

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Інформаційно-аналітичних технологій та менеджменту
(повна назва)

Кафедра Інформатики
(повна назва)

КВАЛІФІКАЦІНА РОБОТА Пояснювальна записка

рівень вищої освіти перший (бакалаврський)

РОЗРОБЛЕННЯ ПРОГРАМНОГО ЗАСТОСУНКУ ДЛЯ ІДЕНТИФІКАЦІЇ ЕМОЦІЙ НА ЗОБРАЖЕННЯХ (тема)

Виконав:

студент 4 курсу, групи ІТІНФ-18-2

Помазан В.В.

(прізвище, ініціали)

Спеціальності 122 Комп'ютерні науки

(код і повна назва спеціальності)

Тип програми освітньо-професійна

Освітня програма Інформатика

(повна назва освітньої програми)

Керівник доц. Творошенко І.С.

(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри

(підпис)

Кобилін О.А.

(прізвище, ініціали)

2022 р.

Харківський національний університет радіоелектроніки

Факультет Інформаційно-аналітичних технологій та менеджменту
(повна назва)Кафедра Інформатики
(повна назва)Рівень вищої освіти перший (бакалаврський)Спеціальність 122 Комп'ютерні науки
(код і повна назва)Тип програми освітньо-професійнаОсвітня програма Інформатика
(повна назва освітньої програми)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

«_____» _____ 2022 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУстудентові Помазану Віктору Вікторовичу
(прізвище, ім'я, по батькові)1. Тема роботи Розроблення програмного застосунку для ідентифікації емоцій на зображеннях

затверджена наказом по університету від 16 травня 2022 року №541Ст

2. Термін подання студентом роботи до екзаменаційної комісії 23 травня 2022 р.

3. Вихідні дані до роботи науково-методична та науково-технічна література, матеріали конференцій, дані інтернет-мережі, бібліотека комп'ютерного зору з відкритим кодом OpenCV, мова програмування Python, середовище розробки Microsoft Visual Studio Code.

4. Перелік питань, що потрібно опрацювати в роботі _____

1. Аналіз існуючих програмних застосунків для ідентифікації емоцій на зображеннях та кадрах відеозйомки.2. Моделювання структури програмного застосунку для ідентифікації емоцій на зображеннях та кадрах відеозйомки.3. Розроблення застосунку для ідентифікації емоцій на зображеннях.

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри) Актуальність проблеми розпізнавання емоцій, постановка задачі, сучасні методи розпізнавання, вибір методу для реалізації, навчання моделі, відео-розпізнавання.

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

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата
Консультант з дотримання діючих стандартів та норм	Доцент Белова Н.В.		

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Терміни Виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	18.04.2022	
2	Аналіз завдання, підбір літератури	13.04.22-16.04.22	
3	Аналіз літератури з досліджуваної проблеми	17.04.22-20.04.22	
4	Аналіз технічних і програмних засобів	21.04.22-25.04.22	
5	Розробка методу	25.04.22-03.05.22	
6	Програмна реалізація	04.05.22-18.05.22	
7	Оформлення пояснювальної записки	19.05.22-26.05.22	
8	Перевірка на плагіат	27.06.22	
9	Рецензування	28.05.22	
10	Підготовка презентації та доповіді	29.05.22-30.05.22	
11	Занесення роботи в електронний архів	31.05.22	
12	Попередній захист кваліфікаційної роботи	31.05.22	

Дата видачі завдання 18 квітня 2022 р.

Студент _____
(підпис)

Керівник роботи _____ доц. Творошенко І.С.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ/ABSTRACT

Пояснювальна записка до кваліфікаційної роботи: 61 с., 27 рис., 1 дод., 38 джерел.

КОМП'ЮТЕРНИЙ ЗІР, РОЗПОЗНАВАННЯ ЗОБРАЖЕНЬ, РОЗПОЗНАВАННЯ ЕМОЦІЙ, КЛЮЧОВІ КАДРИ, МАШИННЕ НАВЧАННЯ, НЕЙРОНА МЕРЕЖА.

Об'єктом роботи є послідовність кадрів із людськими обличчями.

Метою роботи є розроблення програмного застосунку для ідентифікації емоцій на зображеннях.

Використано методи глибокого навчання для виявлення емоцій в кадрі для виділення текстурних ознак емоцій на обличчі. Проведено дослідження методів виділення важливих ознак та класифікації об'єктів на зображеннях. Досліджено метод розробки згорткової нейронної мережі та розроблено алгоритм розпізнавання обличь.

У результаті роботи здійснена програмна реалізація системи для розпізнавання емоцій на зображеннях.

COMPUTER VISION, PICTURE RECOGNITION, EMOTION RECOGNITION, KEY FRAMES, MACHINE LEARNING, NEURAL NETWORK.

The object of the work is a sequence of shots with human faces.

The aim of the work is to develop a software application for the identification of emotions in images.

Deep learning methods were used to identify emotions in the frame to highlight the textural features of emotions on the face. Research has been conducted on methods for identifying important features and classifying objects in images. The method of development of the convolutional neural network is investigated and the algorithm of face recognition is developed.

As a result, the software implementation of the system for recognizing emotions in images was implemented.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів.....	7
Вступ.....	8
1 Аналіз існуючих програмних застосунків для ідентифікації емоцій на зображеннях та кадрах відеозйомки.....	10
1.1 Сучасний стан розвитку програмних застосунків для ідентифікації емоцій на зображеннях та кадрах відеозйомки в Україні та за кордоном.....	10
1.1.1 Огляд основних методів розпізнання та обробки зображень для ідентифікації осіб.....	12
1.1.2 Використання технології розпізнання емоцій у маркетингу..	14
1.1.3 Використання технології розпізнання емоцій у психології...	16
1.1.4 Використання технології розпізнання емоцій у сфері розваг	18
1.2 Аналіз літературних джерел щодо існуючих підходів розроблення програмних застосунків.....	19
1.3 Постановка задачі.....	21
2 Моделювання структури програмного застосунку для ідентифікації емоцій на зображеннях.....	22
2.1 Методи визначення точок на зображеннях.....	22
2.1.1 Детектор ознак FAST.....	23
2.1.2 SIFT (Scale Invariant Feature Transform).....	26
2.1.3 ORB(Oriented Fast and Rotated BRIEF).....	32
2.2 Загальна характеристика згорткової нейронної мережі.....	34
2.2.1 Шари згорткової нейронної мережі.....	35
2.2.2 Сучасні архітектури згорткових нейронних мереж.....	38
2.3 Моделювання структури та наповнення програмного застосунку для ідентифікації емоцій на зображеннях.....	42

	6
3 Розроблення програмного застосунку для ідентифікації емоцій на зображеннях.....	45
3.1 Вибір інструментальних засобів для реалізації поставленої задачі	45
3.2 Етапи розроблення програмного застосунку для ідентифікації емоцій на зображеннях.....	46
3.3 Тестування реалізованого програмного застосунку та аналіз результатів.....	50
3.4 Перспективи подальшої роботи.....	52
Висновки.....	54
Перелік джерел посилання.....	55
Додаток А Знімки процесу розпізнавання емоцій.....	59

**ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ,
СКОРОЧЕНЬ І ТЕРМІНІВ**

CNN – Convolutional Neural Network (згорткова нейронна мережа)

AI – Artificial Intelligence (штучний інтелект)

ШІ – штучний інтелект

ЄС – Європейський Союз

FAST – Features from Accelerated Segment Test

ВСТУП

Емоції представляють психологічний стан людського розуму. Дослідники з різних галузей мають різні думки щодо процесу розвитку емоцій. Філософи вважають, що емоція виникає в результаті субстанціональних (позитивних або негативних) змін в наших особистих ситуаціях або середовищі. Біологи, однак, розглядають нашу нервову та гормональну системи, відповідальними за розвиток емоцій.

Сучасні дослідження з візуалізації мозку показують, що кора і підкірка області лобового мозку відповідають за збудження емоцій. Хоча там є конфлікти в процесі розвитку емоцій, експериментальні психологи показують, що зміна наших зовнішніх або когнітивних станів здійснюється нейрональними сигналами запускає наші гормональні залози, які, у свою чергу, збуджують певні модулі в організмі людини мозку для розвитку почуття емоцій.

Збудження емоцій зазвичай супроводжується проявом у нашому зовнішньому вигляді, наприклад зміни виразу обличчя, голосу, жесту, постави тощо фізіологічні умови. Розпізнавання емоції за зовнішнім її проявом часто призводить до неточних висновків, особливо з двох причин. По-перше, прояв може не відповідати збудженню конкретної емоції. По-друге, вимірювання зовнішнього прояву вимагають приладів високої точності і точність.

Перша проблема є нерозв'язною у випадку, якщо суб'єкти, над якими проводяться експерименти намагаються придушити свої емоції або вдавати, що виявляють помилкові емоції. Припускаючи що предмети сприяють процесу розпізнавання, лише звертаємо увагу друга проблема, яку можна вирішити за допомогою передових приладів [1-14].

Розпізнавання емоцій на обличчі (FER) – це технологія, яка аналізує вирази обличчя як статичні зображення, так і відео, щоб розкрити інформацію про свій емоційний стан [1].

Складність виразів обличчя, потенційне використання технології в будь-якому контексті, і залучення нових технологій, таких як штучний інтелект, значні ризики конфіденційності.

Розпізнавання емоцій – це технологія, яка використовується для аналізу настроїв з різних джерел, наприклад, фотографії та відео [2]. Вона належить до сімейства технологій, які часто називають «афективними».

1 АНАЛІЗ ІСНУЮЧИХ ПРОГРАМНИХ ЗАСТОСУНКІВ ДЛЯ ІДЕНТИФІКАЦІЇ ЕМОЦІЙ НА ЗОБРАЖЕННЯХ ТА КАДРАХ ВІДЕОЗЙОМКИ

1.1 Сучасний стан розвитку програмних застосунків для ідентифікації емоцій на зображеннях та кадрах відеозйомки в Україні та за кордоном

Повноцінне спілкування між людьми неможливо без прояву та аналізу емоцій. Тому, при створенні сучасних людино-машинних систем актуально застосування методів автоматичного розпізнавання образів.

Одним з основних способів розпізнання емоцій між людьми є аналіз візуальної інформації. Тому автоматизація цього процесу вочевидь має бути основана на використанні методів та засобів комп'ютерного зору [3].

Комп'ютерний зір має подвійну мету. З точки зору біологічної науки, комп'ютерний зір спрямований на створення обчислювальних моделей зорової системи людини. З інженерної точки зору, комп'ютерний зір спрямований на створення автономних систем, які могли б виконувати деякі завдання, які може виконувати зорова система людини (і навіть перевершувати її в багатьох випадках). Багато завдань бачення пов'язані з вилученням тривимірної та тимчасової інформації з 2D-даних, що змінюються в часі, наприклад, отриманих однією або кількома телевізійними камерами, і в цілому розуміння таких динамічних сцен.

Звичайно, ці дві цілі тісно пов'язані. Властивості та характеристики зорової системи людини часто надихають інженерів, які розробляють системи комп'ютерного зору. І навпаки, алгоритми комп'ютерного зору можуть дати розуміння того, як працює зорова система людини. У цій роботі приймемо інженерну точку зору [4].

Системи розпізнавання емоцій можна використовувати для того, щоб дізнатися про реакцію аудиторії на фільм. Компанія Disney у 2017 році у співпраці з вченими провела експеримент: встановила камери у кінотеатрі та

підключила алгоритми глибокого навчання для оцінки емоцій глядачів. Система могла передбачити реакцію людей, спостерігаючи лише кілька хвилин. За час експерименту зібрали значний датасет: 68 маркерів з кожного з 3179 глядачів. Усього вийшло 16 мільйонів зображень осіб.

Для такої ж мети відеохостинг YouTube створив свій II під назвою YouFirst. Він дозволяє відеоблогерам та компаніям тестувати контент до випуску на платформі. Користувачі натискають на спеціальне посилання, дають згоду на зйомку відео та дивляться ролик. У цей час нейронна мережа визначає їх реакції та надсилає дані власнику каналу.

Серед російських компаній реакцію на відеоролики може аналізувати, наприклад Neurobotics. Компанія розробила програму EmoDetect, яка розпізнає радість, смуток, здивування, страх, злість, огиду та нейтральний стан. Програма вивчає до 20 локальних ознак особи на стоп-кадрах та серії зображень. Система аналізує рухові одиниці та використовує технологію кодування лицьових рухів FACS. Можна записати відео з вебкамери. Програмний інтерфейс API EmoDetect дозволяє інтегрувати продукт із зовнішніми програмами.

Емоційний III починають застосовувати і в ігровій індустрії. Він допомагає персоналізувати гру та додати більше взаємодії з геймером.

У доповіді Deep Knowledge Analytics відзначають, що при Радянському Союзі математичні школи були доволі сильними, що відрізняється від сьогоденних реалій. Тому, коли справа доходить до першокласних талантів, деякі країни ЄС досягли значних успіхів. Що ж стосується України, то ринок розробки програмного забезпечення у нас стрімко розвивається, як і сфера штучного інтелекту. Все більше постачальників пропонують різні рішення III для бізнесу. За даними Clutch, 28 українських компаній поставляють рішення для штучного інтелекту в порівнянні з 226 постачальниками по всьому світу. Згідно LinkedIn, в Україні понад 2000 розробників програмного забезпечення, які спеціалізуються на штучному інтелекті. Спільнота інженерів штучного інтелекту в Україні зростає з кожним роком.

