

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА»
ФАКУЛЬТЕТ ІНФОРМАЦІЙНО-КОМП'ЮТЕРНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНЖЕНЕРІЇ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

ПОЯСНЮВАЛЬНА ЗАПИСКА

до кваліфікаційної роботи освітнього ступеня «магістр»
за спеціальністю 121 «Інженерія програмного забезпечення»
(освітня програма «Інженерія програмного забезпечення»)

на тему:

**«Розпізнавання емоцій обличчя з використанням згорткової
нейронної мережі»**

Виконав студент групи ЗППЗм-23-1
ТІМОШИК Петро Петрович

Керівник роботи:
ЄФРЕМОВ Юрій Миколайович

Рецензент:
ЧИЖМОТРЯ Олексій Володимирович

Житомир – 2024

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДЕРЖАВНИЙ УНІВЕРСИТЕТ «ЖИТОМИРСЬКА ПОЛІТЕХНІКА»
ФАКУЛЬТЕТ ІНФОРМАЦІЙНО-КОМП'ЮТЕРНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНЖЕНЕРІЇ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

«ЗАТВЕРДЖУЮ»

зав. кафедри інженерії програмного
забезпечення

Тетяна ВАКАЛЮК

«01» жовтня 2024 р.

ЗАВДАННЯ
на кваліфікаційну роботу

Здобувач вищої освіти: **ТІМОШИК Петро Петрович**

Керівник роботи: **ЄФРЕМОВ Юрій Миколайович**

Тема роботи: **«Розпізнавання емоцій обличчя з використанням згорткової нейронної мережі»,**

затверджена Наказом закладу вищої освіти від **«10» жовтня 2024 р. № 521/с**

Вихідні дані для роботи:

Предметом дослідження є система для порівняння методів для розпізнавання емоцій людини.

Об'єктом дослідження засоби розпізнання людських емоцій

Консультанти з магістерської кваліфікаційної роботи із зазначенням розділів, що їх стосуються:

Розділ	Консультант	Завдання видав	Завдання прийняв
1	Єфремов Ю.М.	14.10.2024р.	14.10.2024р.
2	Єфремов Ю.М	14.10.2024р.	14.10.2024р.
3	Єфремов Ю.М	14.10.2024р.	14.10.2024р.

РЕФЕРАТ

Кваліфікаційна магістерська робота полягає в дослідженні, розпізнаванню емоцій обличчя з використанням згорткової нейронної мережі.

У дослідженні представлено оцінка актуальних досліджень та розробки нових рішень у галузі розпізнавання емоцій. Завдання розпізнавання емоцій обличчя полягає у тому, щоб навчити комп'ютерну систему, зокрема нейронну мережу, точно ідентифікувати емоції, які виражаються на людському обличчі.

Магістерська робота складається із вступу, трьох розділів та висновків, списку використаних джерел, що налічує 62 сторінок, 6 ілюстрацій та 48 посилань на зовнішні джерела.

КЛЮЧОВІ СЛОВА: ЕМОЦІЇ, РОЗПІЗНАВАННЯ, ЗГОРТКОВА НЕЙРОННА МЕРЕЖА, МАШИННЕ НАВЧАННЯ, ОБРОБКА ЗОБРАЖЕНЬ, АВТОМАТИЗАЦІЯ.

ABSTRACT

The qualifying master's thesis focuses on research dedicated to facial emotion recognition using a convolutional neural network.

The study presents an evaluation of relevant research and the development of new solutions in the field of emotion recognition. The task of recognizing facial emotions is to train a computer system, specifically a neural network, to accurately identify the emotions expressed on a human face.

The master's thesis consists of an introduction, three chapters, and conclusions, along with a list of references that totals 62 pages, 6 illustrations and 48 citations of external sources.

KEYWORDS: EMOTIONS, RECOGNITION, CONVOLUTIONAL NEURAL NETWORK, MACHINE LEARNING, IMAGE PROCESSING, AUTOMATION.

ЗМІСТ

ВСТУП.....	7
РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ	10
1.1. Огляд наявних меделей, методів та алгоритмів.....	10
1.2. Розпізнавання облич на основі матриці зміни яскравості	16
1.3. Постановка завдання.....	22
Висновки до розділу 1	24
РОЗДІЛ 2. МЕТОДИ ТА АЛГОРИТМИ РОЗПІЗНАННЯ ЕМОЦІЙ.....	25
2.1. Структура згорткові нейронні мережі	25
2.2. Метод зіставлення з еталоном (Template Matching)	27
2.3. Активні моделі зовнішнього вигляду	29
2.4. Метод на основі ключових точок	32
2.5. Метод на основі інформації про текстуру	34
2.6. Локальні бінарні шаблони.....	36
2.7. Розпізнавання облич за допомогою згорткових нейронних мереж.....	38
Висновки до розділу 2	40
РОЗДІЛ 3. ОПИС СИСТЕМИ РОЗПІЗНАННЯ ЕМОЦІЙ	42
3.1 Розробка алгоритмів та проектування системи розпізнавання емоцій	42
3.2. Інструментальні засоби реалізації програмного продукту та вимоги до апаратних засобів.	44
3.3 Розробка системи для аналізу методів для розпізнавання емоцій.....	48
3.4. Порівняння ефективності методів розпізнавання емоцій з використанням створеної системи.....	52
Висновки до розділу 3	54
ВИСНОВКИ.....	56
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	58

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

CNN	– (англ. convolutional neural networks) – згорткові нейронні мережі
NLP	– (англ. natural language processing) – обробка природної мови
ReLU	– (англ. rectified linear unit) – ректифікована лінійна одиниця
SGD	– (англ. stochastic gradient descent) – стохастичний градієнтний спуск
Adam	– (англ. adaptive moment estimation) – адаптивне оцінювання моментів
ResNet	– (англ. residual network) – залишкова мережа
MLL	– (англ. multiple linear regression) – багатократна лінійна регресія
PCA	– (англ. principal component analysis) – метод головних компонент
VGG	– (модель vggnet) – згорткова нейронна мережа, розроблена в Оксфордському університеті.
AAM	– (англ. active appearance models,) – активні моделі зовнішнього вигляду
ASM	– (англ. active shape models) – активні моделі форми
LBP	– (англ. local binary pattern) локальні бінарні шаблони
MLP	– (англ. multilayer perceptron) - багат шарова перцептронна нейронна мережа.
LDA	– (англ. linear discriminant analysis) – лінійний дискримінантний аналіз
ЛБШ	– локальний бінарний шаблон
СКЛід	– (система класифікації на основі лінійної дискримінації)

ВСТУП

Актуальність дослідження. Розпізнавання емоцій обличчя за допомогою нейронних мереж – це галузь штучного інтелекту, яка стрімко розвивається і набуває все більшої актуальності. Ця технологія дозволяє комп'ютерам аналізувати вирази обличчя людини та визначати її емоційний стан. Така здатність відкриває широкі перспективи для застосування в різних сферах життя.

Розпізнавання емоцій дозволяє краще зрозуміти, що відчуває людина в певний момент часу. Це може бути корисним для психологів, соціологів та інших фахівців, які працюють з людьми. Технологія може бути використана для створення більш інтуїтивних інтерфейсів, які здатні адаптуватися до емоційного стану користувача. Це відкриває нові можливості для розробки розумних помічників, ігор та інших інтерактивних систем.

У маркетинга та рекламі розпізнавання емоцій може бути використано для оцінки реакції споживачів на різні продукти та послуги. Це дозволяє створювати більш ефективні маркетингові кампанії. Деякі емоції можуть бути пов'язані з певними психічними розладами. Аналіз виразів обличчя може допомогти в ранній діагностиці таких захворювань.

Системи розпізнавання емоцій можуть бути використані для виявлення ознак агресії або інших небезпечних станів, що може бути корисним в охоронних системах та системах моніторингу громадської безпеки.

Розпізнавання емоцій обличчя за допомогою нейронних мереж – це перспективна технологія, яка має великий потенціал для застосування в різних сферах життя. Незважаючи на існуючі виклики, ця галузь активно розвивається, і в майбутньому ми можемо очікувати появи ще більш точних і ефективних систем.

Вивченням емоцій та їхнього прояву вчені займаються досить давно. Адже емоції є неминучою частиною будь-якої міжособистісної комунікації, виражають ставлення людини до навколишнього світу, ситуації, що склалася навколо неї, до

самої себе. Разом з тим, останнім часом потреба у виявленні людських емоцій ще більше зросла. Насамперед це пов'язано з розширенням сфери застосування завдання розпізнавання емоцій. Нині це і моніторинг стану водія за кермом, і системи відео аналітики «розумного міста», і маркетингові дослідження, і системи безпеки

Емоції можуть бути виражені різними способами: мімікою, голосом, поведінкою, реакціями систем організму [1]. Найбільший інтерес із них становить розпізнавання емоцій людини за виразом її обличчя. Це завдання є досить популярним у цей час з низки причин: такі зображення нескладно отримати, вони містять багато корисної інформації для розпізнавання емоцій, зібрати великий набір даних у вигляді зображень обличчя досить легко (порівняно з іншим матеріалом для розпізнавання: мовленням або зразками почерку)

Згорткові нейронні мережі – це тип штучних нейронних мереж, які особливо ефективні для роботи з даними зображень. Вони імітують роботу зорової кори головного мозку і здатні виявляти та класифікувати різні візуальні образи.

Метою роботи є розробки системи для аналізу розпізнавання емоцій, заснованої на нейромережевих технологіях. У контексті даної роботи розглядаються такі 6 емоцій (базові емоції): гнів, здивування, сум, радість, страх і відраза.

Для здійснення поставленої мети необхідно розв'язати такі завдання.

- Провести теоретичний аналіз досліджуваної галузі.
- Сформувати навчальну вибірку для тренування нейронної мережі.
- Здійснити попереднє оброблення навчальної вибірки.
- Обрати архітектуру нейронної мережі, навчити її та застосувати отриману модель для розпізнавання.
- Обрати найбільш ефективний метод детектування обличчя на зображенні.

Об'єкт дослідження є спектр людських емоцій, які виражаються на обличчі.

Предмет дослідження – фокус конкретних аспектів цього процесу.

Розуміння цих аспектів є важливим для оцінки актуальних досліджень та розробки нових рішень у галузі розпізнавання емоцій. Завдання розпізнавання емоцій обличчя полягає у тому, щоб навчити комп'ютерну систему, зокрема нейронну мережу, точно ідентифікувати емоції, які виражаються на людському обличчі. Це досить складне завдання, оскільки емоції можуть бути дуже різноманітними і виражатися через тонкі нюанси міміки.

Новизна дослідження є надзвичайно динамічним і постійно розвивається. Незважаючи на значні досягнення, ця сфера все ще залишається багатообіцяючою і пропонує безліч напрямків для подальших досліджень. Що робить це дослідження новим і актуальним.

Дослідження розпізнавання емоцій обличчя за допомогою нейронних мереж є динамічною і багатообіцяючою галуззю. Постійний розвиток технологій та нові підходи відкривають широкі перспективи для застосування цієї технології в різних сферах життя.

Теоретичне значення дослідження полягає в тому, що воно не тільки сприяє розвитку штучного інтелекту, але й глибше розуміння когнітивних процесів людини. Це відкриває нові горизонти для наукових досліджень і практичного застосування в різних сферах життя.

Апробація. За темою кваліфікаційної магістерської роботи було опубліковано тези конференції [48]

РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1. Огляд наявних меделей, методів та алгоритмів

В останні роки виникла значна зацікавленість у дослідженнях з області розпізнавання емоцій людини за допомогою інтелектуальних систем. Проведений аналіз різних підходів до цієї задачі виявив різноманітні методи та техніки, які використовуються для ефективного розпізнавання емоцій на зображеннях та відео. Дослідження також відзначають методи покращення ефективності нейронних мереж шляхом оптимізації обробки зображень та використання згорткових нейронних мереж для досягнення високої точності в розпізнаванні емоцій. Ці дослідження спрямовані на подальше розвиток та покращення систем розпізнавання емоцій для різноманітних застосувань, включаючи соціальні мережі, медичні діагностики та інтерфейси машинної взаємодії з людьми.

Шеіх Р. у своїх роботі “A comparative study of emotion recognition methods using facial expressions” [36] проводить аналіз різних методів та підходів до розпізнавання емоцій за допомогою технологій штучного інтелекту.

Використання згорткових нейронних мереж (CNN) є одним з найпоширеніших підходів у розпізнаванні емоцій за допомогою зображень. Ці мережі спроектовані для автоматичного виявлення та аналізу важливих ознак, таких як обличчя, вирази, рухи тощо, що дозволяє їм навчатися розпізнавати емоційні стани на зображеннях.

Другий підхід полягає в аналізі тексту та використанні природної мови (NLP). В цьому випадку дослідники аналізують текстові дані, такі як коментарі, відгуки або соціальні медіа, для виявлення емоційного тону або настрою особи за допомогою методів NLP.

Третій підхід використовує аналіз аудіо та голосу. В даному випадку дослідження зосереджені на виявленні емоцій у голосі людини, використовуючи аналіз тональності, інтонації, швидкості мовлення тощо для розуміння їхнього емоційного стану.

Деякі дослідники поєднують ці підходи для створення комплексних систем розпізнавання емоцій, які використовують інформацію з різних джерел (зображень, тексту, аудіо) для більш точного аналізу емоційних станів. Такі комбіновані підходи можуть покращити точність інтерпретації емоцій і забезпечити більш глибоке розуміння людських емоційних виразів.

