

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Національний аерокосмічний університет ім. М.Є. Жуковського  
«Харківський авіаційний інститут»

Факультет програмної інженерії та бізнесу

Кафедра інженерії програмного забезпечення

## Пояснювальна записка до дипломного проекту

магістра  
(освітній ступінь)

на тему «Експериментальне дослідження нейронних мереж розпізнавання та  
адаптації зображень»

ХАІ.603.667п2.121.166363.200

Виконав: студент 6 курсу групи № 667 п2  
Спеціальність 121 – Інженерія програмного  
забезпечення

(код та найменування)

Освітня програма Хмарні обчислення  
та Інтернет речей

(найменування)

Калугін В.О.

(прізвище й ініціали студента)

Керівник: Вдовітченко О.В.

(прізвище й ініціали)

Рецензент: Філімончук Т.В.

(прізвище й ініціали)

**Міністерство світи і науки України**  
**Національний аерокосмічний університет ім. М. Є. Жуковського**  
**«Харківський авіаційний інститут»**

Факультет програмної інженерії та бізнесу  
(повне найменування)

Кафедра інженерії програмного забезпечення  
(повне найменування)

Рівень вищої освіти другий (магістерський)

Спеціальність 121 – інженерія програмного забезпечення  
(код та найменування)

Освітня програма хмарні обчислення та Інтернет речей  
(найменування)

**ЗАТВЕРДЖУЮ**  
**Завідувач кафедри**

Туркин І.Б.  
(ініціали та прізвище)  
“ \_\_\_\_\_ ” \_\_\_\_\_ 2020 року  
(підпис)

**З А В Д А Н Н Я**  
**НА ДИПЛОМНИЙ ПРОЕКТ СТУДЕНТУ**

**Калугіну Владиславу Олексійовичу**

(прізвище, ім'я, по батькові)

1. Тема дипломного проекту Експериментальне дослідження нейронних мереж розпізнавання та адаптації зображень.

Керівник дипломного проекту Вдовітченко Олександр Валерійович к.т.н., доц.  
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом Університету № \_\_\_\_\_ від “\_\_” \_\_\_\_\_ 2020 року

2. Термін подання студентом проекту 30.11.2020

3. Вихідні дані до проекту виконати експериментальний аналіз проблем методів адаптації та розпізнавання зображень.

4. Зміст пояснювальної записки (перелік питань, які потрібно розробити) аналіз проблем методів адаптації та розпізнавання зображень.

5. Перелік графічного матеріалу: Проблеми аналізу розпізнавання та адаптації зображень – 2 слайди, аналіз та постановка задачі – 2 слайди, моделі і методи адаптації та розпізнавання зображень – 4 слайди, проведення експерименту адаптації та розпізнавання зображень за допомогою нейронних мереж – 6 слайдів, аналіз експериментальних результатів – 3 слайди, висновки – 1 слайд.

## 6. Консультанти розділів проекту

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1-3	Вдовітченко О.В. к.т.н., доц.		

Нормоконтроль \_\_\_\_\_ Постернакова В.А. «\_\_» \_\_\_\_\_ 2020 р.  
(підпис) (ініціали та прізвище)

7. Дата видачі завдання 01.09.2020 р.

## КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проекту	Строк виконання етапів проекту	Примітка
1	Вступний інструктаж. Узгодження типового завдання.	01.09.20	
2	Огляд і аналіз існуючих проблем розпізнавання та адаптації зображень	02.09.20- 08.09.20	
3	Аналіз існуючих методів розпізнавання та адаптації зображень	09.09.20-15.09.20	
4	Планування експериментального дослідження методів розпізнавання та адаптації зображень	19.09.20-29.09.20	
5	Проведення експериментального дослідження методів розпізнавання та адаптації зображень	01.10.20-09.10.20	
6	Висновки з отриманих результатів	10.10.20-14.10.20	
7	Підготовка доповіді.	15.10.20-17.10.20	
8	Доповідь з презентацією.	18.10.20	

Студент \_\_\_\_\_ Калугін В.О.  
(підпис) (прізвище та ініціали)

Керівник проекту \_\_\_\_\_ Вдовітченко О.В.  
(підпис) (прізвище та ініціали)

## РЕФЕРАТ

Пояснювальна записка на дипломну роботу: 80 с., 28 рис., 33 джерел.

Мета досліджень – підвищення ефективності аналізу зображень та адаптації їх з використанням штучного інтелекту.

Для досягнення мети необхідно виконати наступні завдання:

- проаналізувати проблем методи адаптації та розпізнавання зображень;
- проаналізувати методи і засобів адаптації та розпізнавання зображень нейронними мережами;
- проаналізувати існуючі методи адаптації та розпізнавання зображень штучним інтелектом;
- запланувати дослідження методів адаптації та розпізнавання зображень;
- здійснити експериментальні дослідження методів адаптації та розпізнавання зображень;
- провести аналіз результатів отриманих від експериментальних досліджень.

Актуальність досліджень. З кожним роком зростає зацікавленість вирішення більш складних задач розпізнавання об'єктів, що обумовлена автоматизацією, необхідністю образних процесів комунікації в інтелектуальних системах. Тому удосконалення реалізації розпізнавання комп'ютерними системами образів є актуальною.

Об'єктом дослідження є способи організації засобів нейромережевого розпізнавання об'єкта на зображенні з використанням графічних прискорювачів. Предметом дослідження є методи і алгоритми реалізації архітектур згорткових нейронних мереж. Мета і задачі дослідження: створити нейромережеву систему розпізнавання об'єктів на зображеннях, використовуючи згорткову нейронну мережу власної архітектури із використанням ієрархічного класифікатору. Наукова новизна полягає в наступному. Вдосконалено метод розпізнавання об'єктів на зображеннях використовуючи згорткову нейронну мережу власної архітектури, що дозволило суттєво покращити метод класифікації та підвищити показник точності розпізнавання. Практична цінність роботи полягає у можливості застосування отриманих результатів для ефективного використання згорткової нейронної мережі для задачі розпізнавання об'єктів.

У вступі подано загальну характеристику роботи, обґрунтовано актуальність напрямку досліджень, сформульовано мету і задачі досліджень, показано наукову новизну і практичну цінність роботи.

У висновках представлені результати проведеної роботи.

Робота представлена на 80 аркушах, містить посилання на список використаних літературних джерел.

МАШИННЕ НАВЧАННЯ, КОМП'ЮТЕРНИЙ ЗІР, ЗГОРТКОВА  
НЕЙРОННА МЕРЕЖА, РОЗПІЗНАВАННЯ ОБ'ЄКТА, РЕКУРЕНТНА  
НЕЙРОННА МЕРЕЖА КЛАСИФІКАТОР, GPU.

## ABSTRACT

Explanatory note for the thesis: 80 pp., 28 figs., 33 sources.

The purpose of research is to increase the efficiency of image analysis and their adaptation using artificial intelligence.

To achieve this goal you must perform the following tasks:

- analyze the problems of methods of adaptation and recognition of images;
- to analyze methods and means of adaptation and recognition of images by neural networks;
- analyze existing methods of adaptation and recognition of images by artificial intelligence;
- plan research on methods of adaptation and image recognition;
- to carry out experimental researches of methods of adaptation and recognition of images;
- to analyze the results obtained from experimental studies.

Relevance of research. Every year there is a growing interest in solving more complex problems of object recognition, due to automation, the need for figurative communication processes in intelligent systems. Therefore, improving the implementation of image recognition by computer systems is important.

The object of research is the ways of organizing the means of neural network object recognition in the image using graphical accelerators.

The subject of research is methods and algorithms of realization of architectures of convolutional neural networks.

The purpose and objectives of the study: to create a neural network system for recognizing objects in images, using a convolutional neural network of its own architecture using a hierarchical classifier. Conduct network learning experiments using graphic processing units (GPUs) on seven data sets.

The scientific novelty is as follows. The method of recognizing objects in images has been improved using a convolutional neural network of its own architecture, which has significantly improved the method of classification and increased the accuracy of recognition.

The practical value of the work lies in the possibility of applying the obtained results for the effective use of the convolutional neural network for the task of object recognition.

The introduction presents a general description of the work, substantiates the relevance of the research direction, formulates the purpose and objectives of research, shows the scientific novelty and practical value of the work.

The conclusions present the results of the work.

The work is presented on 80 sheets, contains references to the list of used literature sources.

MACHINE LEARNING, COMPUTER VISION, CONVOLUTIONAL NEURAL NETWORK, OBJECT RECOGNITION, CLASSIFIER, GPU.

## РЕФЕРАТ

Пояснительная записка на дипломную работу: 80 с., 28 рис., 33 источников.

Цель исследований - повышение эффективности анализа изображений и адаптации их с использованием искусственного интеллекта.

Для достижения цели необходимо выполнить следующие задачи:

- проанализировать проблем методы адаптации и распознавания изображений;
- проанализировать методы и средств адаптации и распознавания изображений нейронными сетями;
- проанализировать существующие методы адаптации и распознавания изображений искусственным интеллектом;
- запланировать исследования методов адаптации и распознавания изображений;
- осуществить экспериментальные исследования методов адаптации и распознавания изображений;
- провести анализ результатов полученных от экспериментальных исследований.

Актуальность исследований. С каждым годом растет интерес решения более сложных задач распознавания объектов, обусловленная автоматизацией, необходимостью образных процессов коммуникации в интеллектуальных системах. Поэтому совершенствование реализации распознавания компьютерными системами образов является актуальной.

Объектом исследования являются способы организации средств нейросетевого распознавания объекта на изображении с использованием графических ускорителей.

Предметом исследования являются методы и алгоритмы реализации архитектур згортковых нейронных сетей.

Цель и задачи исследования: создать нейросетевом систему распознавания объектов на изображениях, используя згортковых нейронную сеть собственной архитектуры с использованием иерархического классификатора. Научная новизна заключается в следующем. Усовершенствован метод распознавания объектов на изображениях используя згортковых нейронную сеть собственной архитектуры, что позволило существенно улучшить метод классификации и повысить показатель точности распознавания. Практическая ценность работы состоит в возможности применения полученных результатов для эффективного использования сверточного нейронной сети для задачи распознавания объектов. Работа представлена на 80 листах, содержит ссылки на список использованных литературных источников.

МАШИННОЕ ОБУЧЕНИЕ, КОМПЬЮТЕРНОЕ ЗРЕНИЕ,  
РЕКУРРЕНТНАЯ НЕЙРОННАЯ СЕТЬ, СВЕРТОЧНАЯ НЕЙРОННАЯ СЕТЬ,  
РАСПОЗНАВАНИЕ ОБЪЕКТА, КЛАССИФИКАТОР, GPU.

## ЗМІСТ

СПИСОК ТЕРМІНІВ, СКОРОЧЕНЬ ТА ПОЗНАЧЕНЬ .....	71
ВСТУП .....	<b>Ошибка! Закладка не определена.</b>
1 ОСНОВНІ СПОСОБИ ВИРІШЕННЯ ЗАДАЧ РОЗПІЗНАВАННЯ ОБ'ЄКТВ 352	
1.1 Розпізнавання та адаптація зображення методами машинного навчання	12
1.1.1 Основи розпізнавання об'єкта .....	12
1.2 Повнозв'язкові нейронні мережі.....	16
1.3 Методи логічного аналізу .....	17
1.4 Згорткові нейронні мережі .....	19
1.5 Модель штучної нейронної мережі .....	21
1.5.1 Модель штучного нейрона та його функції активації .....	21
1.5.2 Одношарові штучні нейронні мережі .....	23
1.5.3 Багатошарові штучні нейронні мережі.....	24
1.6 Рекурентні нейронні мережі .....	27
1.6.1 Рекурентні нейронні мережі для розпізнавання емоцій на відео .....	27
1.7 Метод навчання з підкріпленням.....	28
1.8 Основні принципи навчання нейронних мереж .....	30
1.8.1 Вибір функції активації при навчанні нейронної мережі .....	31
1.9 Висновки до першого розділу .....	32
2 ПЛАНУВАННЯ ЕКСПЕРИМЕНТАЛЬНИХ ДОСЛІДЖЕНЬ .....	33
2.1 Аналіз проблеми навчання розпізнавання зображень .....	33
2.1.1 Проблема адаптації розпізнавання образів .....	33
2.2 Підходи до вирішення задач .....	37
2.2.1 Геометричний і структурний підходи .....	38
2.2.2 Гіпотеза компактності .....	41
2.2.3 Адаптація та самонавчання .....	41
2.3 Розпізнавання за допомогою послідовної активації детекторів .....	43
2.4 Використання вирішальною зворотного зв'язку при розпізнаванні .....	46
2.5 Висновок до другого розділу.....	50
3 ПРОВЕДЕННЯ ЕКСПЕРИМЕНТАЛЬНОГО ДОСЛІДЖЕННЯ РОЗПІЗНАВАННЯ АДАПТАЦІЇ ЗОБРАЖЕНЬ.....	55
3.1 Методи оцінки ефективності розпізнавання .....	55
3.2 Проведення експерименту.....	57
3.2.1 Обробка результатів роботи автоенкодера.....	59
3.2.2 Обробка результатів розпізнавання зображень . .....	68
3.2.3 Аналіз результатів .....	71
3.3 Висновки до третього розділу.....	74
ВИСНОВКИ.....	76
ПЕРЕЛІК ПОСИЛАНЬ .....	77

## СПИСОК ТЕРМІНІВ, СКОРОЧЕНЬ ТА ПОЗНАЧЕНЬ

**Перцептрон** - (англ. perceptron від лат. perceptio — сприйняття; нім. Perzeptron) — математична або комп'ютерна модель сприйняття інформації мозком (кібернетична модель мозку).