Є багато конференцій, присвячених штучному інтелекту і машинного навчання, таких як AI & Big Data Day, AI Ukraine – Міжнародна конференція з питань штучного інтелекту, BotCamp Kyiv.

Щоб підтримати розвиток штучного інтелекту в Україні у Львівському політехнічному національному університеті у вересні 2018 року запустили нову програму з отримання наукового ступеня в галузі штучного інтелекту.

1.1.1 Огляд основних методів розпізнавання та обробки зображень для ідентифікації емоцій

Для того щоб проаналізувати основні методи розпізнавання та обробки емоцій, необхідно розібратися з таким поняттям як, глибоке навчання.

Глибоке навчання – це різновид машинного навчання, в якому модель навчається для виконання функції класифікації безпосередньо з тексту, зображень або звуку. Глибоке навчання застосовується за допомогою архітектури нейронної мережі. Слово глибокий вказує на кількість шарів у мережі, більше шарів збільшує глибину мережі.

Алгоритм глибокого навчання виконує повторювані завдання, щоразу трохи налаштовуючи його, щоб збільшити результат, щоб комп'ютерна система тренувала виконувати те, що приходить людям природно: навчання на прикладах [5]. Глибоке навчання є основною технологією для кількох додатків, таких як безпілотні автомобілі, воно дозволяє їм розрізняти пішоходів і розпізнавати знак зупинки [6].

Останнім часом з поважної причини глибоке навчання приділяється багато уваги і дає результати, які раніше були неможливими [4]. Моделі глибокого навчання можуть отримати точний стан, який іноді перевершує людські можливості. Вони були навчені з використанням великої кількості ідентифікованих даних, а також нейронних мереж з низкою шарів [4].

Моделі глибокого навчання були навчені з використанням великої кількості ідентифікованих даних і нейронних мереж, які автоматично вивчають функції з даних, тож ручне вилучення функцій не потрібно [7].

Згорткові нейронні мережі (CNN) в даний час є одними з найкращих і яскравих алгоритмів для глибокого навчання з даними зображень. Вони успішно застосовувалися в задачах комп'ютерного зору, а надійність локалізації розпізнавання об'єктів у варіантах зображень підтверджена результатами. Нещодавно було опубліковано ретельне дослідження з автоматичного аналізу експресії [8]. Ці публікації створили набір стандартних алгоритмічних конвеєрів для FER.

Останнім часом FER було розглянуто відповідно до глибокого навчання в [9], але це дуже короткий огляд без ознайомлення з наборами даних FER та технічних подробиць про глибокі FER. Тому в цьому дослідженні проводимо систематичне дослідження глибокого навчання для роботи FER, яке залежить від відео (послідовності зображень) та нерухомого зображення. Прагнучи надати новим дослідникам у цій галузі загальний опис систематизованої архітектури та високої кваліфікації глибокого FER. Він складається з вхідного та вихідного шарів. Прикладами проміжних шарів є згорткові шари, шари повного зв'язку та шари з максимальним об'єднанням. Архітектури CNN відрізняються за кількістю та типом шарів, що застосовуються для конкретного застосування.

Згортковий шар завжди був першим шаром CNN, він витягує ознаки і складається з кількох карт об'єктів. Кожен нейрон на карті об'єктів пов'язаний з невеликою областю, яка називається локальним рецептивним полем, за допомогою набору спільних ваг і єдиного спільного зміщення. Основними двома перевагами та причинами, чому згортковий шар є кращим перед повністю зв'язаним шаром, є спільне використання параметрів; де всі нейрони мають однакову вагу та зміщення в карті ознак, що викликає швидке навчання в результаті значного зменшення кількості параметрів і в кінцевому підсумку допоможе у побудові глибоких мереж [10].

Після згорткового шару шар об'єднання, зазвичай, реалізується з метою зменшення просторової роздільної здатності карт об'єктів, прискорення обчислень та виділення помітних об'єктів. Максимальне об'єднання є найбільш використовуваною технікою для шару об'єднання, де кожна одиниця об'єднання дорівнює найбільшому елементу в сприйнятливому полі.

Повністю зв'язаний (FC) вхідний шар має бути вектором, тому повинні спочатку вирівняти вихідні об'єкти з двох шарів (згортки та об'єднання). Потім можемо передати їх на вихідний шар, де для прогнозування мітки вхідного класу використовуються класифікатор Soft-Max або сигмоїд. Окрім цих мереж, існує також багато поширених похідних архітекторів.

У [9, 10], CNN на основі регіонів (RCNN) був використаний для вивчення функцій для FER. Faster R-CNN використовувався для позначення виразів обличчя шляхом створення пропозицій.

1.1.2 Використання технології розпізнання емоцій у маркетингу

Емоційний маркетинг не є новим. Щоб збільшити продажі та залучити клієнтів, маркетологи вже дуже довгий час використовують хороші розповіді та цікаві оповіді. Ця стратегія обумовлена основним фактом нашого прийняття рішень. Як прийшли до висновку дослідники з Техаського університету в Остіні, раціональний вибір, зумовлений порівнянням продуктів, важливий, «але переважно як виправдання після того, як покупець приймає рішення на основі емоційної реакції».

На практиці це означає, що якщо маркетолог може змусити вас плакати, він також може змусити вас купити. Це може здатися перебільшенням, але це не так. Як пояснює Лоррейн Тучіл з The New York Times, віце-президент Google із глобального маркетингу: «Якщо не змусимо вас плакати, то зазнаємо невдачі».

Тому різдвяні рекламні ролики сповнені емоцій і світла на зміст продукту, іноді навіть не звертаючи уваги на продукти чи послуги. Натомість вони пропонують потужні емоційні заклики, ефективні маркетингові гачки, щоб змусити споживачів відчувати щось про бренд.

Але зі зростанням автоматизації та штучного інтелекту (ШІ) у роздрібній торгівлі клієнти взаємодіють з брендами по-новому. Наприклад, розумні чат-боти змінюють визначення обслуговування клієнтів. Вони розумні, здібні та доступні цілодобово, без вихідних, гарантуючи, що обслуговування клієнтів завжди є роботою номер один. І вони лише одна платформа. Наприклад, носите мобільний телефон з камерою та мікрофоном, щоб записати свій настрій. Але як ці технології будуть розвиватися для задоволення маркетингових потреб? Один із нових підходів передбачає включення штучного інтелекту, керованого емоціями, у цифровий досвід, перенесення маркетингу в нову еру взаємодії та надзвичайної персоналізації.

«Штучний емоційний інтелект або Emotion AI також відомий як технологія розпізнавання емоцій або виявлення емоцій. У дослідженні ринку це зазвичай називають кодуванням обличчя», – пояснює Affectiva, стартап, створений Медіа-лабораторією Массачусетського технологічного інституту. Люди передають свій настрій і почуття за допомогою тонких знаків обличчя, крихітних рухів і змін, які інші люди інтерпретують, не замислюючись про це. Для ШІ це справжня робота.

По суті, кодування обличчя працює так само, як розпізнавання обличчя. Він використовує камеру, щоб «побачити» обличчя людини, а потім відображає риси її обличчя, використовуючи такі помітні точки, як куточки рота, розташування очей або відстань від носа до підборіддя. Його навчили на мільйонах реальних обличь, повільно вивчаючи підказки для виявлення емоційних станів. Тепер, коли його навчили, так би мовити, аналізуючи крихітні відмінності, які з'являються в результаті наших почуттів, Emotion AI Affectiva може розпізнавати «злість, презирство, огиду, страх, радість, смуток і здивування». Поки ще в зародковому стані, ця технологія має вдалий

початок. А коли відео поєднується зі звуком, AI може дізнатися набагато більше про настрій. Наприклад, прислухаючись до «змін у паралінгвістиці мовлення, тону, гучності, темпу та якості голосу», Emotion AI може розрізняти вік, стать та емоції. Аналізуючи ці дані всього за мілісекунди та порівнюючи їх з тим, що він бачить, Emotion AI може виявляти зміни настрою в реальному часі [5].

Отже, зрозуміло, що ця технологія – не просто реклама. І в міру того, як вона буде розвиватися, вона буде вдосконалюватися в тому, що вона робить. Можливо, важливіше те, що думаємо, що він швидко переміститься від чат-ботів до маркетингових платформ, таких як мобільні телефони, відчуваючи настрій користувача, щоб пропонувати продукти або вибирати рекламу в режимі реального часу, оскільки почуття користувача змінюються.

Для маркетологів важливо знати, що відчуває потенційний клієнт. Таке розуміння може забезпечити надзвичайну персоналізацію, значно підвищивши ефективність реклами. Директор із маркетингу IBM Commerce.

Цифрові маркетологи, очевидно, в захваті від можливостей пропозицій ШІ для виявлення емоцій. І ця технологія обіцяє революцію в тому, як бренди охоплюють потенційних клієнтів, революцію, зумовлену все більш розумними комп'ютерами.

1.1.3 Використання технології розпізнання емоцій у психології

Психологія вивчає психічні процеси та поведінку індивідів. Штучна психологія – це дослідження психічних процесів системи штучного інтелекту (AIS), подібної до людей [3]. Йдеться про штучні когнітивні процеси, необхідні для того, щоб штучний інтелект був розумним, навчався, автономним і саморозвиваючимся [4]. У психології є кілька спеціальностей або напрямків вивчення.

Візьмемо, наприклад, когнітивну психологію, яка вивчає, як мозок думає і працює. Сюди входять навчання, пам'ять, сприйняття, мова, логіка [10].

Існує також психологія розвитку, яка розглядає, як індивід пристосовується і змінюється на різних етапах розвитку і що доцільно розглядати для людини на основі розвитку [9].

Існує спортивна психологія, яка розглядає, як впливати на індивідуальну продуктивність і як результативність впливає на особистість. Таким чином, штучна психологія для цілей цієї статті містить штучний розумовий процес, який вважається необхідним для створення інтелектуальних, автономних, саморозвиваючихся, штучних когнітивних систем. AIS має імітувати людські процеси, щоб бути розумним. Зрештою, хіба людина не знаходиться на вершині спектру інтелекту?

Штучна психологія – це теоретична дисципліна, яка була вперше запропонована Деном Кертісом у 1963 році. Ця теорія стверджує, що штучний інтелект наблизиться до рівня складності людського інтелекту, коли система штучного інтелекту відповідає трьом дуже важливим умовам:

– здатність системи зі штучним інтелектом перепрограмувати або саморозвиватися за допомогою процесу самоаналізу та прийняття рішень на основі інформації, доступної для системи, не може забезпечити механізми для вирішення внутрішніх невідповідностей всередині системи без адаптації психологічних конструкцій. до методологій і стратегій AIS, а отже, штучна психологія, за визначенням, потрібна;

– сучасна теорія штучної психології не розглядає особливості того, наскільки складною має бути система для досягнення умов, наведених вище, а лише те, що система є достатньо складною, щоб інтелект не міг бути просто зафіксований розробником програмного забезпечення, і тому цей предмет повинен бути розглядаються через ті самі процеси, через які проходять люди.

Таким же чином, штучна психологія не розглядає питання про те, чи є інтелект справді совістю чи ні.

1.1.4 Використання технології розпізнання емоцій у сфері розваг

Системи розпізнавання емоцій можна використовувати для того, щоб дізнатися про реакцію аудиторії на фільм. Компанія Disney у 2017 році у співпраці з вченими провела експеримент: встановила камери у кінотеатрі та підключила алгоритми глибокого навчання для оцінки емоцій глядачів. Система могла передбачити реакцію людей, спостерігаючи лише кілька хвилин. За час експерименту зібрили значний датасет: 68 маркерів з кожного з 3179 глядачів. Усього вийшло 16 мільйонів зображень осіб.

Для такої ж мети відеохостинг YouTube створив свій ШІ під назвою YouFirst. Він дозволяє відеоблогерам та компаніям тестувати контент до випуску на платформі. Користувачі натискають на спеціальне посилання, дають згоду на зйомку відео та дивляться ролик. У цей час нейромережа визначає їх реакції та надсилає дані власнику каналу.

Серед російських компаній реакцію на відеоролики може аналізувати, наприклад Neurobotics. Компанія розробила програму EmoDetect, яка розпізнає радість, смуток, здивування, страх, злість, огиду та нейтральний стан. Програма вивчає до 20 локальних ознак особи на стоп-кадрах та серії зображень. Система аналізує рухові одиниці та використовує технологію кодування лицьових рухів FACS. Можна записати відео з вебкамери. Програмний інтерфейс API EmoDetect дозволяє інтегрувати продукт із зовнішніми програмами.

Емоційний ШІ починають застосовувати і в ігровій промисловості. Він допомагає персоналізувати гру та додати більше взаємодії з геймером.

Емоції за своєю суттю важко читати, і часто існує розрив між тим, що люди кажуть, що вони відчують, і тим, що вони відчують насправді.

Машина може ніколи не досягти такого рівня розуміння, але хто скаже, як обробляємо емоції, це єдиний шлях. Те, як інтерпретуємо емоції один одного, сповнене упередженості та думок, тому, можливо, ШІ допоможе перейти безпосередньо до того, що стосується наших емоцій.

