

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ ТЕЛЕКОМУНІКАЦІЙ**

**НАВЧАЛЬНО–НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ  
ТЕХНОЛОГІЙ**

**Кафедра інженерії програмного забезпечення**

**Пояснювальна записка**

**до магістерської роботи**

**на ступінь вищої освіти магістр**

**на тему: «РОЗРОБКА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ДЛЯ  
КЛАСИФІКАЦІЇ ТА ПОШУКУ ОБ’ЄКТІВ ЗОБРАЖЕННЯ ЗА ДОПОМОГОЮ  
МАШИННОГО НАВЧАННЯ»**

**Виконав: студент 6 курсу, групи ПДМ–61  
спеціальності 121 Інженерія програмного забезпечення**

**Сьомін Д.А.**

(прізвище та ініціали)

**Керівник**

**Негоденко О.В.**

(прізвище та ініціали)

**Рецензент**

(прізвище та ініціали)

**Київ – 2022**

# ДЕРЖАВНИЙ УНІВЕРСИТЕТ ТЕЛЕКОМУНІКАЦІЙ

## НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

Кафедра Інженерії програмного забезпечення

Ступінь вищої освіти - «Магістр»

Спеціальність підготовки – 121 «Інженерія програмного забезпечення»

**ЗАТВЕРДЖУЮ**

Завідувач кафедри

Інженерії програмного забезпечення

Негоденко О.В.

“ \_\_\_ ” \_\_\_\_\_ 2022 року

### ЗАВДАННЯ

#### НА МАГІСТЕРСЬКУ РОБОТУ СТУДЕНТА

**Сьоміну Денису Андрійовичу**

(прізвище, ім'я, по батькові)

1. Тема роботи: “Розробка інформаційної технології для класифікації та пошуку об’єктів зображення за допомогою машинного навчання”

Керівник роботи: Негоденко О.В. завідувач кафедри, кандидат технічних наук, доцент

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

Затверджені наказом закладу вищої освіти від «11» жовтня 2021 року №170.

2. Срок подання студентом роботи: 24 грудня 2021 рік

3. Вхідні дані роботи

Методи обробки зображень;

Науково-технічна література з питань пов’язаних з програмним забезпеченням щодо розпізнавання образів зображення;

4. Зміст розрахунково-пояснювальної записки(перелік питань, які потрібно розробити).

4.1. Системи розпізнавання та локалізації образів на зображенні;

4.2. Вимоги та оцінка якості системи;

4.3. Опис модельованого алгоритму;

4.4.Опис використаних технологій;

5. Перелік демонстраційного матеріалу (назва основних слайдів)

Мета, об'єкт та предмет дослідження

Актуальність роботи

Аналіз існуючих рішень

Алгоритм навчання нейронної мережі

Структура нейронної мережі та функції активації

Модифікація алгоритму навчання

Практичний результат

Результати моделювання

Висновки

Апробація роботи

6. Дата видачі завдання 29 жовтня 2021 року

**Календарний план**

№ з/п	Назва етапів бакалаврської роботи	Строк виконання етапів роботи	Примітка
1	Підбір науково-технічної літератури	29.10-07.11	Виконано
2	Аналіз захищеності WEB-додатків	08.11-10.11	Виконано
3	Аналіз існуючих вразливостей WEB-додатків	11.11-18.11	Виконано
4	Аналіз нормативного забезпечення в галузі інформаційної безпеки	19.11-28.11	Виконано
5	Розробка комплексної методики пошуку вразливостей	29.11-04.12	Виконано
6	Вступ, висновки, реферат	05.12-12.12	Виконано
7	Розробка обов'язкових демонстраційних матеріалів	29.11-15.12	Виконано
8	Попередній захист роботи	21.12	
9	Здача роботи	24.12	

**Студент**

\_\_\_\_\_

(підпис)

Сьомін Д.А.

(прізвище та ініціали)

**Керівник роботи**

\_\_\_\_\_

(підпис)

Негоденко О.В.

(прізвище та ініціали)





## Реферат

Текстова частина магістерської роботи 77 с., 12 рис., 4 табл., 48 джерел.

РОЗПІЗНАВАННЯ ОБРАЗІВ, НЕЙРОННІ МЕРЕЖІ, ШТУЧНИЙ ІНТЕЛЕКТ, МАШИННЕ НАВЧАННЯ, ІДЕНТИФІКАЦІЯ ОБ'ЄКТІВ, ТЕХНІЧНИЙ ЗІР.

Об'єкт дослідження – автоматизація класифікації та пошуку об'єктів зображення за допомогою машинного навчання

Предмет дослідження – комп'ютерний зір, методи, алгоритми та засоби машинного навчання для систем розпізнавання образів.

Мета – підвищення ефективності процесу автоматизації об'єктів зображення на основі методів машинного навчання.

Методи дослідження: методи розпізнавання та класифікації образів, методи моделювання та теорії ймовірності.

Було проаналізовано декілька існуючих інтегрованих рішень, включаючи системи технічного бачення для розпізнавання образів. Реалізовано та модифіковано алгоритм побудови технічного класифікатора системи зору з урахуванням витрат класифікаційних помилок.

Штучний зір і розпізнавання об'єктів зараз є частиною штучного інтелекту, заснованого на високій обчислювальній потужності сучасних комп'ютерів. Ідея впровадження інформаційних технологій у міське середовище стає все більш популярною, оскільки обчислювальні алгоритми перевершують людину за ефективністю та точністю роботи.

Підвищення ефективності та розширення функціональних можливостей сучасних систем комп'ютерного зору потребує розробки більш ефективних методів обробки візуальної інформації. Основні завдання структурного розпізнавання стосуються удосконалення класифікації інформаційних технологій у просторі опису як наборів дескрипторів ключових точок зображення, а також необхідності оцінки ефективності розпізнавання у використаних вибірках. Особливу увагу було

приділено дослідженню структури даних для багатьох дескрипторів, що безпосередньо впливає на продуктивність систем розпізнавання.

Список основних сфер, де найчастіше використовують комп'ютерний зір:

- системи відеоспостереження в офісах, на виробництві, торгових центрах, на вулицях;
- системи керування автомобілями, що запобігають зіткненням з перешкодою;
- медичні системи аналізу зображень;
- сортування та інші операції у серійному виробництві;
- технології доповненої та віртуальної реальності;
- системи геопозиціонування та картографічні системи;
- системи контролю точності у будівництві та реконструкції;
- аналіз емоційного стану людини;
- читання штрих-кодів у торгівлі та на складських комплексах;
- конвертація паперових книг та документів у цифрові формати.

## ЗМІСТ

Вступ.....	10
<b>1. ОСНОВНІ ПОНЯТТЯ ШТУЧНО ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ ТА НЕЙРОННИХ МЕРЕЖ.....</b>	<b>13</b>
1.1 Історія розвитку нейронних мереж.....	15
1.2 Опис основних компонентів нейромережі.....	17
1.2.1 Нейрон та синопс.....	17
1.2.2 Перцептрон.....	19
1.2.3 Функції активації.....	21
1.3 Класифікація нейронних мереж.....	22
1.4 Машинне та глибоке навчання.....	25
1.5 Комп'ютерний зір.....	28
Висновок.....	31
<b>2. ОГЛЯД ІСНУЮЧИХ МЕТОДІВ ВИРІШЕННЯ ТА ПРОБЛЕМ КОМП'ЮТЕРНОГО ЗОРУ.....</b>	<b>32</b>
2.1 Системи комп'ютерного зору та їх застосування.....	32
2.2 Норми зображення.....	37
2.2.1 Вимоги, що пред'являються до формування ознак зображення.....	37
2.2.2 Основні види ознак та придатність зображення.....	39
2.3 Відстань виявлення об'єктів.....	40
2.4 Способи представлення зображення.....	42
2.4.1 Загальні положення.....	42
2.4.2 Огляд основних видів представлень зображення.....	43
2.5 Огляд і аналіз існуючих методів розпізнавання об'єктів на зображеннях.....	47
2.5.1 Метод порівняння з шаблоном.....	47
2.5.2 Гістограмний аналіз характеристик яскравості зображення.....	48
2.5.3 Метод головних компонент.....	49
2.5.4 Метод опорних векторів.....	50
2.5.5 Алгоритм Віоли-Джонса.....	51
2.5.6 Штучні нейронні мережі.....	52
2.5.7 Згорткові нейронні мережі.....	53
2.6 Огляд методів фільтрації зображення.....	55
2.7 Методи оцінки ефективності розпізнавання.....	58



Висновок .....	61
3. ВИБІР МЕТОДІВ ВИРІШЕННЯ ТА РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ.....	63
3.1 Вибір загальної моделі функціонування програмного засобу .....	63
3.2 Оцінка характеристик зображення.....	65
3.3 Структура нейронної мережі .....	67
3.4 Вибір активаційної функції .....	72
3.5 Порівняльний аналіз якості розпізнавання .....	73
3.10 Результати моделювання.....	75
Висновок .....	76
<b>Висновок.....</b>	<b>77</b>

## Перелік умовних позначень

ШІ – штучний інтелект

ШНМ – штучні нейронні мережі

СТЗ – системи технічного зору

ІС – інформаційна система

ІТ – інформаційні технології

ПЗЗ – прилад із зарядним зв'язком

ЕОМ – електронна обчислювальна машина

ІЧ – інфрачервоний

CBIR (Content-based image retrieval) – пошук зображень за змістом

PCA (principal component analysis) – метод главных компонент

## Вступ

В основі будь-якої організованої діяльності лежить інформація. Характерною особливістю інформаційних ресурсів є їх універсальність, необхідна в кожній сфері людської діяльності. Тож зараз на перший план виходить комп'ютеризація суспільства. Інформаційні технології використовуються в усіх сферах суспільного життя. Кількість інформації, яка потрібна людині в повсякденній роботі, постійно збільшується, що ускладнює структурування та упорядкування даних. Тому в житті сучасної людини використання інформаційних технологій є не розкішшю, а необхідністю, при якій функціонування багатьох об'єктів і систем є надзвичайно складним, а часто й неможливим.

Штучний інтелект (ШІ) сьогодні використовується в багатьох програмах. Майже всі вони постійно розвиваються. За останні роки сучасні комп'ютерні технології зробили великий ривок вперед, головним чином за рахунок підвищення продуктивності великих процесорів і швидкого зниження вартості оперативної та постійної пам'яті. Така ситуація привела до появи додатків, які містять серйозні теоретичні розробки штучного інтелекту.

У сучасному світі системи штучного інтелекту набули великої популярності завдяки своїм багатим можливостям і продуктивності. Багато з сучасних проблем вимагають практичного вирішення з використанням інтелектуальних систем: економіки та комерції, робототехніки, розвідки, математики, біофізики, авіоніки, безпеки, систем безпеки, медицини тощо. Така велика кількість додатків показує, що інтелектуальні системи є одним пакетом вирішення проблем при аналізі та обробці великих обсягів даних, вирішенні завдань різної складності. Штучні нейронні мережі зараз активно використовуються в системах штучного інтелекту.

З самого початку штучні нейронні мережі (ШНМ) використовувалися для виявлення різних застосувань і вивчення можливих принципів роботи мозку. ШНМ використовуються для розпізнавання моделей, передбачення, виявлення залежностей, стиснення даних, перевірки і т. д. Вона успішно використовується в

найширшому діапазоні застосувань, наприклад, в одній з важливих областей практичного застосування ШНМ - розвідувальної місії.

В даний час з розвитком обчислювальних систем та інформаційних технологій зростає популярність систем автоматизації і роботизації, як в промисловості і науці, так і в повсякденному житті. Як наслідок зростає потреба в ефективних методах обробки інформації, що надходить, зокрема відео даних.

Ефективна обробка і використання вхідної інформації дозволяє значно підвищити продуктивність і розширити коло застосування даних систем. Підтвердженням вищесказаного є розвиток і використання робототехніки, систем розпізнання текстів, охоронних систем заснованих на розпізнанні зображень, і т.д.

Останнім часом все більша увага приділяється системам, що використовують машинний зір в якості основного джерела інформації. Що привело до виникнення потреби в нових алгоритмах обробки і розпізнавання зображень. Не дивлячись на це, завдання розпізнавання образів до сих пір не вирішена в повному обсязі.

Дослідження комп'ютерного зору — наукова галузь штучного інтелекту та пов'язаних з ним технологій для зчитування зображень реальних об'єктів, їх обробки та використання даних для автоматизованого вирішення прикладних задач.

Актуальність. Комп'ютерний зір та розпізнавання об'єктів зараз є частиною штучного інтелекту, заснованого на високій обчислювальній потужності сучасних комп'ютерів. Оскільки комп'ютерні алгоритми перевершують людей на робочому місці з точки зору ефективності та точності, ідея впровадження інформаційних технологій у міське середовище стає все більш популярною.

Наукова новизна. Підвищення ефективності та розширення функціональних можливостей сучасних систем комп'ютерного зору зумовлює необхідність розробки більш ефективних методів обробки візуальної інформації. Основні завдання структурного розпізнавання пов'язані з розвитком інформаційних технологій класифікації у просторі описів як множин дескрипторів ключових точок зображень, а також необхідністю оцінювання ефективності розпізнавання на прикладних зразках. В даній роботі пропонується використовувати алгоритм з

врахуванням витрат на помилки класифікації. Особливу увагу приділено вивченню структури даних для багатьох дескрипторів, що безпосередньо впливає на показники функціонування систем розпізнавання.

Мета – підвищення ефективності процесу автоматизації об'єктів зображення на основі методів машинного навчання.

Об'єкт дослідження – автоматизація класифікації та пошуку об'єктів зображення за допомогою машинного навчання.

Предмет дослідження – комп'ютерний зір, методи, алгоритми та засоби машинного навчання для систем розпізнавання образів.

Методи дослідження: методи розпізнавання та класифікації образів, методи моделювання та теорії ймовірності.

Для досягнення поставленої мети необхідно вирішити такі завдання:

1. Створити вимоги до зображення
2. Провести аналіз методів обробки зображення
3. Дослідити аналіз існуючих методів розпізнавання образів
4. Модифікувати алгоритм побудови класифікатора таким чином, щоб він враховував витрати на помилки класифікації
5. Перевірити результативність алгоритму на якість пошуку та класифікації об'єктів зображення

## 1. ОСНОВНІ ПОНЯТТЯ ШТУЧНО ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ ТА НЕЙРОННИХ МЕРЕЖ

Штучний інтелект — розділ інформатики, присвячений моделюванню інтелектуальної діяльності людини. Задачі, що вирішуються методами штучного інтелекту, характеризуються досить складною типізацією. Одним з основних напрямків штучного інтелекту є розробка інтелектуальних інформаційних систем або систем, заснованих на знаннях.

Головною ціллю цих систем є знаходження, дослідження та застосування знань висококваліфікованих фахівців для вирішення складних проблем, що виникають на практиці. Альтернативою попередньому, як ідеологічно, так і практично, є галузь, що використовує нейронні мережі та нейроінформаційні технології.

В даний час штучні нейронні мережі активно використовуються в системах штучного інтелекту.

Нейронна мережа - це різновид штучного інтелекту, яка складається з декількох однотипних елементів. Ці елементи виконують прості функції, і всі процеси, що відбуваються в штучній нейронній мережі, яка дуже схожа на процеси, що відбуваються в нервовій системі живих організмів. Нейронні мережі мають нелінійний характер, у них немає явної залежності, що дозволяє відразу використовувати розроблену технологію (нейромережеву інформаційну модель).

Штучні нейронні мережі (ШНМ): математичні моделі, а також їх програмні чи апаратні реалізації, засновані на принципі організації та функціонування біологічних нейронних мереж: мережі нервових клітин живих організмів. Ця концепція виникла під час вивчення процесів, що відбуваються в мозку, і спроби змоделювати ці процеси. Першу подібну спробу зробили нейронні мережі МакКалоха і Піттса. Після розробки алгоритмів навчання отримані моделі почали використовувати в практичних цілях: у прогнозуванні задач, у розпізнаванні образів, у задачах керування та ін.

З математичної точки зору нейронні мережі є способом розв'язування задач нелінійної оптимізації. Кібернетика використовує теорію нейронних мереж для вирішення проблем адаптивного керування шляхом створення алгоритмів для робототехніки.

ШНМ – це система простих підключених та інтерактивних процесорів (штучних нейронів). Кожен процесор у цій мережі обробляє лише сигнали, які він періодично отримує, і сигнали, які він періодично надсилає іншим процесорам. Будучи підключеними до досить великої мережі з керованою взаємодією, ці локально прості процесори можуть разом виконувати досить складні завдання.

З точки зору машинного навчання нейронна мережа - це особливий випадок методів розпізнавання образів, дискримінаційного аналізу, методів кластеризації та багато чого іншого. З математичної точки зору навчання нейронної мережі являє собою багатопараметричну нелінійну оптимізацію. З кібернетичної точки зору нейронна мережа використовується в діяльності з адаптивного управління і в якості роботизованих алгоритмів. З точки зору технологічного розвитку та комп'ютерного програмування нейронна мережа є одним із способів вирішення проблеми ефективної конкуренції. А з точки зору штучного інтелекту є основою філософської тенденції коннекціонізму та головним напрямком структурного підходу до вивчення можливості моделювання природного інтелекту з використанням комп'ютерних алгоритмів.

Штучні нейронні мережі пропонують перспективні перспективи розвитку, а програмне забезпечення має велику перевагу перед їх використанням. У програмуванні нейронна мережа є способом вирішення проблеми ефективного паралелізму.

Нейронні мережі не програмуються в здоровому сенсі, вони навчаються. Здатність до навчання є однією з головних переваг нейронних мереж перед традиційними алгоритмами. Технічно навчання полягає у знаходженні коефіцієнтів зв'язку між нейронами. У процесі навчання нейронна мережа здатна виявляти складні зв'язки між входами і виходами, а також робити узагальнення. Це

означає, що якщо навчання пройде успішно, мережа зможе повернути правильний результат до бази даних, якої не було в навчальній вибірці.

Прогностична здатність нейронної мережі насамперед пов'язана з її здатністю до узагальнення та ізоляції прихованих відносин між входом і виходом. Після навчання мережа може передбачити майбутнє значення послідовності на основі декількох попередніх значень. Слід зазначити, що передбачення можливо тільки в тому випадку, якщо попередні зміни зумовлюють майбутнє. Наприклад, прогнозування цін на акції на основі винагород минулого тижня може бути ефективним (або, можливо, ні), в той час як прогнозування результатів лотереї завтра на основі даних за останні 50 років майже напевно не дасть результатів.

### 1.1 Історія розвитку нейронних мереж

Штучні нейронні мережі вперше згадуються в 1940-х р. Саме як наукова дисципліна теорія нейронних мереж знайшла відображення в роботах МакКалоха і Піттса в 1943 р. У цій статті стверджувалося, що майже будь-яку логічну або арифметичну функцію можна реалізувати за допомогою найпростішої нейронної мережі.

Однією з основних робіт, якій слід приділити увагу, є моделі Д. Хебба, який у 1949 р. сформулював закон, що став відправною точкою для вивчення нейронних мереж. Хебб був першим, хто припустив, що навчання – це зміна сили синаптичних зв'язків. Теорія Хебба є типовим навчальним випадком, коли система, що тестується, вчиться виконувати необхідне завдання без втручання експериментатора.

Ф. Розенблатт розробив нейронну мережу під назвою перцептрон для класичних об'єктів у 1958 році. Під час тренування перцептрон отримав повідомлення від "вчителя". За допомогою цих повідомлень можна було визнати до якого класу належить об'єкт. Крім того, навчений перцептрон був здатний самостійно класифікувати в несвідомих і упереджених цілях.



Після публікації роботи Д. Хопфілда (1982) інтерес до нейронних мереж значно зріс. Хопфілд, посилаючись на принципи навчання Хебба, показав, що проблеми з нейронами можна звести до узагальнення з серії моделей, розроблених у той час у фізиці неупорядкованих систем.

Потім, у 1986 році, була опублікована стаття Вільямса, Румельхарта та Хінтона, яка відповіла на важливе питання про вивчення багатосарових нейронних мереж. Згодом запропонований Хінтоном алгоритм зазнав багатьох змін.