**Когнітрон** — штучна нейронна мережа на основі принципу самоорганізації.

**Графічні прискорювачі** (англ. graphic processing unit, GPU) - окремий пристрій персонального комп'ютера або ігрової приставки, що виконує графічний рендеринг.

**ЗНМ** – згорткова нейронна мережа.

**Неокогнітрон** — згорткова ієрархічна, багат шарова штучна нейронна мережа, заснована на принципах навчання без учителя.

**Кластер** (англ. cluster) — група однакових або подібних елементів, зібраних разом або близько розташованих один до одного.

**Набір даних** (англ. data set) — колекція однотипних даних, що застосовується в задачах машинної обробки даних.

**Синапс нейрона** (від грецького «synapsis» — з'єднання) — структура, яка дозволяє нейрону (нейронної мережі) проводити сигнал (у випадку штучного нейрону) до іншого нейрону.

**Індуктивний метод** — це процес судження, котрий досягає висновку, що при наявному стані знань є напевно істинний, але не гарантує його.

**Дедуктивний** (англ. deductive, нім. deduktiv) — заснований на дедукції; дедуктивний метод — спосіб дослідження, при якому окремі положення логічно виводяться із загальних положень (аксіом, постулатів, законів).



## ВСТУП

У наш час нейронні мережі - один з напрямків в розробці систем штучного інтелекту. Ідея полягає в тому, щоб максимально близько змоделювати роботу людської нервової системи - а саме, її здатності до навчання і виправлення помилок. У цьому полягає головна особливість будь-нейронної мережі - вона здатна самостійно навчатися і діяти на підставі попереднього досвіду, з кожним разом роблячи все менше помилок. Комп'ютерний зір та машинне навчання вже є одним з найцікавіших та найперспективніших галузей штучного інтелекту. Прискорені темпи розвитку технологій, заміни рутинної людської праці, концепцій «розумного будинку» і «розумного автомобіля», голосового помічника, розвиток інтернету-речей і систем штучного інтелекту визначають цим напрямком особливе місце у науці.

Нейромережа імітує не тільки діяльність, а й структуру нервової системи людини. Така мережа складається з великого числа окремих обчислювальних елементів ( «нейронів»). У більшості випадків кожен «нейрон» відноситься до певного прошарку мережі. Вхідні дані послідовно проходять обробку на всіх шарах мережі. Параметри кожного «нейрона» можуть змінюватися в залежності від результатів, отриманих на попередніх наборах вхідних даних, змінюючи таким чином і порядок роботи всієї системи.

Багато провідних розробників в сфері штучного інтелекту зазначають, що нейронні мережі здатні вирішувати такі ж завдання, як і інші алгоритми машинного навчання, різниця полягає лише в підході до навчання.

Всі завдання, які можуть вирішувати нейронні мережі, так чи інакше пов'язані з навчанням. Серед основних областей застосування нейронних мереж - прогнозування, прийняття рішень, розпізнавання образів, оптимізація, аналіз даних. Також не для кого не секрет, що зараз нейромережі застосовуються повсюдно: Наприклад, багато великих інтернет-сайтів використовують їх, щоб зробити реакцію на поведінку користувачів більш природною і корисною своїй аудиторії. Нейромережи лежать в основі більшості сучасних систем розпізнавання і синтезу мови, а також розпізнавання і обробки зображень. Вони застосовуються в деяких системах навігації, будь то промислові роботи або безпілотні автомобілі. Алгоритми на основі нейромереж захищають інформаційні системи від атак зловмисників і допомагають виявляти незаконний контент в мережі.

У найближчій перспективі (5-7 років), нейронні мережі будуть використовуватися ще ширше. Можна уявити що сільськогосподарський комбайн, виконавчі механізми якого забезпечені безліччю відеокамер. Він робить п'ять тисяч і знімків в хвилину кожної рослини в смузі своєї траєкторії і, використовуючи нейромережу, аналізує - НЕ бур'ян чи це, не уражене воно хворобою або шкідниками. І обробляє кожну рослину індивідуально. Фантастика? Уже не зовсім. А через п'ять років може стати нормою.

Зараз є такі класифікації напрямків розвитку:

- системи розпізнавання і класифікації об'єктів на зображеннях;

- голосові інтерфейси взаємодії для інтернету речей;
- системи моніторингу якості обслуговування в колл-центрах;
- системи виявлення неполадок (в тому числі, що пророкують час технічного обслуговування), аномалій, кібер-фізичних загроз;
- системи інтелектуальної безпеки і моніторингу;
- заміна ботами частини функцій операторів колл-центрів;
- системи відео аналітики;
- самонавчальні системи, оптимізують управління матеріальними потоками або розташування об'єктів (на складах, транспорті);
- інтелектуальні, самонавчальні системи управління виробничими процесами і пристроями (в тому числі, робото технічні);
- поява систем універсального перекладу «на льоту» для конференцій і персонального використання;
- поява ботів-консультантів технічної підтримки або персональних асистентів, за функціями близьким до людини.

Однак до недавнього часу, швидкість роботи нейромереж була занадто низькою, щоб вони могли отримати широке поширення, і тому такі системи в основному використовувалися в розробках, пов'язаних з комп'ютерним зором, а в інших областях застосовувалися інші алгоритми машинного навчання.

Трудомістка і тривала частина процесу розробки нейронної мережі - її навчання. Для того, щоб нейронна мережа могла коректно вирішувати поставлені завдання, потрібно «прогнати» її роботу на десятках мільйонів наборів вхідних даних. Саме з появою різних технологій прискореного навчання і пов'язують поширення нейромереж.

«По-перше, з'явився великий і загальнодоступний масив розмічених картинок (ImageNet), на яких можна вчитися. По-друге, сучасні відео карти дозволяють в сотні разів швидше навчати нейромережі і їх використовувати. По-третє, з'явилися готові, преднавчання нейромережі, що розпізнають образи, на підставі яких можна робити свої додатки, не займаючись тривалою підготовкою нейромережі до роботи. Все це забезпечує дуже потужний розвиток нейромереж саме в області розпізнавання образів.

Чому саме методи розпізнавання зображень дуже важливий напрямок у нейронних мережах? Дві третини всієї сенсорної інформації, яка до нас потрапляє, приходить з зорових органів сприйняття. Більше однієї третини поверхні нашого мозку зайняті двома найголовнішими зоровими зонами - дорсальний зоровий шлях і вентральний зоровий шлях. Спинний зоровий шлях починається в первинній зоровій зоні, в нашому тім'ячку і триває наверх, в той час як вентральний шлях починається на нашому потилиці і закінчується приблизно за вухами. Все важливе розпізнавання образів, яке у нас відбувається, весь сенс, що ми усвідомлюємо, проходить саме там же, за вухами.

Це досить важливо бо часто потрібно для розуміння нейронних мереж. По-перше, всі про це розповідають, а по-друге, справа в тому, що всі області, які

використовуються в нейронних мережах для розпізнавання образів, прийшли до нас саме з вентрального зорового шляху, де кожна маленька зона відповідає за свою строго певну функцію. У далекі 60-ті роки минулого століття, коли тільки починалося вивчення зорових зон мозку, перші експерименти проводилися на тваринах, тому що не було fMRI. Досліджували мозок за допомогою електродів, вживлених в різні зорові зони.

Таким же способом були відкриті ще багато важливих властивості зорових зон, які ми використовуємо зараз в напрямку розпізнавання deep learning. Одна з найважливіших властивостей - це збільшення рецептивних полів наших клітин у міру просування від первинних зорових зон до скроневих часток, тобто більш пізнім зоровим зонам. Рецептивної полі - це та частина зображення, яку обробляє кожна клітинка нашого мозку. У кожній клітині своє рецептивної полі. Ця ж властивість зберігається і в нейронних мережах. До того, як ми навчилися це застосовувати до комп'ютерного зору - в загальному, як такого його не було. Всі ці властивості ми переносимо в нейронну мережу, і ось воно запрацювало.

# 1 ОСНОВНІ СПОСОБИ ВИРІШЕННЯ ЗАДАЧ РОЗПІЗНАВАННЯ ОБ'ЄКТІВ

## 1.1 Розпізнавання та адаптація зображення методами машинного навчання

Машинне навчання і штучний інтелект за останні кілька років стали дуже гарячими темами. В тих чи інших варіантах вони сьогодні є частиною величезної кількості продуктів, і мало хто не задумується над їхнім запровадженням. Приклади застосування ML (Machine Learning) — від автоматичного визначення важливих листів і швидких відповідей в Gmail, створення музики за допомогою машинного навчання до AlphaGo.

Машинне навчання – це спрощена версія процесу навчання, яке відбувається з людиною. Як правило, в машинному навчанні наявний певний набір прикладів, спостережень, реакцій до цих спостережень. Задача полягає у тому, щоб сконструювати такі моделі, які будуть максимально ефективно описувати наявні дані і робити достовірні прогнози.

Машинне навчання відрізняється від звичного нам тим, що ми намагаємося навчити комп'ютер вчитися. Існує 2 типи навчання:

- індуктивне навчання;
- дедуктивне навчання.

Індуктивне навчання: знайоме кожному, адже воно полягає у спостереженні за світом та побудові певних моделей, які пояснюють причини тих чи інших явищ.

Потім такі моделі неодноразово перевіряються, певні з них «виживають» і використовуються, покращуються. А деякі моделі згодом цілком відкидаються.

Дедуктивне навчання: подібне до математики в школі, коли учню дають готові формули і розказують, як застосовувати їх на практиці.

Машинне навчання є індуктивним навчанням, оскільки в основному ми вчимо машину вчитися на прикладах, спостерігати велику кількість прикладів із реального життя, будувати на них моделі, перевіряти їх, застосовувати їх на подальших прикладах.

### 1.1.1 Основи розпізнавання об'єкта

Теорія мистецтв нейронних мереж з'явилася, як спроба змодельювати ЦНС (центральної нервової системи - центральна нервова система) вищих ссавців в 50х роках минулого століття.

З'ясувалося, що моделі штучних нейронних мереж (Далі ІНС, НС) занадто прості і сучасні моделі нейрофізіології за складністю на порядок перевищує моделі НС. Разом з тим з'ясувалося, що теорія ІНС є прекрасним інструментом для вирішення суто математичних задач особливо до Клас трудно-формалізуємих завдань.

Під не формалізуються задачами ми будемо розуміти завдання, для яких завдання сформулювати неможливо.

До числа таких задач відносяться (приклад):

- класифікація;
- кластеризація;
- прогнозування.

Під важко формалізуються задачами задумується такі завдання для яких формулювання існує, але детермінований алгоритм знаходження точного рішення або невідомий, або занадто витратний по ресурсам.

Чим вище розмірність завдання, тим краще працює нейронні мережі і тим гірше класична математика.

Концепції ЦНС вищих ссавців складається з клітин - нейрон пов'язаних один з одним з усілякими клітинами рецепторами (зорові, слухові). Загальна кількість таких клітин близько  $10^{10}$ .

Зв'язок між нейронами здійснюється за допомогою контактів - синапсів.

Сила зв'язку між двома нейронами пропорційна концентрації речовини - нейромедіатр.

Діяльність головного мозку вищих ссавців являє собою перебування в збудженому стані.

Нейрон ставати активним, тоді. коли кількість активних нейронів пов'язаних синапсами перевищує певний поріг.

Концепція навчання. У процесі навчання за допомогою різких зовнішніх стимулів істотно коригується картина світу записана в голові.

Таким чином для здійснення навчання ми повинні мати набір векторів характеризують картину світу і всіх можливих реакції на нашу дійсність.

Такий набір векторів отримав назву - навчальна вибірка.

Чим більше адекватна картина світу записана в нейронах головного мозку, тим менше справжня реакція навколишнього середовища буде відрізнятися від реакції передбачення.

З нейрофізіологічної точки зір навчання може відбуватися двояко, це відмирання зв'язків і зміна сили зв'язку.

Поняття штучного нейрона. Нейроном називається математичний об'єкт, який має  $n$ -входів і 1 вихід. Входи призначені для передачі числової інформації, так само як і вихідна як (показано на рисунку 1.1).