1.2 Аналіз літературних джерел щодо існуючих підходів розроблення програмних застосунків

Визначення емоцій за допомогою кадрів відеозйомки – вкрай містка тема, яка є об'єктом досліджень вже протягом декількох десятиліть. За цей час було написано багато наукових робіт, що зачіпають різні особливості розробки та проблеми, так чи інакше пов'язані з цим напрямом комп'ютерного зору. Практично неможливо осягнути навіть усі найцікавіші та найважливіші праці, присвячені цьому питанню, проте в процесі роботи все ж була опрацьована низка актуальних робіт, перелік яких наведено далі.

У статті [8] розглянуто питання, присвячені аналізу роботи інтелектуальної системи прийняття рішень стосовно моделювання процесів кардіології людини. Проведено ряд тестових прикладів для визначення складності інструментальних засобів, а також показано практичну значимість розробленої системи.

У статті [9] представлені результати наукових досліджень, які проводилися з поставленою метою та вирішують актуальне завдання розробки та вдосконалення методів, моделей та інформаційних технологій оцінки станів складних просторово розподілених об'єктів на основі еволюційних інтервальних нечітких моделей. Запропоновано гібридну модель оцінки просторово розподілених об'єктів, яка інтегрує розроблені нечіткі кольорові мережі Петрі, моделі детермінованих, стохастичних і нечітких баз знань, а також логіку прояв їх взаємодії.

У статті [11] описується описуються фундаментальні будівельні блоки згорткових нейронних мереж і найпопулярніші архітектури згорткових нейронних мереж в історії, в тому числі ті, які досягли найсучаснішої продуктивності щодо стандартних наборів даних розпізнавання та завдань.

У статті [14] описано одну з багатьох проблем, пов'язаних з розпізнаванням емоцій. Метою даної роботи є встановлення деяких основоположних принципів, теорії та концепції, які пов'язані з обробкою емоцій для різних цілей, в основному для задач пов'язаних з психологією.

У джерелі [15] досліджуються різні методи аналізу зображень. Описано методи для пошуку дублікатів зображень, а також методи пошуку схожих зображень, які ґрунтуються на визначені ключових точок.

У статті [16] автором дається дещо особистий і, можливо, упереджений огляд галузі комп'ютерний зір. Спочатку дано визначення комп'ютерному зору і наведемо дуже коротку історію його розвитку. Потім йду окреслення деяких з причин, чому комп'ютерний зір є дуже складним полем дослідження. Зокрема наведено кілька прикладів майбутніх додатків, які, на думку авторів, є дуже перспективними.

У роботі [17] розглядаються реальні сценарії, коли під час розробки, виникає кілька проблем, наприклад, тенденція до нейтралітету, що ускладнює отримання збалансованих наборів даних, або відсутність стандартів для анотації емоційних категорій. Більше того, внутрішня суб'єктивність емоційної інформації збільшує труднощі отримання цінних даних до тренувати алгоритми на основі машинного навчання. У цій роботі розглядалися два різні реальні сценарії: взаємодія людини з людиною в теледебатах і взаємодія людини і машини з віртуальним агентом. З метою порівняння в обох було проведено аналіз емоційної інформації.

У роботі [18] пропонується розробка системи розпізнавання виразів обличчя (FER) на основі глибоких згорток нейронної мережі за допомогою трьох моделей. У цій роботі просте рішення для міміки розпізнавання, яке використовує комбінацію алгоритмів для виявлення обличчя, виділення ознак та класифікації обговорюється.

Запропонований метод використовує моделі CNN з класифікатором SVM та оцінює їх, ці моделі це модель Alex-net, модель VGG-16 і модель Res-Net. Експерименти проводяться на Extended Набори даних Cohn-Kanada (CK+) для визначення точності розпізнавання для запропонованої системи FER.

1.3 Постановка задачі

Актуальність даної роботи полягає у важливості систем розпізнання емоцій, їх широкому спектрі сфер застосування в сучасному світі, а також існуванні ряду пов'язаних із ними проблем та необхідності пошуку оптимальних методів та підходів для розпізнавання.

Об'єктом роботи є послідовність кадрів із людськими обличчями.

Метою роботи є розроблення програмного застосунку для ідентифікації емоцій на зображеннях.

Враховуючи мету роботи, необхідно вирішити такі завдання:

- дослідити методи визначення точок на зображенні;
- проаналізувати методи розпізнання емоцій;
- розробити методіку ідентифікації емоцій;
- виконати етапи розроблення застосунку для ідентифікації емоцій;
- проаналізувати наявні підходи до розпізнання емоцій за певними ознаками;
- провести тестування розробленого застосунку та проаналізувати результати;
- визначити перспективи подальшої роботи.

2 МОДЕЛЮВАННЯ СТРУКТУРИ ПРОГРАМНОГО ЗАСТОСУНКУ ДЛЯ ІДЕНТИФІКАЦІЇ ЕМОЦІЙ НА ЗОБРАЖЕННЯХ

2.1 Методи визначення точок на зображеннях

У кожного зображення є деякі спеціальні точки. Що це таке? Особливі точки – такі точки, за якими можна класифікувати зображення, розпізнати його, якась особливість зображення, унікальність. Як правило – це кутові точки, або ті, де різко змінюється колір, яскравість і т.д.

Потрібно вибирати такі точки, які роблять певний внесок у характеристику зображення, також необхідно вважати особливими такі точки, які з великою ймовірністю будуть знайдені на іншому зображенні. Кожен метод виявлення особливих точок повинен гарантувати інваріантність щодо будь-яких перетворень зображення. Залишилося зрозуміти, яким чином комп'ютер розуміє – які ключові точки різних зображень відповідають один одному.

Ключові точки повинні мати низку особливостей:

- відмінність – кожна точка має явно відрізнятися від інших та бути унікальною у своїй області;
- інваріантність – визначення ключової точки повинно бути незалежним від афінних перетворень;
- стабільність – виділення таких особливостей повинно бути стійким до шумів та модифікацій;
- інтерпретація – ключові точки повинні виділятися так, щоб їх можна було використовувати для аналізу відповідностей і видобування на їх основі необхідної інформації.

Отже, для того щоб знайти фрагменти зображення або ж схожі за змістом ілюстрації, необхідно експериментувати з методами визначення ключових точок, у кожного з яких також свій набір переваг та недоліків.

Основні методи, які використовують під час побудови детекторів та дескрипторів:

- FAST (Features from Accelerated Segment Test);
- SIFT (Scale Invariant Feature Transform);
- ORB (Oriented FAST and Rotated BRIEF);
- AKAZE (Accelerated KAZE);
- BRIEF (Binary Robust Independent Elementary Features);
- BRISK (Binary Robust Invariant Scalable Keypoints).

2.1.1 Детектор ознак FAST

Існує кілька детекторів функцій, і багато з них дійсно хороші. Але якщо дивитися з точки зору застосування в реальному часі, вони недостатньо швидкі. Найкращим прикладом може бути мобільний робот SLAM (одночасна локалізація та відображення), який має обмежені обчислювальні ресурси.

Як рішення цієї проблеми, функції з прискореного сегментного тесту (FAST) – це метод виявлення кутів, який можна використовувати для виділення характерних точок, а потім використовувати для відстеження та відображення об'єктів у багатьох задачах комп'ютерного зору.

Спочатку кутовий детектор FAST був розроблений Едвардом Ростеном і Томом Драммондом і був опублікований у 2006 році. Найбільш багатообіцяючою перевагою кутового детектора FAST є його обчислювальна ефективність.

Крім того, коли використовуються методи машинного навчання, можна досягти кращої продуктивності з точки зору часу обчислень і ресурсів. Завдяки цій високошвидкісній продуктивності кутовий детектор FAST дуже підходить для обробки відео в режимі реального часу (рис. 2.1).

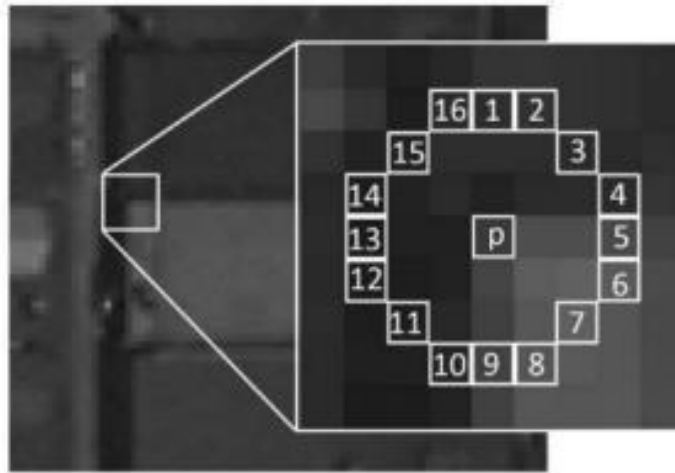


Рисунок 2.1 – Приклад чергової ітерації алгоритму FAST

Алгоритм пояснюється так:

- виберіть піксель p на зображенні, який потрібно ідентифікувати як точку інтересу чи ні. Нехай його інтенсивність буде I_p ;
- виберіть відповідне порогове значення t ;
- розглянемо коло з 16 пікселів навколо досліджуваного пікселя (це коло Брезенхема радіусом 3);
- тепер піксель p є кутом, якщо в колі існує набір з n суміжних пікселів (з 16 пікселів), які світліші за $I_p + t$ або всі темніше за $I_p - t$. (Автори використовували $n=12$ у першій версії алгоритму);
- щоб зробити алгоритм швидким, спочатку порівняйте інтенсивність пікселів 1, 5, 9 і 13 кола з I_p . Як видно з малюнка вище, принаймні три з цих чотирьох пікселів повинні задовольняти пороговому критерію, щоб існувала точка інтересу;
- якщо принаймні три із чотирьох-піксельних значень – I_1, I_5, I_9, I_{13} не вище або нижче $I_p + t$, то p не є точкою (кутом) інтересу. У цьому випадку відкиньте піксель p як можливу точку інтересу. В іншому випадку, якщо принаймні три пікселі знаходяться вище або нижче $I_p + t$, перевірте всі 16 пікселів і перевірте, чи підпадають під критерій 12 суміжних пікселів;
- повторіть процедуру для всіх пікселів на зображенні.

У підході машинного навчання алгоритм FAST порівнює шістнадцять пікселів навколо піксель p точки інтересу (кут), як показано на рисунку 2.1, кутовий детектор визначає піксель p як кут, якщо в колі існує набір n суміжних пікселів, які яскравіші за інтенсивність кут кандидата I_p плюс поріг t або темніші за I_p мінус t . n обрано як 12, тому що це дозволяє проводити високошвидкісний тест, який можна використовувати для виключення дуже великої кількості не кутів. Це визначено у рівнянні:

$$S_{p \rightarrow x} = \left\{ \begin{array}{ll} d \ I_{p \rightarrow x} \leq I_p - t & \text{(darker)} \\ s \ | \ I_p - t < I_{p \rightarrow x} < I_p + t & \text{(similar)}, \\ b \ I_p - t \leq I_{p \rightarrow x} & \text{(brighter)} \end{array} \right\}, \quad (2.1)$$

де I_p – інтенсивність p ;

$I_p \rightarrow x$ – інтенсивність шістнадцяти пікселів, які оточують піксель прикладу та t – це поріг.

Якщо $S_{p \rightarrow x}$ дорівнює d , то піксель належить до темнішої групи; якщо $S_{p \rightarrow x}$ дорівнює s , то піксель належить до подібної групи; якщо $S_{p \rightarrow x}$ дорівнює b , то піксель належить до більш яскравої групи.

Якщо їх 12 суміжні пікселі, які належать до темнішої або яскравішої групи, p класифікується як кут [12].

Алгоритм має кілька обмежень. По-перше, для $n < 12$ алгоритм працює не дуже добре у всіх випадках, тому що при $n < 12$ кількість виявлених точок інтересу дуже велика. По-друге, порядок, у якому запитуються 16 пікселів, визначає швидкість роботи алгоритму.

Інші методи виявлення кутів працюють зовсім інакше, ніж метод FAST, і дивовижним результатом є те, що FAST не виявляє кути на згенерованих комп'ютером зображеннях, які ідеально вирівняні по осях x і y . Оскільки виявлений кут повинен мати кільце з темнішими або світлішими значеннями пікселів навколо центру, яке включає обидва краї кута, чіткі зображення не працюють належним чином.

Це тому, що алгоритм FAST вимагає кільця контрастних пікселів більше ніж на три чверті навколо центру кута. У згенерованому комп'ютером зображенні обидва краї квадрата в куті знаходяться в кільці використуваного пікселя, тому тест на кут не завершується. Обхідним шляхом цієї проблеми є додавання розмиття (за допомогою фільтра Гаусса) до зображення, щоб кути були менш точними, але їх можна було виявити [6].

2.1.2 SIFT (Scale Invariant Feature Transform)

Масштабно інваріантне перетворення ознак (SIFT) – це дескриптор зображення для зіставлення зображень, розроблений Девідом Лоу (1999 рік, 2004 рік). Цей дескриптор, а також пов'язані з ним дескриптори зображень використовуються для великої кількості цілей у комп'ютерному зору, пов'язане з узгодженням точок між різними видами тривимірної сцени та розпізнаванням об'єктів на основі перегляду [19-21].