У 1980-х роках поступово склалася міцна теоретична база, на якій зараз будується більшість мереж. Розроблена теорія широко використовувалася протягом останніх двох десятиліть для вирішення прикладних задач. Почали з'являтися компанії, які розробляли програмне забезпечення для проектування штучних нейронних мереж. У 1990-х роках нейронні мережі почали використовуватися в бізнесі, де вони виявилися надзвичайно ефективними при вирішенні багатьох завдань: від прогнозування попиту на продукт до аналізу кредитоспроможності клієнтів банку.

У 2007 році в Університеті Торонто Джеффри Хінтон розробив алгоритми для поглибленого вивчення нейронних мереж. Для навчання нижніх шарів мережі Хінтон використав машину Больцмана з обмеженими можливостями, яка є рекурсивною і стохастичною нейронною мережею. Після того, як ви ознайомилися з мережею, отримана програма може швидко вирішити проблему.

За словами експертів, які тримають руку на пульсі, очікується значний технологічний розвиток у сфері нейронних мереж та нейрокомп'ютерного моделювання. Останнім часом почало з'являтися багато нових можливостей та праця в цій сфері стає одним з основних факторів розвитку науки, техніки та економіки.

Хоча дослідження нейронного моделювання проводяться понад шістьдесят років, немає жодної області мозку, де б повністю була зрозуміла обробка інформації. Немає навіть нейрона, для якого можна визначити код для передачі інформації у вигляді послідовності імпульсів.

В даний час існує багато конфігурацій нейронних мереж, які відрізняються за принципом роботи та конфігурацій, а саме тому призначені для різних завдань.

Майбутнє комп'ютерних систем буде тісно пов'язане з відкриттями в нейронному моделюванні: як тільки таємниця того, як працює принаймні одна область мозку, буде негайно прояснена багато інших частин.

## 1.2 Опис основних компонентів нейромережі

### 1.2.1 Нейрон та синапс

Штучні нейронні мережі є відносно примітивними електронними моделями, заснованими на нейронній структурі мозку. Основним елементом, який обробляє нейронну мережу, є нейрон. Нейрон (нервова клітина) — це особлива біологічна клітина, яка обробляє інформацію. Підраховано, що в мозку є велика кількість нейронів, кожен з яких має близько  $10^{11}$  з'єднань.

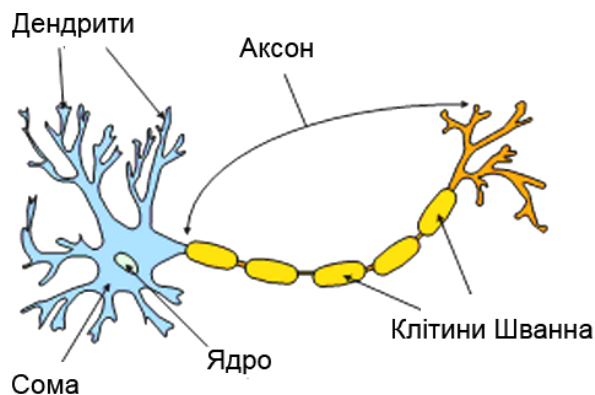


Рис 1. Зображення біологічного нейрону

Штучний нейрон має кілька входів (синаптичних аналогів біологічного нейрона) і один вихід (аксональний аналог). Кожен елемент має вагу, який помножується на значення, отримане з цього елемента. У тілі (клітині) нейрона зважені вхідні дані додаються та отримана кількість буде перетворена за допомогою функції активації (перенесення) нейрона, яка зазвичай є нелінійною. Отже, роботу штучного нейрона можна описати формулами:

$$S = \sum_{i=1}^n w_i x_i + w_0 \quad (1.1)$$

де  $n$  - розмірність вхідного вектора,  $w_i$  - вага  $i$ -го входу нейрона,  $x_i$  - значення, яке надходить на  $i$ -й вхід нейрона;

$$Y = f(S) \quad (1.2)$$

де  $Y$  - вихідне значення нейрона,  $f(S)$  - активаційна функція.

Значення на вході нейрона змінюються в інтервалі  $[0, 1]$ . Іноді нейрон має додатковий вхід  $x_0$  з вагою  $w_0$ , який використовується для встановлення порога чутливості нейрона шляхом переміщення тригерної функції по осі абсцис.

Хоча кожен нейрон окремо виконує дуже просту обробку даних, велика кількість нейронів, що працюють паралельно в нейронній мережі, дозволяє розв'язувати дуже складні залежності даних.

Кожен нейрон складається з двох типів даних: вхідних і вихідних. На першому рівні вхідні дані збігаються з вихідними. В інших випадках вхідні дані отримують підсумкову інформацію з попередніх рівнів і нормалізуються.

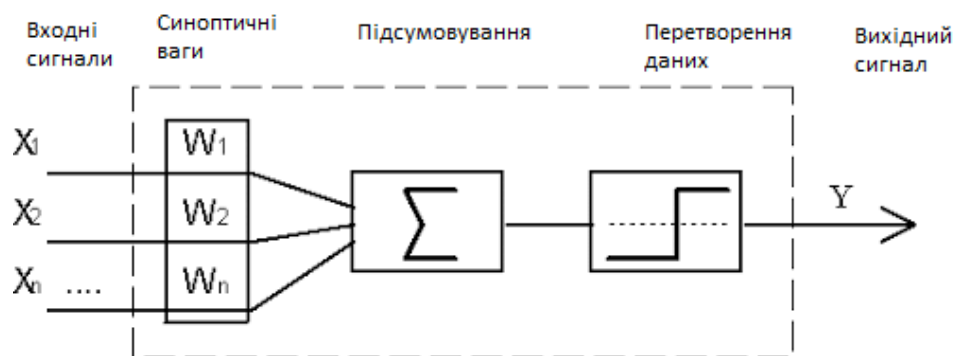


Рис 2. Зображення моделі нейрону

Існує кілька способів класифікації нейронів. По розташуванню нейронів в мережі їх поділяють на:

- вхідні;
- вихідні;
- приховані.

Також нейрони можуть бути класифіковані за видом активаційної функції:

- пороговий - має активаційну функцію у вигляді жорсткого і «м'якого» (лінійного) порога;
- лінійний - активаційна функція має вигляд прямої лінії;
- бінарний - має активаційну функцію, яка забезпечує на виході нейрона два стану - 0 та 1;
- сигмоїдальна - активаційна функція такого нейрона має вигляд плавної S-образної кривої.

Синапс - зв'язок між нейронами, причому кожен синапс має свою вагу. Завдяки цьому вхідні дані видозмінюються при передачі. Під час обробки передана синапсом інформація з великим показником ваги стане переважаючою.

На результат впливають не нейрони, а синапси, які дають сукупність ваги вхідних даних, адже власне самі нейрони постійно виконують абсолютно однакові обчислення. Виставлення ваг здійснюється у випадковому порядку.

### 1.2.2 Перцептрон

Перцептрон є абстрактною моделлю біологічного нейрона. По суті, це дуже простий процесор. Як можна побачити з малюнку, перцептрон приймає кілька довічних входів,  $x_1, x_2, \dots, x_n$  і виробляє один двійковий вихід. Розенблатт запропонував просте правило для обчислення вихідних даних. Він ввів ваги,  $w_1, w_2, \dots, w_n$  - речові числа, що виражають важливість відповідних вхідних даних для виведення.

$$\sum_{j=1}^n w_j x_j \quad (1.3)$$

Вихід нейрона від того, чи є результат роботи суматора – сума звана зваженою, менше або більше деякого порогового значення. Так само, як і ваги, порогове значення є дійсним числом, що є параметром нейрона. Наведемо це в більш точних алгебраїчних термінах, як у формулі:

$$output = \begin{cases} 0, \text{ якщо } \sum_{j=1}^n w_j x_j \leq \text{порогове значення} \\ 1, \text{ якщо } \sum_{j=1}^n w_j x_j > \text{порогове значення} \end{cases} \quad (1.4)$$

Перцептрон - це пристрій, який приймає рішення, зважуючи різні види даних. Тому має бути очевидним, що складна мережа перцептронів може приймати досить складні рішення.

Багатошаровий перцептрон - це клас штучних нейронних мереж прямого поширення, що складаються як мінімум з трьох шарів: вхідного, прихованого та вихідного. За винятком вхідних, всі нейрони використовують нелінійну функцію активації.

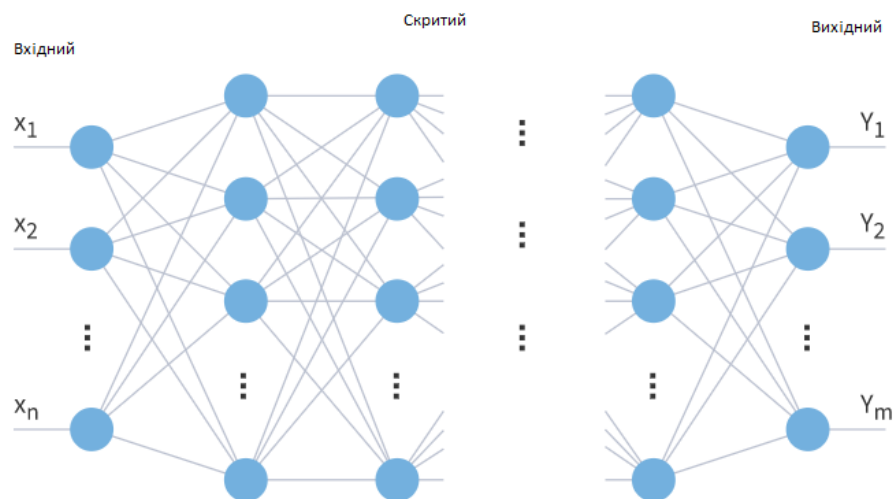


Рис 3. Багатошаровий перцептрон

Навчальні алгоритми автоматично налаштовують ваги та відхилення мережі нейронів у відповідь на зовнішній стимул, без прямої участі програміста. Процес навчання передбачає малу зміну ваг та відхилень, щоб отримати малу зміну на виході, наближаючись до кращого результату класифікації. Навіть малий вплив може викликати різкі зміни, приводячи до непередбачуваних наслідків. Для вирішення цієї проблеми використовують так звані функції активації.

### 1.2.3 Функції активації

Нейрон повністю описується своїми вагами і функцією активації, що позначається  $f(y)$ , де  $y$  - зважена сума. Отримавши вектор  $x$  в якості вхідних даних, нейрон видає деяке число на виході.

Активаційна функція може бути різного виду. Найбільш широко використовуються варіанти наведені в таблиці (за  $s$  позначена зважена сума):

Таблиця 1 – Перелік функцій активації нейронів

Назва	Формула	Область визначення
Порогова	$f(s) = \begin{cases} 0, & s < 0 \\ 1, & s \geq 0 \end{cases}$	(0,1)
Знакова	$f(s) = \begin{cases} 1, & s > 0 \\ -1, & s \leq 0 \end{cases}$	(-1,1)
Сигмоїдальна (логістична)	$f(s) = \frac{1}{1+e^{-s}}$	(0,1)
Полулінійна	$f(s) = \begin{cases} 0, & s \leq 0 \\ s, & s > 0 \end{cases}$	(0; $\infty$ )
Лінійна	$f(s) = s$	( $-\infty$ ; $+\infty$ )
Радіальна базисна	$f(s) = \exp(-s^2)$	(0,1)
Полулінійна з насиченням	$f(s) = \begin{cases} 0, & s \leq 0 \\ s, & 0 < s < 1 \\ 1, & s \geq 1 \end{cases}$	(0,1)
Лінійна з насиченням	$f(s) = \begin{cases} -1, & s \leq -1 \\ s, & -1 < s < 1 \\ 1, & s \geq 1 \end{cases}$	(-1,1)
Гіперболічний тангенс	$f(s) = \frac{e^3 - e^{-3}}{e^3 + e^{-3}}$	(-1,1)
Треугольна	$f(s) = \begin{cases} 1 -  s , &  s  \leq 1 \\ 0, &  s  > 1 \end{cases}$	(0,1)

Однією з найбільш поширених є нелінійна функція з насиченням, так звана логістична функція або сигмоїдальна.

Логістична функція має такі властивості:

- вона є «обмежувальною» функцією, тобто незалежно від аргументу (зваженої суми), вихідний сигнал завжди буде в межах від 0 до 1;

- у всіх точках вона має похідну, і ця похідна може бути виражена через цю ж функцію.

### 1.3 Класифікація нейронних мереж

Штучні інтелектуальні системи мають характерні особливості:

- розвинені комунікативні здібності, які характеризують спосіб взаємодії між комп'ютером користувача з системою. Не виключена можливість звернення до системи з довільним запитом до діалозі з інтелектуальною системою.
- рішення завдань, які не мають конкретного вирішення, а вимагають нестандартного підходу, в залежності від ситуації, існуючих даних і кінцевого результату. Такі завдання ефективно вирішуються за допомогою штучних нейронних мереж.
- здатність до самонавчання - тобто можливості вилучення знань інтелектуальною системою з накопиченого досвіду конкретних ситуацій. Для попереднього навчання системи необхідні оброблені початкові дані.
- відсутність заданого заздалегідь алгоритму вирішення задачі

Таблиця 2. Види штучних інтелектуальних систем

Вид	Тип
Системи з комутативними здібностями	- інтелектуальні бази даних; - природно-мовні інтерфейси; - гіпертекстові системи; - контекстні довідкові системи; - когнітивна графіка;
Експертні системи	- системи класифікації; - системи довизначення; - системи трансформації;

Таблиця 2. Види штучних інтелектуальних систем

Системи самонавчання	- індуктивні системи; - нейронні мережі; - системи на прецедентах; - інформаційні сховища.
Адаптивні системи	- CASE - технології; - компонентна технологія;

Системи можуть вирішувати не одну, а кілька завдань або в процесі вирішення однієї завдання вирішувати і ряд інших завдань.

Інтелектуальні системи також класифікуються за використаними методами. Тут можна виділити м'які, жорсткі і гібридні методи.

М'які обчислення (Soft Computing) - це складна комп'ютерна методологія, заснована на нечіткій логіці, генетичних обчисленнях та імовірнісних обчисленнях.

Жорсткі обчислення - стандартні комп'ютерні обчислення (які не стосуються м'яких).

Гібридні системи - системи, що використовують більш ніж одну комп'ютерну технологію (в разі інтелектуальних систем - технології штучного інтелекту).

Кожна нейронна мережа включає в себе перший шар нейронів, званий вхідним. Цей шар не виконує будь-яких перетворень і обчислень, його завдання в іншому: приймати і розподіляти вхідні сигнали по іншим нейронам. І цей шар єдиний, який є загальним для всіх типів нейромереж, а критерієм для поділу є вже подальша структура:

1. Одношарова структура нейронної мережі. Являє собою структуру взаємодії нейронів, в якій сигнали зі вхідного шару відразу направляються на вихідний шар, який, власне кажучи, не тільки перетворює сигнал, але відразу ж видає відповідь. Як вже було сказано, 1-й вхідний шар тільки приймає та розподіляє сигнали, а потрібні обчислення відбуваються вже в другому шарі. Вхідні нейрони є об'єднаними з основним шаром за допомогою синапсів з різними вагами, що забезпечують якість зв'язків.



2. Багатошарова нейронна мережа. Крім вихідного та вхідного шарів, є ще кілька прихованих проміжних шарів. Число цих шарів залежить від ступеня складності нейронної мережі. Вона більшою мірою нагадує структуру біологічної нейронної мережі. Такі види були розроблені зовсім недавно, до цього всі процеси були реалізовані за допомогою одношарових нейронних мереж. Відповідні рішення мають більші можливості, якщо порівнювати з одношаровими, адже в процесі обробки даних кожен проміжний шар - це проміжний етап, на якому здійснюється обробка і розподіл інформації.

Крім кількості шарів, нейронні мережі можна класифікувати по напрямку розподілу інформації по синапсах між нейронами:

1. Нейромережі прямого поширення (односпрямовані). У цій структурі сигнал переміщається строго по напрямку від вхідного шару до вихідного. Рух сигналу в зворотному напрямку не здійснюється і в принципі неможливо. Сьогодні розробки цього плану поширені широко та на сьогоднішній день успішно вирішують завдання розпізнавання образів, прогнозування та кластеризації.

2. Рекурентні нейронні мережі (із зворотними зв'язками). Тут сигнал рухається і в прямому, і в зворотному напрямку. У підсумку результат виходу здатний повертатися на вхід. Вихід нейрона визначається ваговими характеристиками та вхідними сигналами, плюс доповнюється попередніми виходами, знову повернувшись на вхід. Цим нейромережам властива функція короткочасної пам'яті, на підставі чого сигнали відновлюються і доповнюються під час їх обробки.

У складі нейромережі є три види нейронів: вхідний, вихідний та прихований. У одношаровій структурі прихованих нейронів не буде. Також є одиниці, яких називають нейронами зміщення і контекстними нейронами.

Класифікація штучних інтелектуальних систем охоплює величезну теоретичну базу знань в різних областях науки. Створення та реалізація ШНМ - складний процес від початкового етапу до кінцевого.

## 1.4 Машинне та глибоке навчання

Машинне навчання — це підмножина штучного інтелекту, яка використовує статистичні алгоритми навчання для створення систем, які можуть навчатися та вдосконалюватися автоматично без явного програмування.

У цьому розділі ви можете дізнатися про традиційні програми та навчання, щоб потрапити в лабораторію візуального зору з управлінням, яке дозволяє комп'ютерам розуміти, що відбувається в кадрі, аналізуючи фотографію або відео.

Не рідко комп'ютерно-програмований зір ґрунтувався на методах обчислення таких характеристик зображення, як гострі краї та кути. Ці символи повинні бути визначені вручну програмістом алгоритму, який точно знає, що шукати дані візуально. Потім програміст комбінує функції та формує алгоритм, який дозволяє йому визначити, що включено в зображення.

Алгоритми машинного навчання автоматично створюють математичну модель з використанням великої кількості вибірок даних (навчання), щоб мати можливість приймати рішення, обчислюючи результати без програмування. Символи завжди фіксуються вручну, але як їх поєднувати, алгоритм вивчає шляхом обробки великої кількості спостережуваних або анотованих навчальних даних. Ця техніка називається класичним машинним навчанням.

За допомогою машинного навчання потрібно навчити систему правильно використовувати та розподіляти дані. Результат передається в систему, і цей процес триває до тих пір, поки система не навчиться достатньо, щоб ідентифікувати потрібний об'єкт, наприклад певний тип транспортного засобу. Навчання моделі перетворюється на програму.

Головною перевагою штучного інтелекту перед звичайним програмуванням при створенні програми комп'ютерного зору є можливість до оброблення великої кількості різноманітної інформації та різномасштабних даних. Система здатна обробляти тисячі зображень, зберігаючи при цьому високу точність та продуктивність, а через деякий час людина втомлюється та втрачає концентрацію. Таким чином, використання штучного інтелекту дозволяє значно підвищити

точність програми. Однак чим складніше визначення, тим складніше машині буде досягти бажаного результату.

Глибоке навчання — це розширена версія машинного навчання, де система навчається створювати функції та як об'єднувати ці функції у великі структури правил для досягнення результатів, і це навчання покладається на вхідні дані. Алгоритм може автоматично визначати, які характеристики шукати в навчальних даних, а також має змогу до дослідження глибоких структур малопов'язаних комбінацій.

Основні алгоритми, які використовуються в глибокому навчанні, засновані на тому, як працюють нейрони та як мозок використовує їх для створення знань більш складного рівня, пов'язуючи вихідні нейронні сигнали в глибоку ієрархію. Мозок — це система, де комбінації також складаються з нейронів, що стирає межу між виділенням та комбінацією ознак, що робить ці процеси дещо ідентичними. Вчені змоделювали ці структури як штучні нейронні мережі, які є найпопулярнішим типом алгоритму глибокого навчання. короткий опис нейронних мереж у додатку до цього документа.

Використовуючи алгоритми глибокого навчання, можна створювати складні візуальні детектори та навчати їх виявляти різноманітні, малозв'язні та дуже складні об'єкти.

Причина такої гнучкості полягає в тому, що системи глибокого навчання можуть навчатися на значно більшій кількості даних (а також бути більш різноманітними), ніж традиційні системи машинного навчання. Це робить глибоке навчання ідеальним для складних видів діяльності, де є велика кількість різноманітних даних та важко змоделювати комбінацію ознак.

За наявності вчителя, алгоритми навчання можна розподілити на навчання з вчителем (Supervised Learning), без вчителя (Unsupervised Learning) і з підкріпленням (Reinforcement Learning).

Навчання з вчителем використовують для розпізнавання об'єктів або сигналів.