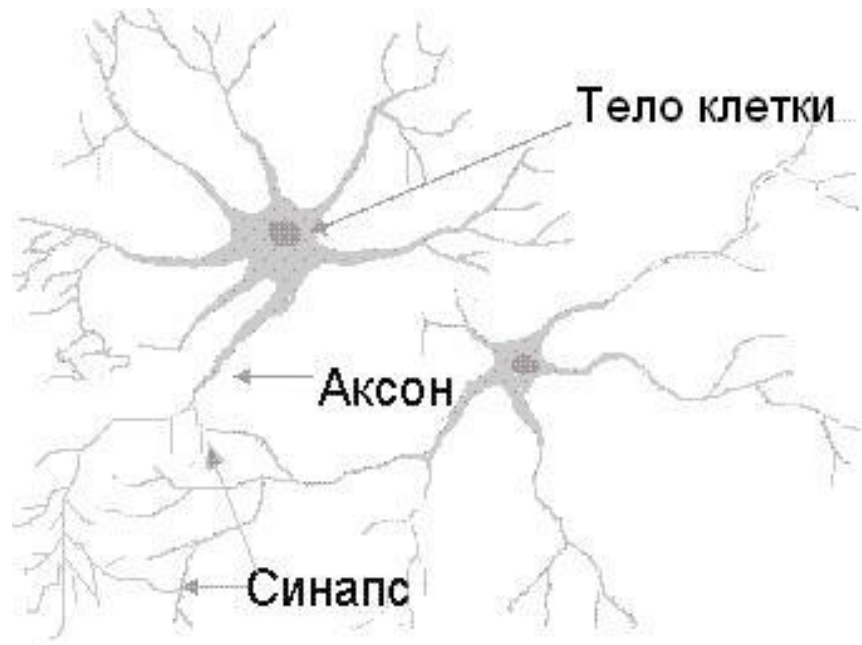


Рисунок 1.1 – Модель біологічного нейрону

Вхідна інформація позначається  $X_i$ . Входи пов'язують штучний нейрон (ІС) з іншими або із зовнішнім джерелом інформації і носять назву зв'язків. З кожною зв'язком пов'язано число, яке називають вагами і позначається -  $W_i$ . (І - номер зв'язку).

Надходить через вхід інформація обробляється в нейроні в 2 етапи. На першому етапі обчислюється зважена сума:  $S = W_1X_1 + W_2X_2 + \dots + W_nX_n + O$ . Число  $O$  - поріг.

Зазвичай, для уніфікації записи поріг розглядається як вага тупий зв'язку, тобто зв'язку яка весь час генерує 1.

Можна представити у вигляді запису  $S = W_0X_0 + W_1X_1 + W_2X_2 + \dots + W_nX_n$ , де  $X_0 = 1$ ,  $W_0 = O$ . (Класичний нейрон).

В узагальненому КН виступають імовірнісні заходи, які число  $O$  - вип. величина  $N(0, G)$ .

Алгебраїчний нейрон це нейрон, де сума  $S$  нелінійна. Вихід нейрона -  $Y$ . визначається як деяка функція від зваженої суми.  $y = f(s)$ ; Функція  $f$  - функція активації. Власне КН класифікуються за їх функцій активації.

Приклади:

- покрокова функція;
- К-крокова;
- сигмоїдальна.

Найчастіше має вигляд:  $f(x) = 1 / (1 + \exp(-x))$ .

Перехід від порогових функцій до сигмоїдальної пов'язаний з тим, що якщо  $S = -\epsilon ps$ , то на вихід дає 0, а якщо  $S = \epsilon ps$ , то 1.

Тобто невелика різниця дає суттєвий результат, а це не є добре. Всі функції активації поділяються на 2 групи:

- симетричні;

– асиметричні.

Іншими словами, розпізнавання образів можна визначити, як віднесення вихідних даних до певного класу за допомогою виділення істотних ознак або властивостей, які характеризують ці дані, із загальної маси несуттєвих деталей.

Задачі МН класифікуються ще на декілька типів по виду вирішуваної проблеми:

- задача кластеризації;
- задача регресії;
- задача зниження розмірності.

Прикладами завдань класифікації є розпізнавання символів, встановлення медичного діагнозу, прогноз погоди, розпізнавання осіб, класифікація документів, тощо.

Найчастіше вихідним матеріалом служить отримане із камери зображення. Процес можна розглядати як процес кодування, що полягає в присвоєнні значення кожної ознаки із простору ознак для кожного класу. Якщо розглянути 2 класи об'єктів: дорослі і діти. В якості значення ознак можна вибрати зріст і вагу як показано на рисунку 1.2.

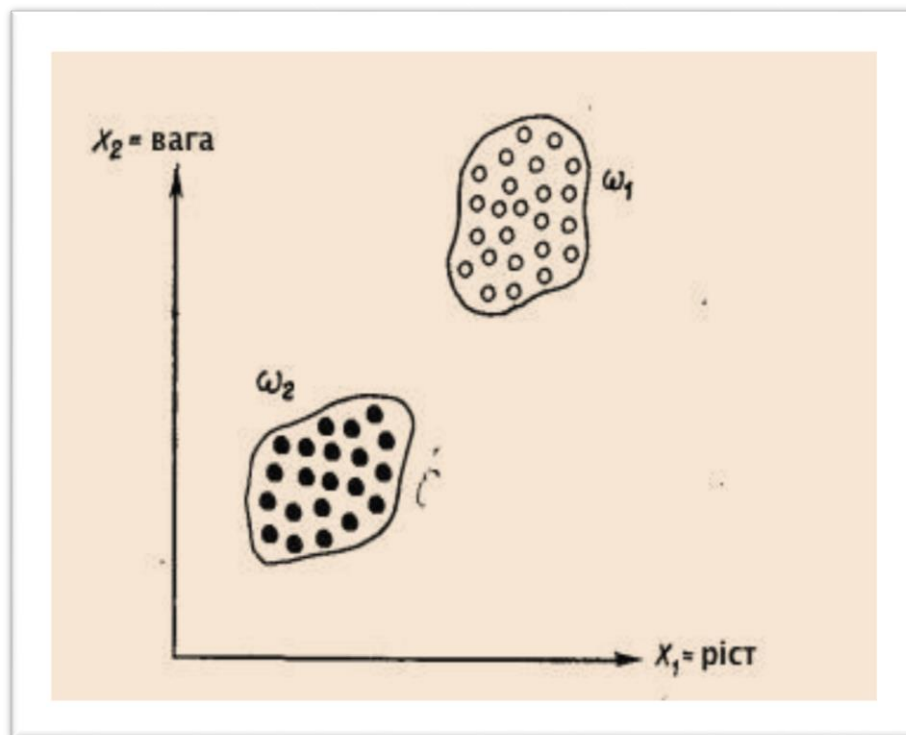


Рисунок 1.2 - Два непересічних класи

З рисунку 1.2 видно, що ці два класи утворюють дві непересічні множини, що можна пояснити обраними ознаками. Проте не завжди є можливість вибрати правильні вимірювані параметри в якості ознак класів. Наприклад, вибрані параметри не підійдуть, щоб створити непересічні класи хокеїстів та волейболістів.

Іншим завданням розпізнавання є виділення із вихідних зображень характерних ознак та/або властивостей. Це завдання можна віднести до попередньої обробки. Ознака повинна являти собою характерну властивість конкретного класу, при цьому загальну для цього класу.

Міжкласові ознаки – це ознаки, що визначають відмінності між класами. Загальні ознаки, що властиві усім класам, не несуть корисної інформації, тому для задачі розпізнавання об'єктів не розглядаються як характерні. Вибрати правильні ознаки - одна із важливих задач побудови систем розпізнавання.

Сукупність нейронів з'єднаних тим чи іншим способом називається штучної нейронної мережею або просто нейронною мережею.

Закон за яким нейрони з'єднані в мережу носить назву структури або топологією мережі.

Безліч нейронів НЕ з'єднаних між собою, але з'єднані з іншими нейронами носить назву шару нейрона.

Мережі бувають двох видів: одношарові, багатошарові. Класичні методи комп'ютерного зору, розпізнавання об'єктів і машинного навчання.

## 1.2 Повнозв'язкові нейронні мережі

Більшість людей під «звичайними» або «класичними» нейромережами розуміє Повнозв'язкові нейронні мережі прямого поширення з зворотним поширенням помилки (показано на рисунку 1.3).

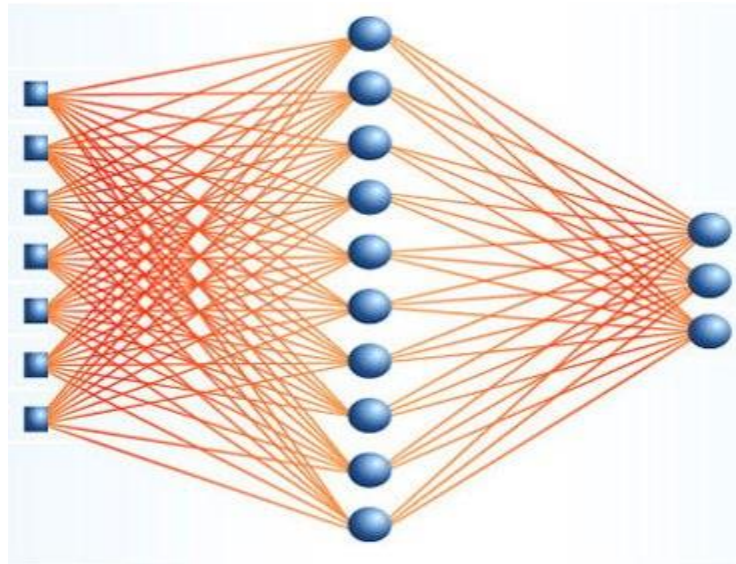


Рисунок 1.3 - Повно зв'язна нейронна мережа



Як випливає з назви в такій мережі кожен нейрон пов'язаний з кожним, сигнал йде тільки в напрямку від вхідного шару до вихідного, немає ніяких рекурсій. Будемо називати таку мережу скорочено ПНС.

Спершу необхідно вирішити як подавати дані на вхід. Найпростіше і майже безальтернативне рішення для ПНС - це висловити двовимірну матрицю зображення у вигляді одновимірного вектора. Тобто для зображення рукописної цифри розміром 28x28 у нас буде 784 входу, що вже не мало. Далі відбувається те, за що нейромережників і їхні методи багато консервативні вчені не люблять - вибір архітектури. А не люблять, оскільки вибір архітектури це чисте шаманство. До сих пір не існує методів, що дозволяють однозначно визначити структуру і склад нейромережі виходячи з опису завдання. На захист скажу, що для трудноформалізуємих завдань навряд чи коли-небудь такий метод буде створений. Крім того існує безліч різних методик редукції мережі (наприклад OBD [1]), а також різні евристики та емпіричні правила. Одне з таких правил свідчить, що кількість нейронів в прихованому шарі має бути хоча б на порядок більше кількості входів. Якщо взяти до уваги що саме по собі перетворення з зображення в індикатор класу досить складне і суттєво нелінійне, одним шаром тут не обійтися. Виходячи з усього вищесказаного грубо прикидаємо, що кількість нейронів в прихованих шарах у нас буде близько 15000 (10 000 у 2-му шарі і 5000 в третьому). При цьому для конфігурації з двома прихованими шарами кількість параметрів і учнів зв'язків буде 10 млн. Між входами і першим прихованим шаром + 50 млн. Між першим і другим + 50 тис. Між другим і вихідним, якщо вважати що є 10 виходів, кожен з яких позначає цифру від 0 до 9. Разом грубо 60 000 000 зв'язків. Я не даремно згадав, що вони настроюються - це значить, що при навчанні для кожної з них потрібно буде обчислювати градієнт помилки.

Ну це добре, що вже тут поробиш, краса штучний інтелект вимагає жертв. Але от якщо замислитися, на думку спадає, що коли ми перетворимо зображення в лінійну ланцюжок байт, ми щось безповоротно втрачаємо. Причому з кожним шаром ця втрата тільки посилюється. Так і є - втрачається топологію зображення, тобто взаємозв'язок між окремими його частинами. Крім того завдання розпізнавання має на увазі вміння нейромережі бути стійкою до невеликих зрушень, поворотів і зміни масштабу зображення, тобто вона повинна отримувати з даних якісь інваріанти, які не залежать від почерку тієї чи іншої людини. Так якою ж має бути нейромережа, щоб бути не дуже обчислювально-складною і, в той же час, більш інваріантною до різних спотворень зображень.

### 1.3 Методи логічного аналізу

Більшість завдань галузі комп'ютерного зору вимагають аналізу внутрішньої структури зображення і морфологічних ознак зображених об'єктів, для цього досліджуються методи і впроваджується логічний аналіз. Методи математичної морфології є результатом теорії та техніки аналізу й обробки

геометричних структур, заснованих на теорії множин, топології та випадкових функціях. Ці методи реалізуються базовими операціями. Операції двійкової морфології є деяким перетворенням впорядкованої множини або підмножини (області зображення) за допомогою структурного елементу. Структурним елементом є двійкове зображення довільного розміру та довільної структури, але найчастіше використовуються симетричні елементи, такі як прямокутник фіксованого розміру або коло фіксованого діаметру. Результатом перетворення також є двійкове зображення [2]. Методи математичної морфології дозволяють видаляти шуми з двійкових зображень, а також реалізують алгоритми пошуку контурів, але на практиці ці методи використовуються в поєднанні з іншими алгоритмами.