Дескриптор SIFT є інваріантним до трансляцій, обертань і перетворень масштабування в області зображення та надійні та помірні перетворення перспективи та варіації освітлення. Експериментально дескриптор SIFT має було доведено, що він дуже корисний на практиці для порівняння зображень і розпізнавання об'єктів у реальних умовах [21].

У своїй оригінальній формульовці дескриптор SIFT містив метод виявлення точок інтересу на рівні сірого зображення, на якому була накопичена статистика локальних напрямків градієнта інтенсивності зображення для узагальнення опису локальних структур зображення в місцевих околицях навколо кожної точки, цікаво, що цей дескриптор слід використовувати для узгодження відповідних точок інтересу між різними зображеннями.

Пізніше SIFT дескриптор також застосовувався до щільних сіток (щільний SIFT), які, як було показано, призводять до кращої продуктивності

в таких завданнях, як категоризація об'єктів, класифікація текстур, вирівнювання зображення та біометричні дані. Детектор SIFT має також розширення від сірого до кольорових зображень і від $2D$ просторових зображень до $2+1D$ просторово-часового відео [22].

Оригінальний детектор SIFT був обчислений на основі інтенсивності зображення навколо цікавих місць в області зображення, які можна назвати точками інтересу, або ключовими точками (рис. 2.2). Ці точки інтересу отримують з екстремумів масштабного простору різниць гаусів (DoG) у межах різниці піраміди Гаусса, які спочатку запропонували Берт і Адельсон (1983 рік) і Кроулі і Стерн (1984 рік).



Рисунок 2.2 – Наслідок роботи детектора SIFT

Піраміда Гаусса будується на основі вхідного зображення шляхом багаторазового згладжування та підвибірки, а різниця піраміди Гаусса обчислюється з різниці між сусідніми рівнями в піраміді Гаусса. Тоді, точки відсотка отримуються з точок, у яких різниця значень Гаусса приймає екстремум відносно як до просторових координат в області зображення, так і до рівня масштабу в піраміді.

Паралельно з побудовою піраміди гаусіанів, будується піраміда різностей гаусіанів, що складається з різноманіття сусідніх зображень у піраміді гаусіанів.

Відповідно, кількість зображень у цій піраміді буде $N+1$ (рис. 2.3) [23].

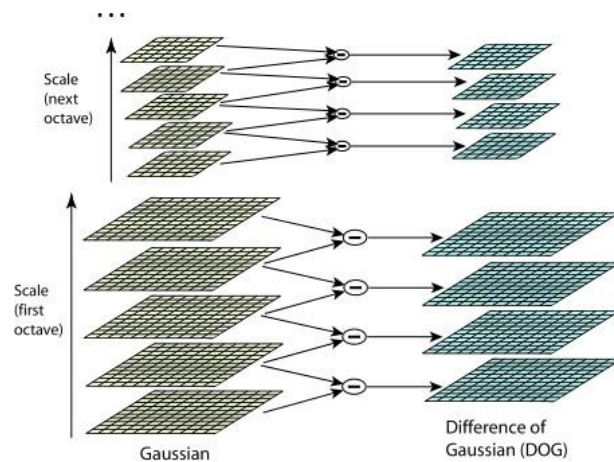


Рисунок 2.3 – Піраміда різниці гаусіанів

Після побудови пірамід діло залишиться за малим. Будем рахувати точку особливою, якщо вона є локальним екстремумом різності гаусіанов. Для пошуку екстремумов будем використовувати метод, схематично зображений на рисунку 2.4.

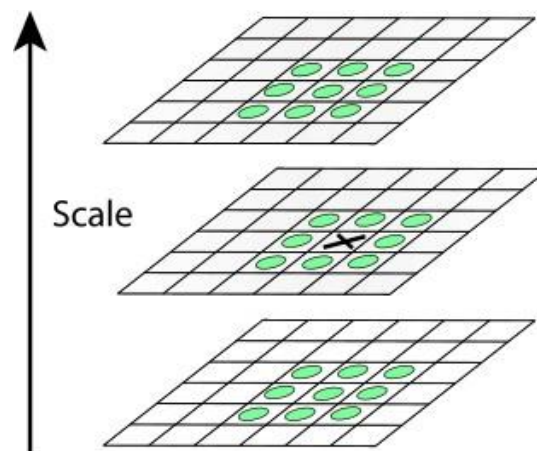


Рисунок 2.4 – Схема методу знайдення ключових точок на зображенні

Уточнення особливих точок. Спочатку визначаються координати особливої точки з субпіксельною точністю. Це досягається за допомогою функції апроксимації DoG багаточленом Тейлора другого порядку, взятого в точці вчисленого екстремума:

$$D(x) = D + \frac{6D^T}{6x}x + \frac{1}{2}x^T \frac{6^2D}{6x^2}x, \quad (2.2)$$

де D – функція DoG;

$X = (x, y, \sigma)$ – вектор смещення щодо точки розкладу.

Перша виробнича DoG – градиент, друга виробнича DoG – матриця Гессе.

Екстремум багаточлена Тейлора знаходиться шляхом вичислення похідної та прирівняння її до нуля. В підсумку отримаємо зміщення точки вчисленного екстремума, щодо точного:

$$\hat{x} = -\frac{6^2D^{-1}}{6x^2} \frac{6D}{6x}, \quad (2.3)$$

Після того, як переконалися, що яка-то точка є ключовою, потрібно вчислити її орієнтацію. Як буде видно далі, точка може мати кілька напрямків.

Управління ключовою точкою обчислюється вихід з направлених градієнтів точок, сусідніх з особою. Всі обчислення градієнтів виробляються на зображенні в піраміді гаусіанів, з великим найбільш близьким до масштабу ключової точки.

Величина і напрямок градієнта в точці (x, y) обчислюється так:

$$m(x, y) = \sqrt{(L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2},$$

$$\theta(x, y) = \tan^{-1}((L(x, y + 1) - L(x, y - 1)) / (L(x + 1, y) - L(x - 1, y))),$$

де m – величина градієнта;

θ – напрямок градієнта.

Тепер перейдемо безпосередньо до дескрипторів. Дане раніше визначення говорить про те, що має робити дескриптор, але не про те, що це таке. В принципі, дескриптором може виступати будь-який об'єкт (аби він справлявся зі своїми функціями), але, зазвичай, дескриптором є певна інформація про околицю ключової точки.

Такий вибір зроблено через кілька причин: на маленькі області менший вплив мають ефекти спотворень, деякі зміни (зміна положення об'єкта на картинці, зміна сцени, перекриття одного об'єкта іншим, поворот) можуть не вплинути на дескриптор зовсім [24-28].

У методі SIFT дескриптор є вектор. Як і напрямок ключової точки, дескриптор обчислюється на гаусіані, найближчому масштабу до ключової точки, і виходячи з градієнтів в деякому вікні ключової точки. Перед обчисленням дескриптора це вікно повертають на кут наряду ключової точки, чим досягається інваріантність щодо повороту.

Для початку подивимося на рисунок 2.5 [29].

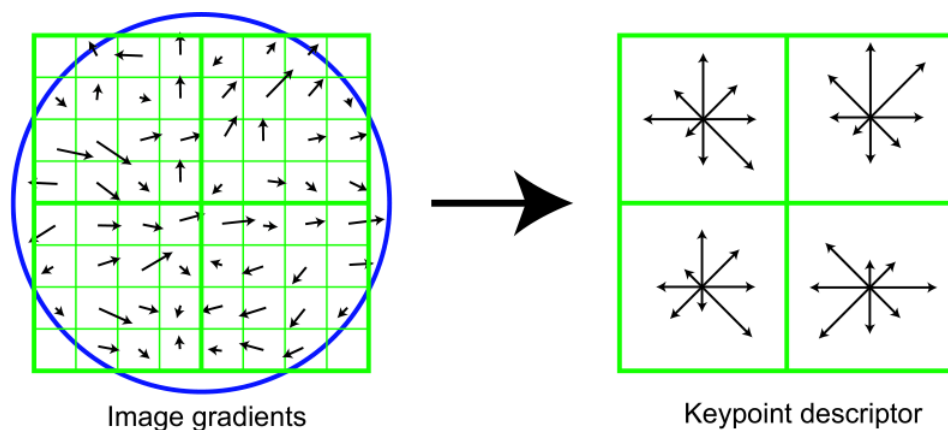


Рисунок 2.5 – Отриманий на основі зображення дескриптор

Тут схематично показана частина зображення (ліворуч) та (праворуч) отриманий на її основі дескриптор. Для початку подивимося ліворуч. Тут можна бачити пікселі, позначені маленькими квадратиками. Ці пікселі беруться з квадратного вікна дескриптора, яке у свою чергу поділено ще на чотири рівні частини (далі називатимемо їх регіонами).

Маленька стрілочка в центрі кожного пікселя позначає градієнт цього пікселя. Цікаво, що центр цього вікна знаходиться між пікселями. Його треба вибрати якомога ближче до точних координат ключової точки. Остання деталь, яку можна побачити – це коло, що означає вікно згортки з ядром гауса (аналогічно вікну для обчислення напрямку ключової точки). Для цього ядра визначається σ , що дорівнює половині ширини вікна дескриптора. Надалі значення кожної точки вікна дескриптора буде примножуватися на значення ядра гауса в цій точці, як на ваговий коефіцієнт.

Тепер подивимося праворуч. Тут можемо бачити схематично зображений дескриптор особливої точки розмірності $2 \times 2 \times 8$.

Перші дві цифри у значенні розмірності – це кількість регіонів по горизонталі та вертикалі. Ті квадрати, які охоплювали певний регіон пікселів на лівому зображенні, праворуч охоплюють гістограми, побудовані на пікселях цих регіонів. Відповідно, третя цифра розмірності дескриптора означає кількість компонент гістограми цих регіонів. Гістограми в регіонах обчислюються так само, як і гістограма напрямків із трьома невеликими але:

- кожна гістограма так само покриває ділянку 360 градусів, але ділить її на 8 частин;
- як ваговий коефіцієнт береться значення гаусового ядра, загального для всього дескриптора (про це вже говорилося);
- як ще одні вагові коефіцієнти беруться коефіцієнти трилінійної інтерполяції.

Кожному градієнту у вікні дескриптора можна приписати три речові координати (x, y, n) , де x – відстань до градієнта по горизонталі, y – відстань по вертикалі, n – відстань до напрямку градієнта в гістограмі (мається на увазі відповідна гістограма дескриптора робить внесок цей градієнт).

За точку відліку приймається нижній лівий кут вікна дескриптора і початкове значення гістограми. За одиничні відрізки беруться розміри регіонів по горизонталі та вертикалі для x та y відповідно, і кількість градусів у компоненті гістограми для n .

Коефіцієнт трилінійної інтерполяції визначається кожної координати (x, y, n) градієнта як $1-d$, де d дорівнює відстані від координати градієнта до середини того одиничного проміжку в який ця координата потрапила. Кожне входження градієнта в гістограму множиться на всі три вагові коефіцієнти трилінійної інтерполяції.

Дескриптор ключової точки складається із усіх отриманих гістограм. Як було зазначено розмірність дескриптора малюнку 32 компоненти ($2 \times 2 \times 8$), але практично використовуються дескриптори розмірності 128 компонент ($4 \times 4 \times 8$).

Отриманий дескриптор нормалізується, після чого всі компоненти, значення яких більше 0,2 урізаються до значення 0,2 і потім дескриптор нормалізується ще раз. У такому вигляді дескриптори готові до використання.

2.1.3 ORB (Oriented FAST and Rotated BRIEF)

Алгоритм зіставлення зображень ORB, зазвичай, поділяється на три кроки:

- виділення характерних точок;
- генерування дескрипторів характерних точок;
- генерування відповідності ознакових точок.

Конкретна блок-схема показана на рисунку 2.6.

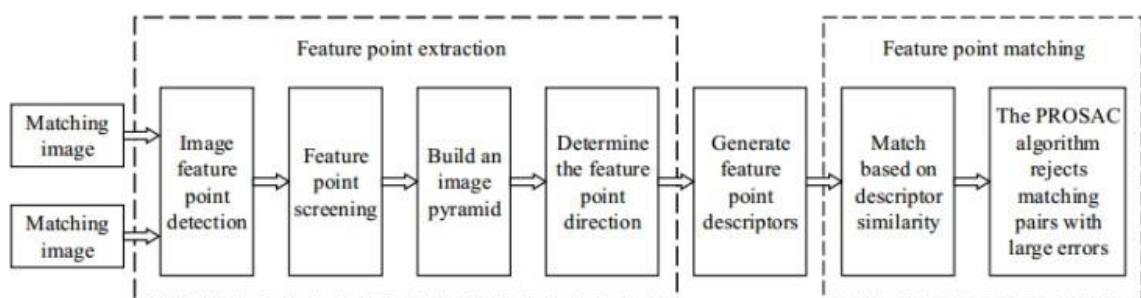


Рисунок 2.6 – Блок-схема алгоритма ORB

Алгоритм ORB використовує покращений алгоритм FAST (функції прискореного тесту сегментів) [30-35] для виявити особливі точки. Ідея полягає в тому, що якщо піксель значно відрізняється від пікселів сусідства то, швидше за все, це буде кутова точка. Процес виявлення виглядає наступним чином: Виявлення функціональної точки зображення.