Оцінка точності розпізнавання в контексті роботи з розпізнавання емоцій є важливим етапом, що включає декілька ключових кроків та методів. Починаючи зі збору набору даних, який містить зображення або відео з емоціями, проходження через модель розпізнавання емоцій, використання різних метрик точності (наприклад, accuracy, sensitivity, specificity) для оцінки результатів, до аналізу помилок моделі та виявлення можливостей для поліпшення алгоритмів.

Для початку, наявність достатньої та репрезентативної колекції даних є важливим аспектом. Цей набір даних може бути зібраний у реальних умовах або симулювати різні сценарії емоцій на зображеннях чи відео.

Після цього, дані проходять через модель розпізнавання емоцій, яка намагається визначити тип та наявність емоцій на кожному зображенні чи відео. Цей етап включає в себе використання алгоритмів машинного навчання, таких як згорткові нейронні мережі, для аналізу інформації та прийняття рішень.

Для оцінки ефективності моделі використовуються різні метрики точності, які дозволяють визначити, наскільки добре модель розпізнає різні емоції. Ці метрики можуть включати точність (accuracy), чутливість (sensitivity) та інші.

Наступний крок - аналіз помилок моделі. Це допомагає виявити ситуації, де модель може допускати помилки або бути менш ефективною, що в свою чергу дозволяє розробити стратегії для поліпшення алгоритмів та збільшення точності розпізнавання емоцій.

У результаті цього процесу можна зробити висновки про ефективність та можливості подальшого вдосконалення методів та підходів до розпізнавання емоцій у дослідженні.

Хаїредін Й. у дослідженні “Emotion Recognition: State of the Art Performance on FER2013” [38] мав за мету покращення ефективності нейронних мереж шляхом зміни зображень.

Сучасні нейронні мережі операціонують з великими обсягами даних, що призводить до високих вимог до обчислювальних ресурсів. Маніпуляція зображеннями може допомогти зменшити розмір даних без втрати важливої інформації, що покращить ефективність обробки.

Деякі завдання в області комп'ютерного зору та обробки зображень можуть бути дуже складними через різноманіття варіантів об'єктів, варіації освітлення та інші фактори. Маніпулюючи зображеннями перед подачею їх на вхід нейронній мережі, можна спростити завдання та покращити її здатність до розпізнавання.

Використання видозміни зображень може оптимізувати використання обчислювальних ресурсів, особливо в ситуаціях з обмеженими обсягами пам'яті чи обчислювальною потужністю. Також це може допомогти знизити рівень шуму на зображеннях, покращити контрастність або іншим чином покращити якість даних, що входять у нейронну мережу. Такий підхід спрямований на забезпечення кращої якості вхідних даних для мережі, що впливає на її загальну ефективність.

У дослідженні розглядаються різні технічні методи маніпулювання зображеннями з метою покращення ефективності нейронних мереж.

Один з підходів - це зменшення розміру зображень перед подачею їх на вхід нейронній мережі. Це може бути доцільним для зменшення обсягу обчислень та збереження ресурсів, особливо при роботі з великими масивами даних.

Деякі методи включають в себе обробку зображень для виділення ключових ознак або зменшення шуму. Це може включати застосування фільтрів для видалення незначущої інформації або підсилення важливих деталей.

Також є методи спрямовані на зменшення шуму в зображеннях, що може покращити якість обробки та точність результатів нейронних мереж.

Технічні засоби можуть включати розробку алгоритмів та програмного забезпечення для автоматизації процесу маніпуляції зображеннями, що дозволяє ефективніше використовувати ресурси та збільшує швидкість обробки.

Дослідження показало, що застосування технічних методів маніпулювання зображеннями може значно покращити ефективність нейронних мереж. Зокрема, зменшення розміру зображень, обробка для виділення ключових ознак та використання спеціальних технік для зменшення шуму допомагають оптимізувати процес обробки даних та підвищити точність результатів.

Це також призводить до економії ресурсів, що особливо важливо при роботі з великими обсягами даних. Крім того, розробка автоматизованих методів маніпулювання зображеннями сприяє швидкості обробки та може знизити витрати на обчислення.

Напрямки подальших досліджень можуть включати розвиток нових методів маніпуляції зображеннями для різних завдань та застосування їх у реальних умовах, що дозволить ще більше покращити роботу нейронних мереж у обробці зображень.

У своєму дослідженні “Fine-grained image analysis for facial expression recognition using deep convolutional neural networks with bilinear pooling” Хусейн С. та Роут Р. [37] розглянули методи навчання деталізованих представлень обличчя для розпізнавання емоцій за допомогою нейронних мереж.

Ціль дослідження полягає у покращенні систем розпізнавання емоцій за обличчям шляхом навчання деталізованих представлень. Для цього використано мережі, які аналізують різні рівні інформації в зображеннях обличчя для кращого розуміння емоційних виразів.

Основні результати дослідження свідчать про те, що такий підхід може значно покращити точність систем розпізнавання емоцій за обличчям. Він дозволяє враховувати дрібні деталі та особливості обличчя, що сприяє кращому розпізнаванню емоцій та забезпечує більш точні результати.

Отримані результати мають важливе значення для розвитку систем штучного інтелекту та комп'ютерного зору, особливо в області розпізнавання емоцій в реальному часі за допомогою інформації з обличчя.

Даний вид мереж здатність адаптуватися до різних рівнів інформації у вхідних даних і враховувати їхню важливість в процесі прийняття рішень.

Вони враховують різні деталі та особливості обличчя, такі як вирази очей, рухи губ, міміка, тон шкіри тощо. Це дозволяє їм краще розуміти вирази обличчя та враховувати дрібні нюанси при розпізнаванні емоцій.

Мережі надають важливість різним частинам обличчя в залежності від їхнього внеску у загальний контекст емоційного виразу. Наприклад, підсилення уваги на областях зі змінними виразами, які часто відображають емоції.

Також вони можуть розглядати обличчя як комплексну систему, де кожна частина має своє значення в контексті загального емоційного виразу. Вони здатні адаптуватися до змін у цьому контексті та змінювати свою увагу відповідно.

Цей вид нейронних мереж має механізми, що дозволяють контролювати рівні уваги на різних рівнях деталізації. Це важливо для збалансованого аналізу деталей та загального контексту обличчя в процесі розпізнавання емоцій.

Робота Ку Д. та Джахала С. “Facial Emotion Recognition using CNN in PyTorch” містить дослідження з процесу розпізнавання емоцій обличчя з використанням згорткових нейронних мереж (CNN) у бібліотеці PyTorch.

Процес розробки мережі для фронтального розпізнавання обличчя за використанням згорткових нейронних мереж (CNN) включає такі етапи:

1. Обробка даних:
 - а. Використання датасету FER2013 з зображеннями обличчя та сімома різними емоціями.
 - б. Розділення даних на тренувальний та валідаційний набори зображень.
 - с. Візуалізація даних для розуміння їхньої структури та властивостей.
2. Архітектура мережі:
 - а. Використання згорткових шарів для виявлення ознак у зображеннях.
 - б. Використання пулінгових шарів для зменшення розмірності даних та виявлення важливих ознак.
 - с. Додавання повнозв'язаного шару для класифікації обличчя на різні емоції.

- d. Використання функції активації ReLU для моделювання нелінійності у мережі.
- 3. Налаштування гіперпараметрів:
 - a. Вибір оптимальних значень для гіперпараметрів, таких як розмір партії (batch size), кількість епох навчання, швидкість навчання (learning rate).
 - b. Використання методів оптимізації, таких як викидання (dropout), пакетна нормалізація (batch normalization) та максимальне згортання (max pooling), для покращення швидкості та точності моделі.
- 4. Навчання та валідація моделі:
 - a. Навчання мережі на тренувальному наборі даних з використанням заданих гіперпараметрів та оптимізатора.
 - b. Оцінка точності моделі на валідаційному наборі для визначення її ефективності та уникнення перенавчання.
- 5. Оптимізація та аналіз результатів:
 - a. Оптимізація моделі шляхом вибору оптимального набору гіперпараметрів та оптимізатора.
 - b. Аналіз результатів навчання для визначення залежності між кількістю епох та точністю моделі.

У цьому дослідженні було розглянуто застосування згорткових нейронних мереж у фронтальному розпізнаванні емоцій на зображеннях обличчя за допомогою бібліотеки PyTorch. Розроблена показала відмінні результати у розпізнаванні семи різних емоцій на датасеті FER2013, до якого входять гнів, огида, страх, радість, сум, здивування та нейтральність.

Процес оптимізації моделі включав налаштування гіперпараметрів, таких як швидкість навчання, розмір партії та кількість епох, а також використання різних оптимізаторів, таких як SGD та Adam. Результати показали, що при належному налаштуванні параметрів, наша модель досягла точності розпізнавання на рівні 60.20% на наборі перевірки.

Дослідження також підтвердило важливість розвитку згорткових нейронних мереж у розпізнаванні емоцій на зображеннях обличчя, а також необхідність

подальших досліджень у напрямку оптимізації та покращення точності моделей для подальшого застосування в реальних умовах.

1.2. Розпізнавання облич на основі матриці зміни яскравості

Нині розроблено велику кількість різноманітних методів розпізнавання емоцій методів розпізнавання емоцій. Однак, у всіх методів існує низка спільних проблем:

- усі методи розпізнавання емоцій орієнтовані на фронтальні зображення;
- особливості оклюзії: наявність бороди, вусів, окулярів або капелюха вводять в оману алгоритми розпізнавання, або взагалі не даючи змоги розпізнати обличчя;
- умови зйомки: до цього пункту можна віднести як якість зйомки самої камери, так і якість освітлення.

Однією з перших успішних спроб систематизувати лицьові емоції є система кодування лицьових рухів (СКЛІД), розроблена Полом Екманом і Воллесом Фрізенем 1978 року. За допомогою СКЛІД можна закодувати вираз обличчя, створюючи модель із певних одиниць дії, а також фіксованого проміжку часу для відтворення того чи іншого виразу обличчя. Система кодування має фіксовані номери дескрипторів, що мають суттєву різницю з одиницями дії.

Основою для СКЛІД є рухові одиниці (ДЕ), які дають змогу закодувати будь-який автоматично можливий вираз обличчя. В прояві емоцій беруть участь не всі м'язи, а лише їх невелика частина. Основу кожної з базових емоцій складають 5-7 ДЕ, які дають змогу закодувати прототипи, які трапляються найчастіше.

Наприклад, формула (1) для здивування:

$$1 + 2 + 5B + 26. \quad (1)$$

Латинські літери від A до E, що стоять після чисел, визначають інтенсивність руху від мінімальної до максимальної відповідно.

Хоча СКЛІД і потребує добре навчених експертів, існує позитивний досвід інтегрування цієї системи в комп'ютери. Але все ж цей метод не набув широкого

застосування в технологіях комп'ютерного зору. Нижче розглянуто методи, що дають змогу домогтися більш високої ефективності розпізнавання емоцій.

У задачах детекції та розпізнавання обличч ключовим моментом є введення міри близькості на зображеннях. При цьому міра близькості за можливості повинна задовольняти аксіоми метрики і, крім того, відповідати особливостям сприйняття зорової системи людини. Під час аналізу зображення обличчя, людина не звертає уваги на спотворення яскравісної складової зображення на наявність тіней і відблисків, аналізує характер зміни яскравості, надаючи більшого значення якісним відмінностям, а не кількісним [5]. Для розв'язання задач детекції та ідентифікації осіб, зручно здійснювати перехід від вихідного зображення I до матриці зміни яскравостей M_I , елементами якої є пари чисел, що відповідають знаками приватних похідних від яскравості вихідного зображення в кожній точці. Вираз переходу за допомогою формули (2):

$$M_{t,j}^I = |sgnI'_x|_{t,j}, |sgnI'_y|_{t,j}|. \quad (2)$$

Подання зображення у вигляді матриці зміни яскравостей забезпечує стійкість алгоритмів розпізнавання до зміни умов освітлення.

Як міру близькості для зображень, що задовольняє описаним вище вимогам, використовується метрика Хеммінга (3) для матриць зміни яскравостей:

$$\rho_w(I, J) = \sum_{t,j} w_{t,j} [M_{t,j}^I \neq M_{t,j}^J], \quad (3)$$

де $w_{t,j} \geq 0$ - вагові коефіцієнти.

На рис 1 представлено шаблон обличчя, який використовують для детекції обличч на зображеннях, і відповідна йому матриця зміни яскравості.

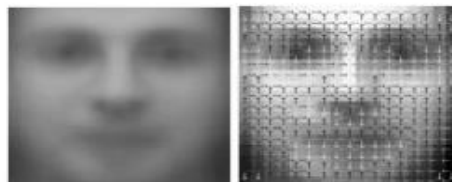


Рис. 1.1. Шаблон обличчя та візуальне представлення відповідної йому матриці зміни яскравості

Назва даної групи методів означає, що для розпізнавання виявлених облич використовується тільки колір чи яскравість пікселів на зображеннях.

Найпростішим схожим методом є порівняння, в якому мірою схожості є відстань між векторами яскравості пікселів зображень. Однак даний метод абсолютно не стійкий до змін освітленості, положення обличчя, масштабування. Більш того, такий підхід має високу обчислювальну складність і абсолютно не придатний для розпізнавання в реальному часі.

Тому, дуже часто використовуються методи, які використовують перехід векторного опису зображень в простору з меншою розмірністю, в яких порівняння набагато ефективніше [23] .