Навчання без вчителя найчастіше використовують коли алгоритми вивчають схожість та можуть виявити відмінність та виявити аномалії, виділяючи, що є незвичайним або несхожим.

Навчання з підкріпленням має широке використання там де необхідно правильно виконати поставлені задачі у зовнішньому середовищі маючи безліч можливих варіантів дії.

За типом застосовуваних алгоритмів можна виділити два види:

1) класичне навчання - відомі та добре вивчені алгоритми навчання, розроблені в основному більше 50-ти років тому. Підходить в першу чергу під завдання роботи з даними: класифікація, кластеризація, регресія і т.п. Застосовують для прогнозування, сегментації і так далі.

2) нейронні мережі та глибоке навчання - найбільш сучасний підхід до машинного навчання. Нейронні мережі мають широке використання там, де необхідне розпізнавання або генерація зображень та відео, складні алгоритми управління або прийняття рішень.

Обидва види навчання є схожими типами алгоритмів, але алгоритм глибокого навчання зазвичай використовує набагато більший набір вивчених комбінацій ознак, ніж класичний алгоритм машинного навчання. Це означає, що аналітичний додаток на базі глибокого навчання може бути гнучкішим та якщо його навчити - зможе виконувати більш складні завдання.

Втім, для окремих задач відеоаналітики може бути досить спеціально оптимізованого класичного алгоритму машинного навчання. Для чітко визначених завдань цей алгоритм може видавати такі самі результати, як і алгоритм глибокого навчання, при цьому вимагаючи менше математичних операцій, що робить це рішення більш економічним та менш енерговитратним. Крім того, для нього необхідно набагато менше даних для навчання та ресурсів для розробки.

## 1.5 Комп'ютерний зір

Розвиток науки супроводжується все більшою складністю досліджуваних систем. Однією з причин є постійне ускладнення технології та організації виробництва, а пов'язана з цим необхідність поглибленого аналізу сутності природних, техніко-економічних, економічних та міжгалузевих явищ. Тому при прогнозуванні ситуації та характеристик об'єктів за допомогою математичних моделей необхідно враховувати велику кількість взаємопов'язаних факторів за умов, коли інформація, доступна досліднику, може мати різні форми та різний ступінь детермінізму.

Штучні нейронні мережі значною мірою запозичують принципи з людського мозку. У галузі штучного інтелекту є багато активних та повноцінних дослідницьких напрямків: розробка природної мови та інтерфейсів машинного перекладу, генерація та розпізнавання мови, навчання та навчання, машинні ігри та творчість, нова комп'ютерна архітектура, інтелектуальна робота та розпізнавання моделей. Розпізнавання зображень охоплює широкий спектр проблем: розпізнавання зображень, символів, текстів, запахів, звуків, шумів.

Розпізнавання образів є однією з найважливіших областей обчислювальної техніки та кібернетики. У цьому розділі йдеться про вирішення широкого кола науково-технічних проблем на основі загальної логічної концепції класифікації. Розпізнавання зображень імітує здатність людини розпізнавати, діагностувати та класифікувати.

Актуальність розпізнавання шаблонів продовжує зростати. Це пов'язано зі складністю технології та організації виробництва, необхідністю обробки великих інформаційних потоків та розширення кола погано формалізованих завдань прийняття рішень.

Завдання розпізнавання образів, такі як завдання автоматизації прийому та обробки даних, аналізу фігур в інформаційних таблицях можна розділити на три групи:

- виділення, моделювання та оцінка суттєвих характеристик (факторів), перетворення характерного простору;
- аналіз закономірностей розподілу елементів кінцевої множини у багатовимірному фазовому просторі;
- поділ простору на класи еквівалентності, віднесення до класів просторових елементів.

Розпізнавання — це завдання перетворення вхідної інформації, за якою необхідно враховувати певні параметри, характеристики розпізнаних зображень, в оригінал, що є висновком про клас, до якого належить розпізнане зображення.

У повсякденному житті людина впорається із завданням розпізнавання настільки легко, що це очевидно. Тим часом спроби змоделювати ці високоінтелектуальні функції на комп'ютері стикаються з серйозними труднощами.

Загалом проблему розпізнавання можна розділити на дві частини: навчання та розпізнавання. Навчання здійснюється шляхом показу окремих предметів із зазначенням їх належності до певного класу. У результаті навчання система розпізнавання повинна набути здатності однаково реагувати на всі об'єкти одного класу і по-різному на всі об'єкти різних класів.

Комп'ютерний зір - набір засобів для цифрового зчитування відеозображень, їх обробки та виведення у вигляді, придатному для практичного використання в реальному часі.

Сьогодні існує багато різних підходів до впровадження систем комп'ютерного зору, починаючи від різних варіацій порівняння отриманого зображення з моделлю до побудови складних тривимірних моделей. Враховуючи широту тематики розпізнавання об'єктів, зупинимося на алгоритмах розпізнавання об'єктів у двовимірних зображеннях.

За допомогою комп'ютерного зору вирішується безліч завдань. До найбільш типовим можна віднести такі завдання: розпізнавання, рух, відновлення зображень. Методи вирішення цих завдань найбільш актуальні до розгляду у цій роботі.

Суттю розпізнавання є виділення на зображенні конкретного об'єкту. У задачі розпізнавання об'єкта може вирішуватись кілька проблем:

- Ідентифікація: розпізнається індивідуальний екземпляр об'єкта.
- Виявлення: відеодані перевіряються на наявність певної умови.
- Пошук зображень за змістом: виділення певного змісту на великому наборі зображень.

Зміст визначається відповідно до вибраних критеріїв пошуку.

- Оцінка положення: визначення куту або позиції певного об'єкту відносно точки, з якої велась зйомка.

- Оптичне розпізнавання символів: пошук, виділення та виявлення символів письмового або друкованого тексту.

Завдання руху полягає у обробці послідовності зображень з метою оцінки швидкості кожної точки зображення.

Завдання відновлення зображень вирішується видаленням шуму з зображення (шум датчика, розмитість об'єкта, що рухається і т. д.).

Подібні системи знайшли широке застосування у багатьох сферах діяльності. Вони стали невід'ємною та основною ланкою систем управління виробничими процесами, що використовуються в промислових роботах при створенні систем автоматичного пілотування транспортних засобів, у системах відеоспостереження та відеореєстрації, у системах доповненої реальності та системах моделювання об'єктів. Найважливіше застосування системи комп'ютерного зору знаходять під час обробки зображень у медицині для постановки медичного діагнозу. В промисловості вони контролюють якість сировини та продукції, протікання технологічного процесу, визначають положення та орієнтацію деталей на конвеєрі.

Загалом для основних систем комп'ютерного зору можна зазначити основні функції:

- Отримання зображень. Цифрові зображення можуть виходити від одного або кількох датчиків зображення. Як датчики можуть застосовуватися різні типи світлочутливих камер, датчики відстані, радары, ультразвукові камери тощо.

- Попередня обробка. Залежно від використовуваного надалі методу обробки зображення повинні відповідати деяким умовам. Для цього їх можуть коригувати:

- видаляти шум, що вноситься датчиком; масштабувати; змінювати контрастність.

- Виділення деталей. Деталі зображення різного рівня складності виділяються з відеоданих: лінії, межі контуру.

- Детектування/Сегментація: виділення на зображенні точок або ділянок, які будуть важливими при подальшій обробці.

- Високорівнева обробка: знаходження позиції об'єкту на зображенні.

## Висновок

Штучний інтелект, здається, межує з вигадкою, але насправді вже вкорінений у нашому житті, навіть якщо ми цього не усвідомлюємо. Штучний інтелект – це програмна та апаратна схема, яка виконує певну інтелектуальну функцію, яку раніше могли виконувати лише люди. У більшості випадків ШІ може справлятися з низкою функцій навіть краще, ніж люди. Головне завдання штучного інтелекту — прийти до логічного завершення. Область машинного навчання охоплює широкий клас проблем розпізнавання образів. Наприклад, це розпізнавання символів, рукописний текст, мова, аналіз тексту.

Завдяки машинному навчанню штучний інтелект може вирішувати не тільки завдання, для яких розробник визначив спеціальний алгоритм, а й подібні завдання. Між ситуаціями та реакціями існує зв'язок, але спочатку воно математично невідомо. Прикладом тренінгу є набір правильних пар ситуація-відповідь.

Завдяки розумному підходу, який поєднує кілька типів машинного навчання, ви можете автоматизувати більшість рутинних процесів. Іншими словами, алгоритми та завдання, підготовлені машинним навчанням, можуть виконати всю роботу. Люди в кінцевому підсумку мають усі творчі речі: стратегії, переговори, закриття бізнесу тощо. Це важливий фактор, оскільки машина не може перевищувати свої межі, а людський мозок може мислити інакше.



## 2. ОГЛЯД ІСНУЮЧИХ МЕТОДІВ ВИРІШЕННЯ ТА ПРОБЛЕМ КОМП'ЮТЕРНОГО ЗОРУ

### 2.1 Системи комп'ютерного зору та їх застосування

В даний час системи технічного зору (СТЗ) є одним з основних засобів за допомогою яких здійснюється розвиток автоматичних систем для управління рухом в умовах, коли обсяг апіорної інформації може бути занадто малий і для вирішення завдань управління необхідно провести аналіз зовнішньої обстановки в режимі реального часу. СТЗ можуть застосовуватися в сучасних космічних, наземних, авіаційних, підводних і надводних мобільних об'єктах. Завдяки досить високої дальності і рівням просторового і колірному дозволу сучасних матричних і лінійних приймачів для оптичного випромінювання СТЗ служать незамінними джерелами інформації при автоматизації вирішення завдань по розпізнавання образів, навігації або наведення.

Сьогодні застосування цифрових відео датчиків з підвищеним дозволом допомогло значно наблизити інформативність створених зорових пристроїв до окремих характеристик людського ока. А самі апаратні можливості, доступні завдяки розвитку електроніки та обчислювальної техніки, так зробили крок вперед, що обчислювальні потужності персональних ЕОМ поступово наближаються до характеристик обчислювальних потужностей, які використовуються для опрацювання зображення в мозку людини.

Системи штучного зору намагаються замінити людей у сферах, пов'язаних зі збором та аналізом візуальної інформації. Компонентами систем штучного зору є: оптична система, перетворювач світлових сигналів, електронний шлях для введення елементів зображення в пам'ять комп'ютера, математичне забезпечення збору та обробки необхідних візуальних даних.

Одним з провідних напрямків в ІТ сьогодні є створення, аналіз, а також застосування методів для вирішення завдань таких як: аналіз, знаходження та оцінювання зображень. Результати різних фундаментальних та прикладних

досліджень часто можуть використовуватися в технічній діагностиці, неруйнівному контролі, віддаленому зондуванні, відстеження екологічного стану, діагностики та прогнозування в медицині, геологічному пошуку, плануванні, прогнозуванні в хімії, автоматизації наукової діяльності. Дана тематика залучає велику кількість дослідників, але більшість робіт в цій галузі, в основному, мають емпіричний характер.

Зазначимо, що ця проблематика несе важливе значення для дослідження інформаційної природи зображень. Цілі та підсумки роботи дуже важливі для створення та аналізу зображень великого призначення, необмежених вузької предметної областю, що працюють із зображеннями, отриманими в великих діапазонах умов спостереження. Розробка нового покоління систем особливо пов'язана з необхідністю автоматизації вибору варіанта зміни зображення в рамках його семантики і інформаційної природи. Одним з центральних напрямків по створенню подібних систем стає автоматизація реалізації їх окремих функцій, прикладом може бути - визначення методів опису та перетворення зображень в рамках їх інформаційної природи.

Перші праці з дослідження алгоритмів комп'ютерного зору почали випускатись ще в середині 1960-1970-х років. У цій «науці» людина змушує систему навчатися виділяти певні характеристики світу без участі людини, тобто машина сама розуміє те, що вона бачить.

Коли у світі почали встановлювати камери (на вулицях, дорогах, будинках тощо), почали створювати програмне забезпечення та пристрої, які дозволяють виділяти різні об'єкти на фотографії: людей, цифри та інші дані, які потрібні людям. Це покращило технологію, яка ідентифікує об'єкти на фотографіях, порівнює їх із моделями та фокусується на полі на основі цих параметрів.

Для створення бортової системи камер створюється цілий клас завдань, які включають в себе складність виявлення, наведення і речей, що знаходяться в полі зору датчика зображення. Для різних видів діяльності, пов'язаних з технічним зором, дані про властивості, що мають інтерес, часто обмежені, але включають приблизні розміри об'єктів та їх нащадків.

Тим часом відеодатчики, встановлені на корпусі мобільного носія, слід в просторі разом з носієм, а також, його орієнтацію в просторі можна поміняти за допомогою пристрою позиціонування. Часто датчик зображення зазнає на собі вплив вібрацій різної природи. Під впливом перерахованих факторів, що складають їм зображення схильні до спотворень. Вони ускладнюють, а найчастіше роблять неможливим взагалі знаходження вирішення багатьох завдань: знаходження рухомих об'єктів, градацію та параметризацію супроводжуваних об'єктів, що знаходяться близько один до одного, аналіз траєкторій супутніх об'єктів. При цьому спотворення, пов'язані з вібраціями корпусу носія, ускладнюють роботу оператора системи, тому іноді виникає завдання стабілізації відслідковується датчика.

Природні явища: дощ, вітер, сніг, туман можуть вносити відчутний відсоток зашумлення та коливання в спочатку статичну сцену, створювати сильні перешкоди, перекриваючи частково поля зору камери. Зазначені нюанси часто роблять алгоритми, чудово працюють в одних умовах, аж ніяк непотрібними для інших.

У підсумку слід зазначити, що, хоч зараз є велика кількість розроблених алгоритмів, сама область ще погано вивчена. У процесі створення системи враховується більшість доступних методів, необхідних для знаходження руху та позначки об'єктів.

Найважливішою особливістю завдання обробки даних, одержуваних з датчиків для визначення об'єктів, що знаходяться на зображенні, стає відсутність для досліджуваних реальних ситуацій або об'єктів якихось адекватних математичних моделей, завдяки яким можна було б провести розрахунки та скласти кількісні або якісні висновки.

Проблема розпізнавання даних багатогранна та проблематична, оскільки зображення мають високу інформаційну ємність та початкову невизначеність (містять будь-які об'єкти і складаються за допомогою різних типів сенсорів), а також високу мінливість (коригування ракурсу або освітлення призводить до зрушень значень відразу у всіх елементах зображення). Також об'єкти, присутні на

зображеннях, складаються зазвичай в складних ієрархічних відносинах типу «приватне-ціле» та «приватне-загальне».

Через величезну кількість вже наявних методів виникає питання про їх схожість, відмінності та результативності застосування в процесі застосування до тієї чи іншої задачі. Основним моментом у процесі підбору методу розпізнавання стає уявлення початкового зображення. Під поданням часто визначають формальну систему, що включає алгоритми для перетворення в явному вигляді описів об'єктів розглянутого класу.

Головними типами уявлень стають:

- признакові;
- низькорівневі;
- структурні;
- контурні.

Вид проміжного представлення даних не стане повноцінною характеристикою методу розпізнавання даних. Найчастіше важливий також підхід до виявлення критерію якості та алгоритм оптимізації даного критерію.

Подання на рівні пікселів стають опорними для будь-яких додатків за обробкою зображень в сучасних ІС. Зображення являє собою групу пікселів, це визначення набуває основного статусу для завдань комп'ютерної обробки зображень. Але потрібно вміти розділяти зміст (саме зображення) від його форми (подання).

Подання на рівні пікселів включають в себе всі доступні дані про що спостерігається сцені, але в формі, абсолютно не придатною для автоматичного аналізу. Найчастіше подібні уявлення майже не використовуються для розпізнавання або порівняння зображень. Це і несе необхідність використання інших уявлень зображень для пошуку в них релевантних даних.

Головною проблемою, що виникає при обробці зображень, стає підвищення якості зображення, наприклад, збільшення розрізнення окремих фрагментів. До факторів, що зменшують якість зображень, відносять:

- Присутні яскравості перешкоди у вигляді шуму;
- Підвищена або знижена освітленість об'єктів зйомки;
- Недостатня різкість при створенні зображення;
- Невеликі розміри деталей, які потрібно розпізнати.

Величезна кількість методів підвищення якості зображень розділяється на дві великі категорії: методи обробки в просторі та методи обробки по частоті. Поняття «просторова область» належить площині растрового зображення, що формується із окремих пікселів, і включає в себе методи, що базуються на прямому зміні пікселів зображення. Методи обробки по частоті базуються на коригуванні сигналу, що створюється шляхом використання в зображенні перетворення Фур'є або подібних перетворень. Частина методів зручно розташовується в просторової області, а для розташування інших більше підходить частотна область. Далі будуть розглянуті просторові методи збільшення якості зображень, що часто пов'язано з проведенням великого обсягу обчислень при реалізації перетворень Фур'є для растрових зображень розміром  $800 \times 600$  пікселів та вище.

Яскравість на фото може значно знижувати розрізнення окремих фрагментів зображення, але важливо розуміти, що для нинішніх систем створення зображення в цифровому вигляді характерно майже повна відсутність будь-яких істотних яскравих перешкод, але отримані за допомогою їх зображення все-таки іноді мають недостатню відмінність частини фрагментів, що може пояснюватися:

- Малою контрастністю зображення, оскільки відомо, що людське око не може розрізнити кордон між фрагментами, яскравість яких відмінна лише на одиницю градації;
- Малою різкістю зображення, що веде до замилювання кордонів різних фрагментів.

Головними джерелами шуму та підвищеної яскравості на цифровому зображенні стають: методика його отримання, оцифровка, а також його передача. Дія сенсорів, що реагують на електромагнітне випромінювання в певному спектральному діапазоні, може залежати як від зовнішніх умов в момент отримання зображення, так і від чутливості самих сенсорів. Наприклад, в момент отримання

зображення фотокамерою з ПЗЗ матрицею, головними чинниками, що визначають величину шуму, стають температура самих сенсорів та рівень освітленості.

При передачі зображення схильні до перешкод, що виникають в каналах зв'язку. Наприклад, при передачі зображення з застосуванням бездротової мережі, воно може піддатися зміні в момент розряду блискавки або інших збурень в атмосфері.

## 2.2 Норми зображення

### 2.2.1 Вимоги, що пред'являються до формування ознак зображення

Найбільш сучасні виробничі, технологічні, а так само офісні системи під час проходження процесу його функціонування використовують інформацію про маркування об'єктів. Можливе використання інформації про маркування різних вантажів, таких як: вагони, контейнери, а також номерні знаки автомобілів і не тільки дає можливість найбільш раціональним чином організувати процес технологічної обробки інформації або проводити облік та контролювання виробів, матеріалів, транспортних засобів. Основним в процесі використання маркувань (нанесення текстоцифрових міток) лежить технологія автоматизованого розпізнавання структурованих символів. Потреба в даній технології була викликана необхідністю створення методів, моделей та систем розпізнавання структурованих символів.

В даний час подібні технології зазвичай реалізуються за допомогою трьох методів: структурний, признакова та шаблонний. Кожен з наведених методів призначений для використання в певних умовах і саме в них дані методи найбільш ефективні.

Однак кожен з розглянутих методів має свої недоліки. Найістотнішим з них є низька стійкість до таких явищ як: зміни масштабу, зсуву, повороти, а також зміна ракурсу та інших можливих спотворень.

Ці недоліки можуть найбільш яскраво проявлятися при масштабній експлуатації програмно-технологічних систем, що використовують в своїй основі

дані методи. У більшості подібних систем розпізнавання структурованих символів характеристики точності здійснюють найбільш різке падіння та вже не дотягують до прийняттого рівня якщо їх піддають спотворенню або перетворенню. Разом з тим технологічні умови отримання інформації при маркуванні не дозволяють повністю усунути ці спотворення.

Вимога ефективності процесу розпізнавання образів є основною вимогою до використовуваної системи ознак даного зображення. Дана вимога має ряд аспектів:

- до ознак пред'являються вимоги обчислювального характеру, які полягають в тому, що існує алгоритм розрахунку ознак, і даний алгоритм є обчислювально ефективним;
- з вимог ефективності виводиться необхідність так само задовольнити вимоги по стійкості або інваріантності образу по відношенню до ряду спотворень об'єкта.