Спочатку виберіть піксель p на зображенні і припустимо його яскравість. Встановіть поріг яскравості T . Потім візьміть піксель p як центр, виберіть 16 пікселів на колі з радіусом 3 і порівняйте значення сірого між пікселем p та іншими пікселями на колі.

Якщо яскравість N послідовних точок на вибраному колі більше $I_p + T$ або менше $I_p - T$, тоді піксель p може розглядати як особливість [31].

Скринінг особливостей. Оскільки розрахунок кутової точки FAST призначений лише для порівняння різниці в яскравості між пікселями, число велике і невизначене і немає напряду інформації. Таким чином, алгоритм ORB покращує вихідний алгоритм FAST, який обчислює значення відповідей Харріса для вихідних кутових точок FAST і сортує їх відповідно до значення сірого кольору і візьміть перші N балів.

Формула розрахунку значення відповіді Харріса така:

$$R = \det(M) - k(\text{trace}(M))^2, \quad (2.4)$$

$$M = \sum w(x, y) \begin{bmatrix} l_x^2 & k l_y \\ l_x l_y & l_y^2 \end{bmatrix}. \quad (2.5)$$

Оскільки оригінальний дескриптор BRIFE не має інваріантності повороту, при цьому легко втратити дані зображення повертається. Тому алгоритм ORB використовує алгоритм Steer BRIEF для обчислення основного напрямку кожної ознаки, щоб дескриптор мав інформацію про напрямки (рис. 2.7).

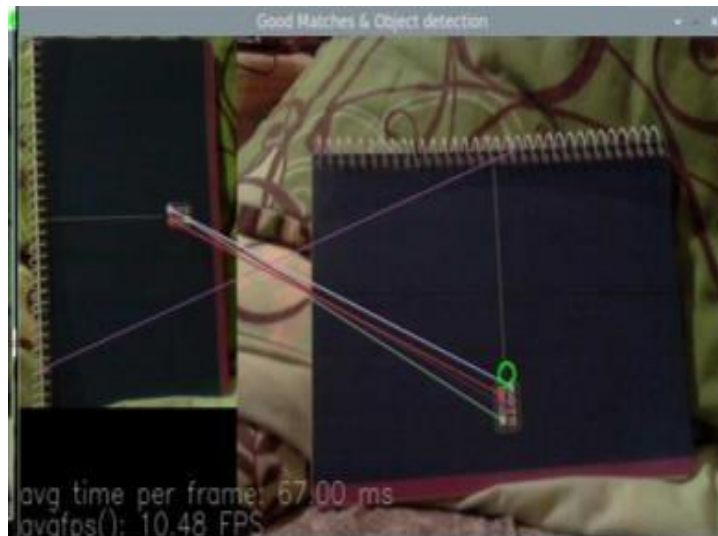


Рисунок 2.7 – Демонстрація виконання методу виявлення об’єктів через особливі точки детектором ORB

2.2 Загальна характеристика згорткової нейронної мережі

У сфері штучного інтелекту спостерігається стрімкий ривок усунення розриву між можливостями людини та машин. І вчені, і любителі, працюючи над численними проблемами у цій галузі, роблять дивовижні речі. Одним із багатьох таких напрямків є Computer Vision (комп’ютерний зір) [36].

Порядок денний у цій галузі полягає в тому, щоб дати можливість машинам бачити світ так, як це роблять люди, сприймати його схожим чином і навіть використовувати знання для множини завдань: розпізнавання зображень та відео, аналіз та класифікація зображень, відновлення медіа, системи рекомендацій, обробка природної мови і т.д.

Досягнення в області Computer Vision і Deep Learning були розроблялися і вдосконалювалися з часом, головним чином завдяки абсолютно конкретному алгоритму – згорткової нейронної мережі.

Згорткова нейронна мережа (Convolutional Neural Network – ConvNet/CNN) – це Deep Learning-алгоритм, який може приймати вхідне зображення, надавати важливість (засвоювані ваги та зміщення) різним

областям/об'єктам у зображенні та може відрізнити одне від одного. Попередня обробка в ConvNet вимагає значно менше в порівнянні з іншими алгоритмами класифікації. У той час як у примітивних методах фільтри сконструйовані вручну, ConvNets, при достатньому навчанні, здатні вивчати ці фільтри/характеристики.

Архітектура ConvNet схожа з архітектурою зв'язності нейронів у людському мозку і була натхненна організацією зорової кори. Окремі нейрони реагують на подразники лише у обмеженій зоні поля зору, відомої як рецептивне поле. Сукупність таких полів накладається, щоби покрити всю зону поля зору.

CNN призначені для обробки даних, які надходять у у формі кількох масивів, наприклад, кольорове зображення, що складається з трьох 2D-масивів, що містять інтенсивність пікселів у триколірні канали. Вони використовують свої згортки фільтри для вилучення інформації із зображень, попередні шари виявляють краї, більш пізні шари можуть виявляти частину об'єктів, тоді навіть більш пізні шари можуть виявляти цілі об'єкти, наприклад як грані або інші складні геометричні фігури. CNN складається з набору шарів, за якими можна групувати їх функціональні можливості три основні типи шарів: згортковий шар, об'єднаний шар і повністю зв'язаний шар.

2.2.1 Шари згорткової нейронної мережі

Згортковий шар. Операція згортки є одним із фундаментальних будівельних блоків згорткової нейронної мережі.

Параметри згорткового шару складаються з набору навчальні фільтри (ядра). Кожен фільтр невеликий у просторі (по ширині та висоті), але простягається через повну глибина вхідного обсягу. Типові розміри фільтрів можуть мають розміри 3×3 , 5×5 , 7×7 .

Третій вимір фільтра відповідає кількості каналів на вході.

Глибина зображення у відтінках сірого дорівнює 1, а кольорового зображення – 3 (RGB) кольорові канали (рис. 2.8).

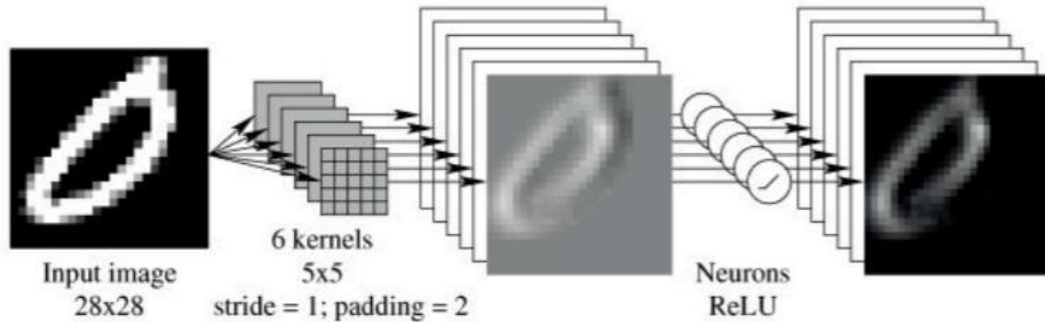


Рисунок 2.8 –Згортковий шар

Під час прямого поширення працює кожен фільтр згортки на вхідному об'ємі по ширині і висота та обчислення крапкового добутку між записами фільтра та входом у будь-якій позиції, за цією операцією слідує нелінійна функція активації (сигмовидна, тангенсіальна, ReLU тощо), результуючі виходи є називаються картами ознак.

Карта об'єктів (також відома як карта активації), дає відповіді фільтра на кожну просторове положення. Приклад шару згортки з подальшою нелінійною активацією показано на рисунку 2.8. Прокладаємо ці карти активації вздовж виміру глибини і виробляємо вихідний обсяг.

Глибина вихідного обсягу являє собою кількість фільтрів, які використовуються в операції згортки. Кожен фільтр вивчає щось своє у введених даних, краях, крапках, кольорах.

Крок згортки – це кількість ітерацій, які використовуємо у вході. Коли крок дорівнює 1, тоді переміщаємо фільтри на один піксель за раз. Коли крок дорівнює 2, тоді фільтри стрибають на 2 пікселі, коли обчислюємо їх. Це буде виробляти менші обсяги випуску в просторі.

Заповнення дозволяє керувати розміром виводу. Застосовуючи згортку до входу, зменшуйте розмір виводу, що призводить до втрати інформації.

Доповнюємо вхідний обсяг нулями навколо кордон. Два загальних варіанти: дійсна згортка та сама згортка. Дійсна згортка означає відсутність заповнення, сама згортка означає, що вихідний розмір залишається таким який самий, як вхідний розмір.

Вихідний розмір обчислюється таким чином:

$$(n + 2p - f) / s + 1, \quad (2.6)$$

де n – кількість фільтрів;

p – кількість заповнення;

f – розмір фільтра;

s – крок.

Об'єднуючий шар. CNN часто використовують роботу рівня об'єднання після шарів згортки, їх функція полягає в зменшенні розмірності, також згадується як підвибірка.

Гіперпараметри шару об'єднання представляють розмір фільтра і кроки. Найчастіше використовується шар об'єднання з фільтром розмір 2 і з кроком 2. Два поширені типи об'єднання шари – максимальне об'єднання та середнє об'єднання, де береться максимальне та середнє значення відповідно.

Максимальне об'єднання використовується частіше, ніж середнє об'єднання. Об'єднуючий шар не має параметрів для навчання [37]. Інтуїція чого максимальне об'єднання означає те, що велике число означає це може бути виявлена функція. Приклад згорткового шару, за яким слідує об'єднуючий шар, показаний на рисунку 2.9.

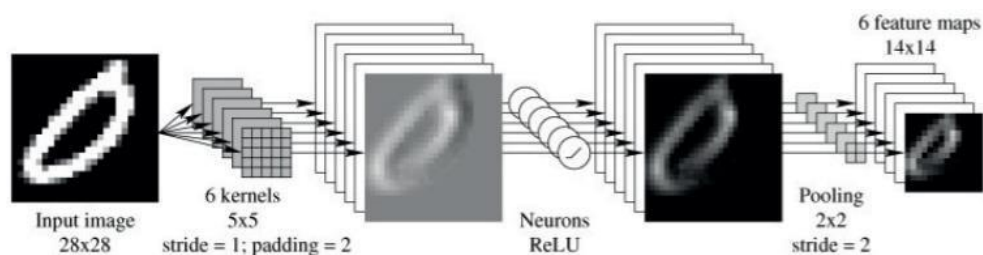


Рисунок 2.9 – Згортковий шар після шару об'єднання

Повністю пов'язаний шар. Після кількох шарів згортки та об'єднання CNN зазвичай закінчується декількома повністю пов'язаними шарами. Тензор, який маємо на виході цих шарів перетворюється у вектор, а потім додаємо кілька шарів нейронної мережі.

Повністю підключені шари зазвичай є останніми кількома шарами архітектури, як показано на рисунку 2.10 – методика регуляризації випадання. Можна наносити в повністю з'єднані шари для запобігання перевантаження. Останній повністю підключений шар в архітектурі містить таку ж кількість вихідних нейронів, що й кількість класів, які підлягають визнанню [2-6].

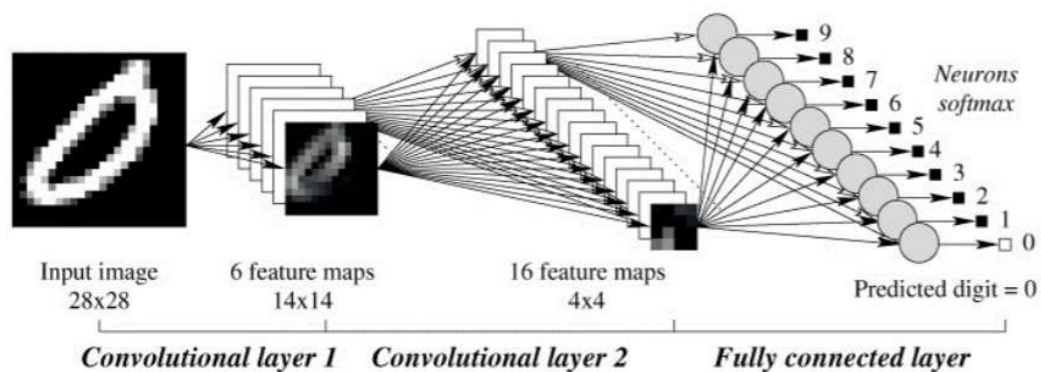


Рисунок 2.10 – Два згорткових шари які йдуть слідом за одним повністю пов'язаним шаром

2.2.2 Сучасні архітектури згорткових нейронних мереж

Нейронні мережі отримують вхідні дані (один вектор) і перетворюють його через серію прихованих шарів. Кожен прихований шар складається з набору нейронів, де кожен нейрон повністю пов'язаний з усіма нейронами в попередньому шарі, і де нейрони в одному шарі функціонують повністю незалежно і не мають спільних зв'язків.

Останній повністю підключений шар називається «вихідним шаром», і в налаштуваннях класифікації він представляє бали класу. Нижче розглянуто основні та найпопулярніші архітектури згорткових нейронних мереж.

GoogLeNet – це 22-шарова глибока згортка нейронна мережа, яка є варіантом Inception Network, глибокої згорткової нейронної мережі, розробленої дослідниками Google.