Eigenfaces - алгоритм, запропонований в 1991 році Метью Терком і Алексом Пентланд [24], який здобув широку популярність в якості першого успішного методу розпізнавання осіб. Основною ідеєю алгоритму є застосування методу головних компонент для знаходження векторів, найкращим чином описують зображення осіб. Використовуючи цей метод можна виявити різні зміни в навчальній вибірці зображень облич і описати цю зміну в базисі декількох ортогональних векторів, які називаються власними (eigenface). Обличчя, мають корисну властивість: зображення, яке відповідає кожному вектору має форму обличчя. Обчислення головних компонент зводиться до обчисленню власних векторів і власних значень коваріантної матриці, яка розраховується із зображення. Отриманий один раз на навчальній вибірці зображень облич набір власних векторів використовується для кодування всіх інших зображень осіб, які представляються зваженої комбінацією цих власних векторів.

Використовуючи обмежену кількість власних векторів можна отримати стислу апроксимацію вхідному зображенню особи, яку потім можна зберігати в базі даних у вигляді вектору коефіцієнтів, який служить одночасно ключем пошуку в базі даних осіб [24]. Сума головних компонент помножених на відповідні власні вектори являється реконструкцією зображення.

Для кожного зображення обличчя обчислюється його головні компоненти. Зазвичай це від 5 до 200 головних компонент. Процес розпізнавання ґрунтується

на порівнянні головних компонент невідомого зображення з компонентами всіх відомих зображень.

При цьому робиться припущення, що зображення облич, які відповідають одній людині, згруповані в кластери у власному просторі. З бази даних вибираються зображення-кандидати, які мають найменшу відстань від вхідного зображення.

Метод власних облич потребує для свого застосування ідеалізування умов таких як, єдині параметри освітленості, нейтральний вираз обличчя, відсутність завад в якості окулярів і бороди. При недотриманні цих умов головні компоненти не будуть відображати міжкласові варіації. Наприклад, при різних умовах освітленості метод власних облич майже не застосовується, оскільки перші головні компоненти переважно відображають зміни освітленості, і порівняння видає зображення, яке має схожий рівень освітленості. При дотриманні ідеалізованих умов точності розпізнавання з використанням методу досягає більше 90%.

Обчислення набору власних векторів відрізняється високою трудомісткістю. Один із способів – верстка зображень по стрічках та стовпчикам; в такій формі представлення зображення має на порядок менший розмір, обчислення і розпізнавання проходить швидше, але відновити початкове зображення неможливо.

Метод головних компонент добре зарекомендував себе в практичних додатках. Однак, в тих випадках, коли на зображенні особи присутні значні зміни в освітленості або виразі обличчя, ефективність методу значно падає.

Переваги методу головних компонент:

Простота реалізації PCA є досить зрозумілим, і його алгоритм можна реалізувати з використанням стандартних бібліотек для обробки даних, таких як NumPy або SciPy.

Придатність для розпізнавання в реальному часі метод дозволяє швидко обробляти зображення, що є критично важливим для застосувань, де потрібно розпізнавати обличчя в режимі реального часу, наприклад, в системах відеоспостереження чи мобільних додатках.

Компактне зберігання даних PCA дозволяє зменшити розмір даних шляхом проектування простору ознак на меншу кількість вимірів, зберігаючи при цьому основну інформацію. Це економить пам'ять і час при обробці.

Швидке розпізнавання завдяки зниженню розмірності даних, процес розпізнавання стає швидшим, оскільки алгоритми можуть працювати з меншою кількістю ознак.

Недоліки:

Але як ви зауважили, у методі головних компонент є й істотні недоліки, особливо при наявності значних змін в освітленості або виразах обличчя:

Чутливість до освітлення PCA не може адекватно обробляти ситуації з різними умовами освітлення, що може призвести до значного зниження точності розпізнавання.

Необхідність попередньої обробки щоб поліпшити ефективність, зображення потрібно попередньо нормалізувати або вирівнювати, що потребує додаткових кроків у підготовці даних.

Втрата деталей при зменшенні вимірності, важлива інформація (яка може допомогти в розпізнаванні) може бути втрачена.

Недоліки викликані тим, що найбільш важливі власні вектори в більшій мірі описують особливості освітлення, ніж характеристики облич, оскільки спочатку метод головних компонент вибирає підпростір з метою апроксимації даних, а не їх класифікації [25].

Fisherfaces - алгоритм, в якому на відміну від методу eigenfaces використовується лінійний дискримінантний аналіз, а саме лінійний дискримінант Фішера. Дія алгоритму заснована на пошуку проекції даних, при якій класи зображень облич максимально роздільні. При використанні методу головних компонент проводиться максимізація розкиду даних по всій базі облич. Ця відмінність дозволяє вирішити проблему високої чутливості до змін освітлення [26].

Метод головних компонент, який являється основою методу Eigefaces, знаходить лінійну комбінацію ознак, яка максимізує загальну дисперсію даних. Це

доволі сильний спосіб представлення даних, але він не враховує класи об'єктів, і може бути втрачено багато відмінної інформації через відкидання компонент, які не пройшли базис. Наприклад, коли в базі є фотографія з різним освітленням, і оскільки відомо, що компоненти, визначені за допомогою методу головних компонент, не завжди містять в собі всю відмінну інформацію, тому зразки різних класів змішались один з одним, і подальша класифікація стає неможливою. Цю проблему рішає алгоритм лінійного дискримінантного аналізу, який є основою методу Fisherfaces.

Лінійний дискримінантний аналіз – метод статистики машинного навчання, який застосовується для пошуку лінійних комбінацій ознак, які найкращим чином розділяють два чи більше класи об'єктів чи подій (тобто об'єкти одного і того ж класу мають шукатись, як можна ближче один до одного в просторі, і при цьому відбувається максимізація відстані між класами.

Даний алгоритм краще розпізнає при різному освітленні, чим EigenFaces, тільки навчання має відбуватись також при різних умовах освітленості.

Переваги:

Стійкість до змін освітленості, завдяки цьому використанню багатьох зображень кожної особи в різних умовах освітлення, LDA може забезпечити кращу стійкість до варіацій освітленості. Це дозволяє моделі краще вчитися на розподілі ознак різних класів.

Оптимізація для класів LDA був спроектований для максимізації відстані між класами, мінімізуючи відстань всередині класу", що робить його ефективним для класифікації. Результируючий простір ознак часто дозволяє досягти кращої сепарації класів, ніж при використанні PCA.

Зменшення розмірності як і PCA, LDA також зменшує розмірність даних, але акцентується на максимізації класової відмінності, що надає більше значення для класифікаційних задач.

Недоліки:

Відсутність інваріантності щодо афінних перетворень: LDA не є стійким до певних афінних трансформацій, таких як обертання, масштабування або зсув зображень. Це може серйозно понизити точність розпізнавання.

Вимоги до навчальних даних LDA є ефективним при різних умовах освітленості, якість навчальних даних залишається критично важливою. Відсутність представлення при різних умовах може призвести до погіршення результатів.

Локальні бінарні шаблони (надалі LBP - Local Binary Pattern) – простий і ефективний оператор перетворення зображень, вперше запропонований в 1996 році для класифікації текстур. [13] Однак, пізніше знайшов застосування і для розпізнавання осіб. [14]

Даний оператор використовує значення яскравостей околиці кожного пікселя зображення і за допомогою функції присвоює кожному пікселю значення, яке описує його околиці. Далі отримане зображення розділяється на підобласті, для кожної з яких розраховується гістограма. Гістограми конкатенуються і порівнюються за допомогою методів машинного навчання. У класичному варіанті використовується метод найближчого сусіда. [14]

Переваги цього методу полягають в простоті реалізації і високій швидкості роботи, яку можна збільшити, використовуючи різні модифікації алгоритму. При цьому алгоритм показує високі результати при розпізнаванні облич і стійкий до монотонних змін освітлення. Все це робить його оптимальним для розпізнавання облич в системах обробки в реальному часі.

1.3. Постановка завдання

1. Дослідження та аналіз існуючих наукових робіт у даній сфері:
 - Проведення огляду літератури, включаючи наукові статті, книги, конференційні тези та інші джерела, що стосуються розпізнавання емоцій обличчя.
 - Аналіз сучасних підходів та методів, що використовуються для розв'язання задачі розпізнавання емоцій через обличчя.

- Виділення основних тенденцій, проблем та досягнень у даній області.
- 2. Збір та обробка набору даних:
 - Вибір та завантаження відповідних даних для навчання та тестування моделі. Дані можуть бути отримані з відкритих джерел або створені самостійно.
 - Перевірка та очищення даних від аномалій та видалення непотрібної інформації.
 - Проведення аугментації даних для підвищення різноманітності та кількості прикладів у наборі даних.
- 3. Проектування та розробка згорткової нейронної мережі:
 - Вибір оптимальної архітектури згорткової нейронної мережі (CNN) для розпізнавання емоцій обличчя. Розгляд можливих варіантів, таких як VGG, ResNet, або власна архітектура.
 - Налаштування гіперпараметрів мережі, таких як кількість шарів, розмір фільтрів, крок зміщення, функції активації та регуляризація.
 - Розробка алгоритмів для тренування та валідації мережі на наборі даних.
- 4. Вимірювання ефективності мережі та оцінка результатів:
 - Навчання згорткової нейронної мережі на тренувальному наборі даних та валідація на тестовому наборі.
 - Вимірювання та аналіз точності, чутливості, специфічності та інших метрик ефективності моделі.
 - Порівняння результатів з іншими методами та дослідженнями у цій області.
- 5. Документування та представлення результатів:
 - Підготовка звіту з описом методології, отриманими результатами, висновками та рекомендаціями.
 - Підготовка презентації для захисту роботи перед науковою аудиторією, включаючи висвітлення ключових результатів та відповіді на запитання.

Ці кроки допоможуть систематично підходити до вирішення завдань у дослідницькій роботі з розпізнавання емоцій обличчя за допомогою згорткової нейронної мережі

Висновки до розділу 1

У першому розділі було проведено детальний аналіз предметної області, що стосується розпізнавання емоцій обличчя. Огляд сучасних моделей, методів та алгоритмів показав, що існує широкий спектр підходів для вирішення цієї проблеми, починаючи з класичних методів обробки зображень до новітніх технологій машинного навчання.

Розглянуто також методи, основані на матриці зміни яскравості, які довели свою ефективність у виявленні особливостей обличчя. Такі матриці дозволяють нам аналізувати зміни у яскравості пікселів, що є критично важливим для точного розпізнавання емоцій.

Завдання, які були поставлені, полягають у впровадженні згорткової нейронної мережі (CNN) для розпізнавання емоцій обличчя. Навчання моделі на різноманітних наборах даних дозволить досягти високої точності в розпізнаванні емоцій, що відкриває перспективи для використання цієї технології в різних сферах, таких як охорона здоров'я, безпека, і людські-комп'ютерні інтерфейси.

У результаті цього розділу було визначено основні вектори дослідження, які слугуватимуть основою для подальшої роботи над темою, що додасть значну цінність існуючим знанням у цій галузі.

РОЗДІЛ 2. МЕТОДИ ТА АЛГОРИТМИ РОЗПІЗНАННЯ ЕМОЦІЙ

2.1. Структура згорткові нейронні мережі

Вступ до згорткових нейронних мереж (CNN)

Згорткові нейронні мережі (CNN) — це спеціальні алгоритми глибокого навчання, які призначені для розпізнавання об'єктів на зображеннях, таких як класифікація, виявлення та сегментація. Вони використовуються в багатьох реальних додатках, наприклад, в автономних машинах, системах відеоспостереження і т.д.

Чому CNN важливі?

CNN мають кілька важливих переваг:

Автономне витягування ознак: Вони можуть самостійно знаходити важливі деталі в зображеннях, без потреби вручну виділяти ці деталі.

Інваріантність до позиції: CNN можуть розпізнавати об'єкти на зображеннях, незалежно від їхнього місця чи масштабу.

Попередньо навчені моделі: Існують моделі CNN, які вже навчені на великих наборах даних (наприклад, VGG, ResNet), і їх можна легко налаштувати для нових завдань, навіть якщо у вас є мало даних.

Універсальність: CNN використовуються не тільки для аналізу зображень, а й у таких сферах, як обробка текстів, аналіз часових рядів або розпізнавання мови.

CNN створювались за принципом роботи людської зорової системи:

Ієрархічна структура: У зоровій корі, як і в CNN, на перших етапах виявляються прості ознаки (наприклад, лінії), а на глибших шарах — складніші, як форми об'єктів.

Місцева зв'язність: Нейрони в нашому мозку реагують тільки на конкретні частини зображення. У CNN це схоже, бо кожен нейрон обробляє лише маленьку ділянку зображення.

Інваріантність до трансляцій: Як і зорові нейрони, CNN може розпізнавати об'єкти на зображеннях, навіть якщо вони змінюють своє місце або розмір.

Основні частини CNN

CNN складаються з кількох основних блоків:

Згорткові шари: Вони шукають конкретні ознаки на зображеннях (наприклад, краї або вигини). Для цього використовуються фільтри, які "проводять" по зображенню і виділяють важливі деталі.

ReLU (активація): Це функція, яка допомагає мережі навчатися на нелінійних залежностях. Вона важлива для того, щоб мережа могла краще розпізнавати різноманітні шаблони.

Шари пулінгу: Ці шари зменшують розміри зображень, залишаючи лише важливі ознаки. Це допомагає мережі працювати швидше та зменшує ризик перенавчання.

Повністю з'єднані шари: В кінці мережі ці шари приймають "сплющене" зображення і роблять остаточне рішення (наприклад, класифікацію).