Так само для ряду практичних завдань вимога інваріантності виявляється досить загальним та має на увазі всі або деякі з таких вимог:

- інваріантність до шумових та динамічних спотворень;
- інваріантність до яркості спотворень (зміни яскравості та контрасту);
- інваріантність до зміни місця розташування об'єкта;
- інваріантність до зміни масштабу об'єкта;
- інваріантність до зміни орієнтації об'єкта (до повороту об'єкта в площині зображення);
- інваріантність до довільних перетворень.

Через проведення попереднього перетворення вихідного зображення досягається інваріантність до даних спотворень. Перша група спотворень це: шумові, а також динамічні зміни зображення. Такі спотворення усуваються шляхом використання методів та алгоритмів відновлення. До другої групи належить яскравість спотворення. Вони можуть бути ефективно усунені шляхом приведення зображення до нормалізованого виду.

### 2.2.2 Основні види ознак та придатність зображення

Результат роботи детектора це безліч особливих точок, для яких потрібно побудувати математичний опис.

Дескриптор приймає на вхід в якості даних зображення, а також набір з особливих точок, які повинні бути виділені на заданому вхідному зображенні. Як вихід для дескриптора стає безліч з векторів ознак для вихідного набору з особливих точок. Так само слід зазначити, що деякі дескриптори здійснюють рішення одночасно двох завдань таких як - пошук особливих точок, а також побудова опису для знайдених особливих точок.

Побудова ознак буде здійснюватися на підставі отриманої інформації про інтенсивність, про колір, а також про текстуру особливої точки. Так само особливі точки можуть представлятися як кутами, ребрами або навіть контуром об'єкта, тому, зазвичай, обчислення виконуються тільки для якоїсь околиці. В ідеальному варіанті хороші ознаки повинні володіти такими властивостями як:

Повторюваність. Якщо дані зображеннях з одного і того ж об'єкта або сцени, які, в свою чергу, можуть бути зроблені з різних точок зору та при різній освітленості, найбільшу кількість ознак має бути продетектировано.

Локальність. Для зниження ймовірності перекриттів всі ознаки повинні бути максимально локальними.

Репрезентативність. Чисельність ознак повинна бути достатня, для того щоб найбільш розумне число ознак детектувати навіть в разі невеликого зображення об'єкта.

Точність. Ознаки повинні бути найбільш точно продетектировано по відношенню до масштаба та форми об'єкта.

Ефективність. Для додатків реального часу критично, якщо процедури обчислення ознак вимагатимуть значних обчислювальних витрат. Тому потрібно оптимізувати мережу на користь їх зменшення.

Часто кажуть, що якість зображення залежить від високої роздільної здатності та високої світлочутливості камери. Незважаючи на те, що важливість цих параметрів не підлягає сумніву, існують й інші фактори, що не менш важливі



при визначенні придатності фотоабо відеозапису. Так, наприклад, відеопотік неперевершеної якості з дорогої камери відеоспостереження може виявитися абсолютно марним через недостатнє підсвічування в нічний час, перенаправлення камери в непотрібний бік чи збою зв'язку із системою.

Перед розгортанням аналітичної програми необхідно ретельно вивчити місце розміщення камери. Щоб відеоаналітика працювала належним чином, камера має бути розташована таким чином, щоб зона спостереження проглядалася повністю та безперешкодно.

Придатність зображення також може залежати від сценарію використання. Якість відео, прийнятне для людського ока, може бути недостатньо для застосування відеоаналітики. Більш того, багато методів обробки зображень, які зазвичай використовуються для поліпшення сприйняття відео людиною, що не рекомендуються при використанні відеоаналітики. До них, в зокрема, належать методи шумоподавлення, технологія широкого динамічного діапазону та алгоритми автоматичного керування експозицією.

Сучасні відеокамери часто оснащені ІЧ-підсвічуванням, завдяки якому вони можуть працювати у повній темряві. У цьому є позитивний момент, адже така функція дозволяє розміщувати камери у важкодоступних місцях та усуває необхідність встановлення додаткового освітлення. Однак, якщо на об'єкті можливі сильні дощі або снігопад, наполегливо рекомендується не покладатися на світло, що походить від камери або з джерела, розташованого дуже близько до неї. Краплі дощу та сніжинки можуть відбивати в камеру занадто багато світла і суттєво утруднити аналітику. З іншого боку, при розсіяному світлі більше шансів, що аналітика буде працювати результативно навіть у погану погоду.

### 2.3 Відстань виявлення об'єктів

Визначення максимальної відстані виявлення об'єкту є важким завданням для програм аналізу на основі штучного інтелекту: значення в метрах або футах не

завжди відповідає дійсності. Зверніть увагу, що на відстань виявлення сильно впливають якість зображення, характеристики розташування, погодні умови та атрибути об'єкта (включаючи колір та яскравість). Звичайно, наприклад, яскраві об'єкти на темному тлі в сонячний день можна виявити далі, ніж темні об'єкти в дощовий день.

Відстань також залежить від швидкості об'єкта. Для отримання з програми аналізу відео даних точних результатів, необхідна велика кількість часу, для розпізнавання образів. Цей час залежить від обчислювальної потужності (частоти кадрів) платформи: чим нижча швидкість, тим довше об'єкт перебуватиме в кадрі, який буде виявлено. Якщо витримка не відповідає швидкості об'єкту, розмитість зображення вплинуть на точність виявлення.

Під час наближення до камери високошвидкісні об'єкти можуть не бути виявлені. Наприклад, легко виявити, що людина тікає досить далеко від камери, а людина, що біжить дуже близько до камери з тією ж швидкістю, може потрапити та зникнути з поля зору так швидко, що програма не зможе захопити його.

Також проблемою аналізу руху є те, що об'єкти входять або виходять з камери безпосередньо. Порівняно з об'єктами, що рухаються по кадру, особливо важко виявити об'єкти, які рухаються повільно і зображення яких показують лише незначні зміни.

Як правило, більша роздільна здатність камери не означає більшу відстань виявлення. Можливості обробки, необхідні для виконання алгоритму машинного навчання, пропорційні розміру вхідних даних. Це означає, що для аналізу повного дозволу камери 4K потрібно як мінімум у чотири рази більше обчислювальної потужності, ніж для камери з роздільною здатністю 1080p. Дуже часто через обмеження у можливостях обробки камерою в додатках на основі ШІ використовують нижчу роздільну здатність, ніж може запропонувати камера або відеопотік.

## 2.4 Способи представлення зображення

### 2.4.1 Загальні положення

В даний час, незважаючи на те, що різного роду зображення все частіше використовуються для подання та обробки вхідної, проміжної і вихідної інформації в багатьох прикладних задачах розпізнавання образів, слабо розвинені засоби вибору алгоритму обробки зображення в залежності від його синтаксичного і семантичного змісту. Найбільша кількість алгоритмів для розпізнавання зображень, які мають обчислювальну ефективність, призначаються для роботи з ознаковими описами або моделями зображень. При побудові такого опису істотна частина інформації, що міститься в зображенні, втрачається. Для того, щоб максимально використовувати при аналізі зображень міститься в них інформацію, необхідно розробити ефективні методи опису зображень та побудови моделей зображень. Для вирішення фундаментального завдання побудови ефективних формальних описів зображень необхідно провести класифікацію та порівняльний аналіз засобів, що дозволяють адекватно описувати зображення та представляти всю інформацію, що міститься в зображенні та необхідну для вирішення поставлених завдань. Вирішення цього завдання відкриває широкі перспективи підвищення ефективності вибору алгоритмів аналізу зображень та автоматизації (часткової або повної) обробки зображень.

Моделлю зображення називається представлення його характеристик у вигляді деякої спеціальної формальної конструкції, зручної для подальшої обробки, розпізнавання. Для побудови моделі зображення використовуються різні математичні об'єкти, у тому числі ознаки і дескриптори зображень, логічні функції, графи, послідовності перетворень зображень та ін. Найбільш часто використовуються числові ознаки. Числовою ознакою зображення називається змінна, що представляє числову характеристику властивостей зображення, наприклад, геометричних, яскравих, структурних або просторових. Формальним описом зображення за допомогою числових ознак називається набір математичних об'єктів (наприклад, чисел, векторів, матриць), що містять деяку інформацію про

зображення, представлену у формі числа або набору чисел, і що допускаються алгоритмом розпізнавання, використовуваним для вирішення завдання. Слід зазначити, що формальний опис зображення може містити як повну інформацію про зображення (таку, по якій можна відновити вихідне зображення), так і часткову (в разі, коли інформації недостатньо для відновлення вихідного зображення з його опису).

Властивості ознаки зображення визначаються як властивостями зображення, так і способом вилучення ознак зображення. Ознака моделі є параметром моделі зображення, тому тип допустимого простору ознак, розкид значень ознаки та методи порівняння значень ознак істотно залежать від обраної моделі зображення, в той же час типи алгоритмів розпізнавання пов'язані з використовуваними типами ознак, оскільки ознаки складають вхідну інформацію алгоритмів розпізнавання. Звідси випливає, що допустимість застосовуваних методів розпізнавання зображень безпосередньо пов'язана з побудованою моделлю зображення.

Необхідно відзначити, що елементами моделі зображення можуть бути не тільки ознаки, а й непохідні елементи, відносини між ними, процедури та ін.

#### 2.4.2 Огляд основних видів представлень зображення

##### Низькорівневе представлення

В рамках зростаючого обсягу мультимедіа колекцій особливу увагу дослідників направлено на алгоритми пошуку зображень за змістом (Content Based Image Retrieval, CBIR), які можуть повністю автоматизувати процес створення індексу. В рамках відсутності будь-яких інших даних про зображення для створення індексу застосовуються низькорівневі характеристики вихідного зображення: яскравість, колір, текстура. За останні кілька років пропонувалося безліч різноманітних моделей опису подібних характеристик, функцій відстані та оцінок схожості зображень. Але якість роботи систем знаходження зображень за змістом все ж не можна називати задовільною. Людина, вибираючи зображення, порівнює їх семантику, смислове оцінку схожості зображень на базі низькорівневих характеристик. Головним напрямком проведених в цій галузі робіт

стає створення алгоритмів, необхідних для зменшення «семантичного розриву» між даними аналізу зображень системою і сприйняттям самого користувача.

Проблемам індексування та пошуку зображень за змістом відведено велику кількість робіт, що може говорити про високу активність дослідників в цій галузі.

Системи пошуку зображень за змістом застосовують різноманітні алгоритми обробки цифрового сигналу для побудови сигнатур - уявлень зображення в рамках конкретного набору параметрів. Майже у всіх пропонованих підходах одна сигнатура може описати одну з характеристик зображення (колір, текстура, і т.п.). Тобто одному і тому ж зображенню може відповідати відразу кілька сигнатур, які описують кожен з характеристик.

У процесі пошуку в рамках запиту часто виступає зображення-зразок. Підсумком такого пошуку стає комбінація зображень, близьких до зображення-замовлення за враженнями людини. Різні системи дозволяють користувачеві створити ескіз того зображення, яке він бажає знайти. Але обидва ці підходи змушують користувача якимось конкретизувати свій запит на мові, що розуміється системою, але не зручному для користувача. Застосування зображення зразка для запиту зобов'язує користувача мати такий зразок, що часто просто неможливо. Розробка ескизу потрібного зображення змушує користувача формулювати запит на мові форм та кольору. А адже користувачеві завжди зручніше складати свій запит на природній мові, вказуючи на словах те, що він хоче знайти на необхідному зображенні. Але зараз немає системи, що дозволяє згенерувати індекс автоматично в рамках змісту зображення і вміє розуміти запит на природній мові.

Більшість дослідників розрізняють відразу кілька рівнів змісту зображень. Колір та яскравість належить до низькорівневого змісту, що знаходять фізичні об'єкти (машина, людина, будинок) - до змісту високого рівня, а текстура часто належить до середнього рівня. У літературі можна знайти величезну кількість підходів до моделювання низькорівневих характеристик зображення та майже всі СВІР побудовані на алгоритмах обробки низькорівневого змісту. Ось уже кілька років дослідники розуміють брак алгоритмів. Важливо взяти до уваги особливості візуального сприйняття людини, його орієнтованість на семантику.

Популярним підходом, покликаними мінімізувати розрив між поданням зображення системою та людиною стає система зворотного зв'язку (relevance feedback). Базуються на ітеративному процесі пошуку, подібні системи накопичують статистику по реакції користувача, яка потім може бути враховуватися при формуванні відповіді системи.

### Контурне представлення

Контурними ознаками називають такі ознаки, обчислення яких проводиться на контурі (замкнутому або незамкнутому). Слід зазначити, що обчислення контурних ознак відбувається в два етапи: 1) виділення пікселів, що належать контуру об'єкта на зображенні; та 2) обчислення власних ознак з використанням виділених пікселів контуру. Як правило, виділений контур є бінарним зображенням, що істотно зменшує час роботи алгоритмів. Контурні ознаки часто використовуються в задачах розпізнавання машинописного і рукописного текстів. Для кожного символу виділяються зовнішній і внутрішній контури, на них обчислюються: ознаки, отримані за допомогою операції поділу на зони, коефіцієнти сплайна, апроксимуючого контур, коефіцієнти Фур'є для опису еліптичних замкнутих контурів та функції від них.

Як додаткові ознаки для розпізнавання рукописних цифр використовується відстань (по вертикалі і по горизонталі) між точками контуру цифри і точками опуклою фігури, описаної навколо цифри. Для апроксимації контурів об'єктів при розпізнаванні осіб використовуються коефіцієнти, що описують криві: коефіцієнти прямої - для апроксимації лінії щоки, параболи - для апроксимації лінії підборіддя і губ. Контури об'єктів використовуються для обчислення характеристик Ейлера, що відображають топологію описуваного об'єкта на зображенні.

Контурні ознаки також часто використовуються при аналізі і класифікації зображень відбитків пальців. У зв'язку зі специфікою таких зображень, на них виділяються контури яскравості переходів, що характеризують власне відбиток, а потім обчислюються ознаки отриманих контурних зображень.

### Структурне подання

Цей клас містить шаблони, представлені символами або структурами. Другий тип — структура даних: набір речень, аргументи яких відповідають частинам об'єкта чи сцени, а передбачення — властивостям частини та просторовим зв'язкам між ними. Структурні описи зазвичай подають у вигляді графа, вершини якого відповідають частинам об'єкта та їх властивостям, а ребра, що з'єднують вершини, відповідають просторовим відношенням.

На жаль, науки, що займаються вивченням інформаційних процесів у мозку людини, сприйняття та обробки зорової інформації живими організмами, не мають даних, необхідних для створення ефективного алгоритму розпізнавання, що моделює процеси зорового сприйняття. Тому алгоритми розпізнавання зображень повинні базуватися на формальному описі.

### Признакове представлення

Модель представляє із себе опис геометричних особливостей зображення з використанням вектора ознак, який характеризується особливостями вхідного зображення. Ознаками зображення можуть бути як вертикальні, так і горизонтальні лінії, криві або з'єднання виду «Т», «У» та ін. Були розроблені спеціалізовані методи для виділення кожної окремої ознаки у вхідному зображенні незалежно від його розташування. Після того як всі методи будуть застосовані створюється вектор ознак, значеннями якого є величини, які характеризують як часто зустрічається кожна ознака на оригінальному документі. Далі в пам'яті зберігається список всіх можливих ознак, пошук яких ведеться на зображенні, і вектор ознак, відповідний вхідного зображення. Кожній ознаці можна поставити у відповідність деяку вагу з тим, щоб ознаки, інваріантні щодо функцій перетворення зображення, надавали значний вплив на зіставлення моделей зображень.

При проходженні процесу зіставлення виходить вектор ознак порівнюється з векторами, що зберігаються в пам'яті, з урахуванням ваг, приписаних кожній ознаці. Визначити міру близькості можна за допомогою різних способів, наприклад, кількість співпадаючих ознак, квадратична різниця між елементами

векторів та ін. Модель, найближча до вихідної моделі в сенсі заданої міри близькості, ставиться їй у відповідність.

## 2.5 Огляд і аналіз існуючих методів розпізнавання об'єктів на зображеннях

### 2.5.1 Метод порівняння з шаблоном

Принцип дії цих методів полягає в порівнянні кожної області зображення з конкретною моделлю, що призводить до взаємної кореляції. Будь-яка функція визначає формули вручну і може відповідати всьому об'єкту або окремим елементам об'єкта. Якщо опорний коефіцієнт перевищує заданий поріг, область відображеного зображення позначається таким чином, що містить потрібне цільове зображення (Рис. 1).

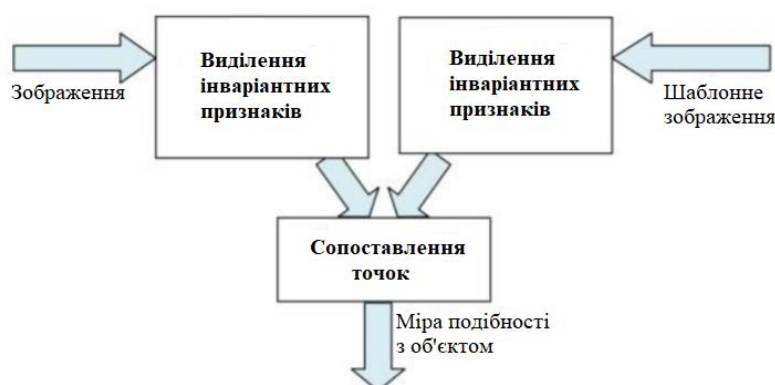


Рис 1. Схема роботи методу порівняння з шаблоном.

Для роботи режиму потрібні два певні кадри: вихідне зображення (наприклад, зображення з камери) та вихідне зображення. Щоб виявити схожу ділянку, потрібно накласти шаблон на знімок і зрушувати по піксельно, зліва направо, зверху вниз. У кожному положенні зображення-шаблону на оригінальному документі обчислюється метрика, яка вже показує, є збіг чи ні.

Кожна позиція на зображенні моделі додається до таблиці подій вимірювання. Формула метрики містить таблицю для кожної з позицій. Після



завершення порівняння продуктивності найкраще відповідність у світі буде низьким або низьким.

Існують різні методи порівняння, такі як метод найменшого типу, реляційний метод і метод перехресних посилань.

Нижче наведено приклад застосування моделі та алгоритму порівняння типів, який використовує метод найменших квадратів для знаходження метрики  $R(x, y)$ .

$$R(x, y) = \sum_{x, y} (T(x', y') - I(x + x', y + y'))^2, \quad (2.1)$$

де  $R(x, y)$ - матриця результатів,  $x', y'$  - поточні координати шаблону,  $x' = 0..w - 1$ ;  $y' = 0..h - 1$ , де  $w$  і  $h$  - це ширина і висота шаблону відповідно.

Першим кроком алгоритму є завантаження зображення моделі в модель. Потім визначається коефіцієнт кореляції між моделлю та вихідним зображенням, вибираються точки максимальної кореляції між зображеннями. Цей метод є найпростішим, але водночас і найбільш неточним методом відображення; Під час моделювання було виявлено багато помилок відображення. Це були перші спроби створення символічних об'єктів.

Недоліками цього методу є низька швидкість і висока чутливість до фотометричних і геометричних спотворень елементів зображень.

### 2.5.2 Гістограмний аналіз характеристик яскравості зображення

Ці методи засновані на припущенні, що частотна характеристика потрібного місця відрізняється від інших частин зображення і що розмір вибіркового пікселя більший.

Для посилення просторових частот та придушення шумів на початковому етапі даного підходу виконується поліпшення контрасту, бінаризація або виділення кордонів на зображенні. Це зробить краї елементів плівки гострішими, а фон – темнішим.

Концепція цих методів полягає в наступному: аналізується зображення та обчислюється середня яскравість пікселів кожного рядка малюнка (іноді

обчислюється у вигляді стовпців). Де б не був розташований цей об'єкт, звичайний розмір пікселя значно відрізняється від інших областей зображення. Гідність даних методів полягає в простій реалізації та високій швидкості роботи.

Недолік даних методів полягає в чутливості до будь-якої області зображення, що має параметри інтенсивності пікселів.

### 2.5.3 Метод головних компонент

К. Пірсон запропонував РСА (Аналіз основних компонентів) у 1901 році. Цей метод використовується в багатьох різних областях, таких як обробка та аналіз зображень, розпізнавання зображень, стиснення даних, матеріалів тощо.