Контурний аналіз - це потужний математичний апарат, що дозволяє описувати, зберігати і знаходити об'єкти які знаходяться в формі зовнішніх контурів. Вище розглядалися фільтри контурів, результатом застосування яких природним чином є контури об'єктів на зображенні без застосування додаткової бінаризації. У контурному аналізі попередні стадії фільтрації і додаткової бінаризації є обов'язковими етапами завдання. Передбачається, що контур містить необхідну інформацію про форму об'єкта, а внутрішні точки до уваги не приймаються, що обмежує область застосування алгоритмів контурного аналізу. Проте отримані з його допомогою контури дозволяють перейти від двовимірного простору образу до простору контурів, в деяких завданнях це значно зменшує складність алгоритму. Методи контурного аналізу інваріантні щодо перенесення, повороту і масштабування зображення об'єкта. Серед методів контурного аналізу можна виділити алгоритми, в яких контур об'єкта відстежується і векторизується; скануючі алгоритми, засновані на перегляді всього зображення і виділення контурних точок без відстеження; алгоритми відстеження контурів на напівтонових зображеннях, дослідження кривизни функцій.

На практиці методи контурного аналізу досить чутливі до умов середовища, що може викликати складність їх використання в реальних умовах для більшості завдань комп'ютерного зору, проте вони корисні в задачах машинного зору, коли умови середовища досить строго визначені. У таких ситуаціях методи контурного аналізу є неперевершеними лідерами по швидкодії, володіючи зрозумілою і простою логікою, що обумовлює зручність їх використання в спеціалізованих областях.

Пошук особливих точок (feature detection). Для вирішення допоміжної прикладної задачі в даній роботі буде застосований один з алгоритмів *feature detection* [2], тому на цій темі варто зупинитися докладніше. Пошук особливих точок є одним з найбільш поширених методів в класичному комп'ютерному зорі. Особливі точки надають унікальні характеристики об'єкта, дозволяючи порівнювати різні зображення одного об'єкта або одного класу об'єктів між собою. Тому на практиці особливі точки найбільш актуальні в задачах, в яких можливим рішенням є обробка серії зображень або відео потоку і подальший аналіз отриманих масивів особливих точок. Алгоритми *feature detection* можна

умовно класифікувати за ступенем стабільності точок при переходах від одного зображення (кадру) об'єкта або класу об'єктів до іншого.

Складність алгоритмів пошуку зростає з необхідним рівнем стабільності шуканих точок.

Результатом роботи таких алгоритмів є безліч особливих точок, зокрема, кутів, для яких необхідно побудувати математичний опис. Формування математичного опису - це завдання дескриптора. Багато дескрипторів одночасно вирішують завдання пошуку особливих точок і побудови описів цих точок за допомогою вбудованих алгоритмів або оригінальним власним способом. Ознаки (описи) будуються на основі інформації про інтенсивність, про кольори та текстуру особливих точок.

#### 1.4 Згорткові нейронні мережі

Вирішення цієї проблеми було знайдено американським вченим французького походження Яном ЛеКуном, натхненним роботами нобелівських лауреатів в галузі медицини Torsten Nils Wiesel і David H. Hubel. Ці вчені досліджували зорову кору головного мозку кішки і виявили, що існують так звані прості клітини, які особливо сильно реагують на прямі лінії під різними кутами і складні клітини, які реагують на рух ліній в одному напрямку. Ян ЛеКун запропонував використовувати так звані згорткові нейронні мережі [2].

Ідея згорткових нейронних мереж полягає в чергуванні згорткових шарів (C-layers), субдискретизуючих шарів (S-layers) і наявності повнозв'язних (F-layers) шарів на виході.

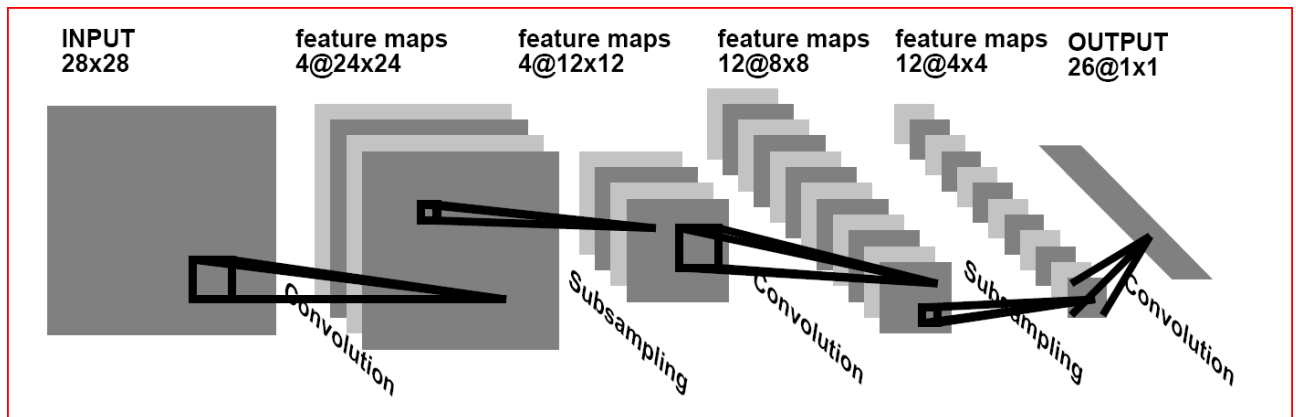


Рисунок 1.4 – Приклад згорткової нейронної мережи

Така архітектура містить в собі 3 основних парадигми:

- локальне сприйняття;
- спільні ваги;
- субдискретизація.

Локальне сприйняття має на увазі, що на вхід одного нейрона подається не все зображення (або виходи попереднього шару), а лише деяка його область. Такий підхід дозволив зберігати топологію зображення від шару до шару.

Концепція поділюваних ваг припускає, що для великої кількості зв'язків використовується дуже невеликий набір ваг. Тобто якщо у нас є на вході зображення розмірами 32x32 пікселя, то кожен з нейронів наступного шару візьме на вхід тільки невелику ділянку цього зображення розміром, наприклад, 5x5, причому кожен з фрагментів буде оброблений одним і тим же набором. Важливо розуміти, що самих наборів ваг може бути багато, але кожен з них буде застосований до всього зображення. Такі набори часто називають ядрами (kernels). Неважко порахувати, що навіть для 10 ядер розміром 5x5 для вхідного зображення розмірами 32x32 кількість зв'язків виявиться рівним приблизно 256000 (порівнюємо з 10 млн.), А кількість параметрів, що настраюються всього 250!

А як же, запитаєте ви, це позначиться на якості розпізнавання? Як не дивно, в кращу сторону. Справа в тому, що таке штучно введене обмеження на ваги покращує узагальнюючі властивості мережі (generalization), що в результаті позитивно позначається на здатності мережі знаходити інваріанти в зображенні і реагувати головним чином на них, не звертаючи уваги на інший шум. Можна подивитися на цей підхід трохи з іншого боку. Ті, хто займався класикою розпізнавання зображень і знає як це працює на практиці (наприклад у військовій техніці) знають, що більшість таких систем будуються на основі двовимірних фільтрів. Фільтр являє собою матрицю коефіцієнтів, звичайно задану вручну. Ця матриця застосовується до зображення за допомогою математичної операції, званої згорткою. Суть цієї операції в тому, що кожен фрагмент зображення множиться на матрицю (ядро) згортки по елементно і результат підсумовується і записується в аналогічну позицію вихідного зображення. Основна властивість таких фільтрів полягає в тому, що значення їх виходу тим більше чим більше фрагмент зображення схожий на сам фільтр. Таким чином зображення згорнуте з якимось ядром дасть нам інше зображення, кожен піксель якого буде означати ступінь схожості фрагмента зображення на фільтр. Іншими словами це буде карта ознак.

Процес поширення сигналу в С-шарі можна проілюструвати на рисунку 1.5:

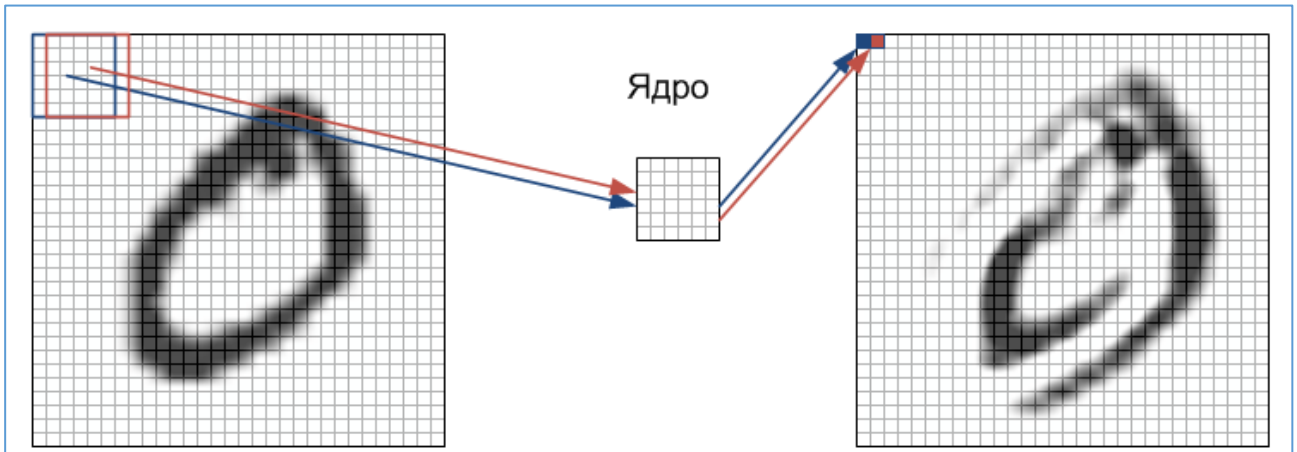


Рисунок 1.5 – Процес поширення сигналу

Кожен фрагмент зображення по-елементно множиться на невелику матрицю ваг (ядро), результат підсумовується. Ця сума є пікселом вихідного зображення, яке називається картою ознак. Тут я опустив той факт, що зважена сума входів ще пропускається через функцію активації (як в будь-який інший нейромережі). Насправді це може відбуватися і в S-шарі, принципової різниці немає. Слід сказати, що в ідеалі не різні фрагменти проходять послідовно через ядро, а паралельно все зображення проходить через ідентичні ядра. Крім того, кількість ядер (наборів ваг) визначається розробником і залежить від того скільки ознак необхідно виділити. Ще одна особливість згорткового шару в тому, що він трохи зменшує зображення за рахунок крайових ефектів.

Суть субдискретизації і S-шарів полягає в зменшенні просторової розмірності зображення. Тобто вхідне зображення грубо (усередненням) зменшується в задану кількість разів. Найчастіше в 2 рази, хоча може бути і не рівномірний зміна, наприклад, 2 по вертикалі і 3 по горизонталі. Субдискретизація потрібна для забезпечення інваріантності до масштабу.

Чергування шарів дозволяє складати карти ознак з карт ознак, що на практиці означає здатність розпізнавання складних ієрархій ознак.

Зазвичай після проходження декількох шарів карта ознак вироджується в вектор або навіть скаляр, але таких карт ознак стає сотні. У такому вигляді вони подаються на один-два шари повно-мережі. Вихідний шар такої мережі може мати різні функції активації. У найпростішому випадку це може бути тангенціальна функція, також успішно використовуються радіальні базисні функції.

## 1.5 Модель штучної нейронної мережі

Штучна нейронна мережа є концептуальною моделлю біологічної нейронної мережі і складається з пов'язаних між собою різним чином шарів штучних нейронів, які організують загальну активну структуру і функціонально впливають на роботу один одного. У більшості архітектур

штучних нейромереж активність нейрона визначається перетворенням зовнішнього сумарного впливу інших нейронів на даний нейрон.

З моменту свого зародження технології штучних нейронних мереж розвивалися досить відокремлено від класичних методів, нерідко докорінно змінюючи уявлення про предмет і проблематику теорії машинного навчання і розпізнавання об'єктів, залишаючи значний вплив на теоретичний, термінологічний і методологічний апарати цих дисциплін.

Через деякий час після розвитку базових моделей штучних нейронних мереж, відбувся значний поділ науки про нейромережі на види топологій архітектури мереж і методи навчання мереж. У більшості архітектур штучних нейронних мереж функції активації нейронів фіксовані, а ваги синапсів є параметрами мережі. Деякі входи нейронів є зовнішніми входами сукупної мережі, а деякі виходи нейронів - виходами сукупної мережі. Завдання нейромережі полягає в перетворенні вхідного вектора у вихідний вектор, що здійснюється вагою і топологією мережі.