Перенавчання і боротьба з ним у CNN

Перенавчання — це проблема, коли мережа надто добре вивчає навчальні дані і починає "погано" працювати на нових, невідомих даних. Це може трапитись через те, що модель занадто складна і вчить навіть непотрібні деталі.

Як боротися з перенавчанням?

Dropout: Це метод, коли випадковим чином вимикаються частини мережі під час навчання, що змушує решту нейронів шукати нові ознаки.

Нормалізація: Це процес, який допомагає стабілізувати навчання і зменшити перенавчання, нормалізуючи входи до мережі.

Раннє зупинення навчання: Якщо модель почала показувати погані результати на перевірочних даних, навчання зупиняється, щоб уникнути перенавчання.

Аугментація даних: Це техніка, коли ми штучно збільшуємо різноманітність даних (наприклад, повороти чи зміни розміру зображень), щоб допомогти моделі краще узагальнювати інформацію.

Ці методи дозволяють зробити CNN більш стійкими до перенавчання та забезпечити хорошу роботу на нових даних.

Згорткові нейронні мережі є потужним інструментом для розпізнавання зображень і не тільки, і завдяки їхній здатності до автоматичного навчання та адаптації, вони значно полегшують завдання комп'ютерного зору і глибокого навчання.

2.2. Метод зіставлення з еталоном (Template Matching)

У цьому підході процес розпізнавання розбивається на частини, що відповідають окремим рисам обличчя. Кожна фотографія, що надходить на вхід розпізнавальної системи, має являти собою фронтальне зображення обличчя людини з певною для конкретної бази даних кількістю масок, що представляють основні для ідентифікації регіони обличчя (наприклад, очі, ніс, рот). Обличчя (наприклад, очі, ніс, рот і нижня частина обличчя). Крім того, розташування даних масок мають бути однаково нормалізовані (наприклад, щодо положення очей) для всіх зображень у базі даних.

Під час процесу розпізнавання, коли частини вхідного зображення по черги порівнюються з частинами зображення, що зберігається в базі, використовується вектор, що відображає результат порівняння в балах (один бал за кожен рис обличчя, що збіглася) і обчислюваний шляхом нормалізованої взаємної кореляції (втім, методи порівняння можуть бути різними). Після чого вхідне зображення класифікується відповідно до максимально набраними балами. Є також деякі різновиди даного підходу, наприклад, із еталонами, що змінюються в процесі порівняння еталонами [6].

В інженерній психології вирішальне значення для розуміння закономірностей сприйняття набуло питання про механізми виділення релевантних ознак і скорочення надмірності образу ситуації. Образ ситуації включає всі елементи робочого простору, характеристики об'єкта, технічних засобів інформаційного забезпечення та управління, зовнішніх умов, образи себе та інших

учасників спільної діяльності. Для різних актуальних завдань критерії релевантності повинні розрізнятися, зміна завдань повинна відображатися у зміні набору релевантних ознак в структурі перцептивного образу. Теорії уваги як селективного процесу не забезпечують пояснення релевантності чинності неоднозначності критеріїв селекції в перцептивних діях.

Концепція еталонів (*theory of template matching*) дозволяє пояснити механізм скорочення надмірності перцептивного образу; вона припускає, що розпізнавання стимулу здійснюється в результаті порівняння викликаного ним нервового патерну з зберігається в пам'яті його ж внутрішньої репрезентацією (еталоном, шаблоном). При позитивному результаті порівняння (збігу викликаного і еталонного патернів) відбувається впізнавання і впізнання стимулу.

Схема організації ознак - еталон - формується в ході практичної діяльності з предметом. Для ситуацій, що характеризуються повторюваністю цілей і умов, схема набуває характеру зразка, фіксуючого спосіб організації складових образу і забезпечує ефективне рішення перцептивних завдань. У ситуаціях, що характеризуються відсутністю, множинністю або суперечливістю еталонів, суб'єкт стикається з необхідністю маніпулювання чином, його реорганізації в напрямку забезпечення відповідності структур образу і уявлення про мету.

Маніпулювання чином як атрибутивна характеристика перцепції припускає введення уявлень про ймовірнісної структурі еталонів і дозволяє зняти з обговорення питання про необхідність безлічі еталонів для розпізнавання безлічі змін об'єкта. Концептуальні пояснювальні моделі сучасних експериментальних досліджень когнітивних процесів реконструюють основні положення концепції еталонів для опису більш широкого кола особливостей пізнавальних процесів, використовуючи поняття "когнітивна схема".

У дослідженнях професійної діяльності концепція еталонів добре пояснює відмінність образів об'єктів у представників різних професій і високу ступінь їх подібності у професіоналів однієї професії [12].

Становлення професіонала на стадіях його навчання і професійної адаптації може бути інтерпретовано як процес формування еталонів, що забезпечують

оптимальність оцінки ситуації та ефективність відповідних їм дій. Формування навичок професійної діяльності включає "згортання" процесу аналізу зовнішньої ситуації - виділення релевантних ознак, їх збереження в системі взаємозв'язків (аналогічно формуванню рухового навичку). Перцептивний еталон забезпечує "згортання" когнітивної складової навичку, її збереження як цілісної константної одиниці та детермінанти виконавчих операцій.

Концепція еталонів може бути використана також для пояснення труднощів і помилок в нестандартних ситуаціях. Так, було показано, що помилки лікарів-рентгенологів при діагностиці рідкісних патологій пов'язані з особливостями перцептивного "еталона" рентгенограми і відповідного процесу її "читання". Невдачі або зниження ефективності діагностики відмов і порушень у роботі технічних пристроїв можуть бути пояснені з залученням еталонної концепції та мінімізовані за рахунок навчальних програм, спрямованих на формування перцептивних еталонів.

2.3. Активні моделі зовнішнього вигляду

Активні моделі зовнішнього вигляду (Active Appearance Models, AAM) - це статистичні моделі зображень, які шляхом різного роду деформацій можуть бути підігнані під реальне зображення деформацій можуть бути підігнані під реальне зображення. Цей тип моделей у двовимірному варіанті був запропонований Тімом Кутсом і Крісом Тейлором у 1998 році. Спочатку активні моделі зовнішнього вигляду застосовували для оцінювання параметрів зображень облич.

На рис. 2.1 показано результат адаптації активної моделі зовнішнього вигляду до зображення обличчя. Синя сітка показує початковий стан моделі, а червона те, що вийшло.

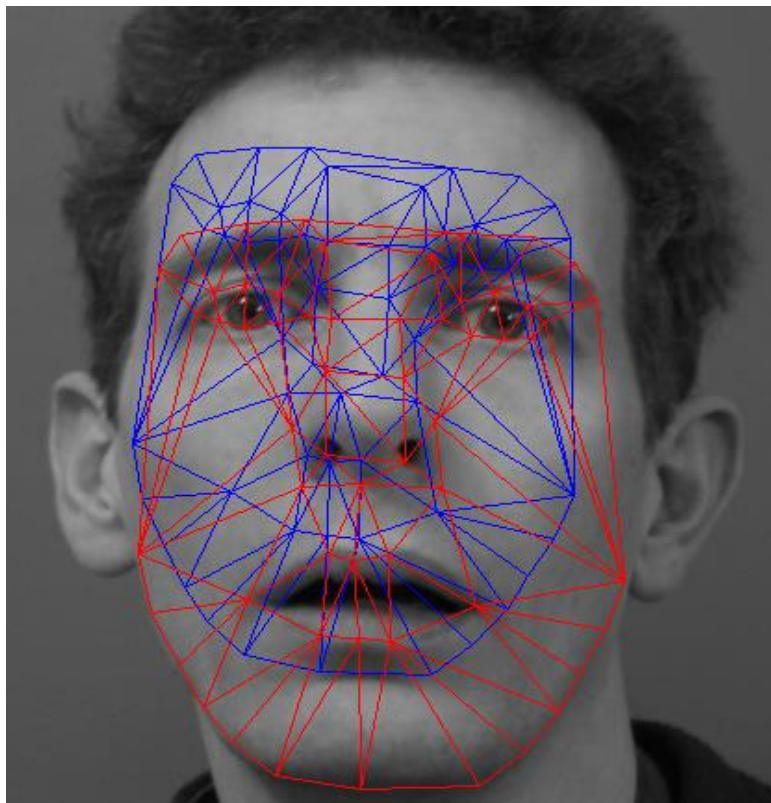


Рис. 2.1 Активна модель зовнішнього вигляду

Протягом останніх років математичний апарат активних моделей зовнішнього вигляду активно розвивався і на даний момент можна виділити 2 підходи до побудови подібних моделей: класичний (той, що був запропонований Кутесом спочатку) та на основі так званої зворотної композиції (запропонований Метьюсом та Бейкером у 2003 році [27]).

Розглянемо спочатку загальні частини двох підходів. В активних моделях зовнішнього вигляду моделюються два типи параметрів: параметри, пов'язані з формою (параметри форми), і параметри, пов'язані зі статистичною моделлю зображення або текстурою (параметри зовнішнього вигляду). Перед використанням модель має бути навчена на безлічі заздалегідь розмічених зображень. Розмітка зображень проводиться вручну або в напівавтоматичному режимі, коли за допомогою будь-якого алгоритму знаходяться наближені розташування міток, а потім вони уточнюються експертом. Кожна мітка має свій номер і визначає характерну точку, яку повинна буде знаходити модель під час адаптації до нового зображення. Приклад подібної розмітки (база облич XM2VTS) показано на малюнку нижче.

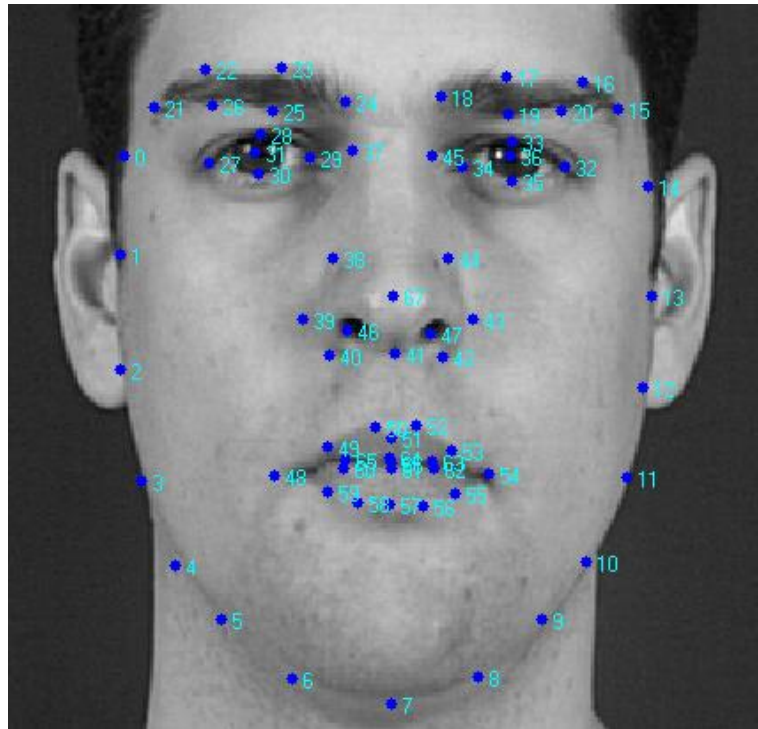


Рис.2.2. Модель активного зовнішнього вигляду з 68 мітками

У представленому прикладі на рис.3 позначено 68 міток, що утворюють форму моделі активного зовнішнього вигляду. Ця форма позначає зовнішній контур обличчя, контури рота, очей, носа, брів. Такий характер розмітки дає змогу надалі визначити різні параметри обличчя за його зображенням, які можуть бути використані для подальшої обробки іншими алгоритмами. Наприклад, це можуть бути алгоритми ідентифікації особистості, аудіо-візуального розпізнавання мови, визначення емоційного стану суб'єкта.

Процедура навчання активних моделей зовнішнього вигляду починається з нормалізації положення всіх форм для того, щоб компенсувати відмінності в масштабі, нахилі та зміщенні. Для цього використовується так званий узагальнений Прокрустов аналіз. Тут ми не наводитимемо його докладний опис, а читач, який цікавиться, може ознайомитися з відповідною статтею у Вікіпедії. Ось який вигляд має множина міток до і після нормалізації [28].

Активна модель зовнішнього вигляду містить два типи параметрів: параметри, пов'язані з формою (параметри форми), і параметри, пов'язані зі статистичною моделлю пікселів зображення або текстурою (параметри

зовнішнього вигляду). Перед використанням модель має бути навчена на безлічі заздалегідь розмічених зображень.

Розмітка зображень проводиться вручну. Кожна мітка має свій номер і визначає характерну точку, яку повинна буде знаходити модель під час адаптації до нового зображення.

Процедура навчання ААМ починається з нормалізації форм на розмічених зображеннях з метою компенсації відмінностей у масштабі, нахилі та зміщенні. Для цього використовується так званий узагальнений Прокрустовий аналіз.

З усієї множини нормованих точок потім виділяють головні компоненти з використанням методу РСА.

Далі з пікселів усередині трикутників, утворених точками форми, формується матриця, така що, кожен її стовпець містить значення пікселів відповідної текстури. Варто зазначити, що використовувані для навчання текстури можуть бути як одноканальними (градації сірого), так і багатоканальними (наприклад, простір кольорів RGB або інше). У випадку багатоканальних текстур вектори пікселів формуються окремо по кожному з каналів, а потім виконується їх конкатенація. Після знаходження головних компонент матриці текстур модель ААМ вважається навченою.