Метод РСА спрямований на зменшення розміру даних з мінімальною втратою даних, що дуже важливо в області аналізу та обробки зображень. Цей метод виконує ортогональне лінійне перетворення вхідного вектора  $X$ , що містить взаємопов'язані компоненти, у менший вектор  $Y$ , що містить незв'язані змінні. Неістотні змінні називаються головними компонентами. Вектор  $X$  містить усі зразки навчальних зображень. Вектори властивостей отримують розв'язуванням такого рівняння:

$$D = \Phi^T K \Phi, \quad (2.2)$$

де  $D$  - діагональна матриця власних чисел;  $K$  - матриця коваріації для вектора  $X$ ;  $\Phi$  - матриця власних векторів.

З матриці власних векторів  $\Phi$  вибираються  $M$  найбільших власних чисел, з яких формується підматриця  $\Phi_M$ . Таким чином, отримаємо вираз:  $y = \Phi_M^T \tilde{x}$ , де  $\tilde{x} = x - \bar{x}$  - нормалізований вектор з нульовим математичним очікуванням.

При виборі перших  $M$  компонент векторний простір розбивається на власний простір, що містить головні компоненти.

По перше розраховуються головні характеристики для навчання об'єктів. Характеристики об'єктів проходять декілька рівнів фільтрації, менш важливі відпадають, а простір для параметрів звужується. З цього випливає, що об'єкти на зображеннях можуть відображатися як мінімальний набір параметрів.

Завдяки класифікації об'єктів на зображення розраховується евклідова відстань між основними компонентами досліджуваного зображення та зваженою комбінацією основних компонентів навчальної послідовності. Якщо основні компоненти зображення ближче до основних компонентів елемента в навчальному набору, вважається, що зображення містить цей елемент.

Недоліком цього методу є слабка стійкість об'єктів у різних умовах зйомки до різного шумового спотворення та відображення на знімках. Ефективне використання цього методу вимагає, щоб зображення були зроблені в однакових умовах зйомки.

Гідність методу PCA полягає в забезпеченні зменшення розмірності даних, що є важливим чинником в роботі з великими базами зображень.

#### 2.5.4 Метод опорних векторів

Метод опорних векторів (Support Vector Machine, SVM) був запропонований російськими вченими В. Вапніка і А. Червоненкіса. даний метод є лінійним класифікатором і застосовується для вирішення завдань класифікації та регресійного аналізу.

Застосування методу SVM до вирішення задачі класифікації об'єктів на зображеннях можна описати таким чином. Дано безліч точок в просторі, які згруповані за двома класами. Для того щоб розділити безліч точок на класи можна провести розділяючу гіперплощину. З точки зору класифікації необхідно знайти оптимальну гіперплощину, відстань від якої до кожного класу буде максимальною. Чим більше відстань від розділяючої гіперплощини до кожного класу, тим менше помилка класифікації. Вектори, розташовані в просторі ближче всього до гіперплощини, називаються опорними векторами.

Необхідно знайти класифікуючу функцію  $f(x)$ , яка приймає різні значення для векторів різних класів  $(x_1, y_1), \dots, (x_m, y_m)$ . Для визначення даної функції використовується навчальний набір даних. Функція, при якій очікувана помилка класифікації приймає мінімальне значення, є найбільш оптимальною функцією. Класифікуюча функція має вид:

$$f(x) = \text{sign}(\langle w|x \rangle + b), \quad (2.3)$$

де  $x$  - об'єкт, що належить простору;  $w$  - вектор нормалі до гіперплощини;  $\langle \rangle$ , - скалярний добуток;  $b$  - допоміжний параметр.

Об'єкти, для яких функція  $f(x)$  приймає значення рівні 1 або -1, відносяться до різних класів. Необхідно вибрати параметри  $w$  та  $b$  таким чином, щоб відстань до кожного класу було максимальною.

На практиці часто виникають випадки, коли дані неможливо розділити лінійним способом. Для вирішення такої ситуації виконується перетворення початкового простору  $x$  в простір вищої розмірності за допомогою спеціального відображення  $\varphi: R^n \rightarrow X$ . Відображення  $\varphi$  вибирається таким чином, щоб в отриманому просторі безліч даних було лінійно нероздільні. Основна складність полягає в тому, що з збільшенням розмірності простору, збільшується складність процесу класифікації.

Перевага методу SVM полягає в тому, що об'єкти впорядковуються за невеликим набором даних.

За допомогою цього методу можна створити класифікатор, який може зменшити максимальну оцінку очікуваної помилки розподілу, навіть для невідомих елементів, які не існують у вихідному наборі даних. Опорний векторний підхід, безпосередньо до проблеми розпізнавання образів, полягає в отриманні надбудови в просторі малювання, яка є кордоном між простором поверхневих зображень і безликих зображень.

Недоліком цього методу є те, що в процесі класифікації використовуються лише деякі зразки з області поширення. Перевага полягає в тому, що класифікація виконується методом еталонного вектора в невеликому загальному наборі даних.

### 2.5.5 Алгоритм Віоли-Джонса

У 2001 році Р. Viola і М. J. Jones запропонували алгоритм Віоли-Джонса, заснований на AdaBoost. Цей алгоритм дозволяє аналізувати об'єкти на зображеннях на основі властивостей волосся.

Ознаки Хаара – це прямокутні ділянки, що складаються з кількох суміжних частин. Різні класи компонентів мають унікальні властивості, які можуть бути виражені розподілом властивостей ознак Хаара (Рис 2).

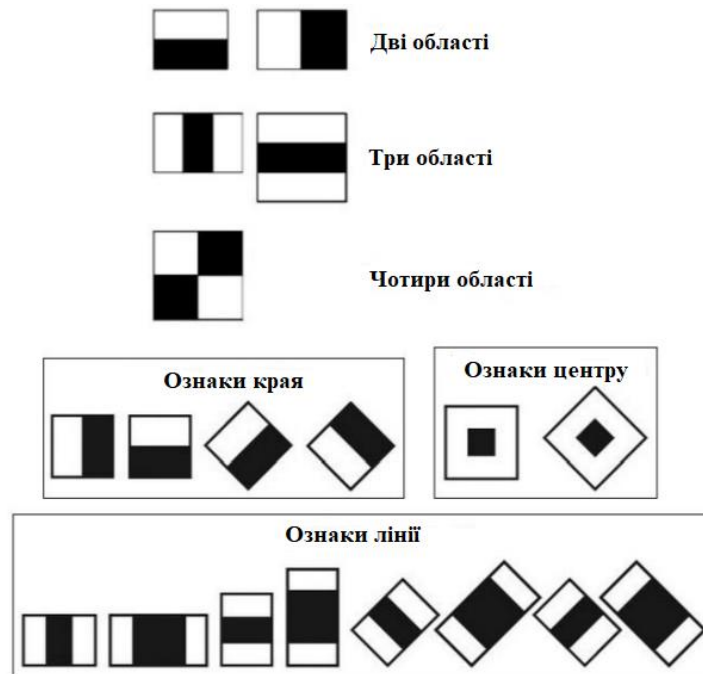


Рис 2. Ознаки Хаара

У кожній області зображення обчислюється кілька тисяч варіантів розташування ознак, залежно від їхнього економічного становища і масштабу. Величина щільності пікселів додається в чорних і білих областях штрихів волосся, а потім обчислюється різниця між сумами.

Цей алгоритм показує дуже високу точність виявлення та високу швидкість. Недоліком є довгий час навчання.

### 2.5.6 Штучні нейронні мережі

Штучна нейронна мережа (ШНМ) — це математична модель програмного або апаратного забезпечення, яка працює за принципом біологічних нервових клітин в організмі людини.

ШНМ мають ієрархічну структуру та складаються з простих інтерактивних процесорів, нейронів, згрупованих у шари. Кожен нейрон має вхідні та вихідні зв'язки, які ведуть до нервових клітин у сусідніх шарах.

Якщо вхідне зображення спотворене або зміщено, вихідний шар буде зміщено таким же чином. Ця функція забезпечує стійкість ШНМ до вхідних спотворень.

Переваги використання ШНМ для детектування об'єктів на зображеннях:

- Структура ШНМ добре підходить для обробки двовимірних даних.
- Нашаровуючись один на одного рецептивні поля забезпечують взаємозв'язок просторово залежних областей зображення.
- Підвищена стійкість до проєкційних спотворень вхідних даних, шумів, зміни масштабу.

Завдяки своїм перевагам на сьогоднішній день ШНМ активно використовуються для виявлення осіб та інших об'єктів на зображеннях, а також відео послідовностях.

### 2.5.7 Згорткові нейронні мережі

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).

ЗНМ являє собою особливий клас багат шарового перцептрона, який володіє двовимірною структурою і добре підходить для обробки зображень з високим ступенем інваріантності до зсуву, поворотам, масштабування та іншим спотворенням вхідних даних.

Структура ЗНМ являє собою послідовність з двох типів шарів: згорткові та вибіркові. Кожен шар складається з кількох шарів нервових клітин (карти ознак).

У шарі згортки кожен нейрон спілкується з невеликою групою нейронів попереднього шару (локальне поле прийому). У складному шарі локальні рецептивні поля нейронів частково перекриваються за принципом бляшки. Значення нейронів в локальному рецептивному полі множать на матрицю

синаптичних коефіцієнтів та результат записують у відповідний нейрон шару згортки.

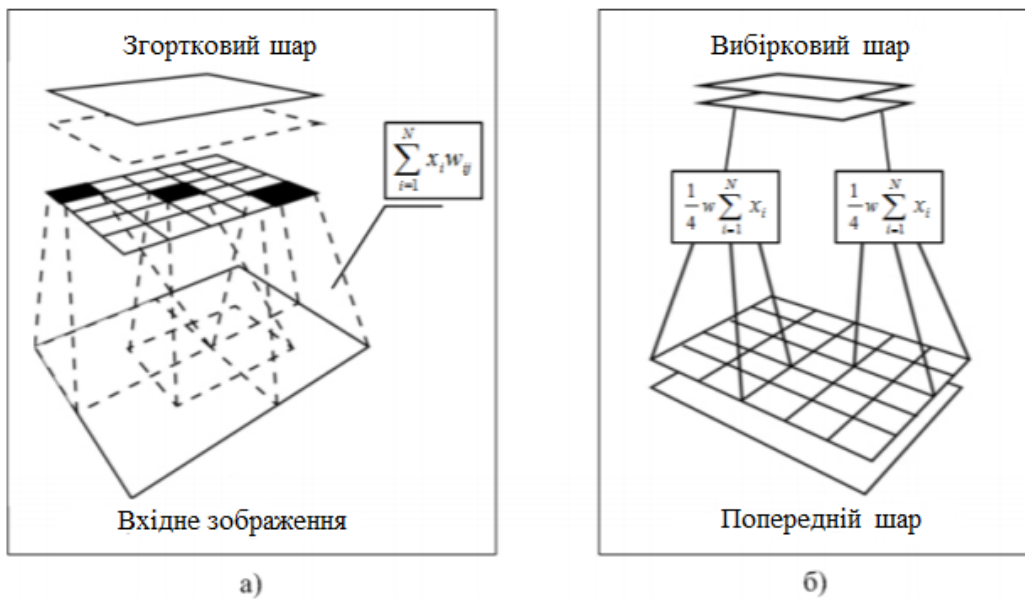


Рис 3. Функціонування шарів в ЗНМ

За згортковим шаром слід шар вибірки, який забезпечує часткову інваріантність нейронної мережі для вимірювання вхідного зображення. Кількість рівнів на рівні підвибірки зазвичай така ж, як і на попередньому рівні. У шарі підвибірки місцеві поля прийому не перекриваються і мають фіксований розмір нейронів 2x2. Кожен нейрон у цьому шарі обчислює середнє з чотирьох своїх вхідних даних, множить його на синаптичні ваги та передає результат функції активації. Таким чином тестовий шар ділить пропорції на площини попереднього шару в 2 рази.

Як варіант, розмір шарів зменшується, але їх кількість збільшується. Перемикання перемикача дозволяє створювати різні карти характеристик, що дозволяє ЗНМ визначати більш складні ієрархічні характеристики. Поступово карта на кресленні руйнується у векторі, коли проходить через різні рівні. Останні шари ЗНМ покривають класичний перцептрон, складаються з нормальних нейронів та призначені для класифікації окремих ознак.

Важливою особливістю ЗНМ є використання загальної вагової матриці для всіх рівнів нейронів. Такий підхід дозволяє використовувати менш підходящі параметри (ваги) з великої кількості зв'язків і прискорює процес навчання. Тому шари ЗНМ є фільтрами, кожен з яких шукає окремі атрибути у вхідному зображенні. Це дозволяє згортковій нейронній мережі запам'ятовувати зображення просторово залежних областей. Процес навчання визначає характеристики, вибрані з конкретного рівня.

Багато дослідників використовували ЗНМ у різноманітних конфігураціях для вирішення проблем розпізнавання та розпізнавання об'єктів. Однак більшість з них не ефективні у вирішенні таких проблем з наступних причин:

- Великий розмір вхідного зображення призводить до збільшення кількості нейронів і синаптичної ваги нейронної мережі. Це збільшує обчислювальні зусилля та час, необхідні для процесу навчання.
- Висока чутливість до вхідного спотворення, шуму, збільшення та кутів.

Все це вимагає застосування додаткових алгоритмів, тим самим збільшуючи обчислювальну складність і час виконання завдання.

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

## 2.6 Огляд методів фільтрації зображення

У задачах комп'ютерного зору фільтрація найчастіше використовується для попередньої обробки зображення перед аналізом його внутрішніх морфологічних ознак, але зустрічаються і завдання, в яких достатнім і бажаним буде використання тільки фільтрації як такої (наприклад, в задачах машинного зору).

### Бінаризація по порозі, вибір області гісто-грами

Для RGB зображення та зображення в градаціях сірого порогом є значення кольору. Вибір порога, по якому відбувається бінаризація, багато в чому визначає процес бінаризації. Зазвичай бінаризація здійснюється за допомогою адаптивного алгоритму, що вибирає поріг. Таким алгоритмом може бути вибір математичного



очікування, моди або піків гістограми. При роботі з гістограмою бінаризація ефективна для сегментації кольорів.

### Перетворення Фур'є

Перетворення Фур'є майже не використовується при обробці зображень у чистому вигляді, оскільки для аналізу зображень одновимірного перетворення зазвичай не вистачає та виникає необхідність використання куди більш ресурсоємного двовимірного перетворення. Цей метод застосовується тільки у разі якщо необхідний аналіз спектра, оскільки використання згортки цікавої області з вже готовим фільтром виявляється швидше та простіше в реалізації. Тим не менш, одновимірне перетворення Фур'є застосовується при компресії зображень.

### Фільтри частот

Найпростіший приклад фільтра низьких частот – фільтр Гауса, фільтра високих частот – фільтр Габора. Для кожної точки зображення вибирається вікно, в рамках якого виконується добуток вихідних даних з фільтром того ж розміру (згортка). Такий підхід набув досить широкого поширення в практиці, дозволяючи виділяти на зображенні цікаву інформацію, редукуючи зайву інформацію. Зокрема, підхід поширений як одна з реалізацій швидкого шумоподавлення в багатьох галузях науки і техніки.

### Вейвлет-перетворення

Для згортки з сигналом (областю зображення) використовується якась характеристична функція, що називається вейвлетом, що визначається як вейвлет-перетворення. Вейвлети – це сімейства функцій, у яких всі функції виходять з однієї у вигляді її зрушень та розтягувань по осі часу. Існує набір класичних функцій, що використовуються у вейвлет-аналізі. До них належать вейвлет Хаара, вейвлет Морле, вейвлет МНат, вейвлет Добеші. На практиці вейвлет-аналізом називається пошук довільного патерна на зображенні за допомогою згортки зображення з моделлю цього патерна. Класичні вейвлети використовуються для стиснення чи класифікації зображень.

### Обчислення показників кореляції

Обчислення кореляції, що лежить в основі вейвлет-перетворення, саме по собі є незамінним інструментом в системах комп'ютерного зору та часто використовується у своєму природному вигляді, наприклад, для знаходження зрушень або оптичних потоків (кореляція відеопотоку). На основі різницевого корелятора реалізується найпростіший детектор зсуву.

### Фільтри функцій

Цей підхід використовує математичні фільтри для виявлення простих математичних функцій у зображенні та створює інтегроване зображення (простір пам'яті), в якому для кожної точки вихідного зображення створюється декілька функцій генератора. Класичним прикладом є загальне перетворення Хаффа, застосоване до двійкових зображень, яке дозволяє знайти в зображенні будь-яку параметризовану функцію. Модифіковане перетворення Хаффа дозволяє шукати будь-який формат, але його використання в обробці зображень пов'язане з недостатньою стабільністю: висока чутливість до бінарної якості і низька швидкість дозволяють шукати кращі альтернативи цьому методу. Аналогічно перетворенню Хаффа для прямих ліній, до небінарних зображень застосовується перетворення Радона, що забезпечує значні переваги в продуктивності порівняно з великими наборами даних.

### Фільтри контурів

Фільтри контурів дуже корисні при вирішенні завдань комп'ютерного зору у разі якщо необхідна реалізація обробки складних об'єктів на зображенні. Фільтрування контурів реалізує ряд алгоритмів, таких як: оператор Кенні, що є найчастіше використовуваним алгоритмом пошуку контурів, оператор Собеля, оператор Лапласа, оператор Прюїтт, оператор Робертса. Алгоритм Кенні реалізує поняття придушення немаксимумів, що означає, що пікселями кордонів оголошуються пікселі, у яких досягається локальний максимум градієнта у напрямку вектора градієнта. У комп'ютерному зорі та обробці зображень при пошуку меж об'єктів крім особливих окремих випадків важко визначити детектор, який би працював значно краще, ніж детектор Кенні.

## Інші фільтри

Модифікації вищезазначених фільтрів дозволяють вирішити переважну кількість завдань фільтрації на рівні класичних методів, але крім них існують десятки більш спеціалізованих алгоритмів фільтрації, на практиці використовуються значно рідше. Найбільш цікавими є ітераційні фільтри, такі як активна модель зовнішнього вигляду, курвлет-перетворення, що виникли на основі класичних вейвлет-функцій та аналізу в полі перетворення Радона, бімлет-перетворення, що реалізує частково завдання фільтрації, частково – завдання аналізу, що дозволяє знаходити контури об'єктів. Ці фільтри мають рідкісну специфіку та в силу своєї вузької спрямованості не будуть розглядатися в рамках даної роботи докладніше.

## 2.7 Методи оцінки ефективності розпізнавання

Оцінка ефективності роботи системи розпізнавання є ключовим етапом дослідження, що дозволяє чисельно виміряти різні аспекти якості розробленої системи. Для отримання чисельного значення оцінки широко використовуються як загальні методи математичної статистики, та специфічні показники, що застосовуються для оцінки алгоритмів машинного навчання.

Однією з найпростіших метрик оцінки ефективності є відсоткова частка коректно розпізнаних зображень. Коректним розпізнаванням вважається отримання на виході алгоритму класу, відповідного попередньо заданого класу. Для оцінки використовується вибірка, спроектована аналогічно навчальній вибірці, але містить зображення, яких алгоритм було мати доступ у процесі навчання. Для цього, як правило, вихідна загальна вибірка поділяється на дві нерівні частини (Розмір тестової вибірки при цьому може різнитися, і складати 20-30% розміру загальної вибірки). Тоді якщо функція  $c(x_i)$  є функцією-індикатор коректності розпізнавання того зображення.

$$c(x_i) = \begin{cases} 1, & h(x_i) = y_i \\ 0, & \text{в іншому випадку} \end{cases}, \quad (2.4)$$

тоді частка коректно розпізнаних зображень оцінюється наступним чином:

$$Q = \frac{\sum_{i=1}^n c(x_i)}{n}, \quad (2.5)$$

де  $n$  – кількість зображень у тестовій вибірці. Таким чином, оцінка набуває значення на відрізку  $[0, 1]$ , де 1 відповідає коректному розпізнаванню всіх зображень вибірки.