#### 1.4.1 Модель штучного нейрона та його функції активації

Штучний нейрон характеризується своїм поточним станом. Тут можна провести аналогію з нервовими клітинами головного мозку, які можуть бути пошкоджені або неправильно працювати. Він володіє групою синапсів - односпрямованих вхідних зв'язків, з'єднаних з виходами інших нейронів, присутня така частина, як аксон – вихідний зв'язок штучного нейрона, з якого сигнал (збудження або гальмування) надходить на синапси наступних нейронів. Загальний вигляд штучного нейрона наведено на рисунку 2.1. Штучний нейрон спочатку імітує властивості, що дані біологічному нейрону. Тут безліч вхідних сигналів, позначених  $x_1, x_2, \dots, x_N$ , надходить на штучний нейрон. Ці вхідні сигнали, в сукупності позначаються вектором  $X$ , приходять до синапсів біологічного нейрона.

Кожен синапс характеризується величиною синаптичного зв'язку або його вагою  $w_i$  [3].

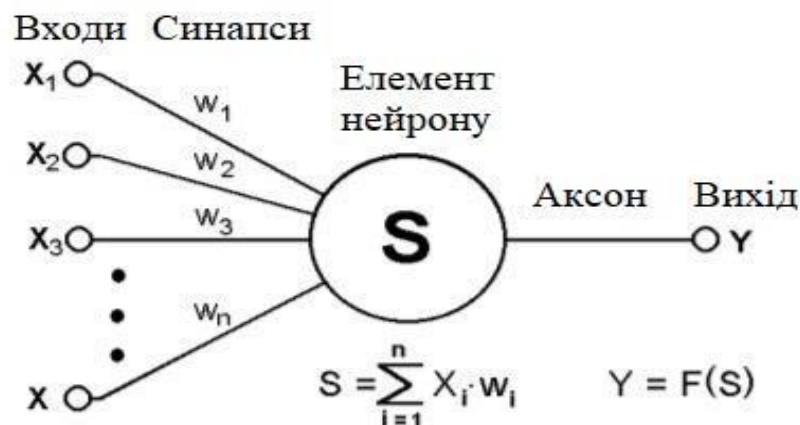


Рисунок 1.6 - Модель штучного нейрона

Для реалізації нелінійності при активації нейрона, його активність, крім різних видів суматорів і систем ваг на входах, визначається функцією одного аргументу - функцією активації. Нейрон в цілому реалізує скалярну функцію векторного аргументу, а вихідний сигнал нейрона визначається видом функції активації і може бути дійсним або цілим. Функція активації застосовується до зваженої суми постсинаптичних сигналів на вході нейрона. Таким чином, активність нейрона повністю визначається його параметрами - вагами і його функцією активації. Існує безліч передавальних функцій, що застосовуються на практиці при розробці нейронних мереж, деякі з них служать для реалізації нелінійності системи. Вибір тієї чи іншої функції активації часто залежить від умов завдання і структури мережі. Деякі з розглянутих нижче функцій застосовуються тільки в застарілих системах або в навчанні, але вважаються класичними і згадуються щоразу при вивченні штучних нейронних мереж як видно з рисунка 1.6.

Порогова функція Хевісайда є найпростішою кусочно-лінійною передавальною функцією. Ця функція використовувалася в класичному перцептроні, в даний час використовується в основному з метою навчання теорії нейронних мереж.

Лінійна функція активації. При використанні нескладної кусочно-лінійної функції сигнал на виході нейрона лінійно пов'язаний зі зваженою сумою сигналів на його вході. В даний момент лінійна функція на практиці також використовується дуже рідко.

Сигмоїдальна функція активації. Сигмоїд є монотонно зростаючою всюди диференційованою S-образною нелінійною функцією з насиченням. Забезпечує посилення слабких сигналів і запобігає насиченню сильних сигналів. Є однією з найбільш поширених передавальних функцій, часто використовується в нейронних мережах і сьогодні. Введення сигмоїдальних функцій було зумовлено недостатньою гнучкістю класифікаторів на основі порогових передавальних функцій і дозволило перейти від жорсткої однорозрядної логіки до більш гнучкої поведінки і адаптивної параметризації нейронних мереж. Прикладом сигмоїдальної функції є логістична передавальна функція [3].

Функція гіперболічного тангенса відрізняється від логістичної кривої тим, що її область значень лежить в інтервалі  $(-1; 1)$ , що в деяких випадках може спростити завдання навчання нейромереж.

*ReLU (Rectified Linear Unit)*. У неглибоких нейромережах використовуються нелінійні функції активації. Часто зустрічаються різновиди сигмоїдальних і тангенціальних функцій є нелінійними, але на практиці при навчанні глибоких нейромереж такі функції можуть привести до проблем із загасанням або збільшенням градієнтів. Функція ReLU є випрямленою лінійною функцією і на даний момент вважається набагато більш простим і ефективним з точки зору обчислювальної складності варіантом передавальної функції. Похідна цієї функції дорівнює або 0, або 1, від чого її застосування запобігає розростанню і загасанню градієнтів, і призводить до зменшення ваг, що позитивно позначається на обчислювальній здатності нейромережі.

Передавальна функція ReLU (1.1) є одним з останніх успіхів в області методів налаштувань глибоких нейронних мереж.

$$f(x) = \max(0, x) \quad (1.1)$$

де  $x$  - вхід нейрона.

Сьогодні існує сімейство різних модифікацій ReLU, які вирішують проблеми надійності цієї функції при проходженні через нейрон великих градієнтів: Leaky ReLU, Parametric ReLU, Randomized ReLU [3].

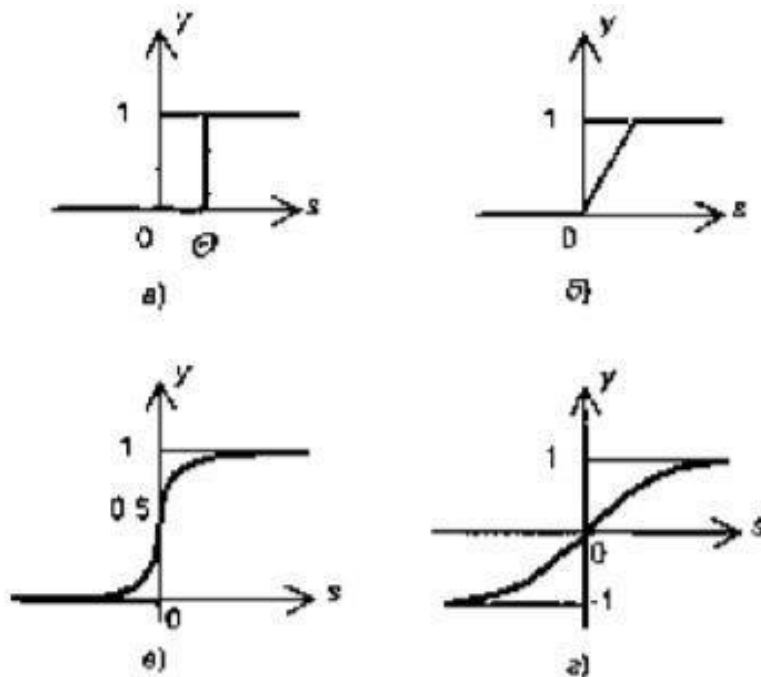


Рисунок 1.7 - Приклади функцій активації:

де а - функція одиничного стрибка,  
 б - лінійний поріг,  
 в - логістична функція,  
 г - гіперболічний тангенс.

### 1.5.2 Одношарові штучні нейронні мережи

Можливість простого розпізнавання доступна одному нейрону, проте з'єднання тих же нейронів в мережі - запорука кращого результату. Одношаровою є мережа, що складається з сукупності нейронів, утворення яких формує шар, як зображено на рисунку 1.8. Ліві вершини схеми зображеного нейрону служать для розподілу сигналів, що подаються на вхід. За цими вершинами не закріплено жодних обчислень, тому вони не вважаються шаром і позначені колами, щоб відрізнити їх від обчислювальних нейронів, що позначаються квадратами. Існує сполучення окремою вагою кожен елементу з



множини входів  $X$  з кожним штучним нейроном. За нейроном закріплена робота подачі зваженої суми входів в мережу. З метою спільності штучні та біологічні мережі мають відсутні з'єднання. Існують з'єднання між виходами і входами елементів в самому шарі. Ваги представлені елементами матриці  $W$ . Матриця має  $n$  рядків і  $m$  стовпців, де  $n$  - число входів, а  $m$  - число нейронів. Наприклад,  $w_{12}$  – це вага, що зв'язує перший вхід з другим нейроном. Таким чином, обчислення вихідного вектора  $Y$ , зводиться до матричному множенню  $Y = XW$  [3].

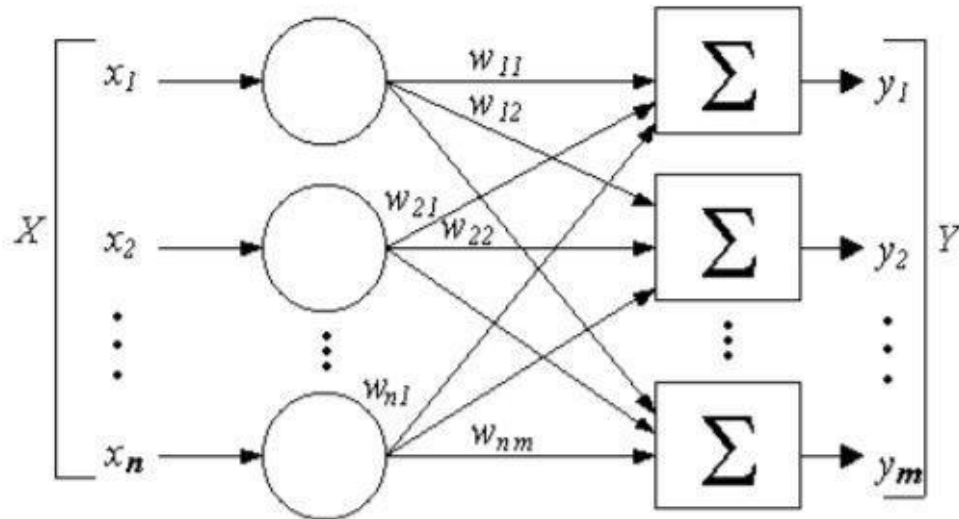


Рисунок 1.8 - Проста одношарова штучна нейронна мережа

### 1.5.3 Багатошарові штучні нейронні мережі

Відмінності обчислювальних процесів в нейронних мережах часто обумовлені способом взаємозв'язків нейронів. За сукупністю критеріїв на сьогоднішній день багатошарові архітектури можна розділити на статичні і динамічні. Кожен з класів архітектур нейронних мереж може включати безліч підкласів, реалізуючи різні підходи, нижче будуть наведені основні з них.

До статичних архітектур відносять мережі прямого поширення, в яких реалізована одно напрямлений зв'язок між шарами, відсутні динамічні елементи і зворотній зв'язок, а вихід навченої нейромережі однозначно визначається входом і не залежить від попередніх станів мережі.

Статичні штучні нейронні мережі прямого поширення:

- перцептрон;
- нейронна мережа Кохонена;
- когнітрон та неокогнітрон;
- сучасна згорткова нейронна мережа.

На противагу статичним архітектурам, існують динамічні архітектури штучних нейромереж, що реалізують рекурентну структуру з використанням зворотних зв'язків, завдяки чому стан мережі в кожний момент часу залежить

від попереднього стану. Рекурентні нейромережі як правило базуються на багатошаровому перцептроні.

### 1.5.3.1 Перцептрон, як модель розпізнавання інформації

Таким чином з'являється така модель як – перцептрон (зображено на рисунку 1.9), який також утворений за образом і подобою нашого мозку. Найпростіший елемент нагадує клітку мозку - нейрон. Має входні елементи, які згідно зі стандартом розташовуються зліва направо, зрідка від низу до верху. Зліва це входні частини нейрона, праворуч вихідні частини нейрона. Найпростіший перцептрон здатний виконувати тільки найпростіші операції. Для того, щоб виконувати більш складні обчислення, нам потрібна структура з великою кількістю прихованих шарів.

В одношаровому перцептроні входні елементи безпосередньо пов'язані з вихідними за допомогою системи ваг, зв'язки S - A організовані за принципом однозначної відповідності. Одношаровий перцептрон є окремим випадком класичного елементарного перцептрону, найпростішою мережею прямого поширення - лінійним класифікатором, і має безліч принципів обмежень, таких як неможливість реалізації функції XOR.

Простий перцептрон складається з 1го нейрона (один шар) з n входами і порогової функції активації

Оскільки виходами мережі є +1 або -1. Те перцептрон ефективний для вирішення завдання класифікації 2х класів. Якщо вихід +1, то поданий на вхід вектор належить до I-класу, інакше - II класу.

Спочатку на 0-кроці навчання, ваги входу перцептрону задаються випадковим чином. Суть навчання полягає в зміні ваг. Для здійснення процедури навчання для будь-якої нейронної мережі в тому числі і для простого перцептрону до початку формується навчальна вибірка складається з векторів, які називаються навчальними векторами, причому кожен навчальний вектор складається їх 2х частин.

Ті значення, які подаються на вхід, те, що повинно бути з нашої точки зору на виході мережі коли на вхід мережі подані компоненти з 1й частини. Загальний вигляд навчального вектора:  $(x_1, x_2, \dots, x_n, \{+1, -1\})$ .