Ефективність методу: до 88%.

2.4. Метод на основі ключових точок

Метод на основі ключових точок є потужним інструментом для розпізнавання емоцій обличчя, а також для аналізу форми та текстури обличчя. Цей алгоритм можна поділити на два основних етапи: отримання координат ключових точок та класифікація емоцій на їхній основі.

Етап 1: Отримання координат ключових точок

Вхідними даними для методики є набір зображень обличчя, при цьому кожному зображенню відповідає файл розмітки, що містить координати ключових точок. Ці ключові точки представляють важливі ознаки обличчя, такі як кути губ,

очі, ніс та інші значущі деталі, які можуть змінюватися в залежності від емоційного стану особи.

Формування статистичних моделей

На основі отриманих координат ключових точок будуються дві статистичні моделі:

Модель форми — параметрична лінійна модель, яка описує можливі варіанти положення ключових точок. Формою називається вектор координат ключових точок, який відображає структуру обличчя (6).

$$s = (x_1, y_1, x_2, y_2, \dots, x_n, y_n)^T. \quad (6)$$

Залучаючи цю модель, можливо виявити загальні шаблони в статевих та індивідуальних рисах обличчя, а також проаналізувати, як ключові точки змінюються під час прояву різних емоцій.

Модель текстури схожа на модель форми, але її мета — описати варіації інтенсивності пікселів, що забезпечують текстурну інформацію про обличчя. Текстурою називається вектор усіх пікселів усередині зовнішнього контуру форми (7).

$$s = (t, t_2, \dots, t_n)^T. \quad (7)$$

Вона важлива для виявлення змін у кольорі шкіри, відтінках, а також для детекції емоційних відгуків, які можуть бути пов'язані зі зміною текстури обличчя (наприклад, почервоніння при збудженні чи збліднення при страху).

Етап 2: Кусочно-афінне перетворення

Оскільки на різних зображеннях кількість пікселів усередині зовнішнього контуру форми може відрізнятися, перед створенням моделі текстури всі зображення обличчя приводяться до єдиної усередненої форми. Цей процес здійснюється через кусочно-афінне перетворення:

Триангулювання точок: Першим кроком є триангулювання набору ключових точок, щоб поділити форму обличчя на трикутники. Це дозволяє зберегти геометричну інформацію, а також структурну цілісність зображення.

Афінне перетворення: Кожен симплекс (трикутник), отриманий на першому етапі, трансформується в нові координати за допомогою звичайного афінного

перетворення. Це забезпечує сплющене представлення обличчя, при якому всі картинки з різними розмірами та співвідношеннями форм можуть бути приведені до стандартного формату для подальшого аналізу.

Класифікація емоцій

На основі отриманих моделей (форм та текстури) алгоритм може здійснювати класифікацію емоцій. Спочатку ключові точки та текстурні ознаки використовуються для формування набору рис, які потім вводяться в класифікатор (можливо, із застосуванням методів машинного навчання, таких як SVM, нейронні мережі тощо).

Ефективність методу

Метод на основі ключових точок демонструє вражаючу ефективність у розпізнаванні емоцій. Це високий показник, який свідчить про надійність і точність використаної методології. Висока ефективність досягається завдяки використанню статистичних моделей, які адаптуються до варіацій у формах і текстурах обличчя, а також через оптимізацію даних, завдяки чому метод фasetно враховує зміни емоцій.

Ефективність методу - 89,4%.

2.5. Метод на основі інформації про текстуру

Існує можливість визначення (вручну або комп'ютерними засобами) виразу обличчя людини навіть за відсутності інформації про положення і форму його ключових елементів.

Інакше кажучи, передбачається, що емоції на обличчі кодуються не тільки тільки положенням ключових точок, а й іншими ознаками. Перевірити це припущення, втім, не складає труднощів. Для цього достатньо перетворити обличчя, що виражають емоції, до усередненої форми, а потім спробувати визначити початковий вираз. На малюнку 4 зображено низку осіб, до яких було застосовано таке перетворення.

Якщо придивитися, то можна помітити, що всі ключові точки брів, очей, носа, губ, а також підборіддя перебувають в ідентичному положенні. Ці зображення отримано шляхом тріангуляції за ключовими точках і застосування кусочно-афінного перетворення до знімків облич, що виражають емоції. Незважаючи на вирівнювання положення ключових елементів обличчя, емоції на ньому все ще легко вгадуються за зморшками, тінями та іншими елементами текстури, що повністю підтверджує гіпотезу.

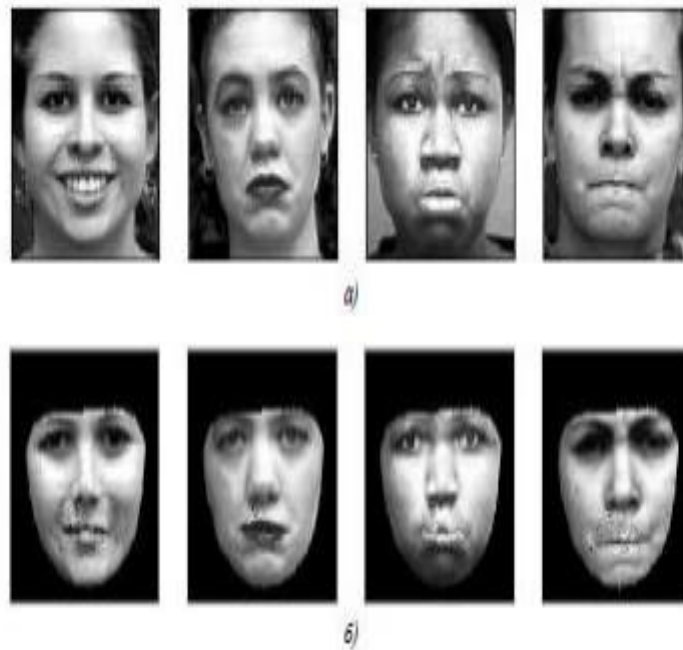


Рис. 2.3. - Приклади приведення до середньої форми:

a - вихідні зображення облич; *б* - відповідні зображення, приведені до середньої форми

Як і в методі на основі положення ключових точок, перетворення до усередненої форми дає змогу сформувавши для кожного зображення відповідний вектор текстури (8):

$$t = (t_1, t_2, \dots, t_n)^T, \quad (8)$$

де n - кількість пікселів усередині зовнішнього контуру форми.

Цей вектор можна безпосередньо використовувати як вектор ознак під час класифікації. Як класифікатор можна застосовувати SVM.

У підсумку отримуємо такий алгоритм.

На етапі навчання:

1. зображення піддаються попередньому обробленню, тобто за допомогою моделей активного образу на них визначають положення ключових точок, а потім кусочно-афінним перетворенням обличчя на зображеннях приводиться до усередненої форми.

2. Пікселі отриманих зображень разом із відповідними мітками емоцій використовуються для навчання методом SVM.

На етапі застосування:

1. Перетворення зображень класифікується за допомогою раніше навченої моделі;

2. Перетворені зображення класифікують за допомогою навченої раніше моделі.

Ефективність методу: 86,3%.

2.6. Локальні бінарні шаблони

Локальні бінарні шаблони являють собою бінарні коди певної розрядності, що використовуються для класифікації в комп'ютерному зорі. Локальні бінарні шаблони вперше були запропоновані в 1996 р. Для аналізу текстури напівтонових зображень. При цьому подальші дослідження показали, що ЛБШ інваріантні до змін в умовах освітлення і поворотів зображення.

Локальний бінарний шаблон (ЛБШ) являє собою опис околиці пікселя зображення в двійковому поданні. Для обчислення базового ЛБШ у деякій точці зображення використовується вісім пікселів її околиці, а значення інтенсивності центрального пікселя приймається за поріг. Пікселі зі значенням інтенсивності, більшим або рівним пороговому значенню, приймають значення, що дорівнюють одиниці, інші приймають значення, що дорівнюють нулю. Таким чином, результатом операції є восьмирозрядний бінарний код, який описує околицю цього пікселя.

Для більш гнучкого аналізу текстурних особливостей зображення використовуються кругова околиця та білінійна інтерполяція значень значень

інтенсивностей пікселів, які дають змогу побудувати розширений ЛБШ з довільним числом точок P і радіусом R .

Гістограма ЛБШ. Локальний бінарний шаблон обчислюється для кожного пікселя зображення, а потім будується гістограма, в якій кожному рівномірному коду ЛБШ відповідає окремий стовпець. Також формується ще один додатковий стовпець, який містить інформацію про всі нерівномірні шаблони.

Зображення обличч можуть бути представлені як набір всіляких локальних особливостей, які добре описуються за допомогою ЛБШ. Однак гістограма, побудована для всього зображення в цілому, кодує лише наявність тих чи інших локальних особливостей, але при цьому не містить жодної інформації про їхнє розташування на зображенні. Для врахування просторового розташування зображення розбивається на підобласті, у кожній з яких обчислюється своя гістограма ЛБШ (рис. 2.4.). Шляхом конкатенації цих гістограм будується загальна гістограма, що враховує як локальні, так і глобальні особливості зображення.

За такого підходу для кращого вилучення ознак виконується варіювання параметрів оператора ЛБШ і числа розбиття зображення на подобласти [8].

Ефективність методу - залежно від підходу до обчислення міри відмінності гістограм ЛБШ ефективність може досягати 86%.

У 1998 році дослідники Y. LeCun, L. Bottou, Y. Bengio і P. Haffner запропонували вид нейронних мереж, що працюють за принципом зорової системи людини, які були названі згортковими нейронними мережами (ЗНМ).



Рис. 2.4. Розбиття зображення обличчя на подобласти

2.7. Розпізнавання облич за допомогою згорткових нейронних мереж

ЗНМ являє собою особливий клас багатоварового персептрона, який має двовимірну структуру і добре підходить для оброблення зображень з високим ступенем інваріантності до зміщення, поворотів, масштабування та інших спотворень вхідних даних.

Структура ЗНМ являє собою послідовність із двох типів шарів: згорткові та підвибіркові. Кожен шар складається з набору площин (карт характеристик), які своєю чергою складаються з нейронів.

Кожен нейрон згорткового шару має зв'язок із невеликою групою нейронів попереднього шару (локальне рецептивне поле). Локальні рецептивні поля нейронів згорткового шару частково накладаються один на одного за принципом черепиці. Значення нейронів із локального рецептивного поля множаться на матрицю синаптичних коефіцієнтів, а результат записується у відповідний нейрон згорткового шару (рис. 2.5.).

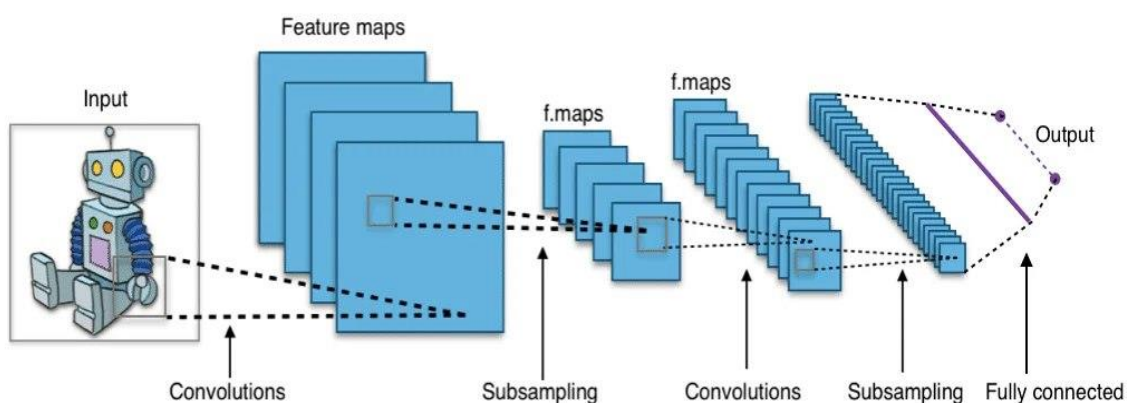


Рис. 2.5. Функціонування шарів у ЗНМ

Слідом за згортковим шаром розташовується підвибірковий шар, який забезпечує часткову інваріантність нейронної мережі до зміни масштабу вхідного зображення. Кількість площин підвибіркового шару зазвичай така сама, як і в попередньому шарі. В підвибірковому шарі локальні рецептивні поля не перетинаються одне з одним і мають фіксований розмір одне з одним і мають фіксований розмір 2×2 нейрона. Кожен нейрон цього шару обчислює середнє значення своїх чотирьох входів, множить їх на синаптичні коефіцієнти й отриманий результат передає через функцію активації. Таким чином, підвибірковий шар зменшує розмірність площин попереднього шару вдвічі.

Послідовно чергуючись одна за одною, розміри площин зменшуються, але їхня кількість збільшується. Чергування шарів дає змогу формувати різні карти характеристик, що наділяє ЗНМ здатністю до ідентифікації складніших ієрархічних ознак. Поступово під час проходження кількох шарів карта ознак вироджується у вектора. Останні кілька шарів ЗНМ являють собою класичний перцептрон, складаються зі звичайних нейронів і призначені для класифікації виділених ознак.