Розглянутий показник є найбільш узагальненим та підходить для безлічі завдань розпізнавання з обмеженою кількістю класів. У задачах, де кількість класів не фіксована, і завдання розпізнавання представляє собою завдання ідентифікації об'єкта певної категорії серед множини інших, потенційно необмежених категорій, замість неї застосовуються такі показники як точність та повнота оцінки. Їх використання дозволяє розрізнити помилково-позитивні (класифікатор прийняв позитивне рішення щодо зображенню, що не містить шуканого об'єкта) та помилково-негативні (класифікатор не розпізнав об'єкт на зображенні, де він був присутній) помилки розпізнавання, або помилки першого та другого роду. Таким чином, точність оцінки в межах класу є частка зображень, що дійсно належать даному класу щодо всіх зображень які система віднесла до цього класу. Повнота системи – це частка знайдених класифікатором зображень, що належать класу щодо всіх зображень цього класу у тестовій вибірці. Інакше,

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN}, \quad (2.6)$$

де  $P$  – точність,  $R$  – повнота,  $TP$  – число істинно-позитивних рішень,  $TN$  - число істинно-негативних рішень,  $FP$  - число помилково-позитивних рішень,  $FN$  - число помилково-негативних рішень.

Показники точності та повноти широко використовуються в області обробки інформації, як правило, розраховуються спільно. При цьому існує кілька методик зіставлення двох показників:

- Кожен показник враховується індивідуально,

– для випадків, коли між показниками може спостерігатися залежність, проводиться оцінка одного з показників при фіксації іншого (наприклад, оцінка точності при рівні повноти 0.75),

– обидва показники можуть бути скомбіновані в один.

Серед прикладів агрегованих метрик для третьої методики трапляються такі показники як міра, коефіцієнт кореляції Меттьюза, регресійні коефіцієнти  $\Delta P$  та  $\Delta P'$ . Також, для деяких цілей розглядається зважене арифметичне середнє показника точності та значення, зворотного показником повноти, і навпаки. З погляду теорії ймовірності ці показники можуть інтерпретуватися так: точність відповідає ймовірності того, що випадково обране з множини позитивно розпізнаних зображень дійсно розпізнано коректно, при цьому повнота є ймовірністю того, що випадково обране із загального зображення буде коректно класифіковано алгоритмом (позитивно чи негативно). Так, в залежності від додатка задачі, до продуктивності методу розпізнавання можуть бути пред'явлені вимоги, що стосуються максимізації повноти (для випадків, коли деяка кількість помилково-позитивних рішень допускається), так і збалансованого значення двох показників або заходи, вираженої таким чином:

$$F = 2 \frac{RP}{P + R}, \quad (2.7)$$

Таким чином, міра являє собою гармонійне середнє значення показників точності та повноти.

Крім перерахованих показників, важливу роль в оцінці ефективності методів розпізнавання грають такі показники, як розмір довірчого інтервалу для вибірки зображень, оцінка кількості ітерацій та часу роботи алгоритму під час навчання, необхідні досягнення збіжності, темп збіжності тощо.

Для оцінки розміру вибірки, що забезпечує статистично достовірні результати, використовується метод розрахунків довірчих інтервалів. Для параметра  $\theta$  — числа зображень вибірки, що є випадковою величиною  $X$  з рівнем довіри  $p$ , таку, що  $X = (x_1, x_2, \dots, x_n)$  інтервал представляється за допомогою таких

меж  $l(x_1, x_2, \dots, x_n)$  і  $u(x_1, x_2, \dots, x_n)$ , які є реалізаціями випадкових величин  $L(X_1, X_2, \dots, X_n)$ ,  $U(X_1, X_2, \dots, X_n)$ , таких, що:

$$P(L \leq \theta \leq U) = p, \quad (2.8)$$

при цьому граничні точки довірчого інтервалу  $l$  та  $u$  собою довірчі межі. Рівень довіри  $p$  вибирається виходячи з специфіки поставленого завдання та вимог, що пред'являються до системи обробки інформації (а також пов'язаних із нею ризиків).

Досягнення збіжності під час навчання визначається як етап, коли  $\Delta J = J_i - J_{i-1} < a$ , тобто зміні функції ціни моделі, що навчається, на черговому кроці оптимізації не перевищує деякого фіксованого малого значення, вибір якого може здійснюватися з урахуванням завдання. Існує набір методик оптимізації, що дозволяють варіювати розміри цього порогового значення виходячи з темпів збіжності на конкретній ділянці області визначення функції, заданої моделлю, що навчається.

## Висновок

Важлива відмінність, властива процедурам обробки та аналізу зображень порівняно із завданнями розпізнавання або інтерпретації вже сегментованого образу, полягає в тому, що обробка інформації на практичних випадках зазвичай пов'язана з процедурою локалізації образів на зображенні. Саме реалізація процедури пошуку об'єкта пов'язана з небезпекою лавиноподібного зростання необхідної кількості обчислень. Беручи до уваги значний розвиток та збільшення потужностей сучасних комп'ютерних систем, такі обсяги обчислень займають секунди машинного часу, що далеко виходить за межі вимог систем виявлення реального часу, призначених для таких завдань, як навігація та наведення, інспекційний контроль, обробка документів, що машиночитуються, тощо.

Перераховані у данному розділі методи мають свої особливості, що дозволяє застосовувати їх до різних завдань. Для покращення їх якості та зменшення обчислювальних витрат, а також витрат пам'яті можна використовувати методи

відбору найінформативніших ознак та заповнення пропущених значень з погляду поставленого завдання.

Порівняльний аналіз методів розпізнавання зображень показав, що незважаючи на різноманітність підходів до навчання та виділення класифікуючих ознак, більшість методів не бере до уваги проблему розпізнавання образів з врахуванням витрат на помилки класифікації. Існуюча низка методів, створених для класифікації об'єктів зображення, наприклад нейронні мережі та констеляційні моделі, здатний вирішувати завдання інваріантного розпізнавання, проте демонструє неспроможність до локалізації об'єктів на зображенні та визначення параметрів їх просторового розташування, що є важливим для цілого ряду прикладних завдань обробки інформації.

Зазначені недоліки можуть бути усунені за допомогою побудови моделі, що включає в себе інформацію про витрати на помилки класифікації та використання відповідних алгоритмів навчання та розпізнавання. Як новий підхід для навчання такої моделі пропонується застосувати алгоритм навчання за допомогою потоку даних, що дозволяє отримувати та оброблювати інформацію про помилки класифікації та розв'язати проблему розпізнавання образів на зображенні.

### 3. ВИБІР МЕТОДІВ ВИРІШЕННЯ ТА РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

#### 3.1 Вибір загальної моделі функціонування програмного засобу

У той час як потреба в системах, здатних точно ідентифікувати образи, в тому числі й зорові, зростає, в даний час немає альтернативи до систем розпізнавання на основі комбінації методів, які нейтралізують існуючі недоліки.

В умовах, коли ми володіємо лише нечисленними апріорними відомостями про образи, що розпізнаються, при побудові розпізнаючої системи найкраще використовувати навчальну процедуру. Велика складність завдання, велика кількість вхідних даних вимагають гнучкої захисної системи, яка здатна аналізувати грандіозну кількість інформації не традиційним способом. Система розпізнавання на основі нейромереж могла б у перспективі вирішити багато проблем.

Першою перевагою у використанні нейромереж при виявленні вторгнень є гнучкість, яку ці мережі надають. Нейромережа здатна аналізувати дані, навіть якщо ці дані є неповними чи спотвореними, має можливість проводити аналіз даних у нелінійному режимі. Більше того, здатність обробляти дані від великої кількості джерел у нелінійному режимі є особливо важливою, оскільки деякі образи досить складні за своєю структурою та в перспективі існує можливість обробки за декількома важливими параметрами. Оскільки завдання розпізнавання вимагає достатньої продуктивності комп'ютерів, швидкість обробки в нейромережі може бути достатньою для вирішення в реальному часі, що може використовуватися в динамічних системах, в яких потрібне розпізнавання та класифікація будь-яких образів.

Характерно, що вихідні дані нейромережі виражаються у формі ймовірності, нейромережа надає можливість прогнозування (пророкування). Система виявлення вторгнень на основі нейронної мережі ідентифікує ймовірність того, що окрема подія, або серія подій вказують на параметри образу, який можна віднести до будь-якого класу не дивлячись на відмінності між поданим чином та способом, що



зберігається в базі. У міру того, як нейромережа «набирається досвіду», вона покращуватиме свою здатність визначати клас образу. Однак найважливіша перевага нейромереж полягає в здатності нейромереж «вивчати» нові характеристики та параметри образів та ідентифікувати елементи, які не схожі на ті, що спостерігалися раніше.

Нейромережа може бути навчена розпізнавати відомі класи образів із високим рівнем точності. У той час як це є дуже цінною здатністю, оскільки деякі класи мають ряд постійних ознак, за якими можна точно визначити приналежність до даного класу, мережа також має здатність використовувати ці знання для ідентифікації класів, які неточно відповідають характеристикам попередніх образів.

Оскільки здатність штучної нейромережі ідентифікувати клас належності тієї чи іншої образу повністю залежить від точності навчання системи, навчальні дані та використовувані методи навчання є найважливішими. Порядок навчання вимагає великої кількості даних для того, щоб переконатися, що результати є статистично значущими.

Однак, основна незручність застосування нейромереж для детектування вторгнення - це природа «чорної скриньки» нейромережі. На відміну від експертних систем, які мають жорстко закодовані правила аналізу подій, нейромережі адаптовані на аналіз даних за результатами навчання. Вага та активаційні функції вузлів зазвичай заморожуються після того, як мережа досягла допустимого рівня успіху в ідентифікації подій.

Хоча аналіз зображень досягає досить хорошого значення, базовий рівень його часто не відомий. Проблема «чорної скриньки» дуже характерна для нейромережевих додатків.

### 3.2 Оцінка характеристик зображення

#### Оцінка розмитості зображення

Розмитість зображення – один із найважливіших показників його якості, багато в чому визначальний придатність зображення до подальшої обробки. Розмитість зображення – це рівень чіткості кордону між двома сусідніми ділянками зображення з різною оптичною густиною (яскравістю).

$$S_i = \operatorname{tga} = \frac{G}{w}, \quad (3.1)$$

де  $i$  – кількість крайових пікселів на зображенні,  $w$  – ширина перепаду,  $G$  – це різниця між значеннями яскравостей пікселів, позначених  $a$  та  $b$ .

$$S_i = \operatorname{tga} = \frac{I(a) - I(b)}{w}, \quad (3.2)$$

Алгоритм включає наступні кроки

1. Виділити крайові пікселі за допомогою детектора країв.
2. У циклі, для кожного крайового пікселя, визначити напрямок градієнта.
3. У циклі, для кожного крайового пікселя, вздовж напрямку градієнта знайти найближчі пікселі з мінімальним  $a$  та максимальним  $b$  значеннями яскравостей та відстань між ними.
4. Обчислити міру розмитості для кожного крайового пікселя  $S_i$ , підставивши  $w$ ,  $a$  і  $b$ .
5. Максимальне значення з усіх отриманих є мірою розмитості зображення  $S$ :  
 $S = \max\{S_i\}$ .

#### Оцінка контрасту зображення

Ще одним параметром, який визначає якість зображень, є контраст. Контраст – це градаційна характеристика чорно-білого чи кольорового зображення за різницею у світлоті (насиченості кольору) його найбільш яскравих і найбільш темні ділянки.

Оскільки зображення має складний набір характеристик, це породжує необхідність з розмаїття окремих комбінацій елементів зображення. При цьому всі елементи вважаються рівнозначними, і контраст кожної їхньої пари обчислюється

за формулою. Пікселі для порівняння можна вибирати у різний спосіб. Найпростіший спосіб - порівняння сусідніх пікселів у горизонтальному та вертикальному напрямку.

$$C_{ij} = \frac{L_i - L_j}{L_i + L_j}, \quad (3.3)$$

де  $L_i, L_j$  – яскравість елементів зображення. Далі, застосовуючи правило підсумовування контрастів, обчислюють набір величин, що визначають сприйняття кожної пари елементів зображення. Проводячи усереднення матриці локальних контрастів, набувають сумарний контраст. Отриманий результат може бути використаний як один із параметрів оцінки візуальної якості зображення.

Експериментально було встановлено, що оптимальне з погляду суб'єктивного сприйняття, зображення має нормальний розподіл яскравостей його елементів. Для зручності подальших розрахунків був використаний критерій нормального розподілу. За ступенем відхилення реального розподілу яскравостей від нормального проводилася оцінка якості зображення. Крім кількісної оцінки якості зображення, даний метод дозволяє отримати інформацію про наявність та вагове співвідношення яскравих градацій зображення. Результати оцінки якості зображення, отримані за даним методом, добре корелюють з суб'єктивною оцінкою візуальної якості зображення.

Якщо (3.1) замість  $G$  підставити (3.3), вийде комбінована оцінка різкості і яскравість зображення.

Визначення абсолютної кількісної оцінки якості зображення – це дуже складне, трудомістке і досі не вирішене до кінця завдання. На практиці найчастіше використовують порівняльні кількісні заходи якості зображень.

#### Оцінка різкості зображень

Однією з кількісних порівняльних оцінок якості зображення може бути служити різкість. У запропонованому алгоритмі різкість вимірювалася шляхом обчислення міри ексцесу двовимірного спектра Фур'є.

#### Алгоритм визначення різкості зображення

1. Виділити прямокутну область зображення.

2. Виділену область інтересу перетворити на частотну область.
3. Обчислити міру ексцесу двовимірного спектра Фур'є отриманої області.

Ексцес визначається як відношення четвертого центрального моменту та квадрата другого центрального моменту (стандартне відхилення четвертого ступеня).

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} = \frac{\mu_4}{\sigma^4}, \quad (3.4)$$

де  $\mu_2$  – другий центральний момент,  $\mu_4$  – четвертий центральний момент,  $\sigma$  – стандартне відхилення.

Оскільки високих просторових частот на різкому зображенні більше, ніж на розмитом, отже, двомірний розподіл спектру Фур'є для різкого зображення буде ширшим. Більш різкому зображенню райдужної оболонки відповідає велика величина ексцесу.

#### Згладжування зображення

Шум на зображенні суттєво впливає на результати виявлення ребер, тому потрібно його відфільтрувати. Для згладжування застосовується фільтр Гауса для згортки із зображенням. Рівняння для ядра фільтра Гауса має розмірність  $(2k + 1) * (2k + 1)$  та задається наступною формулою:

$$H_{ij} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(i - (k + 1))^2 + (j - (k + 1))^2}{2\sigma^2}\right), 1 \leq i, j \leq (2k + 1), \quad (3.5)$$

Важливо розуміти, що вибір розміру ядра Гауса має значний вплив на роботу детектора. Чим більший розмір ядра, тим нижче чутливість детектора до шуму. Крім того, помилка локалізації виявлення краю трохи збільшуватиметься зі збільшенням розміру ядра фільтра. У більшості випадків розмір  $5 \times 5$  є оптимальним, проте це залежить від конкретної ситуації.

### 3.3 Структура нейронної мережі

Серед різних структур нейронних мереж однією з найбільш відомих є багат шарова структура, в якій кожен нейрон в кожному шарі з'єднаний з усіма

аксонами нейронів попереднього шару, або, в разі першого шару, до тих пір, з усіма входами нейромережі. Вони мають назву - повнозв'язкові. Якщо мережа має тільки один шар, базове навчання, алгоритм цілком очевидний, оскільки правильні вихідні стану нейронів шару широко відомі, а також регулювання синаптичних з'єднань призначене для мінімізації помилки виведення мережі.

За цим принципом будується, наприклад, алгоритм навчання одношарового перцептрона. У багат шарових мережах оптимальні вихідні значення нейронів всіх шарів, крім останнього, як правило, не відомі, і двох або більше шаровий перцептрон вже неможливо навчити, керуючись тільки величинами помилок на виходах мережі.

Один з варіантів вирішення цієї проблеми - розробка наборів вихідних сигналів, що відповідають вхідним, для кожного шару нейронної мережі, що, звичайно, є дуже трудомісткою операцією і не завжди можливо.

Другий варіант — це динамічне регулювання синаптичного зважування, яке зазвичай передбачає вибір найслабших ланок і модифікацію їх так чи інакше з невеликою кількістю, зберігаючи при цьому модифікації, які зменшували вихідну помилку всієї мережі. Очевидно, незважаючи на свою простоту, цей метод вимагає великих рутинних обчислень.

Третій - поширення сигналів помилки від виходів мережі до її входів, у напрямку, зворотному до прямого поширення сигналів у звичайному режимі роботи. Він має назву процедури зворотного поширення. Надалі саме цей варіант буде використовуватися.

Згідно з методом найменших квадратів, цільовою функцією помилки нейромережі, що мінімізується, є величина:

$$E(w) = \frac{1}{2} \sum_{j,p} (y_{j,p}^N - d_{j,p})^2, \quad (3.6)$$

де  $y_{j,p}^N$  - реальний вихідний стан нейрона  $j$  вихідного шару  $N$  нейронної мережі при подачі її входи  $p$ -го образу;  $d_{j,p}$  - ідеальний (бажаний) вихідний стан цього нейрона.

Підсумовування виконується для кожного з нейронів вихідного шару та для всіх зображень, оброблюваних мережею. Зменшення відбувається за рахунок падіння градієнта, що означає, що вагові коефіцієнти коригуються наступним чином:

$$\Delta w_{ij}^n = -\eta * \frac{\partial E}{\partial w_{ij}}, \quad (3.7)$$

Тут  $w_{ij}$  - ваговий коефіцієнт синаптичного зв'язку, що з'єднує  $i$ -тий нейрон шару  $n-1$  з  $j$ -им нейроном шару  $n$ ,  $\eta$  - коефіцієнт швидкості навчання,  $0 < \eta < 1$ .

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial y_j} * \frac{dy_j}{ds_j} * \frac{\partial s_j}{\partial w_{ij}}, \quad (3.8)$$

Тут під  $y_j$ , як і раніше, мається на увазі вихід нейрона  $j$ , а під  $s_j$  – виважена сума його вхідних сигналів, тобто аргумент активаційної функції. Так як множник  $\frac{dy_j}{ds_j}$  є похідною цієї функції за її аргументом, з цього випливає, що похідна активаційна функція має бути визначена на всій осі абсцис. У зв'язку з цим функція одиничного стрибка та інші активаційні функції з неоднорідностями не підходять для нейромережі. Вони застосовуються такі гладкі функції, як гіперболічний тангенс чи класичний сигмоїд з експонентою. У разі гіперболічного тангенсу

$$\frac{dy}{ds} = 1 - s^2, \quad (3.9)$$

Третій множник  $\frac{\partial s_j}{\partial w_{ij}}$ , очевидно, дорівнює виходу нейрона попереднього шару  $y_i^{(n-1)}$ . Що стосується першого множника (3.8), він легко розкладається наступним чином:

$$\frac{\partial E}{\partial y_j} = \sum_k \frac{\partial E}{\partial y_k} * \frac{dy_k}{ds_k} * \frac{\partial s_k}{\partial y_j} = \sum_k \frac{\partial E}{\partial y_k} * \frac{\partial y_k}{\partial s_k} * w_{jk}^{(n+1)}, \quad (3.10)$$

Тут підсумовування  $k$  виконується серед нейронів шару  $n+1$ , ввівши нову змінну

$$\delta_j^n = \frac{\partial E}{\partial y_j} * \frac{dy_j}{ds_j}, \quad (3.11)$$

отримаємо рекурсивну формулу для розрахунків величин  $\delta_j^n$  шару  $n$  із величин  $\delta_k^{(n+1)}$  більш старшого шару  $n+1$ .

$$\delta_j^n = \left[ \sum_k \delta_k^{n+1} * w_{jk}^{(n+1)} \right] * \frac{dy_j}{ds_j}, \quad (3.12)$$

Для вихідного ж шару

$$\delta_l^n = (y_l^n - d_l) * \frac{dy_l}{ds_l}, \quad (3.13)$$

Теперь можемо записати (3.7) в розгорнотому вигляді:

$$\Delta w_{ij}^n = -\eta * \delta_j^n * y_i^{n-1}, \quad (3.14)$$

Також вводимо функцію, яка несе інформацію про витрати класифікації. Для  $k$ -го атрибуту вона визначається як  $ICF_k = \frac{(2^{\Delta I_k - 1})}{(C_{ij+1})^\alpha}$ , де  $0 \leq \alpha \leq 1$ ,  $\Delta I_k$  – приріст інформації, пов'язані з розбиттям по  $k$ -му атрибуту,  $C_{ij}$  – витрати, пов'язані з класами, приклади яких брали участь у розбитті.

