survey. IEEE Transactions on Affective Computing, 04. 2018. P. 1–10.

2. M. Abadi., A. Agarwal., P. Barham., E. Brevdo., Z. Chen., C. Citro., G. S. Corrado. Tensorflow: Large-scale machine learning on heterogeneous systems:Tensorflow.org , 2015. URL:http://download.tensorflow.org/paper/whitepaper2015.pdf (Last accessed 24.04. 2021)

3. Convolutional neural networks.URL:https://en.wikipedia.org/wiki/Convolutional_neural_network(Last accessed 23.04.2021)

4. K.Simonyan., A.Zisserman. Very deep convolutional Networks for Large-Scale Image Recognition. International Conference on Learning Representations, 2015. 30 p.

5. P. Lucey., J. F. Cohn., T. Kanade., J. Saragih., Z. Ambadar., I. Matthews. The Extended Cohn-Kanade Dataset (CK+): A complete facial expression dataset for action unit and emotion-specified expressions. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition – Workshops , San Francisco, CA, 2010. P. 94–101.

6. M. F. Valstar., M. Pantic. Induced Disgust, Happiness and Surprise: an Addition to the MMI Facial Expression Database. Proc. Int'l Conference Language Resources and Evaluation, 01, 2010, P. 65–70.

7. M. Lyons., S. Akamatsu., M. Kamachi., J. Gyoba. Coding facial expressions with gabor wavelets. Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998. P. 200–205.

8. O.Russakovsky., J. Deng., H. Su., J. Krause., S. Satheesh.ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV), 2015. P.211– 252.

9. Transfer Learning. URL:https://en.wikipedia.org/wiki/Transfer_learning (Last accessed 24.04. 2021)

10. C. Szegedy., W. Liu., Y. Jia., P. Sermanet, S. E. Reed, D. Anguelov.,

D. Erhan. Going Deeper with Convolutions. 2015 IEEE Conference on Computer Vision, Boston, MA, 2015. P. 1–9.

11. K. He, X. Zhang., S. Ren., J. Sun. Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016. P. 770–778.

12. O.M.Parkhi., A.Vedalbi., A.Zisserman. Deep Face Recognition. Proceedings of the British Machine Vision Conference 1,2015. 24 p.

13. CIFAR. URL: CIFAR-10 and CIFAR-100 datasets (toronto.edu) (Last accessed 24.04.2021)

14. Wikipedia.Emotion recognition.URL:https://en.wikipedia.org/wiki/Emotion_recognition (Last accessed 24.04. 2021)

15. Wikipedia.Affectiva.URL:https://en.wikipedia.org/wiki/Affectiva (Last accessed 24.04. 2021)

16. C. Szegedy., V. Vanhoucke., S. Ioffe., J. Shlens., Z. Wojna. Rethinking the Inception. Architecture for Computer Vision. Computing Research Repository (CoRR), Las Vegas,NV, 2016. P. 2818–2826.

17. A. Mollahosseini., D. Chan., M. Mahoor. Going Deeper in Facial Expression Recognition using Deep Neural Networks. IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, 2016. P. 1–10.

18. S. Oullet. Real-time emotion recognition for gaming using deep convolutional network features. Architecture for Computer Vision. Computing Research Repository (CoRR), 2014. P.15-35.

19. A.Krizhevsky., I.Sutskever., G.Hinton.ImageNet with Deep Convolutional Neural Networks. Neural Information Processing Systems, 2012. P.25-27.

20. D.Hamester., P.Barros., S. Wermter.Face expression recognition with a 2-channel Convolutional Neural Network. International Joint Conference on Neural Networks, Killarney, 2015. P. 1–8.

21. Galeone. Convolutional Autoencoders. URL:https://pgaleone.eu/neurallnetworks/2016/11/24/convolutionalautoencoders/(Last accessed 24.04. 2021)

22. Dlib.net. URL: http://dlib.net (Last accessed 24.04. 2021)

23. Gavriushenko, M., Kaikova, O., & Terziyan, V. (2020). Bridging Human and Machine Learning for the Needs of Collective Intelligence Development. *Procedia Manufacturing*, *42*,2020. P. 302–306.