Важливою властивістю ЗНМ є використання загальної матриці вагових коефіцієнтів для всіх нейронів у межі площини. Цей підхід дає змогу використовувати меншу кількість параметрів, що налаштовуються (вагових коефіцієнтів) за великої кількості зв'язків і підвищити швидкість процесу навчання. Таким чином, площини ЗНМ являють собою фільтри, кожен з яких здійснює пошук індивідуальних характерних ознак вхідного зображення. Це дає змогу згортковій нейронній мережі запам'ятовувати взаємозв'язок просторово залежних областей зображення. Характерні ознаки, що витягуються тією чи іншою площиною, визначаються в процесі навчання.

Якщо вхідне зображення спотворене або зміщене, то на виході площини буде аналогічно зміщений результат. Завдяки цій властивості забезпечується стійкість ЗНМ до спотворень вхідних даних.

Недолік використання ЗНМ полягає в складності налаштування оптимальних параметрів: кількість шарів, площин, нейронів, розмір рецептивного поля тощо.

Переваги використання ЗНМ для детектування об'єктів на зображеннях:

- структура ЗНМ добре підходить для обробки двовимірних даних.
- рецептивні поля, що нашаровуються одне на одне, забезпечують взаємозв'язок просторово залежних областей зображення.
- підвищена стійкість до афінних і проєкційних спотворень вхідних даних, шумів, зміни масштабу. Завдяки своїм перевагам на сьогодні СНС активно використовують для виявлення облич та інших об'єктів на зображеннях і відео послідовностях.

Результати: відсоток розпізнавання за найкращого результату - 97,4%.

Висновки до розділу 2

У другому розділі були розглянуті ключові методи та алгоритми розпізнавання емоцій, що ґрунтуються на сучасних техніках комп'ютерного зору. Розпочавши з основної структури згорткових нейронних мереж (CNN), було показано, як їх архітектура дозволяє ефективно виявляти та класифікувати емоції на основі вхідних зображень. CNN забезпечують значну перевагу в обробці візуальної інформації завдяки своїй здатності автоматично вчитися на ознаках.

Метод зіставлення з еталоном (Template Matching) надав базову довідку про традиційні підходи, які, попри свою простоту, можуть бути менш ефективними в умовах змінних освітлювальних умов та виразів осіб. Наступний розділ про активні моделі зовнішнього вигляду охоплює динамічні моделі, що дозволяють представити варіації форм обличчя, підвищуючи точність розпізнавання.

Досліджені методи на основі ключових точок і інформації про текстуру також показали свою ефективність у виявленні емоцій, особливо в комбінації з локальними бінарними шаблонами (LBP), які забезпечують надійний інструмент для кодування текстурних ознак обличчя.

Особливу увагу було приділено розпізнаванню облич за допомогою згорткових нейронних мереж, де з використанням глибоких навчальних структур досягається висока точність і стабільність у класифікації емоцій. Результати

досліджень вказують на те, що комбінація різних методів може суттєво покращити результативність систем розпізнавання емоцій.

Таким чином, у цьому розділі виявлено широкий спектр методів, які можуть бути інтегровані для підвищення ефективності розпізнавання емоцій обличчя. Подальші етапи дослідження мають на меті не лише розробку комплексної моделі, але й експериментальну перевірку її працездатності в реальних умовах.

РОЗДІЛ 3. ОПИС СИСТЕМИ РОЗПІЗНАННЯ ЕМОЦІЙ

3.1 Розробка алгоритмів та проектування системи розпізнавання емоцій

Задача полягає у розробці моделі, яка здатна класифікувати емоції обличчя на основі зображень. Основні емоції, які зазвичай розпізнаються, включають:

- Гнів
- Відраза
- Страх
- Радість
- Сум
- Подив
- Відсутність емоцій

Для розпізнавання емоцій зображень обличчя зазвичай використовують згорткові нейронні мережі (CNN). Основні компоненти архітектури CNN: вхідний шар, згорткові шари, пулінг, dense шар та вихідний шар.

Вхідний шар відповідає зображенню обличчя, яке може мати різні розміри. Зазвичай зображення масштабуються до стандартних розмірів.

Згорткові шари відповідають за виявлення ознак (фільтрація). Зазвичай використовуються кілька згорткових шарів з різними фільтрами, що дозволяє моделі вловлювати різноманітні деталі, такі як контури, текстури та специфічні ознаки обличчя.

Пулінг шари (наприклад, макс-пулінг) використовуються для зменшення розмірності вхідних даних і зменшення обчислювальних витрат, а також для підвищення стійкості моделі до незначних зсувів і спотворень зображень.

Dense шари: йдуть після згорткових та пулінгів і відповідають за остаточну класифікацію.

Вихідний шар використовує активаційну функцію (наприклад, softmax) для визначення ймовірностей кожної з емоцій.

Попередня обробка даних містить в собі їх збір, аугментацію та нормалізацію. Для навчання моделі використовують набори даних, які містять зображення облич з відповідними мітками емоцій. Щоб збільшити обсяг навчальних даних і поліпшити стійкість моделі, застосовують техніки аугментації, такі як обертання, віддзеркалення, масштабування та зміна яскравості. Піксельні значення зображень зазвичай нормалізують, щоб покращити швидкість та якість навчання.

Навчання моделі складається з наступних етапів:

Розділення даних. Набір даних розділяється на тренувальний, валідаційний та тестовий набори, що дозволяє контролювати якість моделі.

Вибір оптимізатора: Використовують оптимізатори, такі як Adam або SGD, для налаштування ваг моделі.

Функція втрат: Зазвичай використовується крос-ентропійна втрата для багатокласової класифікації.

Епохи та пакетний розмір: Модель навчається протягом декількох епох з певним розміром пакетів, що впливає на швидкість та якість навчання.

В результаті навчання отримується модель для якої проводять операцію оцінку її ефективності. Це відбувається з використанням наступних засобів:

Метрики: Для оцінки продуктивності моделі використовують такі метрики, як точність, відчутність, специфічність, F1-міра.

Матриця плутанини: Дозволяє візуалізувати, які емоції класифікуються правильно, а які — ні.

Тестування на нових даних: Для перевірки загальної здатності моделі до розпізнавання емоцій на нових, раніше не бачених зображеннях.

3.2. Інструментальні засоби реалізації програмного продукту та вимоги до апаратних засобів.

Раніше було розглянуто методи виявлення та розпізнавання об'єктів. За підсумком необхідно розробити програму, що дає змогу виявляти обличчя і розпізнавати емоції.

Було визначено такі завдання:

- програма повинна виявляти обличчя і розпізнавати емоції в реальному часі;
- як емоції для розпізнавання визначено такі базові емоції: нейтрально, усмішка, смуток, здивування відраза, злість;
- для виконання завдання розпізнавання емоцій програма повинна мати два режими: режим навчання і режим тестування (розпізнавання емоцій).

З урахуванням перелічених завдань було визначено такі інструменти:

- мова програмування Python;
- метод Віолі-Джонса як метод виявлення обличч;
- метод головних компонентів;
- використання багат шарового персептрона (MLP) для розпізнавання емоцій на виявленому обличчі;
- бібліотека OpenCV, що містить у собі вище перераховані алгоритми.

Але про все по порядку.

Мова Python є, мабуть, найпростішою у вивченні та найприємнішою у використанні з мов програмування, які приємною у використанні з мов програмування, що набули широке поширення. Програмний код мовою Python легко читати і писати, і, будучи лаконічним, він не виглядає загадковим. Python – дуже виразна мова, що дає змогу вмістити додаток у меншу кількість рядків, ніж на це знадобилося б в інших мовах, таких як C++ або Java.

Python є крос-платформною мовою: зазвичай одна й та сама програма мовою Python може запускатися і в Windows, і в UNIX-подібних системах, таких як Linux, BSD і Mac OS, для чого достатньо просто скопіювати файл або файли, що складають програму, на потрібний комп'ютер; при цьому навіть не буде потрібно

виконувати «збірку», або компілювання програми. Звичайно, можна написати мовою Python програму, яка буде використовувати деякі характерні особливості конкретної операційної системи, але така необхідність виникає вкрай рідко, оскільки практично вся стандартна бібліотека мови Python і більшість бібліотек сторонніх виробників забезпечують повну крос-платформеність.

Однією з основних переваг мови Python є наявність повної стандартної бібліотеки, що дає змогу забезпечити завантаження файлу з Інтернету, розпакування архіву або створення веб-сервера за допомогою написання кількох рядків програмного коду. На додаток до неї існують тисячі додаткових бібліотек сторонніх виробників, серед яких одні забезпечують складніші та потужніші можливості, ніж стандартна - наприклад, бібліотека для організації мережових взаємодій Twisted і бібліотека для вирішення обчислювальних задач NumPy; а інші надають функціональність, яка занадто вузькоспеціалізована, щоб її можна було включити в стандартну бібліотеку - наприклад, пакет моделювання SimPy.

Python може використовуватися для програмування в процедурному, в об'єктно-орієнтованому і, меншою мірою, у функціональному стилі програмування, хоча в глибині душі Python - об'єктно-орієнтована мова програмування

TensorFlow — це відкритий фреймворк для машинного навчання, розроблений компанією Google. Він надає потужні інструменти для побудови та навчання нейронних мереж, що робить його одним з найпопулярніших і найживаніших фреймворків у сфері глибинного навчання.

Основні особливості TensorFlow

TensorFlow підтримує різноманітні платформи (від мобільних пристроїв до розподілених обчислень у хмарі), що дозволяє легко розгортати моделі на різних пристроях.

TensorFlow пропонує високоабстрактний інтерфейс, такий як Keras, що полегшує створення та налаштування моделей. Keras включено до TensorFlow, що робить його ще більш доступним для розробників.

TensorFlow має вбудовану підтримку для обробки на графічних процесорах (GPU) та тензорних процесорах (TPU), що пришвидшує навчання великих моделей.

TensorFlow підтримує як статичні графи (TensorFlow 1.x), так і динамічні (TensorFlow 2.x), що дозволяє розробникам обирати найбільш зручний для них підхід. У TensorFlow 2.x введено концепцію "eager execution", яка спрощує розробку, дозволяючи виконувати операції негайно.

Розробники можуть легко створювати та інтегрувати свої власні функції, а також використовувати попередньо навчені моделі з TensorFlow Hub.

Keras — це високорівнева бібліотека для машинного навчання, що дозволяє легко створювати та тренувати нейронні мережі. Спочатку Keras була незалежним проектом, але з 2017 року вона стала частиною TensorFlow, що забезпечило їй ще більшу популярність і інтеграцію з іншими компонентами екосистеми TensorFlow.

Основні характеристики Keras

Keras має інтуїтивно зрозумілий API, що дозволяє швидко створювати моделі нейронних мереж. Це робить її ідеальним вибором для новачків, які тільки починають знайомитися з глибинним навчанням.

Keras базується на концепції модульності. Розробники можуть легко комбінувати різні шари, оптимізатори та функції втрат для створення кастомізованих моделей.

Keras підтримує різноманітні типи нейронних мереж, включаючи повнозв'язкові, згорткові, рекурентні та багато інших. Це дозволяє розробникам створювати моделі для різних застосувань, таких як обробка зображень, обробка тексту та часові ряди.

Хоча Keras найбільш часто використовується з TensorFlow, вона також може працювати з іншими фреймворками, такими як Theano і Microsoft Cognitive Toolkit (CNTK).

Keras має вбудовані функції для навчання, валідації та оцінки моделей, що значно спрощує ці процеси.

Keras легко інтегрується з TensorBoard для візуалізації процесу навчання та моніторингу метрик моделі.

Навчання моделей нейронних мереж відбувалося за допомогою вебсервісу Kaggle з використанням прискорювача Nvidia Tesla P100.

Бібліотека OpenCV

OpenCV - бібліотека комп'ютерного зору з відкритим вихідним кодом. Бібліотека написана на C і C++ і працює на комп'ютерах під управлінням Linux, Windows, Mac OS X. Так само активно розвиваються інтерфейси бібліотеки для Python, Ruby, Matlab та інших мов програмування.

Бібліотека OpenCV була розроблена з метою підвищення обчислювальної ефективності та з ухилом на додатки реального часу. OpenCV написана з використанням оптимізованого C і може використовувати багатоядерні процесори. Надалі, якщо можливість купівлі бібліотеки IPP (Integrated Performance Primitives), яка складається з процедур із низькорівневою оптимізацією для різних алгоритмічних областей. OpenCV буде автоматично використовувати бібліотеку IPP під час виконання програми.

Однією з основних цілей OpenCV є надання простого у використанні інтерфейсу, який дасть змогу людям досить таки швидко будувати складні додатки, що використовують комп'ютерний зір. Бібліотека OpenCV містить понад 500 функцій, які охоплюють багато галузей комп'ютерного зору, такі як: інспекція фабричної продукції, медицина, безпека, користувальницький інтерфейс, калібрування камери, стереозріз і робототехніка. І все це завдяки тому, що комп'ютерний зір і машинне навчання часто йдуть «рука об руку», до того ж OpenCV повністю включає в себе бібліотеку загального призначення MLL (Machine Learning Library). MLL бібліотека орієнтована на розпізнавання статичних образів і кластеризацію. MLL дуже корисна для завдань комп'ютерного зору, які складають основу OpenCV, але вона досить-таки узагальнена, щоб вирішувати конкретні проблеми машинного навчання

3.3 Розробка системи для аналізу методів для розпізнання емоцій

Для навчання та валідації було обрано набір даних на вебсервісі Kaggle. Дані складаються з зображень розміром 48x48 пікселів. Обличчя були автоматично відцентровані таким чином, щоб вони були більш-менш в центрі та займали приблизно однакову кількість простору на кожному зображенні. Завдання полягає в категоризації кожного обличчя на основі емоції, яка відображається в виразі обличчя, у одну з семи категорій (0=Гнів, 1=Відраза, 2=Страх, 3=Щастя, 4=Сум, 5=Подив, 6=Нейтральна).