Параметр  $\alpha$  дозволяє варіювати ступенем «прагнення» алгоритму вибору атрибутів, із якими пов'язані менші витрати. Якщо  $\alpha = 0$ , а  $ICF = 1$ , то витрати не враховуються. Якщо  $\alpha = 1$ , то має місце максимальний вплив витрат на процес побудови дерева. Регулюючи значення цього параметра, дослідник досягає раціональної чутливості алгоритму до витрат класифікації.

Іноді для надання процесу корекції ваги деякої інерційності, що згладжує різкі стрибки при переміщенні по поверхні цільової функції (3.14), доповнюється значенням зміни ваги на попередній ітерації.

$$\Delta w_{ij}^n(t) = -\eta * (\mu * \Delta w_{ij}^n(t-1) + (1 - \mu) * \delta_j^n * y_i^{n-1}), \quad (3.15)$$

де  $\mu$  – коефіцієнт інерційності,  $t$  – номер поточної ітерації. Таким чином, повний алгоритм навчання нейромережі за допомогою процедури зворотного розповсюдження будується так:

Подати на входи мережі один із можливих образів та в режимі звичайного функціонування, коли сигнали поширюються від входів до виходів, розрахувати значення останніх. Нагадаємо, що

$$S_j^n = \sum_{i=0}^M y_i^{n-1} * w_{ij}^n, \quad (3.16)$$

де  $M$  - число нейронів у шарі  $n-1$  з урахуванням нейрона з постійним вихідним станом  $+1$ , що задає зміщення;  $y_i^{(n-1)}=x_{ij}^{(n)}$  -  $i$ -тий вхід нейрона  $j$  шару  $n$ .

$$j(n) = f(sj(n)), \quad (3.17)$$

де  $f$  - сигмоїд

$$q(0) = Iq, \quad (3.18)$$

де  $Iq$  –  $q$ -та компонента вектора вхідного образу. Розрахувати  $\delta^{(n)}$  для вхідного шару за формулою (3.13).

Розрахувати за формулою (3.14) або (3.15) зміни ваг  $\Delta w(N)$  шару  $N$ .

Розрахувати за формулами (3.12) та (3.14) (або (3.12) та (3.15)) відповідно  $\delta^{(n)}$  та  $\Delta w^{(n)}$  для всіх інших шарів,  $n=N-1, \dots, 1$ . Скоригувати всі ваги

$$w_{ij}^n(t) = w_{ij}^n(t-1) + \Delta w_{ij}^n(t), \quad (3.19)$$

Якщо помилка мережі суттєва, перейти до шагу 1. В іншому випадку - кінець.

Нейромережі на кроці 1 по черзі у випадковому порядку пред'являються всі тренувальні образи, щоб мережа, не забувала один під час запам'ятовування інших.

З виразу (14) слід, що коли вихідне значення  $y_i^{(n-1)}$  прагне до нуля, ефективність навчання помітно знижується. При двійкових вхідних векторах в середньому половина вагових коефіцієнтів не коригуватиметься, тому область можливих значень виходів нейронів  $[0,1]$  бажано зрушити в межі  $[-0.5, +0.5]$ , що досягається простими модифікаціями логістичних функцій. Наприклад, сигмоїд з експонентою перетворюється на вигляд

$$f(x) = -0.5 + \frac{1}{1 + e^{-ax}}, \quad (3.20)$$

Тепер торкнемося питання ємності, тобто числа образів, що пред'являються на її входи, які вона здатна навчитися розпізнавати. Для мереж із кількістю шарів більше двох, він залишається відкритим. Вихідним і одним прихованим шаром, детерміністська ємність мережі  $C_d$  оцінюється так:



$$\frac{N_w}{N_y} < Cd < \frac{N_w}{N_y} * \log\left(\frac{N_w}{N_y}\right), \quad (3.21)$$

$N_w$  - число ваг, що підлаштовуються,  $N_y$  - число нейронів у вихідному шарі.

Для розділення набору вхідних зображень кожен логічний рівень — «1» і «0» — матиме окремий клас. Тепер можна закодувати 4 класи на два виходи тощо. Однак результати роботи мережі, організованої таким чином, – не дуже надійні. Для підвищення надійності класифікації доцільно ввести надлишок, присвоївши кожному вихідному шару нейрон або, ще краще, більше, кожен з яких вчиться визначати належність образу до класу зі своїм ступенем надійності, наприклад високий, середній та низький. Такі мережі дозволяють класифікувати вхідні зображення на незрозумілі набори. Ця властивість наближає ці нейронні мережі до умов життя.

### 3.4 Вибір активаційної функції

Сигмоїда являється гладкою та подібна до ступінчастої функції. Розглянемо її переваги. По-перше, сигмоїда — нелінійна за своєю природою, а комбінація таких функцій робить також нелінійну функцію. Ще одна перевага такої функції - вона не бінарна, що робить активацію аналоговою, на відміну від ступінчастої функції. Для сигмоїди також характерний гладкий градієнт.

Сигмоїда дійсно виглядає підходящою функцією для класифікаційних завдань. Вона прагне привести значення однієї зі сторін кривої (наприклад, до верхнього при  $x=2$  і нижньому при  $x=-2$ ). Така поведінка дозволяє знаходити чіткі межі під час передбачення.

Сьогодні сигмоїда є однією з найчастіших активаційних функцій у нейромережах. Але й вона має недоліки, на які варто звернути увагу. При наближенні до кінців сигмоїди значення  $Y$  мають тенденцію реагувати на зміни в  $X$ . Це означає, що градієнт у таких областях набуває маленьких значень. А це, у свою чергу, призводить до проблем із градієнтом зникнення.

$$f(x) = \frac{1}{1+e^{-s}}, \quad (3.22)$$

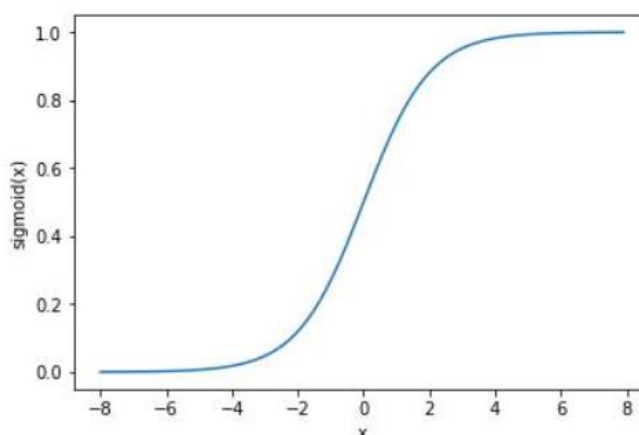


Рисунок 1. Графік функції – сигмоїд.

Вибір даної функції обумовлено наступними чинниками:

- симетричні активаційні функції типу сигмоїда забезпечують швидшу збіжність, ніж стандартна логістична функція;
- функція має безперервну першу похідну;
- функція має просту похідну, яка може бути обчислена через її значення, що дає економію у обчисленнях.

### 3.5 Порівняльний аналіз якості розпізнавання

Запропонований алгоритм був апробований експериментальним шляхом з використанням зображень, які отримують, наприклад, з камер спостереження.

Для класу розглянутих алгоритмів при розпізнаванні зображень доволі часто використовують згорткові або штучні нейронні мережі. Але тільки в тому випадку, коли на навчання подаються чисті зображення. Якщо параметричні алгоритми навчати лише чистим зображенням або лише зображенням у шумі, то результати їхнього розпізнавання можуть мати значну похибку при різних типах зображення. Для багатьох алгоритмів трохи відрізняються результати розпізнавання під час навчання, тому досить часто у навчанні використовують як чисті так і зображенням у шумі.

Було проведено порівняння наступних методів: метод Віоли-Джонса та запропонований алгоритм детекції об'єктів з врахуванням витрат на помилки класифікації. Для порівняння використовувалася база зображень, отримана з загальнодоступної навчальної тестової бази та складається з 2 частин: навчальний набір та тестуючий набір. Навчальний набір складається 12567 зображень, а тестуючий набір складається з 1742 зображень.

Експеримент проходив у 2 етапи: на першому етапі використовувався навчальний набір. На другому етапі використовувалася вибірка, для тестування. Детекція осіб виконувалася класифікатором Віоли-Джонса та запропонованим алгоритмом. Результати порівняння алгоритмів представлені на графіках, зображених на рис. 2 та рис. 3, червоним кольором відображено запропонований алгоритм, синім – Віоли-Джонса.

Порівняння класифікаторів було виконано за допомогою побудови ROC-кривих (receiver operating characteristic curve), що виражає співвідношення рівня вірних та хибних виявлень. ROC-крива є параметрично заданою функцією порога прийняття рішень, на якій по осі X відкладаються ймовірності помилково-позитивних рішень, а по осі Y - ймовірності істинно-позитивних рішень. При цьому розраховувався чисельний показник площі під ROC-кривою Area Under Curve.

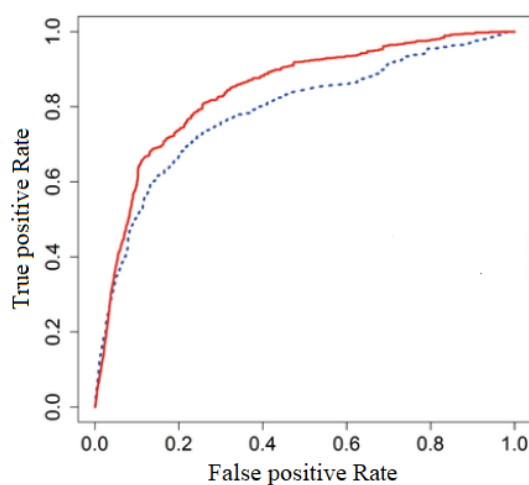


Рис 2. Roc curve

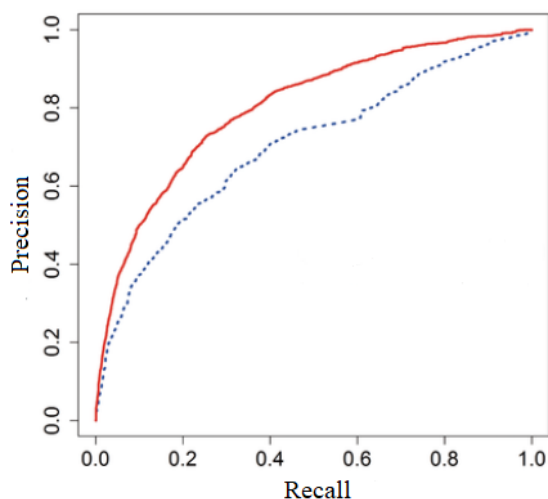


Рис 3. Precision-recall curve

### 3.10 Результати моделювання

В результаті моделювання було з'ясовано, що ефективність класифікатора на базі запропонованої моделі нейромережі становить 91% і є максимальною для даного класифікатора. При цьому класифікатор Віоли-Джонса на цій же базі дорівнює 86%, що суттєво нижче (Рис 2).

Результати другого експерименту показали наступне: на тестовій вибірці для класифікатора Віоли-Джонса залишився на значенні 86%, з відхиленнями в соті частки. Різниця може бути пояснена меншою кількістю зображень у вибірці, що тестувалася. Більший інтерес викликають результати роботи класифікатора нейромережі, якого склав 89,5%, що вище за значення, отриманого першим класифікатором. Цей факт дозволяє зробити висновок ефективність застосування запропонованого алгоритму в реальних системах розпізнавання образів. Можливість онлайн навчання без втручання людини дозволяє адаптувати алгоритм під умови зйомки, що змінюються, і якість вступників.

Недоліком запропонованого алгоритму є збільшення обчислювальних витрат під час роботи класифікатора, оскільки цей класифікатор більш вимогливий до ресурсів. Під час роботи з відеопотоком цей недолік компенсується використанням детектора руху. Тоді пошук образу виконується не у всьому зображенні, а лише в області руху, що знижує обчислювальне навантаження.

## Висновок

Очевидно, що кількісні заходи якості зображення необхідні для проектування та оцінки систем відтворення зображень. Ці заходи внесуть істотні зміни в усунення складності та неточності сучасних методів обробки зображення у формі суб'єктивної експертизи. Крім того, існують методи оптимізації систем обробки зображень, які можуть бути розроблені на основі кількісних вимірювань.

У розробці кількісних оцінок якості зображення досягнуто значних успіхів. Проте запроваджені критерії є досить досконалими. Більшість спроб знайти прийнятні оцінки якості зображення відноситься до окремих випадків. Пропонується конкретна оцінка, заснована на фізіологічних умовах та часто просто підходить для аналізу або розрахунку, після чого оцінюється її характеристика. Створення більш точних оцінок якості зображення вимагає більш глибокого вивчення характеристик зорової системи людини.

Розглянута нейромережа має кілька «вузьких місць». По-перше, у процесі навчання може виникнути ситуація, коли великі позитивні чи негативні значення вагових коефіцієнтів змістять робочу точку на сигмоїдах багатьох нейронів до області насичення. Малі величини похідної від логістичної функції приведуть у відповідність до (3.12) та (3.13) до зупинки навчання, що паралізує нейромережу.

По-друге застосування методу градієнтного спуску не гарантує, що буде знайдено глобальний, або не локальний мінімум цільової функції. Ця проблема пов'язана ще з однією, а саме - з вибором величини швидкості навчання. Доказ збіжності навчання у процесі зворотного поширення ґрунтується на похідних, тобто збільшенні ваги, отже швидкість навчання має бути нескінченно малою, проте навчання відбуватиметься повільно. Крім того, для виключення випадкових потраплянь у локальні мінімуми іноді, після того, як значення вагових коефіцієнтів стабілізується, короткочасно сильно збільшують, щоб почати градієнтний спуск з нової точки. Якщо повторення цієї процедури кілька разів приведе алгоритм в той самий стан мережі, можна більш-менш впевнено сказати, що знайдено глобальний максимум, а не якийсь інший.

## Висновок

В результаті виконання магістерської роботи підвищено ефективність процесу автоматизації об'єктів зображення на основі методів машинного навчання.

При виконанні роботи було виконано наступні задачі.

1. Було проведено критичний аналіз існуючих методів оптичного розпізнавання образів та виявлено недоліки. Як правило, проблемою є низька стійкість алгоритмів до впливу зовнішніх умов, що ускладнюють якість розпізнавання: рівень освітлення; якість зображення; нахили зображення. При цьому алгоритми, стійкі до зовнішніх умов, найчастіше виявляються вимогливими до апаратних обчислювальних ресурсів, що утруднює застосування їх у системах, що працюють у реальний час.

2. Розглянуто алгоритми на основі машинного навчання, які активно застосовуються у практичних завданнях комп'ютерного зору. Розглянуто ключові моменти розробки моделі на основі штучних нейронних мереж, обрано цільову функцію та метод оптимізації при навчанні моделей.

3. Проаналізовано вихідний масив даних, що використовується для навчання моделей. Складено план заходів щодо регуляризації процесу навчання. Розглянуто ряд практик з реалізації та організації процесу навчання даних моделей найефективнішим чином.

4. Розроблена модель на основі штучних нейронних мереж для вирішення поставленого практичного завдання. Вибрано метрику оцінки моделей, ґрунтуючись на точності їх пророцтв при роботі з незнайомими даними, іншими словами, здатність до узагальнення. Робота моделі була оцінена з точки зору точності рішення поставленого завдання, в результаті чого обрана модель, що показала кращу продуктивність.

## Список використаної літератури

1. Fedorenko, F. Real-time object-to-features vectorisation via Siamese neural networks / F. Fedorenko, S. Usilin // Proceedings of the Ninth International Conference on Machine Vision (ICMV 2016). — SPIE, 2017. — Vol. 10341. — 103411R. (SCOPUS)
2. Minkina, A. Generalization of the Viola-Jones method as a Decision Tree of Strong Classifiers for Real-time Object Recognition in Video Stream / A. Minkina, D. Nikolaev, S. Usilin, V. Kozyrev // Proceedings of the Seventh International Conference on Machine Vision (ICMV 2014). — SPIE, 2015. — Vol. 9445. — 944517. (SCOPUS)
3. Usilin, S. Visual Appearance Based Document Image Classification / S. Usilin, D. Nikolaev, V. Postnikov, G. Schaefer // Proceedings of the 17th IEEE International Conference on Image Processing (ICIP 2010). — IEEE, 2010. — P. 2133-2136. (SCOPUS, WoS)
4. Usilin, S. Guilloche Elements Recognition Applied to Passport Page Processing / S. Usilin, D. Nikolaev, D. Sholomov // Proceedings of the 8th Open German-Russian Workshop «Pattern Recognition and Image Understanding» OGRW-8-2011. — Nizhny Novgorod : Lobachevsky State University of Nizhny Novgorod, 2011. — P. 303- 306. 1
5. H. Golnabi and A. Asadpour, “Design and application of industrial machine vision systems,” *Robotics and Computer-Integrated Manufacturing*, vol. 23, no. 6, pp. 630–637, 2007.
6. O. Russakovsky et al., “Imagenet large scale visual recognition challenge,” *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.
7. K. B. Bulatov and D. V. Polevoy, “Reducing overconfidence in neural networks by dynamic variation of recognizer relevance.” in *ECMS*, 2015, pp. 488–491.
8. Iba, W. Induction of One-Level Decision Trees / W. Iba, P. Langley // *ML92: Proceedings of the Ninth International Conference on Machine Learning*, Aberdeen, Scotland, 1–3 July 1992. — 1992. — P. 233–240.

9. Roth, D. A SNoW-based face detector / D. Roth, M. Yang, N. Ahuja // *Adv. Neural Inf. Process. Syst.* — 2000. — Vol. 12. — P. 855–861.
10. Schneiderman, H. A statistical method for 3D object detection applied to faces and cars / H. Schneiderman, T. Kanade // *IEEE Conference on Computer Vision and Pattern Recognition.* — 2000. — Vol. 1. — P. 746–751.
11. Sung, K.K. Example-based learning for view-based human face detection / K.K. Sung, T. Poggio // *IEEE Trans. Pattern Anal. Mach. Intell.* — 1998. — Vol. 20. — № 1. — P. 39–51.
12. Rowley, H.A. Neural network-based face detection / H.A. Rowley, S. Baluja, T. Kanade // *IEEE Trans. Pattern Anal. Mach. Intell.* — 1998. — Vol. 20. — № 1. — P. 23–38.
13. Huang, C.-C. An Extended Set of Haar-like Features for Bird Detection Based on AdaBoost / C.-C. Huang, C.-Y. Tsai, H.-C. Yang // *Signal Processing, Image Processing and Pattern Recognition SE – 17* / ed. Kim T. et al. — Springer Berlin Heidelberg, 2011. — Vol. 260. — P. 160–169.
14. Francke, H. Real-Time Hand Gesture Detection and Recognition Using Boosted Classifiers and Active Learning / H. Francke, J. Ruiz-del-solar, R. Verschae // *Lect. Notes Comput. Sci.* — 2007. — Vol. 4872. — P. 533–547.
15. Ojala, T. A comparative study of texture measures with classification based on feature distributions / T. Ojala, M. Pietikäinen, D. Harwood // *Pattern Recognit.* — 1996. — Vol. 29. — № 1. — P. 51–59.
16. Yang, Z. Demographic Classification with Local Binary Patterns / Z. Yang, H. Ai // *Proceedings of the 2007 International Conference on Advances in Biometrics.* — Berlin, Heidelberg: Springer-Verlag, 2007. — P. 464–473.
17. Chawla N. C4.5 and imbalanced datasets: investigating the effect of sampling method, probabilistic estimate, and decision tree structure / N. Chawla // *ICML 2003 Workshop on Imbalanced Datasets.*
18. Vinciotti V. Scorecard construction with unbalanced class sizes / V. Vinciotti, D.J. Hand // *Journal of Iranian Statistical Society.* — 2002. — Vol. 2 — P. 189–205.