Архітектура GoogLeNet, представлена в ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC14), вирішує задачі комп'ютерного зору, такі як класифікація зображень і виявлення об'єктів.

Сьогодні GoogLeNet використовується для інших завдань комп'ютерного зору, таких як виявлення та розпізнавання облич, змагання тощо.

Усі згортки, в тому числі всередині модулів Inception, використовують випрямлену лінійну активацію.

Розмір сприйнятливих полів в нашій мережі 224×224 в кольорі RGB простір з нульовим середнім. «Зменшити № 3×3 » та «Зменшити № 5×5 » означає кількість фільтрів 1×1 у зменшувачому шарі використовувався до згорток 3×3 і 5×5 .

Можна побачити кількість фільтрів 1×1 в проєкційному шарі після вбудований максимальний пул у стовпці pool proj (рис. 2.11).

Усі ці редуційні / проєкційні шари використовують як випрямлену лінійну активацію.

Мережа була розроблена з обчислювальною ефективністю і практичністю, так що висновки можна виконувати на окремих пристроях, включаючи навіть ті з обмеженими обчислювальними ресурсами, особливо з низьким обсягом пам'яті.

Мережі GoogLeNet навчалися за допомогою розподіленої системи машинного навчання DistBelief [7-12] з використанням невеликої кількості моделей і паралелізму даних.

type	patch size/ stride	output size	depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	pool proj	params	ops
convolution	7×7/2	112×112×64	1							2.7K	34M
max pool	3×3/2	56×56×64	0								
convolution	3×3/1	56×56×192	2		64	192				112K	360M
max pool	3×3/2	28×28×192	0								
inception (3a)		28×28×256	2	64	96	128	16	32	32	159K	128M
inception (3b)		28×28×480	2	128	128	192	32	96	64	380K	304M
max pool	3×3/2	14×14×480	0								
inception (4a)		14×14×512	2	192	96	208	16	48	64	364K	73M
inception (4b)		14×14×512	2	160	112	224	24	64	64	437K	88M
inception (4c)		14×14×512	2	128	128	256	24	64	64	463K	100M
inception (4d)		14×14×528	2	112	144	288	32	64	64	580K	119M
inception (4e)		14×14×832	2	256	160	320	32	128	128	840K	170M
max pool	3×3/2	7×7×832	0								
inception (5a)		7×7×832	2	256	160	320	32	128	128	1072K	54M
inception (5b)		7×7×1024	2	384	192	384	48	128	128	1388K	71M
avg pool	7×7/1	1×1×1024	0								
dropout (40%)		1×1×1024	0								
linear		1×1×1000	1							1000K	1M
softmax		1×1×1000	0								

Рисунок 2.11 – Характеристика архітектури GoogLeNet [20]

Хоча використано лише реалізацію на основі ЦП, приблизна оцінка припускає, що мережу GoogLeNet можна навчитиконвергенція з використанням кількох високоякісних графічних процесорів протягом тижня.

Основним обмеженням є використання пам'яті. Наше навчання використано асинхронний стохастичний градієнтний спуск з імпульсом 0,9, фіксований графік швидкості навчання (зменшення швидкості навчання на 4% кожні 8 епох) [13-15].

Методи вибірки зображень істотно змінилися протягом місяців, що передують змаганню, і вже конвергентні моделі тренувалися з іншими параметрами, іноді в поєднанні зі зміненими гіперпараметрами, наприклад як відсівання та швидкість навчання.

Тому це важко дати чіткі вказівки щодо найбільш ефективного єдиного способу для навчання цих мереж. Щоб ще більше ускладнити справу, деякі моделі в основному навчалися на менших відносних культурах, інші на більші, натхненні.

Тим не менш, один випадок, який було підтверджено, що він працює дуже добре, включає вибірку фрагментів зображення різного розміру, розмір яких рівномірно розподілений від 8% до 100% області зображення зі співвідношенням сторін, обмеженим інтервалом $[3/4, 4/3]$. Також виявили, що фотометричні спотворення [15-20].

Залишкова мережа (ResNet) стала переможцем ILSVRC 2015, вона має загалом 152 шари. ResNet побудований із залишкового блоку, який показаний на рисунку 2.12, шляхом укладання залишкових блоків разом, кожен залишковий блок має два шари згортки 3×3 , періодично подвійні кількість фільтрів і зменшити дискретизацію просторово за допомогою кроку.

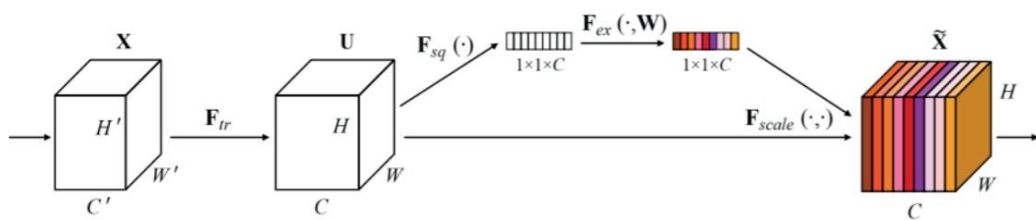


Рисунок 2.12 – Залишковий блок ResNet

ResNet має спеціальні з'єднання пропуску та використання пакетної нормалізації після кожної згортки шар. Більш глибокі моделі важче оптимізувати, рішенням є використання skip-з'єднання, що дозволяє взяти активація з одного шару і подача його на інший шар. Використання цього дозволяє тренувати дуже глибокі мережі та уникнути проблеми зникнення та вибуху градієнта. Зменшують кількість параметрів, як це роблять ResNets.

ResNet є першим архітектурою, яка має кращу продуктивність, ніж людська продуктивність.

CNN краще, ніж інші методи глибокого навчання в додатках комп'ютерного зору, він дає найкращу якість зображення проблеми з розпізнаванням і навіть перевершує людей певні випадки.

Ось чому у цій роботі для розпізнання емоцій буде використовуватися саме ця нейронна мережа.

2.3 Моделювання структури та наповнення програмного застосунку для ідентифікації емоцій на зображеннях

Розвиток комп'ютерного зору та машинного навчання робить розпізнавання емоцій набагато більш точним та доступним для широкої аудиторії. Розпізнавання емоцій за виразом обличчя (facial expression recognition, face computing) є підобластю технології обробки зображень.

У найбільш широкому сенсі «читання» осіб дозволяє компаніям виявляти емоції людей, які проходять повз їхні камери. Його можна використовувати в маркетингу, медицині, робототехніці – загалом у будь-якій галузі, де потрібне глибоке розуміння емоційної реакції людини на певні дії. Технологія розпізнавання осіб може застосовуватися в системах безпеки, включаючи контроль доступу, автентифікацію, верифікацію платежів, а також під час інтерв'ю або допитів [21].

Emotion AI в системах розпізнавання зчитує вирази обличчя за допомогою будь-якого оптичного датчика, наприклад звичайної вебкамери або камери смартфона, визначаючи обличчя на відео в режимі реального часу, на запису або зображеннях. Алгоритми комп'ютерного зору фіксують основні точки людського обличчя: очі, кінчик носа, брови, куточки рота і відстежують їх рух для розшифровки емоцій.

Зіставляючи зібрані дані із зразками з бази зображень, програма для розпізнавання може визначати почуття людини щодо поєднання виразів його обличчя. Просунуті рішення в галузі емоційного штучного інтелекту, такі як Affectiva або Kairos, вміють відслідковувати такі показники емоцій: радість, смуток, гнів, зневага, огида, страх і подив.

Додаткові програмні функції можуть включати ідентифікацію та верифікацію по зображенню особи, визначення віку та статі, етнічної приналежності, наявність на фото кількох людей та багато іншого [22-24].

Класифікація зображень передбачає виділення ознак із зображення для спостереження за деякими закономірностями в наборі даних. Використання ANN з метою класифікації зображень в кінцевому підсумку буде дуже дорогим з точки зору обчислень, оскільки параметри, які можна навчати, стають надзвичайно великими.

Наприклад, якщо є зображення кішки розміром 50×50 , і хочемо навчити нашу традиційну ANN на цьому зображенні, щоб класифікувати його на собаку або кішку, параметри, які можна навчати, стають – $(50 \times 50) \times 100$ пікселів зображення, помножені на прихований шар + 100 зміщення + 2×100 вихідних нейронів + 2 зміщення = 2,50,302 [25-31].

Використовуємо фільтри при використанні CNN. Фільтри бувають різних типів відповідно до їх призначення (рис. 2.13).

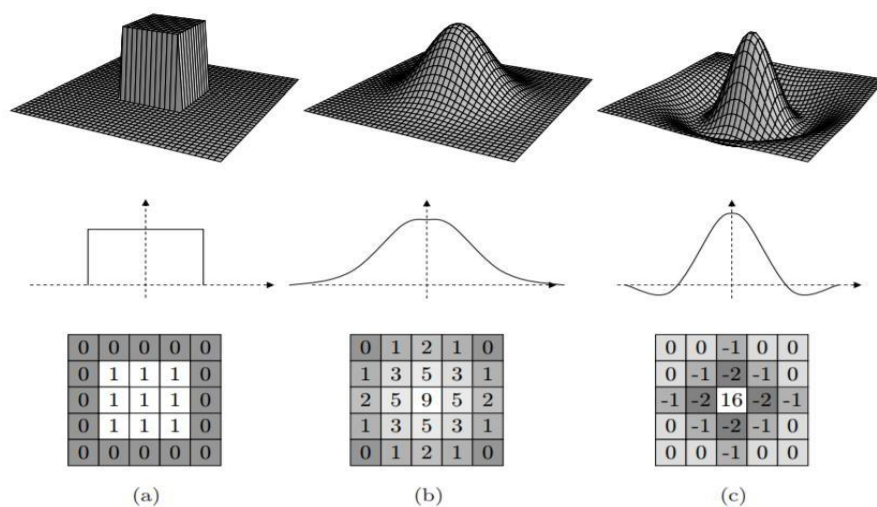


Рисунок 2.13 – Приклади різних фільтрів та їх ефекти

Фільтри допомагають нам використовувати просторову локальність певного зображення, забезпечуючи локальний шаблон зв'язку між нейронами. Згортка в основному означає поточкове множення двох функцій для отримання третьої функції.

Тут одна функція – це матриця пікселів зображення, а інша – фільтр. Ковзаємо фільтром по зображенню і отримуємо крапковий добуток двох матриць [32-33].

Алгоритм розпізнання емоцій на зображеннях за допомогою нейронної мережі CNN:

Крок 1. Обрати датасет який Вас цікавить, або розробити свій датасет для вирішення власної проблеми класифікації зображень. В даніх роботі використовується датасет fer2013.

Крок 2. Підготувати набір даних для навчання.

Крок 3. Створення даних про тренування. Навчання – це масив, який міститиме значення пікселів зображення та індекс.

Крок 4. Нормалізація X і перетворення міток у категоріальні дані.

Крок 5. Розділити X і Y для використання в CNN.

Крок 6. Визначення, компіляція та навчання моделі CNN.

CNN краще, ніж інші методи глибокого навчання в застосунках комп'ютерного зору, він дає найкращу якість зображення проблеми з розпізнаванням і навіть перевершує людей певні випадки. Початковий модуль разом із залишковим мереж, покращив продуктивність CNN і представив нові можливості [34-37].

3 РОЗРОБЛЕННЯ ПРОГРАМНОГО ЗАСТОСУНКУ ДЛЯ ІДЕНТИФІКАЦІЇ ЕМОЦІЙ НА ЗОБРАЖЕННЯХ

3.1 Вибір інструментальних засобів для реалізації поставленої задачі

У ролі допоміжної бібліотеки для розроблення застосунку було використано OpenCV – бібліотеку комп'ютерного зору та обробки зображень з відкритим кодом, яка має надзвичайно потужне функціональне наповнення.

У бібліотеці є мноржина функцій та алгоритмів для вирішення різноманітних задач, пов'язаних із роботою з зображеннями, серед яких:

- засоби зчитування, зберігання та запису зображень;
- колірні та геометричні перетворення зображень;
- ядра згортки;
- відстежування руху;
- робота з гістограмами;
- розпізнання об'єктів;
- машинне навчання;
- визначення ключових точок та інше.

Мовою програмування було обрано Python, одну з найпопулярніших мов програмування в світі, особливо, коли мова йде про штучний інтелект, або машинне навчання.

Основними перевагами Python є низький порог входження, ця мова дуже проста в освоєнні і оволодіти нею на базовому рівні можуть люди за мінімальним знанням концепцій програмування.

Також, як було вказано вище, Python дуже підходить для розробки програм зі штучним інтелектом через велику кількість зручних бібліотек для розробки штучного інтелекта. Наприклад: Matplotlib, Numpy, Scikit-learn, Tensorflow.

Також ця мова ідеально підходить для роботи з великими обсягами даних, має найширший набір пакетів і розширень для автоматизації роботи.

Незважаючи на свою доступність, Python – це серйозна мова. Вона вирішує промислові проблеми. Інженерія може масштабуватися завдяки своєму багатому макету, гарній структурі та модульності.