Файл **train.csv** містить два стовпці: "emotion" та "pixels". Стовпець "emotion" містить числовий код від 0 до 6, що відповідає емоції, присутній на зображенні. Стовпець "pixels" містить рядок, обгорнутий в лапки, для кожного зображення. Вміст цього рядка — це значення пікселів, розділені пробілами, у порядку рядків. Файл **test.csv** містить лише стовпець "pixels", і ваше завдання — передбачити стовпець "emotion".

Навчальний набір складається з 28,709 прикладів. Публічний тестовий набір складається з 3,589 прикладів.

Перед початком навчання моделі дані зчитувалися з відповідного csv файлу та розбивалися за призначенням у 3 категорії: навчання, валідація та тестування.

```
data = pd.read_csv('/kaggle/input/challenges-in-representation-learning-facial-expression-recognition-challenge/icml_face_data.csv')  
train_data = data[data[' Usage'] == 'Training']  
val_data = data[data[' Usage'] == 'PublicTest']  
test_data = data[data[' Usage'] == 'PrivateTest']
```

Дані зображень зберігаються у вигляді одновимірного масиву, тому була впровадження функція для їх конвертацію у двовимірний масив (48 на 48 пікселів з 1 каналом)

```
def prepare_data(data):
```

```

        return data['pixels'].apply(lambda x:
np.array(x.split(' ')).reshape(48, 48,
1).astype('float32'))

```

Також була впроваджена функція для обробки вхідних даних та аугментації з метою зменшення фактору перенавчання.

```

def preprocess_image(image):
    image = tf.image.random_flip_left_right(image)
    image = image / 255.0
    return image

```

В результаті використання створених функцій були створенні 3 датасети:

```

X_train_dataset =
tf.data.Dataset.from_tensor_slices(X_train.tolist())
X_train_dataset = X_train_dataset.map(preprocess_image)

X_val_dataset =
tf.data.Dataset.from_tensor_slices(X_val.tolist())
X_val_dataset = X_val_dataset.map(preprocess_image)

X_test_dataset =
tf.data.Dataset.from_tensor_slices(X_test.tolist())
X_test_dataset = X_test_dataset.map(preprocess_image)

```

Ці датасети в подальшому використовувалися для створення конвеєрів даних:

```

def prepare_dataset(dataset, shuffle = True):
    if shuffle:
        dataset = dataset.shuffle(buffer_size = 1024)
    dataset = dataset.batch(128, drop_remainder = False)

```

```

        return dataset.prefetch(4)

    y_train_dataset =
tf.data.Dataset.from_tensor_slices(y_train)
    y_val_dataset =
tf.data.Dataset.from_tensor_slices(y_val)
    y_test_dataset =
tf.data.Dataset.from_tensor_slices(y_test)

    # об'єднання зображень та відповідних емоцій у спільний
дасет
    train_dataset = tf.data.Dataset.zip((X_train_dataset,
y_train_dataset))
    train_dataset = prepare_dataset(train_dataset)

    val_dataset = tf.data.Dataset.zip((X_val_dataset,
y_val_dataset))
    val_dataset = prepare_dataset(val_dataset)

    test_dataset = tf.data.Dataset.zip((X_test_dataset,
y_test_dataset))
    test_dataset = prepare_dataset(test_dataset, shuffle =
False)

```

Для змінення контролю швидкості навчання моделі впровадж навчання було впроваджено відповідну функцію:

```

def lr_scheduler(epoch, lr):
    return lr - 1e-5

```

На основі проведеного аналізу актуальних досліджень задачі розпізнавання емоцій було створено різні варіації глибоких нейронних мереж. Вони створювалися з використанням засобів фреймворку TensorFlow

```
model = tf.keras.models.Sequential([
    tf.keras.layers.Conv2D(64, (3, 3),
activation='relu', input_shape=(48, 48, 1)),
    tf.keras.layers.BatchNormalization(),
    tf.keras.layers.MaxPooling2D(pool_size=(2, 2)),
    tf.keras.layers.Dropout(0.2),

    tf.keras.layers.Conv2D(128, (3, 3),
activation='relu'),
    tf.keras.layers.BatchNormalization(),
    tf.keras.layers.MaxPooling2D(pool_size=(2, 2)),
    tf.keras.layers.Dropout(0.2),

    tf.keras.layers.Conv2D(256, (3, 3),
activation='relu'),
    tf.keras.layers.BatchNormalization(),
    tf.keras.layers.MaxPooling2D(pool_size=(2, 2)),
    tf.keras.layers.Dropout(0.2),

    tf.keras.layers.Flatten(),
    tf.keras.layers.Dense(512, activation='relu'),
    tf.keras.layers.BatchNormalization(),
    tf.keras.layers.Dense(7, activation='softmax')
])
```

Навчання та компіляція моделі відбувалася за допомогою наступних функцій

```
model.compile(optimizer = 'adam', loss =
'sparse_categorical_crossentropy', metrics=['accuracy'])
```

```

history = model.fit(train_dataset, epochs = 100,
validation_data = val_dataset,
callbacks=[tf.keras.callbacks.LearningRateScheduler(lr_scheduler)])

```

Для аналізу ефективності створених глибоких нейронних мереж використовувалися матриці плутанини та метод порівняння метрик створених моделей. Ці засоби були реалізовані наступним чином:

```

test_preds = model.predict(test_dataset).argmax(axis=1)

emotions = {0: 'Angry', 1: 'Disgust', 2: 'Fear', 3: 'Happy', 4: 'Sad', 5: 'Surprise', 6: 'Neutral'}
conf_mat = confusion_matrix(y_test, test_preds)
fig, ax = plot_confusion_matrix(conf_mat = conf_mat,
                                show_normed = True,
                                show_absolute = False,
                                class_names = emotions.values(),
                                figsize=(8, 8))

report = classification_report(y_test, test_preds)
print(report)

```

3.4. Порівняння ефективності методів розпізнавання емоцій з використанням створеної системи

На основі проведеного порівняльного аналізу результатів нейронних мереж з різними архітектурами глибокої нейронної мережі

Класифікатор на її основі розпізнавав емоції з наступною точністю:

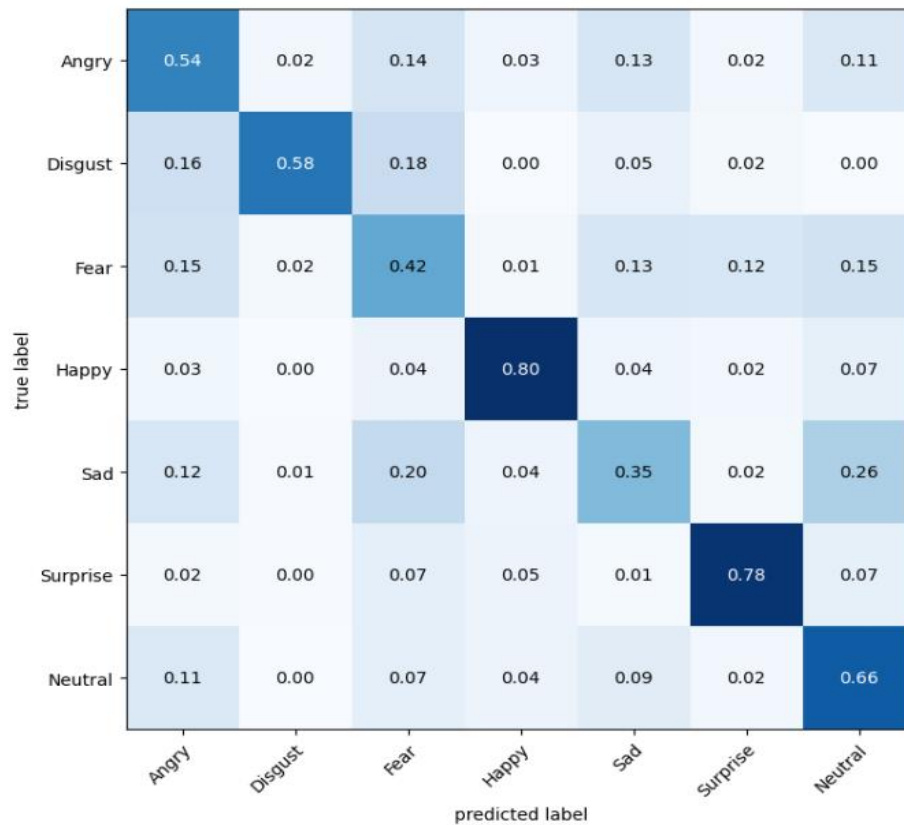


Рис. 3.1

Для покращення ефективності були застосовані техніки агментації даних.

Застосування масштабування та відзеркалення зображень дозволило отримати наступні результати:

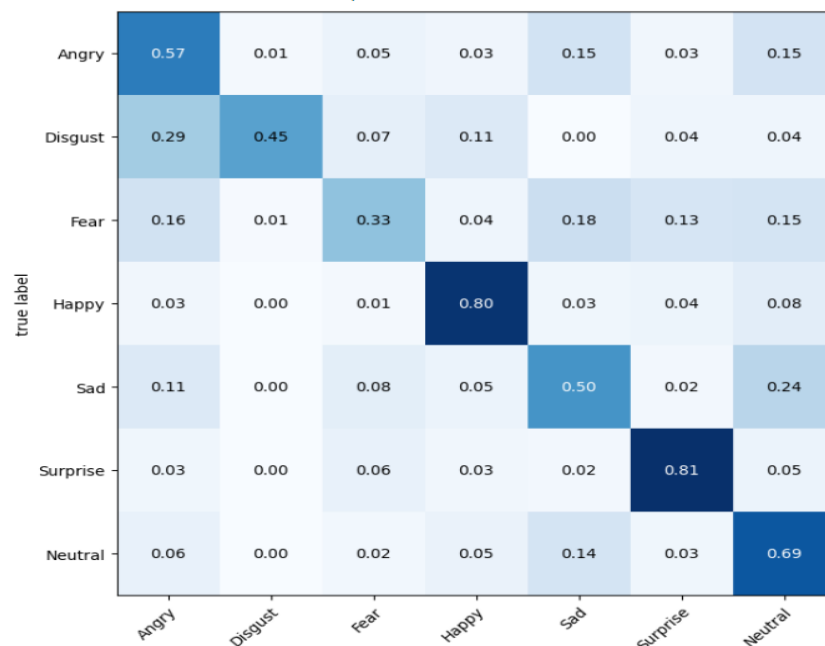


Рис. 3.2

Також був проведений аналіз функцій оптимізацій, які застосовувалися у подібному дослідженні й експериментальним методом було визначено, що функція оптимізації adam дає забезпечує кращі результати (особливо якщо зменшувати параметр learning rate у процесі навчання). Також було визначено, що найбільш оптимальною кількістю епох є 100. (кращі результати порівняно з меншою кількістю епох і менш виражене перенавчання порівняно з більшою кількістю).

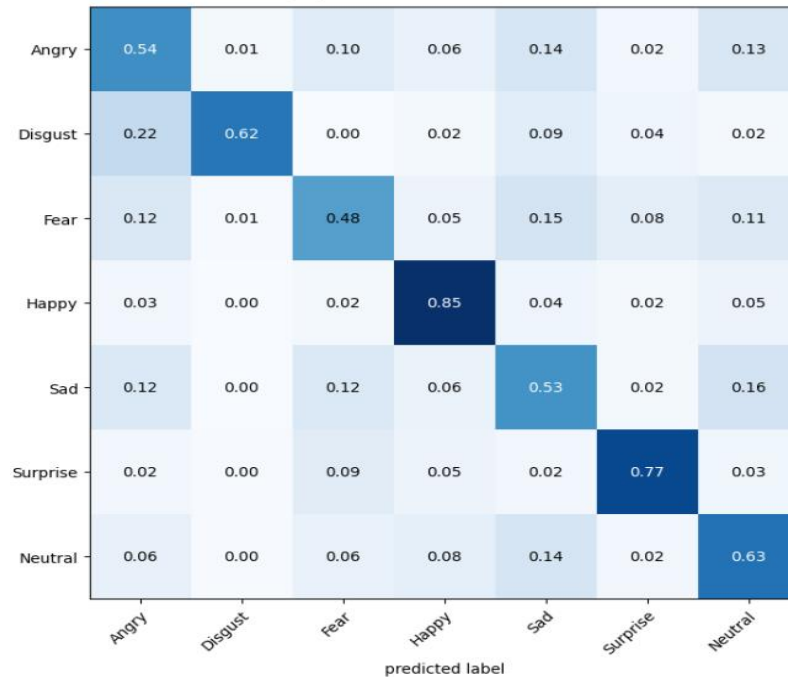


Рис. 3.3.

Висновки до розділу 3

В даному розділі детально охарактеризовано процес розробки системи розпізнавання емоцій, що включає різні аспекти, починаючи з проектування алгоритмів і закінчуючи вимогами до апаратного забезпечення.

Розробка алгоритмів та проектування системи розпізнавання емоцій дозволила визначити ключові компоненти архітектури, які забезпечують ефективний аналіз обличчя та класифікацію емоцій. Зокрема, було акцентовано увагу на інтеграції згорткових нейронних мереж, що демонструють високу

продуктивність завдяки своїй здатності автоматично витягувати характеристики з вхідних зображень.