19. Weiss G.M. Cost-Sensitive Learning vs. Sampling: Which is Best for Handling Unbalanced Classes with Unequal Error Costs? / G.M. Weiss, K. McCarthy, B. Zabar // Proceedings of the 2007 International Conference on Data Mining, CSREA Press, 2007. – P. 35-41.
20. McCarthy K. Does Cost-Sensitive Learning Beat Sampling for Classifying Rare Classes? / K. McCarthy, Zabar B., Weiss G.M. // Proceedings of the First International Workshop on Utility-Based Data Mining (at KDD-05), ACM Press, 2005. – P. 69-75.
21. Elchan Ch. The Foundations of Cost-Sensitive Learning / Ch. Elchan // Proc. of the 17th International Joint Conference on Artificial Intelligence, 2001. – P. 973-978.
22. Kubat M. Addressing the curse of imbalanced training sets: one-sided selection / M. Kubat, S. Matwin // In: Proc. 14th International Conference on Machine Learning, 1997. – P. 179-186.
23. SMOTE: Synthetic Minority Over-sampling Technique / N. Chawla, K. Bowyer, L. Hall, W. Kegelmeyer // Journal of Artificial Intelligence Research. – 2002. – Vol. 16. – P. 341-378.
24. Kniaz V. V. Real-time optical flow estimation on a GPU for a skied-steered mobile robot // Proc. SPIE 9897, Real-Time Image and Video Processing. - 2016. - DOI:10.1117/12.2227556.
25. Hussain M., Chen D., Cheng A., Wei H., Stanley D. Change detection from remotely sensed images: From pixel-based to object-based approaches // ISPRS J. Photogrammetry and Remote Sensing. - 2013. - V. 80. - P. 91–106.
26. Lukas J., Fridrich J., Goljan M. Digital camera identification from sensor pattern noise // IEEE Trans. Information Forensics and Security. - 2006. - V. 1(2). - P. 205–214.
27. Smolin V. Some Ideas of Informational Deep Neural Networks Structural Organization // Advances in Neural Networks: Proc. Springer. - 2016. - P. 572–582.

28. Vizilter Yu., Gorbatsевич V.S., Rubis A., Zheltov S. Shape-Based Image Matching Using Heat Kernels and Diffusion Maps // Intern. Arch. Photogrammetry and Remote Sensing. Spatial Inf. Sci. - 2014. - V. XL-3. - P. 357–364.
29. Vizilter Y.V., Rubis A.Y., Zheltov S.Y., Vygolov O.V. Change detection via morphological comparative filters // ISPRS Ann. Photogrammetry, Remote Sensing and Spatial Information Sci. 2016. 23rd ISPRS Congress, 12–19 July, Prague, Czech Republic. - 2016. - V. III-3. - P. 279–286. - DOI:10.5194/isprs-annals-III-3-279-2016.
30. Volkov V. Segmentation and Extraction of Extensive Objects on Digital Images // Proc. 2009 Intern. Conf. Image Processing, Computer Vision and Pattern Recognition, IPCV2009. Las Vegas, Nevada, USA: CSREA Press. - 2009. - V. 2. - P. 656–662.
31. Knyaz V. A. Scalable Photogrammetric Motion Capture System ‘Mosca’ // ISPRS — International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. - 2015. - V. XL-5/W6. - P. 43–49.
32. Tang, R., Fritsch, D. Correlation analyses of camera self-calibration in close range photogrammetry. Photogrammetric Record. - 2013. – № 28 (141). P. 86-95.
33. Stamatopoulos, C., Fraser, C. Calibration of long focal length cameras in close range photogrammetry. Photogrammetric Record. 2011. – № 26 (135). P. 339-360.
34. Zhang Z. A Flexible New Technique for Camera Calibration / Zhang Zhengyou // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 2000. - December. - T. 22. - P. 1330–1334 [Электронный ресурс]. URL: <https://www.microsoft.com/en-us/research/publication/a-flexible-new-technique-for-camera-calibration/> (дата обращения: 20.12.2020).
35. Lepetit, V., Moreno-Noguer, F., & Fua, P. (2009). An accurate O (n) solution to the PnP problem. IJCV. №81 (2). - P. 155-166.

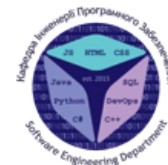
36. Kneip L. UPnP: An Optimal  $O(n)$  Solution to the Absolute Pose Problem with Universal Applicability / Kneip L., Hongdong L., Yongduek S. // *Computer Vision – ECCV 2014*. - Cham: Springer International Publishing. - 2014. - P.127–142. - DOI:10.1007/978-3-319-10590-1\_9.
37. Josephson K. Pose estimation with radial distortion and unknown focal length / Josephson K., Byrod M. // *2009 IEEE Conference on Computer Vision and Pattern Recognition*. – June, 2009. - P. 2419–2426. – DOI: 10.1109/CVPR.2009.5206756.
38. Chandraker, M., Agarwal, S., Kriegman, D., Belongie, S. Globally optimal algorithms for stratified autocalibration. *IJCV*. – 2010. - № 90 (2010). – P. 236–254.
39. Orghidan R. Camera calibration using two or three vanishing points / R. Orghidan, J. Salvi, M. Gordan, B. Orza // *2012 Federated Conference on Computer Science and Information Systems (FedCSIS)*. - 2012. - Sept. - P.123–130.
40. Kong H., Audibert J.-Y., Ponce J. Vanishing point detection for road detection// *Proc. IEEE Conf. Comput. Vis. Pattern Recog.* – 2009. - P. 96–103.
41. Moghadam P., Starzyk J. A., Wijesoma W. S. Fast vanishingpoint detection in unstructured environments// *IEEE Trans. on Image Processing*. – 2012. - V. 21. - № 1. - P. 425–430.
42. Zhai M., Workman S., Jacobs N. Detecting vanishing points using global image context in a non-manhattan world// *Proc. IEEE Conf. Comput. Vis. Pattern Recog.* – 2016. - P. 5657–5665.
43. Josephson K. Pose estimation with radial distortion and unknown focal length / Josephson K., Byrod M. // *2009 IEEE Conference on Computer Vision and Pattern Recognition*. - 2009. - June. - P.2419–2426. - DOI: 10.1109/CVPRW.2009.5206756.
44. Wang Y. An Analysis of the Viola-Jones Face Detection Algorithm / Wang Y.// *Image Processing On Line*, 2014. — 06. — V. 4. — P. 128—148. - DOI: 10.5201/ipol.2014.104.

45. Khan, F.A.: M. Ur Rehman, A. Khalid, M. Ali, M. Imran, M. Nawaz, et al. An intelligent data service framework for heterogeneous data sources// Journal of Grid Computing. – 2019. - №17. - P. 577–589.
46. Fischler A.M. Random Sample Consensus: A Paradigm for Model Fitting with Applications To Image Analysis and Automated Cartography / A. Fischler A. M., Bolles C. R./ Communications of the ACM, 1981. - 06. - T. 24. - P. 381–395. – DOI: 10.1145/358669.358692.
47. Canny J. A Computational Approach to Edge Detection / Canny J. // IEEE Transactions on Pattern Analysis and Machine Intelligence. - V. PAMI-8. - No.6. – P.679-698. - DOI: 10.1109/TPAMI.1986.4767851.
48. Duda R.O. Use of the Hough Transform to Detect Lines and Curves in Pictures / Duda R.O, Hart P.E. // Communications of the ACM. – January, 1972. – DOI: 10.1145/361237.361242.

Додаток А  
Презентація



ДЕРЖАВНИЙ УНІВЕРСИТЕТ ТЕЛЕКОМУНІКАЦІЙ  
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ  
ТЕХНОЛОГІЙ



Кафедра інженерії програмного забезпечення

## Магістерська робота

«РОЗРОБКА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ДЛЯ КЛАСИФІКАЦІЇ ТА ПОШУКУ  
ОБ'ЄКТІВ ЗОБРАЖЕННЯ ЗА ДОПОМОГОЮ МАШИННОГО НАВЧАННЯ»

Виконав: студент групи ПДМ-61 Сьомін Денис Андрійович

Керівник: завідувач кафедри, кандидат технічних наук, доцент, Негоденко Олена  
Василівна

Київ - 2021

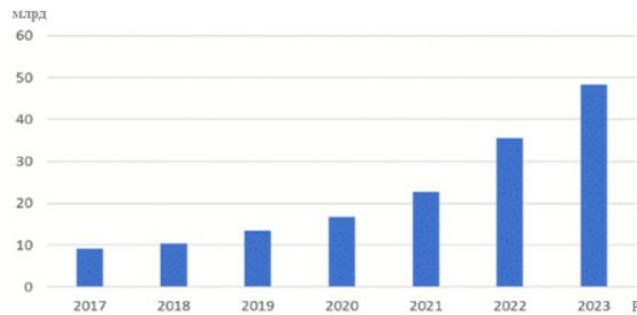
### МЕТА, ОБ'ЄКТ ТА ПРЕДМЕТ ДОСЛІДЖЕННЯ

**Мета роботи:** підвищення ефективності процесу автоматизації об'єктів зображення на основі методів машинного навчання

**Об'єкт дослідження:** автоматизація класифікації та пошуку об'єктів зображення за допомогою машинного навчання.

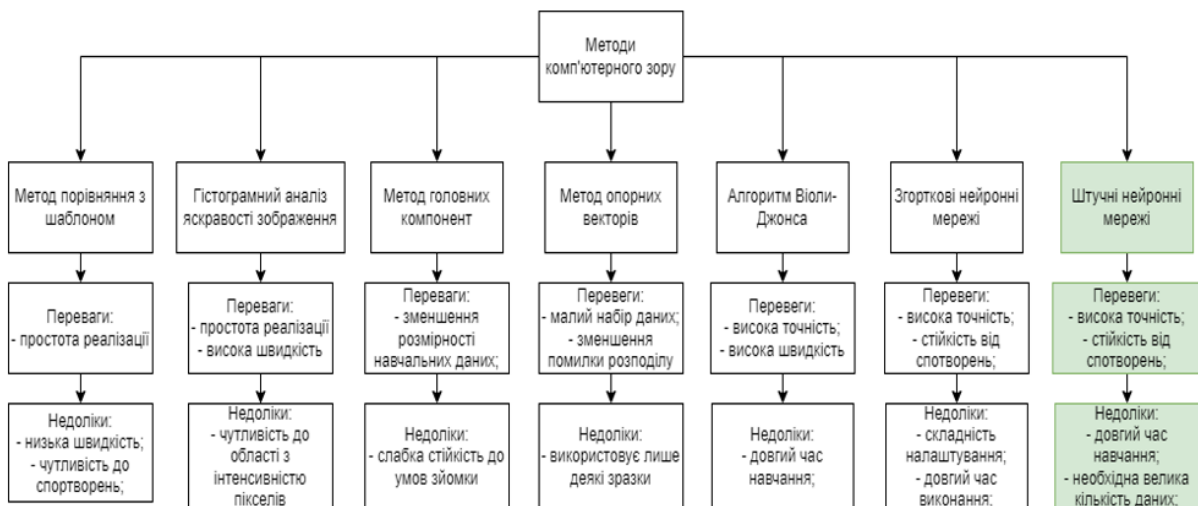
**Предмет дослідження:** комп'ютерний зір, методи, алгоритми та засоби машинного навчання для систем розпізнавання образів.

## АКТУАЛЬНІСТЬ РОБОТИ



3

## АНАЛІЗ ІСНУЮЧИХ РІШЕНЬ



4

## АЛГОРИТМ НАВЧАННЯ НЕЙРОННОЇ МЕРЕЖІ

Крок 1. Задається крок навчання  $\alpha$  ( $0 < \alpha < 1$ ) та бажана середньоквадратична помилка мережі  $E_m$ .

Крок 2. Ініціалізуються випадковим чином вагові коефіцієнти  $w_{ij}$  та порогові  $b_j$  к значення НР.

Крок 3. Подаються послідовно образи з навчальної вибірки на вхід нейронної мережі. При цьому для кожного образу виконуються такі дії:

3.1. Виробляється фаза прямого поширення вхідного образу нейронної мережі. Обчислюється вихідне значення всіх нейронів  $Y_j$  к .

3.2. Обчислюються помилки  $X_j$  нейронів вихідного та прихованого шарів.

3.3. Здійснюється зміна вагових коефіцієнтів та порогів нейронних елементів для кожного шару нейронної мережі.

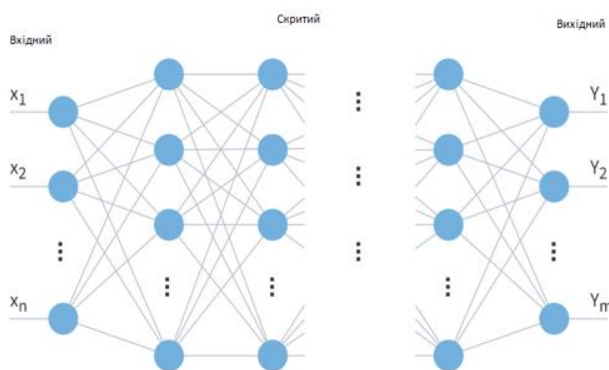
Крок 4. Обчислюється сумарна помилка нейронної мережі

Крок 5. Якщо  $E > E_m$ , то відбувається перехід до кроку 3, інакше виконання алгоритму завершується



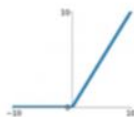
5

## СТРУКТУРА НЕЙРОННОЇ МЕРЕЖІ

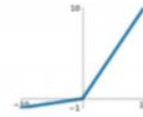


## ФУНКЦІЇ АКТИВАЦІЇ

**ReLU**  
 $\max(0, x)$

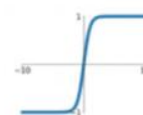


**Leaky ReLU**  
 $\max(0.1x, x)$

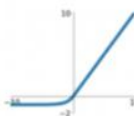


**Maxout**  
 $\max(w_1^T x + b_1, w_2^T x + b_2)$

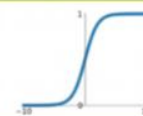
**tanh**  
 $\tanh(x)$



**ELU**  
$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



**Sigmoid**  
$$\sigma(x) = \frac{1}{1 + e^{-x}}$$



6

## МОДИФІКАЦІЯ АЛГОРИТМУ НЕЙРОННОЇ МЕРЕЖІ

Вводимо функцію, яка несе інформацію про витрати класифікації.

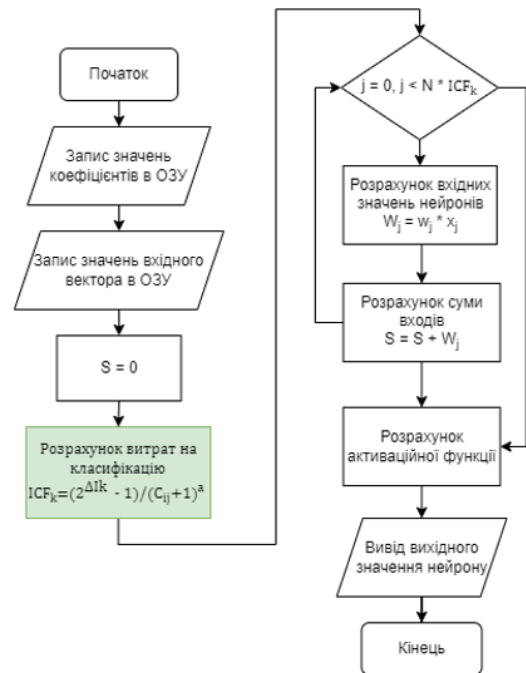
Для  $k$ -го атрибуту вона визначається як

$$ICF_k = \frac{(2^{\Delta I_k - 1})}{(C_{ij+1})^\alpha}, \quad (1)$$

де  $0 \leq \alpha \leq 1$ ,

$\Delta I_k$  – приріст інформації, пов'язані з розбиттям по  $k$ -му атрибуту,

$C_{ij}$  – витрати, пов'язані з класами, приклади яких брали участь у розбитті.



7

## ПРАКТИЧНИЙ РЕЗУЛЬТАТ

№	Достовірність	Об'єкт	Правильність	Точність	Відгук
1	89%	Машина	Так	1/1 = 1.0	1/16 = 0.06
2	33%	Машина	Так	2/2 = 1.0	2/16 = 0.13
3	87%	Людина	Так	3/3 = 1.0	3/16 = 0.2
4	82%	Людина	Так	4/4 = 1.0	4/16 = 0.25
5	17%	Сумка	Ні	4/5 = 0.8	-
6	64%	Лавочка	Так	5/6 = 0.83	5/16 = 0.31
7	49%	Людина	Так	6/7 = 0.86	6/16 = 0.38
8	52%	Велосипед	Ні	6/8 = 0.75	-
9	20%	Сумка	Ні	6/9 = 0.67	-
10	37%	Людина	Так	7/10 = 0.7	7/16 = 0.44
11	73%	Машина	Так	8/11 = 0.73	8/16 = 0.5
12	69%	Посажена рослина	Так	9/12 = 0.75	9/16 = 0.56
13	66%	Машина	Так	10/13 = 0.77	10/16 = 0.63
14	33%	Машина	Так	11/14 = 0.79	11/16 = 0.7
15	54%	Посажена рослина	Так	12/15 = 0.8	12/16 = 0.75
16	-	Посажена рослина	Не знайдено	-	12/16 = 0.75
17	-	Машина	Не знайдено	-	12/16 = 0.75
18	-	Посажена рослина	Не знайдено	-	12/16 = 0.75
19	-	Машина	Не знайдено	-	12/16 = 0.75



Точність алгоритму розраховується за формулою:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (1)$$

де  $TP$  – число правильно знайдених об'єктів, а  $FP$  – число помилково знайдених об'єктів.

Відгук алгоритму розраховується за формулою:

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (2)$$

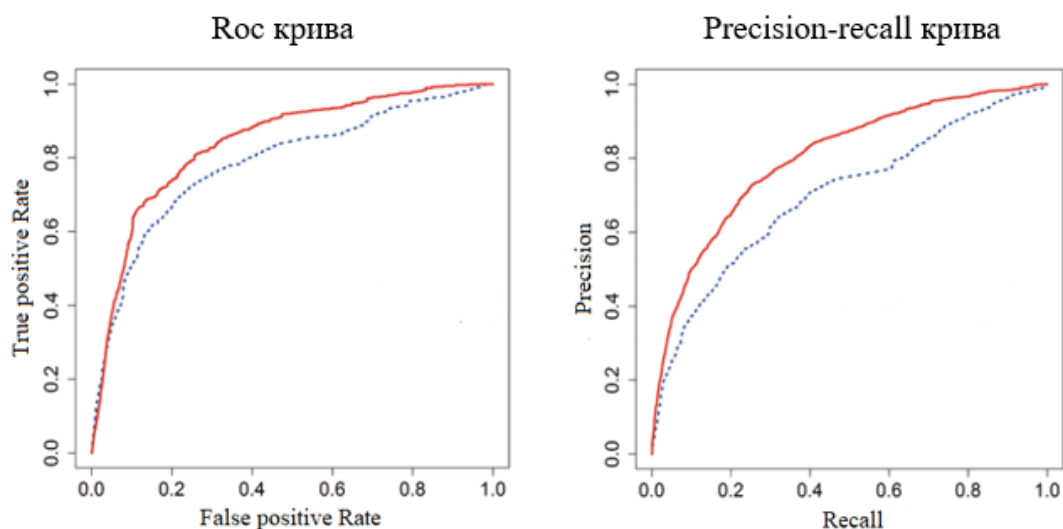
де  $TP$  – число правильно знайдених об'єктів, а  $FN$  – число не розпізнаних областей.

Використовуючи формулу наведену ви було розраховано, що точність класифікації склала 80%, а відгук склав 75%.

8



## РЕЗУЛЬТАТИ МОДЕЛЮВАННЯ



ROC-крива є параметрично заданою функцією порога прийняття рішень, на якій по осі X - відкладаються ймовірності помилково-позитивних рішень, по осі Y - ймовірності істинно-позитивних рішень.

Червоною лінією відображено запропонований алгоритм, синій – алгоритм Віоли-Джонса.

9

## ВИСНОВКИ

1. Було проведено критичний аналіз існуючих методів оптичного розпізнавання образів та виявлено недоліки. Як правило, проблемою є низька стійкість алгоритмів до впливу зовнішніх умов, що ускладнюють якість розпізнавання: рівень освітлення; якість зображення; нахили зображення.
2. Розглянуто алгоритми на основі машинного навчання, які активно застосовуються у практичних завданнях комп'ютерного зору. Розглянуто ключові моменти розробки моделі на основі штучних нейронних мереж, обрано цільову функцію та метод оптимізації при навчанні моделей.
3. Проаналізовано вихідний масив даних, що використовується для навчання моделей. Складено план заходів щодо регуляризації процесу навчання. Розглянуто ряд практик з реалізації та організації процесу навчання даних моделей найефективнішим чином.
4. Розроблена модель на основі штучних нейронних мереж для вирішення поставленого практичного завдання. Вибрано метрику оцінки моделей, ґрунтуючись на точності пророцтв при роботі з незнайомими даними або здатності до узагальнення.

10