Взагалі друга частина може бути порожньою, в цьому випадку говорять, що навчання відбувається без учителя. Якщо є - з учителем. Вектора навчальної вибірки подаються на вхід мережі і відповідно до поданих векторами змінюються ваги  $W_i$ ,  $i = 1, n$  в ході процедури навчання.

Багатошаровий перцептрон організовується з додатковими прихованими шарами A-елементів, розташованими між S-елементами і R-елементами. Принципова складність завдань, що вирішуються багатошаровим перцептроном, є найвищою для класу перцептронів. Навчання елементарного і багатошарового перцептрона полягає в зміні вагових коефіцієнтів зв'язків A – R. Перцептрон здатний працювати в режимі розпізнавання або узагальнення.

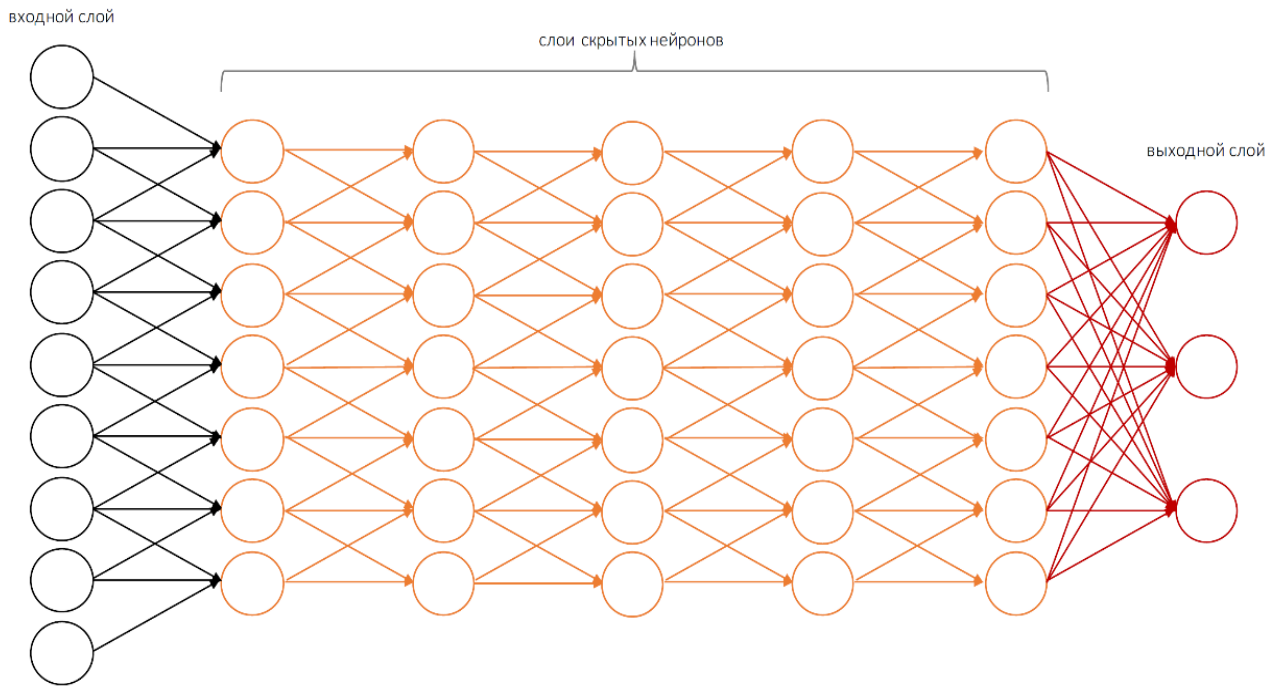


Рисунок 1.9 – Багатошаровий перцептрон

У разі комп'ютерного зору нам потрібно ще більше прихованих шарів. І тільки тоді система буде осмислено розпізнавати те, що вона бачить.

Зараз це відбувається не так. Розбиваємо наше зображення або на пікселі, або на якісь патчі: 2x2, 3x3, 5x5, 11x11 пікселів - як зручно творцям системи, в якій вони служать вхідним шаром в нейронну мережу.

Сигнали з цих вхідних шарів передаються від шару до шару за допомогою синапсів, кожен з шарів має свої певні коефіцієнти. Отже, потрібно передавати від шару до шару, від шару до шару, поки не вийде, що розпізнали особа.

### 1.5.3.2 Когнітрон та неокогнітрон, як основа згорткових нейромереж

Когнітрон був розроблений на основі будови біологічної зорової кори, має ієрархічну принципово багатошарову архітектуру. Нейрони між шарами когнітрону пов'язані тільки локально, і кожен шар реалізує різні рівні узагальнення: вхідні шари сприймають прості образи, такі як лінії, великі однорідні ділянки, їх орієнтацію і локалізацію в просторі вхідних даних, в той час як глибокі шари сприймають складніші абстрактні структури, незалежні від локалізації та інших простих ознак образу.

Когнітрон [3] організовується з ієрархічно пов'язаних збудливих і гальмуючих шарів. Співвідношення збуджуючих і гальмуючих сигналів на вході нейрона визначає його стан збудження. Існують спрощені моделі когнітрону, що будуються з одновимірних шарів, але спочатку когнітрон конструювався як каскад двовимірних шарів.

Пресинаптичний простір сигналів визначає виходи попереднього шару, постсинаптичний простір - входи наступного шару або площини. Нейрон когнітрону сприймає не весь постсинаптичний простір сигналів, а тільки його частину, реалізується принцип локальної зв'язності. Область пресинаптичного простору сигналів, що утворюють постсинаптичний простір сигналів, що впливають на стан даного нейрона, називається його локальним рецептивним полем.

Рецептивні поля близьких один до одного постсинаптичних нейронів, звані зонами конкуренції, перекриваються, тому активність даного пресинаптичного нейрона позначається на все більше поширення області постсинаптичних нейронів наступних шарів ієрархії. Розміри зон конкуренції обумовлюють кількість сприймання в просторі області ознак.

Неокогнітрон є прямим розвитком ідеї, що лежать в основі когнітрону і точніше моделює структуру зорової кори головного мозку і є класифікатором, здатним до кращого розпізнавання об'єктів. Кожен шар неокогнітрона складається з площини простих S-нейронів і площини складних C-нейронів, що також організують локальну зв'язність.

Локальне рецептивне поле на площині S-нейронів наступного шару формується пресинаптичними сигналами площини C-нейронів попереднього шару. Локальні ознаки способу сприймаються S-нейронами, а спотворення локальних ознак компенсуються C-нейронами. В результаті цього процесу кожен шар після вхідного має своїм входом все більш узагальнену картину, утворену C-нейронами попередніх шарів. З кожним рівнем глибини первинні прості ознаки визначаються у більш складних з'єднаннях. Площину S-нейронів можна розглядати як один нейрон, ваги якого визначають ядро згортки, що застосовуються до попереднього шару у всіх можливих позиціях. Всі C-нейрони реагують на образ, відповідний ядру згортки, в їх рецептивному полі, тому він визначається інваріантно до його локалізації.

Сучасні глибокі згорткові нейронні мережі засновані на ідеях, що лежать в основі неокогнітрона, і сьогодні застосовуються для вирішення широкого кола завдань: від промислових, корпоративних та дослідницьких до повсякденно побутових, включаючи завдання, які вирішуються мобільними пристроями.

## 1.6 Рекурентні нейронні мережі

Recurrent neural networks використовуються тоді, коли недостатньо розпізнавати тільки зображення. У тих випадках, коли важливо дотримуватися послідовність, потрібен порядок того, що відбувається, необхідно використовувати звичайні рекурентні нейронні мережі. Це застосовується для розпізнавання природної мови, для обробки відео, навіть використовується для розпізнавання зображень.

Що таке рекурентні нейронні мережі? Це приблизно те ж саме, що і звичайні нейронні мережі, але зі зворотним зв'язком. Зворотній зв'язок

необхідний для того щоб передавати на вхід нейронної мережі або на якийсь із її шарів попередній стан системи.

Коли потрібно обробити емоції. Навіть в усмішці - однієї з найпростіших емоцій - є кілька моментів: від нейтрального виразу обличчя до того моменту, коли у нас буде повна посмішка. Вони йдуть один за одним послідовно. Щоб це добре розуміти, потрібно вміти спостерігати за тим, як це відбувається, передавати те, що було на попередньому кадрі в наступний крок роботи системи.

### 1.6.1 Рекурентні нейронні мережи для розпізнання емоцій на відео

Приклад рекурентної нейронної мережі представлений на рисунку 1.10.

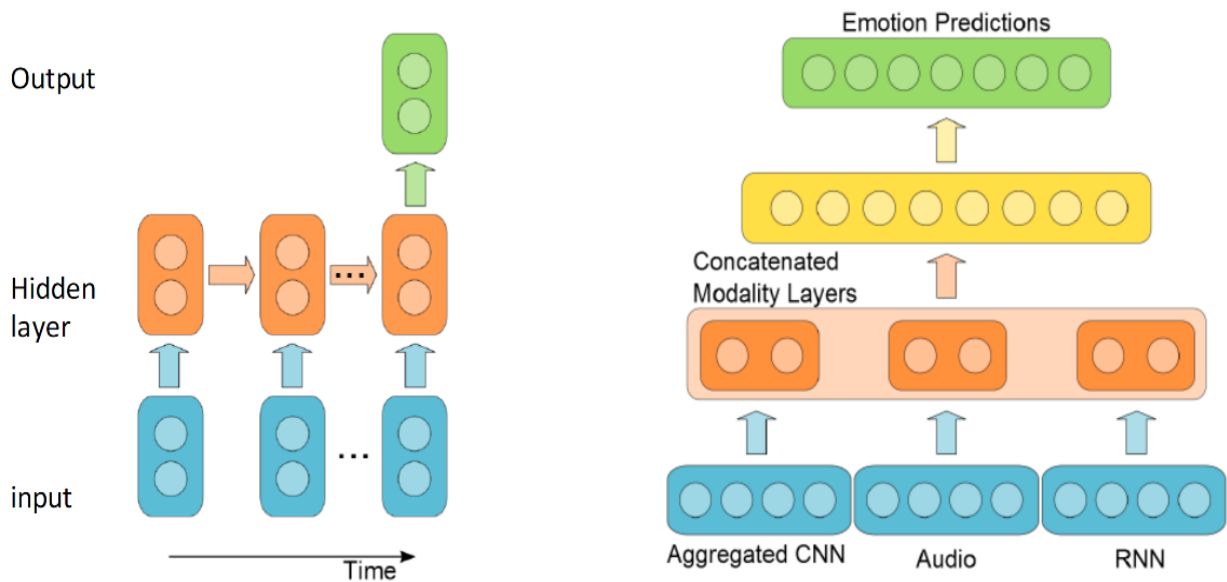


Рисунок 1.10 – Рекурентна нейронна мережа

У 2005 році на змаганні Emotion Recognition in the Wild спеціально для розпізнавання емоцій команда з Монреалю представила рекурентну систему, яка виглядала дуже просто. У неї було всього кілька згорткових шарів, і вона працювала виключно з відео. Останнього разу вони додали також розпізнавання аудіо і агрегували по кадрові дані, які виходять з convolutional neural networks, дані аудіо сигналу з роботою рекурентної нейронної мережі (з поверненням стану) і здобули перше місце на змаганні.

### 1.7 Метод навчання з підкріпленням

Для того щоб можна було почати навчання нашої мережі потрібно визначитися з тим, як вимірювати якість розпізнавання. У нашому випадку для цього будемо використовувати найпоширенішу в теорії нейронних мереж функцію середньоквадратичної помилки (СКО, MSE) [3]:

У цій формулі  $E_p$  - це помилка розпізнавання для  $p$ -ої навчальної пари,  $D_p$  - бажаний вихід мережі,  $O(I_p, W)$  - вихід мережі, що залежить від  $p$ -го входу і вагових коефіцієнтів  $W$ , куди входять ядра згортки, зміщення, вагові коефіцієнти  $S$ - і  $F$ - шарів. Завдання навчання так налаштувати ваги  $W$ , щоб вони для будь-якої навчальної пари  $(I_p, D_p)$  давали мінімальну помилку  $E_p$ . Щоб порахувати помилку для всієї навчальної вибірки просто береться середнє арифметичне по помилках для всіх навчальних пар. Таку усереднену помилку позначимо як  $E$ .

Для мінімізації функції помилки  $E_p$  найефективнішими є градієнтні методи. Розглянемо суть градієнтних методів на прикладі найпростішого одновимірного випадку (тобто коли у нас всього один вага). Якщо ми розкладемо в ряд Тейлора функцію помилки  $E_p$ , то отримаємо такий вираз:

Тут  $E$  - все та ж функція помилки,  $W_c$  - деяке початкове значення ваги. Зі шкільної математики ми пам'ятаємо, що для знаходження екстремуму функції необхідно взяти її похідну і прирівняти нулю. Так і вчинимо, візьмемо похідну функції помилки по вагам, відкинувши члени вище 2го порядку.