Як середовище розробки було обрано Microsoft Visual Studio Code – це редактор коду, розроблений для Windows, Linux та macOS. Крім того, VSC також поєднує в собі простоту редактора та потужні інструменти розробника, такі як налагодження, вбудоване управління Git, підсвічування синтаксису та багато іншого. Безкоштовний кросплатформовий редактор з відкритим вихідним кодом, який працює у Windows, Linux та macOS, тому можна працювати незалежно від платформи, на якій пристрій заснований.

VS Code підтримує мову за замовчуванням, яка буде відповідати певному файлу, але також можете змінити мову. Для цього натисніть індикатор мови, розташований праворуч від рядка стану, і з'явиться список вибору мовного режиму. Можна вибрати різні мови для поточного файлу.

VS Code поставляється з вбудованим налагоджувачем, який також є однією з основних функцій. Це допомагає прискорити цикл редагування, компіляції та налагодження. Однак за замовчуванням він поставляється тільки з відладчиком, який підтримує NodeJS, який може налагоджувати все, що було перенесено в JavaScript, але знову ж таки можна використовувати розширення для інших середовищ виконання.

3.2 Етапи розроблення програмного застосунку для ідентифікації емоцій на зображеннях

Зрозроблюваний застосунок складається з 2 частин – частина, що відповідає за навчання моделі, та частина, в якій відбувається класифікація нових даних.

Вирішено реалізувати можливість роботи як з даними з відеокамери. Це стосується як процесу навчання, так і процесу розпізнавання.

Починається робота програми з того, що на екрані перед користувачем з'являється 2 вікна. В одному запис з відеокамери, в іншому жива статистика по виявленим у людині емоціям.

Також в ході розробки за допомогою бібліотек мови програмування Python, було створено CNN (згорткову нейронну мережу). Спочатку тренувальне зображення проходить через усю мережу. На цьому етапі ваги встановлені довільно, тому на виході мережа визначить трохи ознак і не зможе точно відповісти, що зображено на зображенні. Тут можна оцінити функцію втрат – наскільки відрізняється фактичний вихід алгоритму прогнозованого.

Далі ваги нейронів коригують, починаючи з найближчих до виходу. Це дозволяє зменшити функцію втрат і називається методом зворотного розповсюдження помилки. Один прохід датасета в обох напрямках називають епохою. Універсальної кількості епох немає. Це ітеративний процес, який впливає швидкість навчання. Висока швидкість означає, що ваги змінювали сильніше, і для налаштування алгоритму потрібно менше часу. Але тоді алгоритм може бути недостатньо точним.

І навпаки, занадто багато епох призведе до перенавчання, коли алгоритм запам'ятає приклади датасетів, але зможе узагальнювати дані. Завдання навчання – створити систему, здатну будувати прогнози на основі реальних даних, яких не було в тренувальному датасеті.

Варто також зазначити, що для оптимізації роботи застосування на кожному зображенні виділяється тільки одне обличчя, яке займає найбільшу площу (очікується, що в такому випадку людина знаходиться ближче до камери). На рисунку 3.1 зображено фрагмент коду для знаходження облич на камері.

Після знаходження обличчя на камері починається процес виявлення на обличчі емоцій. На основі датасету fer2013 починається процес навчання нейронної мережі.

```

cv2.namedWindow('your_face')
camera = cv2.VideoCapture(0)
while True:
    frame = camera.read()[1]
    frame = imutils.resize(frame,width=300)
    gray = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)
    faces = face_detection.detectMultiScale(gray,scaleFactor=1.1,
    minNeighbors=5,minSize=(30,30),
    flags=cv2.CASCADE_SCALE_IMAGE)

    canvas = np.zeros((250, 300, 3), dtype="uint8")
    frameClone = frame.copy()
    if len(faces) > 0:
        faces = sorted(faces, reverse=True,
        key=lambda x: (x[2] - x[0]) * (x[3] - x[1]))[0]
        (fx, fy, fw, fh) = faces
        roi = gray[fy:fy + fh, fx:fx + fw]
        roi = cv2.resize(roi, (64, 64))
        roi = roi.astype("float") / 255.0
        roi = img_to_array(roi)
        roi = np.expand_dims(roi, axis=0)
        preds = emotion_classifier.predict(roi)[0]
        emotion_probability = np.max(preds)
        label = EMOTIONS[preds.argmax()]
    else: continue

```

Рисунок 3.1 – Фрагмент коду для знаходження облич на камері

На початковому етапі нейронна мережа є ненавченою (неналаштованою).

У загальному сенсі під навчанням розуміють послідовне пред'явлення образу на вхід нейромережі, з навчального набору, потім отримана відповідь порівнюється з бажаним виходом, у нашому випадку це 1 – образ представляє особу, мінус 1 – образ представляє фон (не особа), отримана різниця між очікуваним відповіддю та отриманим є результатом функції помилки (дельта помилки). Потім цю дельту помилки необхідно поширити на пов'язані нейрони мережі.

Таким чином, навчання нейронної мережі зводиться до мінімізації функції помилки, шляхом коригування вагових коефіцієнтів синаптичних зв'язків між нейронами. Під функцією помилки розуміється різниця між отриманою відповіддю та бажаною. Для нейронів вихідного шару відомі їх фактичні та бажані значення виходів. Тому налаштування ваг зв'язків для таких нейронів є простим.

Однак для нейронів попередніх шарів налаштування не настільки очевидне. Фрагмент коду із створенням згорткової нейронної мережі для даної роботи зображено на рисунках 3.2 та 3.3


```
def simple_CNN(input_shape, num_classes):

    model = Sequential()
    model.add(Convolution2D(filters=16, kernel_size=(7, 7), padding='same',
                            name='image_array', input_shape=input_shape))
    model.add(BatchNormalization())
    model.add(Convolution2D(filters=16, kernel_size=(7, 7), padding='same'))
    model.add(BatchNormalization())
    model.add(Activation('relu'))
    model.add(AveragePooling2D(pool_size=(2, 2), padding='same'))
    model.add(Dropout(.5))
    model.add(Convolution2D(filters=32, kernel_size=(5, 5), padding='same'))
    model.add(BatchNormalization())
    model.add(Convolution2D(filters=32, kernel_size=(5, 5), padding='same'))
    model.add(BatchNormalization())
    model.add(Activation('relu'))
    model.add(AveragePooling2D(pool_size=(2, 2), padding='same'))
    model.add(Dropout(.5))
    model.add(Convolution2D(filters=64, kernel_size=(3, 3), padding='same'))
    model.add(BatchNormalization())
    model.add(Convolution2D(filters=64, kernel_size=(3, 3), padding='same'))
    model.add(BatchNormalization())
    model.add(Activation('relu'))
    model.add(AveragePooling2D(pool_size=(2, 2), padding='same'))
    model.add(Dropout(.5))
```

Рисунок 3.2 – Частина коду для створення згорткової нейронної мережі

```
model.add(Convolution2D(filters=128, kernel_size=(3, 3), padding='same'))
model.add(BatchNormalization())
model.add(Convolution2D(filters=128, kernel_size=(3, 3), padding='same'))
model.add(BatchNormalization())
model.add(Activation('relu'))
model.add(AveragePooling2D(pool_size=(2, 2), padding='same'))
model.add(Dropout(.5))
model.add(Convolution2D(filters=256, kernel_size=(3, 3), padding='same'))
model.add(BatchNormalization())
model.add(Convolution2D(filters=num_classes, kernel_size=(3, 3), padding='same'))
model.add(GlobalAveragePooling2D())
model.add(Activation('softmax', name='predictions'))
return model
```

Рисунок 3.3 – Частина коду для створення згорткової нейронної мережі

Наступним кроком роботи програми є виведення на екран діаграми, яка характеризує емоції, які зараз на обличчі (рис. 3.4). Тобто, програма показує не одну емоцію, а їх спектр, які вона виявила на обличчі.

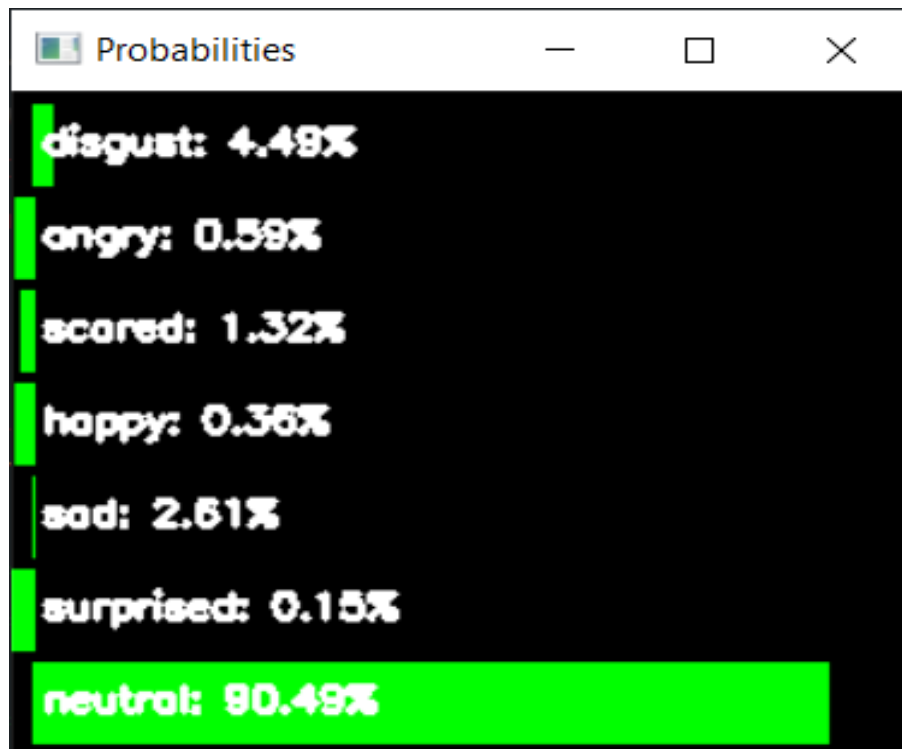


Рисунок 3.4 – Діаграма емоційного спектру в даний момент часу

3.3 Тестування реалізованого програмного застосунку та аналіз результатів

У тестуванні застосунку загалом взяло участь п'ятеро людей. Обличчя кожного з них було розпізнане програмним застосунком, та був зроблений аналіз емоцій. Аналіз був проведений на основі датасету fer2013.

Треба зауважити, що якість тестування залежить не тільки від дієздатності застосунку, але і від якості пристрою для захоплення обличчя, тобто вебкамери. Оскільки тестування проводилося на різних пристроях, і точність виявлення емоцій різна тільки через те, що на різних пристроях может бути різна чіткість захоплення зображення (рис. 3.5 – 3.7).

Процес виявлення емоцій для кожного обличчя відбувався у різних умовах, різному освітленні і т.д. Декілька знімків процесу тестування наведено у додатку А.

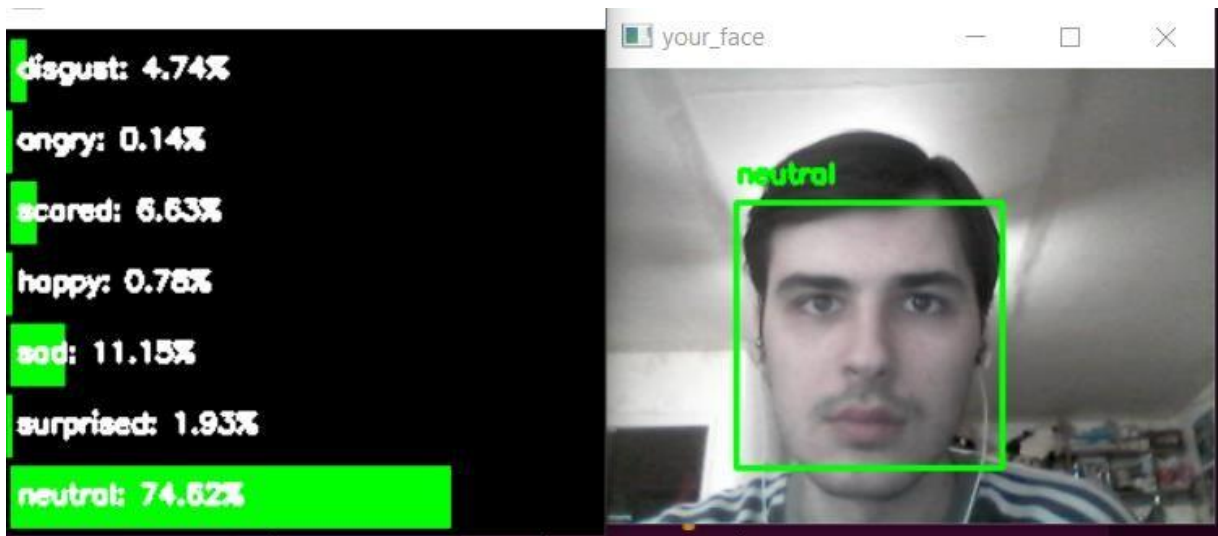


Рисунок 3.5 – Процес визначення емоції

З наведеного прикладу видно, що коли система отримує на вхід обличчя, вона вірно визначає його межі, потім відбувається процес обробки зображення і визначення емоцій.