У розділі також були розглянуті інструментальні засоби реалізації програмного продукту, в тому числі обране програмне середовище, бібліотеки для глибокого навчання, а також вимоги до апаратних засобів, потрібних для оптимальної роботи системи. Ці вимоги включають потужність графічних процесорів, об'єм оперативної пам'яті та обсяг дискового простору, який необхідний для зберігання навчальних наборів даних і моделей.

Додатково було розроблено систему для аналізу методів розпізнавання емоцій, що дозволила порівняти їх ефективність. Порівняння показників точності, швидкості обробки та стійкості різних методів дало змогу виявити сильні та слабкі сторони кожного з них.

Узагалі, результати розділу підтверджують, що запропонована система демонструє значну потенційну продуктивність у галузі розпізнавання емоцій, здатну адаптуватися до реальних умов використання. Наступні етапи дослідження фокусуватимуться на подальшій оптимізації системи та проведенні тестування з реальними користувачами для вдосконалення її функціональності.

ВИСНОВКИ

У результаті досліджень над темою "Розпізнавання емоцій обличчя з використанням згорткової нейронної мережі" вдалося визначити основні етапи та завдання, необхідні для досягнення поставленої мети.

Теоретичний аналіз досліджуваної галузі дозволив зрозуміти сучасні тенденції у розпізнаванні емоцій, виявити основні методи, інструменти та виклики, з якими стикаються дослідники.

Успішне формування навчальної вибірки для тренування нейронної мережі є критично важливим аспектом, оскільки якість та обсяги даних суттєво впливають на результати навчання.

Здійснене попереднє оброблення навчальної вибірки, такого як нормалізація й аугментація даних, дозволило покращити загальну якість вхідних даних та забезпечити стабільніше навчання моделі.

Обрання найбільш відповідної архітектури згорткової нейронної мережі стало наступним важливим етапом. Після проведення експериментів з різними архітектурами, вдалося виділити модель, яка демонструвала найкращі результати у задачі розпізнавання емоцій.

Крім того, вибір ефективного методу детектування облич на зображеннях суттєво підвищив точність системи.

Загалом, результати проведеної роботи підтверджують ефективність застосування згорткових нейронних мереж для задачі розпізнавання емоцій обличчя. Отримані дані свідчать про перспективність подальших досліджень у даній галузі, а також можливість впровадження розроблених алгоритмів у реальні програми та системи.

На основі всього вище сказаного створено систему для аналізу та порівняння методів розпізнавання емоцій людини запропонованих у розглянутих наукових працях.

При обробці даних проведено детальне попереднє опрацювання зображень, що включало нормалізацію, аугментацію даних та підбір оптимальних параметрів для підвищення якісної характеристики моделі.

Таким чином, проведене дослідження не лише виявило важливі аспекти розпізнавання емоцій, але й проклало шлях для подальших розробок у цій актуальній галузі.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Тимошик Ю.А., Орленко С.П. Алгоритм розпізнавання обличчя людей на базі згорткової нейронної мережі. Алгоритм розпізнавання обличчя людей на базі згорткової нейронної мережі. Міжвідомчий науково-технічний збірник «Адаптивні системи автоматичного управління» № 1' (32) 2018, С. 166-173.
2. P. Viola and M.J. Jones, «Robust real-time face detection», International Journal of Computer Vision, vol. 57, no. 2, 2016., pp.137–154
3. A. Adranova Neural Network User Authentication by Geometry of the Auricle. Recent Developments in Data Science and Intelligent Analysis of Information Proceedings of the XVIII International Conference on Data Science and Intelligent Analysis of Information, June 4–7. 2018, Kyiv, Ukraine, pp. 11-19.
4. Bapiev, Ya.M., Aichanov, B.H., Tereykovsky, Ya.A., Tereykovskaya, L.A., Korchenko, AA: Deep neural networks in cyber attack detection systems. International Journal of Civil Engineering and Technology (IJCET), vol. 8, pp. 1086–1092 (2017)
5. O. Arriaga, P. Plöger, Matias Valdenegro. Real-time Convolutional Neural Networks for Emotion and Gender Classification. ArXiv, 2017. P.13-18.
6. D. Duncan, G. Shine, C. English. Facial Emotion Recognition in Real Time. 2016. P. 24-31.
7. Vishwakarma, B., Dange, P., Chavan, A., Chavan, A., Galiyal, H.: Розпізнавання обличчя та виразу обличчя для сліпих людей. Міжн. рез. J. Eng. технол. 04 (2017)
8. Stanciu, L., Blidariu, F.: Emotional states recognition by interpreting facial features. In: The 6th IEEE International Conference on E-Health and Bioengineering, Grigore T. Popa University of Medicine and Pharmacy, Sinaia, Romania, 22–24 June 2017.
9. Korchenko A. Neyrosetevyie modeli, metodyi i sredstva otsenki parametrov bezopasnosti Internetorientirovannyih informatsionnyih sistem: monografiya / A. Korchenko, I. Tereykovskiy, N. Karpinskiy, S. Tyinyimbaev. – K. : TOV «Nash Format». – 2016. – 275 s.

10. Tarasenko V. P. Metod zastosuvannya produktsiinykh pravyl dlia podannia ekspertnykh znan v neiromerezhevykh zasobakh rozpoznavannia merezhevykh atak na kompiuterni systemy / V. P. Tarasenko, O. H. Korchenko, I. A. Tereikovskiy // Bezpeka informatsii. – 2013. – T. 19, № 3. – S. 168-174.
11. Tereykovskaya L. Prospects of neural networks in business models [Text] / L. Tereykovskaya, O. Petrov, M. Aleksander // TransComp 2015. 30 November – 3 December, 2015, Zakopanem, Poland. – P. 1539–1545.
12. FaceReader [Электронный ресурс]: URL: <https://www.noldus.com/human-behavior-research/products/facereader> (дата обращения 02.10.2024).
13. FaceReader Noldus Information Technology. URL: <https://www.noldus.com/human-behaviorresearch/products/facereader> (дата звернення: 13.10.2024).
14. Zhengbing H. The Analysis and Investigation of Multiplicative Inverse Searching Methods in the Ring of Integers Modulo M / H. Zhengbing, I. A. Dychka, M. Onai, A. Bartkowiak // Intelligent Systems and Applications, 2016, 11, P. 9-18.
15. Аналіз існуючих підходів до розпізнавання облич. URL: <https://habr.com/ru/company/synesis/blog/238129/> (дата звернення: 14.11.2024)
16. Victor Wiley, Thomas Lucas. Computer Vision and Image Processing: A Paper Review. International Journal Of Artificial Intelegence Research. 2018. Vol. 2. P. 28-36.
17. Computer Vision: What it is and why it matters. URL: https://www.sas.com/en_us/insights/analytics/computer-vision.html (дата звернення: 12.11.2024).
18. Paweł Tarnowski, Marcin Kołodziej, Andrzej Majkowski, Remigiusz J. Rak. Emotion recognition using facial expressions. Procedia Computer Science. 2017. Vol. 108. P. 1175-1184.
19. Howse J. OpenCV: Computer Vision Projects with Python. IEEE Xplore. 2016. P. 541.
20. He, K., Zhang, X., Ren, S., and Sun, J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. ArXiv e-prints, February 2015.

21. Yurko I.V., Aldobaeva V.N. Research of main problems, related to the recognition and identification of faces from video recordings and improving the operation of face recognition algorithms from video recordings in real time. Problems of modern science and education. 2018. P. 28.
22. Borylo P., Face Occurrence Verification Using Haar Cascades - Comparison of Two Approaches. Communications in Computer and Information Science. 2021. P. 149.
23. S. Ioffe, C. Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In International Conference on Machine Learning, pages, 2015. P. 448–456
24. L. Wiskott, J.-M. Fellous, N. Kruger, C. Von der Malsburg. Face recognition by elastic bunch graph matching. IEEE Xplore. 2022. Vol. 1. P. 129-132.
25. Chandran S., Washeef A., Somar M., Debasis M. Facial Expressions: A Cross Cultural Study / Emotion Recognition: A Pattern Analysis Approach. Wiley, 2016. 89 P.
26. What are Recurrent Neural Networks? URL: <https://www.ibm.com/cloud/learn/recurrent-neural-networks> (дата звернення: 21.11.2024).
27. Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression. Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, USA, 2020, P. 94-101
28. Зінов'єв Є., Арсенюк І. Дослідження методів розпізнавання емоцій за допомогою нейронних мереж // Матеріали XLIX науково-технічної конференції підрозділів ВНТУ. Вінниця, 2020. URL: <https://ir.lib.vntu.edu.ua/bitstream/handle/123456789/29505/8981.pdf?sequence=3>
29. A. Gudi. Recognizing semantic features in faces using deep learning. arXiv preprint arXiv:1512.00743, 2015.
30. Яровий А. А. Розпізнавання мімічних мікровиразів обличчя людини на основі Time Delay Neural Network / Яровий А. А., Кашубін С. Г., Кулик О. О.,

Липкань І. М. // Вісник Хмельницького національного університету. Технічні науки. – 2015. – № 1. – С. 122 – 126.

31. Інтелектуальна система нейромережевого розпізнавання мімічних мікроекспресій обличчя людини / Кашубін С., Яровий А.: Збірник праць IX Міжнародної науково-практичної конференції [Інтернет — Освіта — Наука (ІОН-2019)], (Вінниця, 14 –17 жовтня 2019 р.) – Вінниця, ВНТУ, 2019. – с. 60 –62.

32. Мельник В., Мельник Е. та Шульга Б.. Ідентифікація людських емоцій за допомогою нейронної мережі Keras і TensorFlow. Комп'ютерно-інтегровані технології: освіта, наука, виробництво, (36), 2019, С. 109-122. <https://doi.org/10.36910/6775-2524-0560-2019-36-11>

33. Офіційна документація TensorFlow. URL: <https://www.tensorflow.org/> (дата звернення: 18.12.2024)

34. Мельник К.В., Мельник В.М., Коптюк Ю.Ю. Дослідження методів нейромережевого розпізнавання зображень. Комп'ютерно-інтегровані технології Наука: освіта, наука, виробництво, Луцьк, 2019. № 35. – с. 161-165.

35. Гончаров Д.А. Програмне забезпечення розпізнавання акустичної інформації в сенсорних мережах / Магістерська робота на здобуття ступеня магістра за спеціальністю 123 «Комп'ютерна інженерія» в спеціалізованих комп'ютерних системах. Київ, КПІ. 2018 р. – 122 с. URL: http://ela.kpi.ua/bitstream/123456789/25805/1/Matviiiv_magistr.pdf.

36. А. А. Яровий, С. Г. Кашубін, О. О. Кулик. Розпізнавання міміки обличчя людини. / Системи технічного зору та штучного інтелекту з обробкою та стратифікацією зображень. Вінниця. – ВНТУ, 2015. – с.76-83.

37. Bartlett M.S., Littlewort G., Frank M., Lainscsek C., Fasel S., Movellan J. Recognizing facial expression: machine learning and application to spontaneous behavior. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'15), 20–25 June 2015. San Diego, 2015. P. 568–573.

38. V.Dumoulin, F. Visin A guide to convolution arithmetic for deep learning. ArXiv, 2016. 31 p.

39. D. Pitaloka., A. Wulandari, T. Basaruddin, D. Liliana. Enhancing CNN with Preprocessing Stage in Automatic Emotion Recognition. *Procedia Computer Science*, 2017. 124 p.
40. T. Rawlinson, A. Bhalerao, L. Wang. Principles and Methods for Face Recognition and Face Modelling // *Handbook of research on computational forensics, digital crime and investigation: methods and solutions*. IGI Global, 2017. PP. 53-78.
41. Rainfall–runoff modelling using Long ShortTerm Memory (LSTM) networks / Kratzert F., Klotz D., Brenner C., Schulz K., Herrnegger M. *Hydrology and Earth System Sciences*. 2018. Vol. 22, № 11. P. 6005-6022.
42. Data Preprocessing in Machine Learning. URL: <https://www.v7labs.com/blog/data-preprocessing-guide> (дата звернення: 03.12.2024).
43. F. Yu, H. Chen, X. Wang. BDD100K : A Diverse Driving Dataset for Heterogeneous Multitask Learning. *IEEE/CVF Conference on Computer Vision and Pattern Recognition : conference materials*, Seattle, 13-19 June 2020. Seattle, 2020. 263P.
44. M. Braun, S. Krebs, F. Flohr. EuroCity Persons : A Novel Benchmark for Person Detection in Traffic Scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2019. Vol. 41, No 8. 1844 P.
45. A.Kuznetsova, H. Rom., N. Alldrin. The Open Images Dataset V4: Unified Image Classification, Object Detection, and Visual Relationship Detection at Scale. *International Journal of Computer Vision*. 2020. Vol. 128, No 4. 1956P.
46. W. Fang, L. Wang, P. Ren. Tinier-YOLO: A Real-time Object Detection Method for Constrained Environments. *IEEE Access*. 2019. Vol. 8. 1935 P.
47. Cheikh R. A comparative study of emotion recognition methods using facial expressions [Електронний ресурс] / R. Cheikh, I. Falih // *arxiv.org*. – 2022. – Режим доступу до ресурсу: <https://arxiv.org/pdf/2212.03102>.
48. Тімошик П.П., ЄфремовЮ.М.. Розпізнавання емоцій на обличчі за допомогою згорткових нейронних мереж. *Комп'ютерні технології: інновації, проблеми, рішення: тези VII Всеукраїнської науково-технічної конференції, м. Житомир, 02-03 грудня 2024 р. - Житомир: Державний університет «Житомирська політехніка», 2024.*