З виразу випливає, що вага, при якому значення функції помилки буде мінімальним можна обчислити з наступного виразу який (видно на формулі 1.2).

$$W_{min} = W_c (d^2 E(W_c) / dW^2)^{-1} dE(W_c) / dW \quad (1.2)$$

де  $W_c$  - деяке початкове значення ваги,

$E$  - все та ж функція помилки,

$W$  - вихід мережі.

Тобто оптимальну вагу обчислюється як поточний мінус похідна функції помилки по вазі, поділена на другу похідну функції помилки. Для багатовимірного випадку (тобто для матриці ваг) все одно, тільки перша похідна перетворюється в градієнт (вектор приватних похідних), а друга похідна перетворюється в гессіан (матрицю друге приватних похідних). І тут можливі два варіанти. Якщо ми опустимо другу похідну, то отримаємо алгоритм найшвидшого градієнтного спуску. Якщо все ж захочемо враховувати другу похідну, то здивуємося від того скільки обчислювальних ресурсів нам потрібно, щоб порахувати повний гессіан, а потім ще і звернути його. Тому зазвичай гессіан замінюють чимось простішим. Наприклад, один з найвідоміших і успішних методів - метод Левенберга-Марквардта (ЛМ) замінює гессіан, його апроксимацією за допомогою квадратного якобіана. Подробиці тут розповідати не буду.

Але що важливо знати про ці два метода, так це те, що алгоритм ЛМ вимагає обробки всієї навчальної вибірки, тоді як алгоритм градієнтного спуску може працювати з кожною окремо взятою навчальною вибіркою. В останньому випадку алгоритм називають стохастичним градієнтом. З огляду на, що наша база містить 60 000 навчальних зразків нам більше підходить стохастичний

градієнт. Ще однією перевагою стохастичного градієнта є його менша схильність потрапляння в локальний мінімум в порівнянні з ЛМ.

Існує також стохастична модифікація алгоритму ЛМ, про яку, можливо, згадаю пізніше.

Представлені формули дозволяють легко обчислити похідну помилки по вагам, що знаходяться в вихідному шарі. Обчислити помилку в прихованих шарах дозволяє широко відомий в ІІ метод зворотного поширення помилки.

Навчання з підкріпленням дуже часто використовується останнім часом, але не отримав такого широкого розголосу, як попередні два типи - це deep reinforcement learning, навчання з підкріпленням, приклади наведені на рисунку 1.11.

Справа в тому, що в попередніх двох випадках використовуються бази даних. Використовується як правило або дані з осіб, які дані з картинок, які дані з емоціями з відеороликів. Якщо цього всього немає, якщо не виходить це відзняти, як навчити робота брати об'єкти? Це робиться автоматично. Інший приклад: складати великі бази даних в комп'ютерних іграх складно, та й не потрібно, можна зробити набагато простіше.

В робототехніці ситуація така. Тут є кілька складнощів. По-перше, не так багато баз даних. По-друге, потрібно координувати відразу три системи: сприйняття робота, його дії за допомогою маніпуляторів і його пам'ять - то, що було зроблено в попередньому кроці і як це було зроблено.

Справа в тому, що жодна нейронна мережа, навіть deep learning на даний момент, не може впоратися з цим завданням досить ефективно, тому deep learning тільки виключно шматочки того, що потрібно зробити роботам.

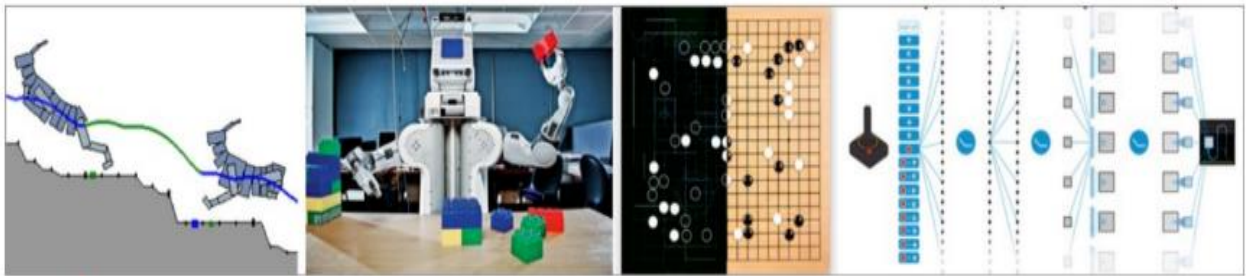


Рисунок 1.11 – Приклади використання навчання з підкріпленням

Бази даних - це якийсь відгук середовища, який потрібно накопичити для того, щоб мати можливість навчити робота щось робити надалі. Надалі роботи будуть навчатися на цій множині станів системи.

## 1.8 Основні принципи навчання нейронних мереж

Розумні алгоритми вже вміють знаходити і розпізнавати обличчя, визначати головну частину картинки, дізнаватися різні предмети. А нейронні мережі пішли далі і навіть можуть самостійно створювати твори мистецтва.

Які дані знадобляться? Необхідно з реплік в діалозі до поточного моменту, а також, можливо, якийсь інший контекстної інформації про діалог - наприклад, де і коли цей діалог відбувається - передбачити, якою має бути наступна репліка. Тобто - передбачити відповідь.

Процес навчання штучний нейронних мереж розглядається як налаштування архітектури і ваг зв'язків між нейронами (параметрів) для ефективного виконання поставлених перед мережею завдань. Існує два великих класи навчання: клас детермінованих методів і клас стохастичних методів.

Щоб вирішувати таке завдання за допомогою deep learning, нам добре б мати корпус з діалогами. Цей корпус краще б був великим, тому що deep learning з маленькими текстовими корпусами. Добре б, щоб діалоги були на потрібні теми. Тобто якщо треба зробити бота, який буде обговорювати з вами ваші почуття або говорити про погоду, то такі діалоги мають бути в діалоговому корпусі. Тому корпус діалогів зі службою підтримки інтернет-провайдера нам у вирішенні проблеми навряд чи підійде.

До класу детермінованих методи входять методи, в основі яких лежить ітеративна корекція параметрів мережі, в ході поточної ітерації яка ґрунтується на поточних параметрах. Основним детермінованим методом і найпоширенішим методом навчання мереж сьогодні є метод зворотнього поширення помилки.

До класу стохастичних методів входять методи, що змінюють параметри мережі випадковим чином і зберігають тільки ті зміни параметрів, які призвели до поліпшення результатів роботи. Стохастичні алгоритми навчання реалізуються за допомогою порівняння помилок.

Навчання з учителем. Під час навчання мережі з учителем кожному прикладу з навчальної вибірки відповідає вектор, що характеризує однозначну правильну відповідь, що подається відразу на вихід мережі в обхід всієї її архітектури.

Після отримання власного результату мережі, алгоритм порівнює результуючий вектор з правильною відповіддю, на основі чого відбувається корекція подальших помилок. Правило корекції помилки, на якому заснований метод зворотнього поширення помилки, є класичним прикладом навчання з учителем.

Навчання без вчителя у застосуванні до штучний нейромереж реалізується природним чином в процесі навчання, коли автоматичне налаштування параметрів мережею призводить до появи однакових результатів її функціонування при досить близьких вхідних значеннях, що на практиці можна порівняти зі зниженням розмірності даних в результаті ітераційного методу головних компонент.

Правило Хеба, засноване на гіпотезі про посилення зв'язків між біологічними нейронами в разі їх одночасного збудження, і методи навчання при змаганні нейронів шляхом порівняння інтенсивності їх реакції є класичними прикладами методів навчання без учителя.



Метод зворотного поширення помилки (Backprop). Метод є класичним методом навчання з учителем, заснований на правилі корекції помилок і був розроблений як метод навчання багаточарового перцептрона. Основна ідея полягає в поширенні сигналів помилки після її обчислення на виході мережі в напрямку, зворотному прямому поширенні сигналів під час звичайного обчислювального процесу, від виходів мережі до її входів. При зворотному проході синаптичні ваги налаштовуються з метою мінімізації помилки. Фактично, реалізується стохастичний градієнтний спуск, відбувається рух в багатовимірному просторі ваг до мінімуму помилки в сторону, протилежну градієнту. В процесі навчання циклічно знаходяться рішення на однокритеріальні завдання оптимізації.

Для можливості реалізації методу передавальна функція нейронів повинна бути диференційована. Оскільки метод є модифікацією класичного методу градієнтного спуску, то може розглядатися як градієнтний спуск по поверхні помилки.

### 1.8.1 Вибір функції активації при навчанні нейронної мережі

Синапси здійснюють зв'язок між нейронами і множать вхідний сигнал на число, що характеризує силу зв'язку - вага синапсу. Суматор виконує додавання сигналів, що надходять по синаптичеським зв'язків від інших нейронів і зовнішніх вхідних сигналів. Перетворювач реалізує функцію одного аргументу, виходу суматора, в деяку вихідну величину нейрона. Ця функція називається функцією активації нейрона. Нейрон в цілому реалізує скалярну функцію векторного аргументу. У загальному випадку вхідний сигнал і вагові коефіцієнти можуть приймати дійсні значення. Вихід визначається видом функції активації і може бути як дійсним, так і цілим. Синаптичні зв'язки з позитивною вагою називають збудливими, з негативними вагами - гальмують.

Таким чином, нейрон повністю описується своїми вагами і функцією активації  $F$ . Одержавши набір чисел (вектор) в якості входів, нейрон видає деяке число на виході.

Активаційна функція може бути різного виду [61]. Найбільш широко використовуються варіанти які показані нижче.

Одними з найбільш поширених функцій є:

- лінійна;
- нелінійна з насиченням - логістична функція або сигмоид;
- гіперболічний тангенс.

Лінійна функція найкращим чином відповідає суті даного завдання. Її областю визначення є діапазон  $(-\infty, \infty)$ . Це дозволяє, використовуючи цінові характеристики товару на вході, отримувати характеристичні значення будь-якої величини на виході, рівні їх фактичної сумі. Слід зазначити, що сігмоїдної функція диференційована на всій осі абсцис, що широко використовується в багатьох алгоритмах навчання.

## 1.9 Висновки до першого розділу

У першому розділі коротко розглянуто теоретичні відомості про машинне навчання, також описано два основних типів навчання. Наведено приклади використання у житті. Технологія створення машин, що мають здатність бачити лежить в основі комп'ютерного зору. Ця дисципліна відноситься до штучних систем, котрі можуть виконувати завдання розпізнавання. У вищенаведеному розділі наведена основна теорія розпізнавання об'єктів, що лежить в основі класифікації. Описано про теорію розпізнавання образів, як розділ дисциплін. Розглянуто основи розпізнавання об'єктів, виділені основні проблеми, наведені приклади класичних задач та методи якими можна їх вирішувати, але поки без глибокої характеристики. Розписано всі основоположні терміни з ілюстраціями і прикладами їх.

Технологія створення машин, що мають здатність бачити лежить в основі комп'ютерного зору. Ця дисципліна відноситься до штучних систем, котрі можуть виконувати завдання розпізнавання. У вищенаведеному розділі наведена основна теорія розпізнавання об'єктів, що лежить в основі класифікації.

Описано про технологію комп'ютерного зору, та умовно розділено на три групи методи вирішення завдань комп'ютерного зору. Огляд методів вирішення задач комп'ютерного зору за групами обґрунтував переваги нейромережевого методу над усіма запронованими. Було розглянуто такі методи як: фільтрація, логічний аналіз. Вирішення задач за допомогою штучних одно та багатошарових нейронних мереж. Таких як: перцептрон, когнітрон та неокогнітрон. Також описано про рекурентні нейронні мережи, і чому у деяких випадках треба використовувати саме їх.

Принцип роботи нейронів штучної мережі дозволяє виявити його активність, що визначається його параметрами - вагами і його функцією активації.

Труднощі при спотвореннях, такі як зміна образу у розмірах і поворот зображення, вирішуються при моделюванні неокогнітрона, що використовує якісно нову архітектуру. В основу архітектури неокогнітрона покладена організація зорової системи живих істот. Виділено два класи навчання: клас детерменірованих методів і клас стохастичних методів. Надано їх характеристики. Також описано навчання яке базується на порівнянні помилок. Це навчання з учителем та навчання без учителя. Наведено приклади та правила на яких засновано ці методи.

## **2 ПЛАНУВАННЯ ЕКСПЕРИМЕНТАЛЬНИХ ДОСЛІДЖЕНЬ**

### **2.1 Аналіз проблеми навчання розпізнавання зображень**

Сприйняття світу - одне із загадкових властивостей живого мозку, що дозволяє розібратися в нескінченному потоці сприйманої інформації і зберігати орієнтацію в океані розрізнених даних про зовнішній світ.

Сприймаючи зовнішній світ, ми завжди проводимо класифікацію сприйманих відчуттів. Розбиваємо їх на групи схожих, але не тотожних явищ. Наприклад, незважаючи на суттєву різницю, до однієї групи належать всі букви А, написані різними почерками, або все звуки, відповідні одній і тій же ноті, взятої в будь-якій октаві і на будь-якому інструменті, а оператор, керуючий технічним об'єктом, на ціле безліч станів об'єкта реагує однієї і тієї ж реакцією.