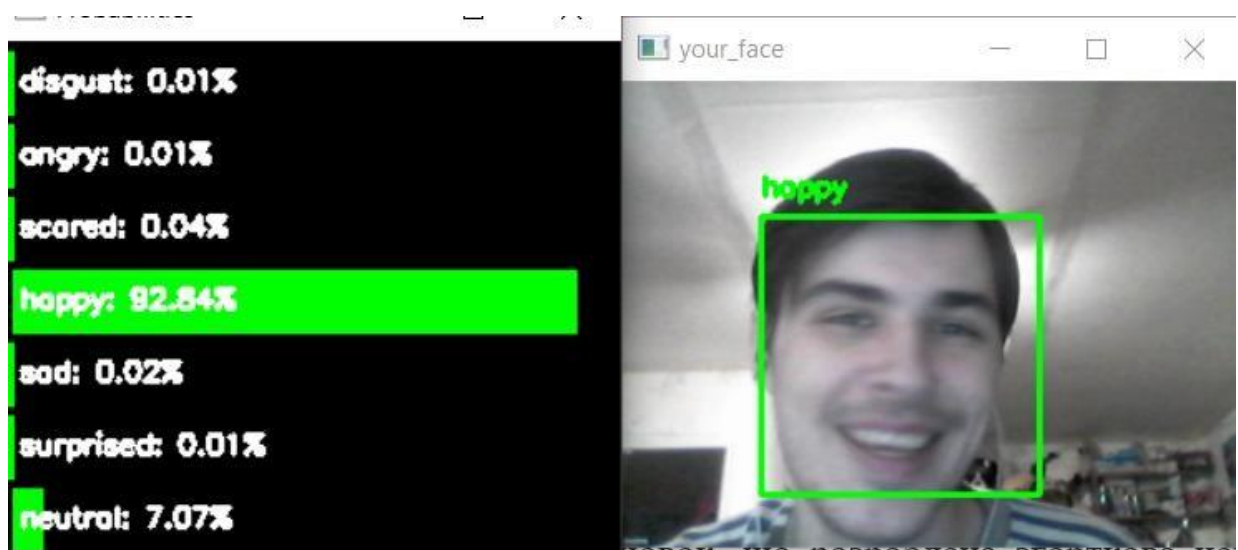


Рисунок 3.6 – Процес визначення емоції

Якщо казати про загальні результати роботи, то, проаналізувавши процес випробування застосунку, можна сказати наступне. В тих випадках, коли розпізнавання обличчя відбувалося в тих самих умовах, які є ідеальними для будь-яких операцій з будь-якою відеокамерою, тобто, ідеальне освітлення, потужний пристрій для виводу зображення на екран застосунку

практично завжди вірно визначав емоції людини, навіть з невеликим повтором голови, й лише в деяких випадках були складності з виявленням емоцій, скоріше пов'язані з недотриманням тих умов, що є описані вище в деяких кадрах давав невірні відповіді. Проте, коли значно змінювались умови освітлення, то точність розпізнання сильно падала й система здебільшого видавала хибні результати.



Рисунок 3.7 – Процес визначення емоцій в темряві

З цього можна зробити висновок, що розроблена згорткова нейрона мережа є дуже чутлива до освітлення, саме через те, що у датасеті, який використовували усі зображення є чорнобілими. Однак з рівномірним освітленням (бажано світлим), проблема неточності розпізнання більш пов'язана з особливостями прояву емоцій в кожній людині.

3.4 Перспективи подальшої роботи

Розроблений застосунок задовільно виконує поставлену задачу, модель може навчатися на еталонних зображеннях, які отримані з конкретного датасету, де багато емоцій можна просто описати одним словом, кривлянням.

Це один з найкращих датасетів для того, щоб навчати нейронну мережу розпізнавати емоції. Також, треба зауважити, що, застосунок хоч і виявляє емоції, і виявляє їх досить точно, але він не враховує, що емоції вкрай суб'єктивна річ, і те, що на обличчі однієї людини може казатись невдоволенням, або, навпаки, задоволення, насправді може бути просто нейтральною емоцією, яка для людини є звичайною у спокійному, беземоційному стані. Це також застосунком не враховується.

Тестування застосунку показало непогані результати, але й визначило слабкі місця, для дослідження й, в перспективі, усунення яких необхідно проводити масштабні вивчення.

Отримані результати показують, що реалізований метод в цілому є перспективним і заслуговує на дальший розвиток. Підхід визначення текстурних ознак обличчя дає стійкі ознаки для класифікації і багато можливостей для розширення.

Серед можливих покращень может бути більша суб'єктивізація емоцій для тренування та зробити сам процес розпізнання більш точним. Але, не дивлячись на всі недоліки, кінцева мета була виконана, застосунок працює так, як і було заплановано на стадії планування розробки.

ВИСНОВКИ

У рамках кваліфікаційної роботи був розроблений і реалізований застосунок для розпізнання емоцій на зображеннях.

Виконано всі поставлені задачі, а саме:

– досліджено методи визначення ключових точок на зображеннях, розібрано поняття ключових точок та кутів, надано огляд деяких найвідоміших методів знаходження ключових точок, визначено їх переваги та недоліки, розглянуто проблеми, пов'язані з визначенням ключових точок;

– проаналізовані методи розпізнання емоцій;

– розроблено методику ідентифікації емоцій на зображеннях;

– визначено необхідні інструментальні засоби для створення застосунку;

– виконано всі етапи розроблення застосунку для ідентифікації емоцій на зображеннях, на кожному етапі здійснено різноманітні програмні та алгоритмічні оптимізації, реалізовано гнучкий метод роботи застосунку, його взаємодії з користувачем;

– проведено тестування розробленого застосунку, за результатами якого було визначено особливості роботи реалізованого методу, умови використання та його ефективність, яка виявилася на даному етапі задовільною, але водночас надавала розуміння необхідності ширших випробувань та можливого подальшого доопрацювання;

– визначено перспективи подальшої роботи, які виходять з отриманих результатів тестування та теоретичних знань в досліджуваній області.

Результати роботи апробовано у вигляді тез доповідей під час XVIII Міжнародної науково-практичної конференції «Advancing in research, practice and education», 10-13 травня 2022 р., Флоренція, Італія [38].

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Гороховатський В.О., Творошенко І.С. (2021) Методи інтелектуального аналізу та оброблення даних: навч. посібник. Харків: ХНУРЕ, 92 с.
2. Gorokhovatskyi V., Rusakova N., and Tvoroshenko I. (2020) The application of image analysis methods and predicate logic in applied problems of magnetic monitoring, *Telecommunications and Radio Engineering*, 79(20), pp. 1801-1811.
3. Кобилін О.А., Творошенко І.С. (2021) Методи цифрової обробки зображень: навч. посібник. Харків: ХНУРЕ, 124 с.
4. Tvoroshenko I., Ahmad M.A., Mustafa S.K., Lyashenko V., and Alharbi A. R. (2020) Modification of Models Intensive Development Ontologies by Fuzzy Logic, *International Journal of Emerging Trends in Engineering Research*, 8(3), pp. 939-944.
5. Кучеренко Е.И., Творошенко И.С. (2003). Процессы принятия решений в сложных системах на основе нечетких интервальных представлений. *Вісник Національного технічного університету «ХПІ»*. Тематичний випуск: Системний аналіз, управління та інформаційні технології. Х.: НТУ «ХПІ», 1(7), С. 79-86.
6. Творошенко И.С. (2010) Анализ процессов принятия решений в интеллектуальных системах, *Системи обробки інформації*, 2, С. 248-253.
7. Daradkeh Y.I., Tvoroshenko I., Gorokhovatskyi V., Latiff L.A., and Ahmad N. (2021) Development of Effective Methods for Structural Image Recognition Using the Principles of Data Granulation and Apparatus of Fuzzy Logic, *IEEE Access*, 9, pp. 13417-13428.
8. Tvoroshenko I., and Gorokhovatskyi V. (2022) The Application of Hybrid Intelligence Systems for Dynamic Data Analysis, *International Journal of Engineering and Information Systems*, 6(2), pp. 40-48.

9. Гороховатський В.О., Творошенко І.С. (2022) Аналіз багатовимірних даних за описом у формі множини компонент: монографія. Харків: ХНУРЕ. 124 с.
10. Flah P. (2015) *Machine learning. The science and art of building algorithms that extract knowledge from data*, Moscow, Russia: DMK Press, 400 p., (in Russian).
11. Lyashenko V., Mustafa S.K., Tvoroshenko I., and Ahmad M.A. (2020) Methods of Using Fuzzy Interval Logic During Processing of Space States of Complex Biophysical Objects, *International Journal of Emerging Trends in Engineering Research*, 8(2), pp. 372-377.
12. Творошенко І.С. (2021) Технології прийняття рішень в інформаційних системах: навч. посібник. Харків: ХНУРЕ, 120 с.
13. Ahmad M.A., Tvoroshenko I., Baker J.H., Kochura L., Lyashenko V. (2020) Interactive Geoinformation Three-Dimensional Model of a Landscape Park Using Geoinformatics Tools, *International Journal on Advanced Science, Engineering and Information Technology*, 10(5), pp. 2005-2013.
14. Daradkeh Y.I., Gorokhovatskyi V., Tvoroshenko I., Gadetska S., and Al-Dhaifallah M. (2021) Methods of Classification of Images on the Basis of the Values of Statistical Distributions for the Composition of Structural Description Components, *IEEE Access*, 9, pp. 92964-92973.
15. Mikel de Velasco, M., Justo, R., & Inés Torres, M. (2022). Automatic identification of emotional information in Spanish TV debates and human-machine interactions.
16. Nahla Nour, Mohammed Elhebir, Serestina (2020) Face Expression Recognition using Convolution Neural Network (CNN) Models.
17. O. Vinyals, A. Toshev, S. Bengio, and D. Erhan (2015) "Show and tell: A neural image caption generator," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3156-3164.
18. A. Krizhevsky, I. Sutskever, and G.E. Hinton (2012) "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, pp. 1097-1105.

19. Tvoroshenko, I. (2019). Development of models of spatial analysis of status of interactive processes of complex systems.
20. Scherer, K.R. (1984) Approaches to Emotion. Chapter: On the Nature and Function of Emotion: A Component Process Approach; Scherer, K.R., Ekman, P., Eds.; Taylor and Francis Group: New York, NY, USA.
21. Tvoroshenko I., and Zarivchatskyi R. (2020) Analysis of existing methods for searching object in the video stream, *Abstracts of VI International Scientific and Practical Conference «About the problems of science and practice, tasks and ways to solve them» (October 26-30, 2020). Milan, Italy*, pp. 500-505.
22. T.S. Huang. (2018) Computer Vision: Evolution and Promise.
23. A. Toshev, Ch. Szegedy (2014) “DeepPose: Human Pose Estimation via Deep Neural Networks,” *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1653-1660.
24. Творошенко, І. С., & Табашник, В. А. (2018). Розробка просторової моделі геоінформаційної підтримки людей з обмеженими можливостями, що пересуваються на інвалідних колясках, у місті Харків. *Збірник наукових праць Харківського національного університету Повітряних Сил*, (1), 122-128.
25. Fukushima, K. (1980) “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,” *Biological Cybernetics*, pp. 193–202
26. Tvoroshenko I., and Tkachenko D. (2020) Mechanisms of image classification based on descriptors of local features, *Abstracts of IV International Scientific and Practical Conference «Integration of scientific bases into practice» (October 12-16, 2020). Stockholm, Sweden*, pp. 443-448.
27. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014) “Dropout: a simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, Volume 15, pp. 1929-1958.
28. LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (2015) “Deep learning,” *nature* 521.7553, pp. 436.

29. Russakovsky, Olga et al. (2015) "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, 115, pp. 211-252.
30. Tvoroshenko I., and Dziubenko M. (2020) Modern methods of analysis of the movement scheme using video detection of vehicles, *Abstracts of V International Scientific and Practical Conference «Study of modern problems of civilization» (October 19-23, 2020). Oslo, Norway*, pp. 422-428.
31. Crowder, J. A. (July 2011) "The Artificial Prefrontal Cortex: Artificial Consciousness," *International Conference on Artificial Intelligence, ICAI'11*.
32. Nass, C., and Moon, Y. (2018) "Machines and mindlessness: Social responses to computers," *Journal of Social Issues*, 56(1), 81-103.
33. Ashcroft, M. (1997) "Human Memory and Cognition". Prentice Hall Professional.
34. Szegedy, C., Wei Liu, Yangqing Jia, Sermanet et al. (2015) "Going deeper with convolutions," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
35. Treen, G. (n.d.) (2017). Methods for faster feature matching using the scale-invariant feature transform.
36. Ahmad M.A., Gorokhovatskyi V., Tvoroshenko I., Vlasenko N., Mustafa S.K. (2021) The Research of Image Classification Methods Based on the Introducing Cluster Representation Parameters for the Structural Description, *International Journal of Engineering Trends and Technology*, 69(10), pp. 186-192.
37. Daradkeh Y.I., Gorokhovatskyi V., Tvoroshenko I., and Al-Dhaifallah M. (2022) Classification of Images Based on a System of Hierarchical Features, *Computers, Materials & Continua*, 72(1), pp. 1785-1797.
38. Помазан В. (2022) Аналіз технологій ідентифікації, розпізнавання та оброблення емоцій людини на зображеннях, *Abstracts of XVIII International Scientific and Practical Conference «Advancing in research, practice and education» (May 10 – 13, 2022). Florence, Italy*, pp. 645-648.