Характерно, що для складання поняття про групу сприйняття певного класу досить ознайомитися з незначною кількістю її представників. Дитині можна показати всього один раз будь-яку букву, щоб він зміг знайти цю букву в тексті, написаному різними шрифтами, або дізнатися її, навіть якщо вона написана в навмисне спотвореному вигляді. Це властивість мозку дозволяє сформулювати таке поняття, як зображення.

Зображення мають характерною властивістю, що виявляється в тому, що ознайомлення з кінцевим числом явищ з одного і того ж безлічі дає можливість дізнаватися як завгодно велике число його представників. Прикладами образів можуть бути: річка, море, рідина, музика Чайковського, вірші Маяковського та інше. В якості способу можна розглядати і деяку сукупність станів об'єкта управління, причому вся ця сукупність станів характеризується тим, що для досягнення заданої мети потрібно однаковий вплив на об'єкт. Образи мають характерні об'єктивними властивостями в тому сенсі, що різні люди, які навчаються на різному матеріалі спостережень, здебільшого однаково і незалежно один від одного класифікують одні і ті ж об'єкти. Саме ця об'єктивність образів дозволяє людям усього світу розуміти один одного. Здатність сприйняття зовнішнього світу в формі образів дозволяє з певною вірогідністю дізнаватися нескінченне число об'єктів на підставі ознайомлення з кінцевим їх числом, а об'єктивний характер основного властивості образів дозволяє моделювати процес їх розпізнавання. Будучи відображенням об'єктивної реальності, поняття образу настільки ж об'єктивно, як і сама реальність, а тому це поняття може бути саме по собі об'єктом спеціального дослідження.

#### **2.1.1 Проблема адаптації розпізнавання образів**

Одним з найцікавіших властивостей людського мозку є здатність відповідати на безліч станів зовнішнього середовища кінцевим числом реакцій. Може бути, саме ця властивість дозволило людині досягти вищої форми існування живої матерії, що виражається в здатності до мислення, т.д.

Активному відображенню об'єктивного світу у вигляді образів, понять, суджень і т.д. Тому проблема адаптації розпізнавання образів виникла при вивченні фізіологічних властивостей мозку .

Розглянемо приклад завдань з області адаптації розпізнавання образів.

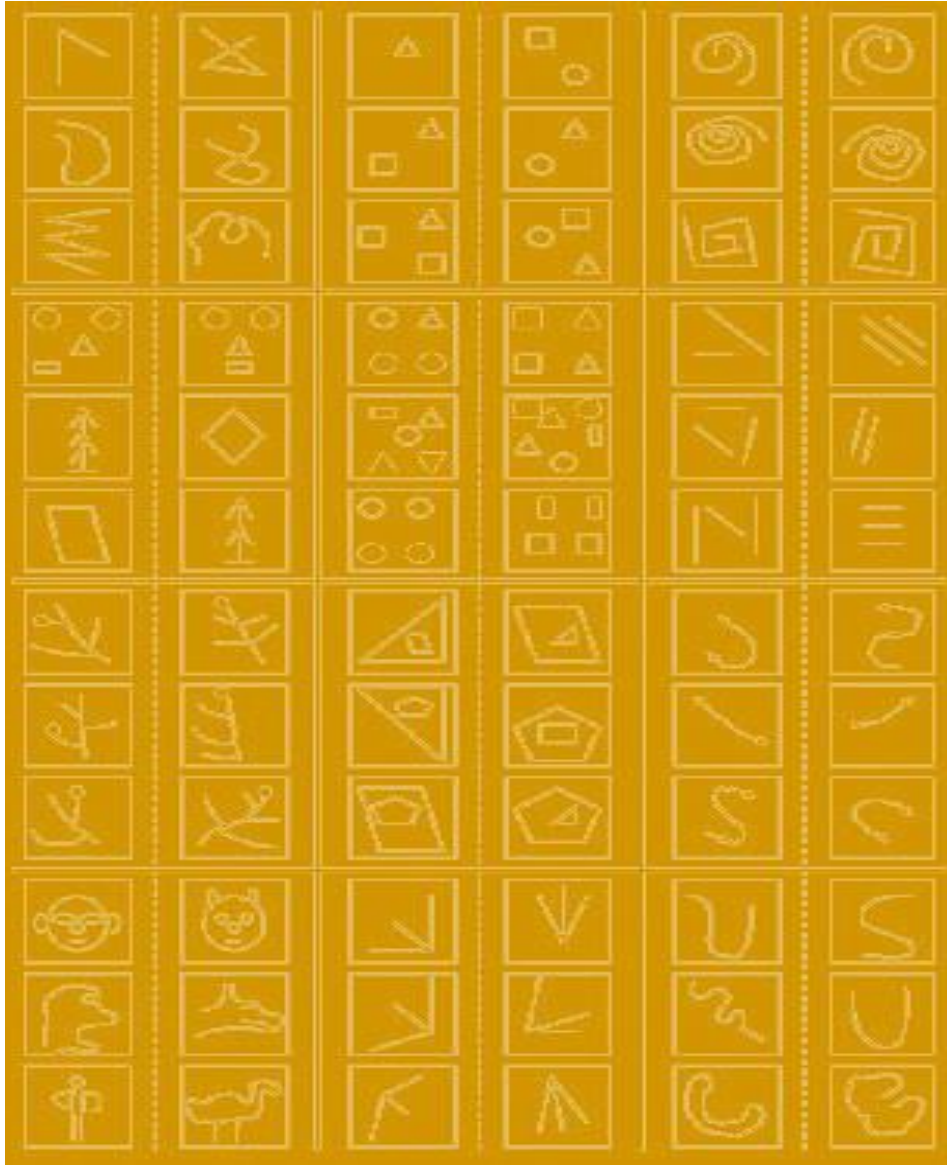


Рисунок 2.1 – Приклад зображення для розпізнання

Проблема адаптації розпізнавання образів цікава як з прикладної, так і з принципової точки зору. З прикладної точки зору рішення цієї проблеми важливо перш за все тому, що воно відкриває можливість автоматизувати багато процесів, які до сих пір пов'язували лише з діяльністю живого мозку.

Тут представлені 12 завдань, в яких слід відібрати ознаки, за допомогою яких можна відрізнити ліву тріаду картинок від правої. Рішення даних завдань вимагає моделювання логічного мислення в повному обсязі. В цілому проблема розпізнавання образів складається з двох частин: навчання та розпізнавання.

Коло завдань, які можуть вирішуватися за допомогою систем, що розпізнають, надзвичайно широкий. Сюди відносяться не тільки задачі розпізнавання зорових і слухових образів, а й завдання розпізнавання складних процесів і явищ, що виникають, наприклад, при виборі доцільних дій керівником підприємства або виборі оптимального управління технологічними, економічними, транспортними або військовими операціями. У кожній з таких завдань аналізуються деякі явища, процеси, стану зовнішнього світу, усюди далі звані об'єктами спостереження. Перш ніж почати аналіз будь-якого об'єкта, потрібно отримати про нього певну, будь-яким способом впорядковану інформацію наприклад як зображено приклад аналізу розпізнавання облич через камеру смартфона на рисунку 2.2. Така інформація є характеристикою об'єктів, їх відображення на безлічі сприймають органів розпізнає. Але кожен об'єкт спостереження може впливати по-різному, в залежності від умов сприйняття.

Наприклад, будь-яка буква, навіть однаково написана, може в принципі як завгодно зміщатися відносно сприймають органів. Крім того, об'єкти одного і того ж образу можуть досить сильно відрізнятися один від одного і, природно, по-різному впливати на сприймають органи.



Рисунок 2.2 – Аналіз об'єктів через камеру

Навчання здійснюється шляхом показу окремих об'єктів із зазначенням їх приналежності того чи іншого способу. В результаті навчання розпізнає система повинна придбати здатність реагувати однаковими реакціями на всі об'єкти одного образу і різними - на всі об'єкти різних образів. Дуже важливо, що процес навчання повинен завершитися тільки шляхом показів кінцевого числа об'єктів без будь-яких інших підказок. Як об'єкти навчання можуть бути або картинки, або інші візуальні зображення (літери), або різні явища зовнішнього світу, наприклад звуки, стану організму при медичному діагнозі,

стан технічного об'єкта в системах управління і ін. Важливо, що в процесі навчання вказуються тільки самі об'єкти і їх приналежність образу. За навчанням слід процес розпізнавання нових об'єктів, який характеризує дії вже навченої системи. Автоматизація цих процедур і становить проблему навчання розпізнаванню образів. У тому випадку, коли людина сама розгадує або придумує, а потім нав'язує машині правило класифікації, проблема розпізнавання вирішується частково, так як основну і головну частину проблеми (навчання) людина бере на себе.

Принципове значення проблеми тісно пов'язане з питанням, який все частіше виникає в зв'язку з розвитком ідей кібернетики: що може і що принципово не може робити машина? Якою мірою можливості машини можуть бути наближені до можливостей живого мозку? Зокрема, чи може машина розвинути в собі здатність перейняти у людини вміння виробляти певні дії в залежності від ситуацій, що виникають у навколишньому середовищі? Поки стало ясно тільки те, що якщо людина може спочатку сам усвідомити своє вміння, а потім його описати. Вказати, чому він виробляє дії у відповідь на кожне стан зовнішнього середовища або як (по яким правилом) він об'єднує окремі об'єкти в образи, то таке вміння без принципових труднощів може бути передано машині. Якщо ж людина володіє умінням, але не може пояснити його, то залишається тільки один шлях передачі вміння машині - навчання прикладами.

Кожне відображення будь-якого об'єкта на сприймаючі органи системи, що розпізнає, незалежно від його положення щодо цих органів, прийнято називати зображенням об'єкта, а безлічі таких зображень, об'єднані певними загальними властивостями, являють собою образи. При вирішенні завдань управління методами розпізнавання образів замість терміна "зображення" застосовують термін "стан".

Стан - це певної форми відображення вимірюваних поточних (або миттєвих) характеристик об'єкта, що спостерігається. Сукупність станів визначає ситуацію. Поняття "ситуація" є аналогом поняття "образ". Але ця аналогія неповна, тому що не всякий образ можна назвати ситуацією, хоча будь-яку ситуацію можна назвати чином.

Ситуацією прийнято називати деяку сукупність станів складного об'єкта, кожна з яких характеризується одними і тими ж або схожими характеристиками об'єкта. Наприклад, якщо в якості об'єкта спостереження розглядається деякий об'єкт управління, то ситуація об'єднує такі стани цього об'єкта, в яких слід застосовувати одні і ті ж дії, що управляють. Якщо об'єктом спостереження є військова гра, то ситуація об'єднує всі стани гри, які вимагають, наприклад, потужного танкового удару при підтримці авіації.

Вибір вихідного опису об'єктів є однією з центральних завдань проблеми адаптації розпізнавання зображень. При вдалому виборі вихідного опису (простору ознак) завдання розпізнавання може виявитися тривіальною і, навпаки, невдало вибране вихідне опис може привести або до дуже складної подальшої переробки інформації, або взагалі до відсутності рішення.

Наприклад, якщо вирішується завдання розпізнавання об'єктів, що відрізняються за кольором, а в якості вихідного опису обрані сигнали, одержувані від датчиків ваги, то завдання розпізнавання в принципі не може бути вирішена.

Типовим прикладом недостатнього рівня безпеки впровадженої системи віддаленого розпізнавання документів може служити досвід фінансових організацій, які видали шахраям в віддаленому режимі кредити за копіями скомпрометованих документів. Причиною служить недостатньо досконала система віддаленої верифікації користувача, що робить можливим взяти кредит по зображенню паспорта, купленому в даркнеті.

Вже давно як провайдери послуг, так і клієнти розуміють, що в умовах розвитку інформаційних технологій одно факторної аутентифікації користувача на підставі зв'язки "логін-пароль" явно недостатньо. Точно так же недостатньо і аутентифікації виключно за даними засвідчують документів або їх образів.

Витік зв'язки "логін-пароль" стала вже звичним кейсом в області порушень інформаційної безпеки. Таким же звичайним явищем стає витік даних посвідчення особи або його образу у вигляді скан або фото. Саме тому їх використання в сервісах як єдиний спосіб підтвердження особи сьогодні практично не застосовується, тому що являє реальну загрозу досить легкого несанкціонованого доступу не тільки до персональних даних користувача, а й отриманням контролю його облікового запису.

## 2.2 Підходи до вирішення задач

Перш ніж розпізнати зображення, необхідно його виділити або сегментувати. Сегментація як процедура пошуку однорідних, наприклад, по яскравості, кольору або текстурою, областей досить важка і до кінця не вирішена задача.

Для реалізації зазначеної проблеми використовуються два способи досліджень, дозволяють досягти бажаного результату. Перший передбачає декомпозицію проблеми, об'єкта на окремі фрагменти з визначенням не тільки фрагментів, але і зв'язків між ними. Природно для отримання математичної моделі опис життєвого циклу використання методів і інструментальних засобів при створенні системи вкрай важко і можливо застосування для окремих не зв'язаних фрагментів. Другий пропонує семантичну сегментацію моделі зображення як (показано на рисунку 2.3).



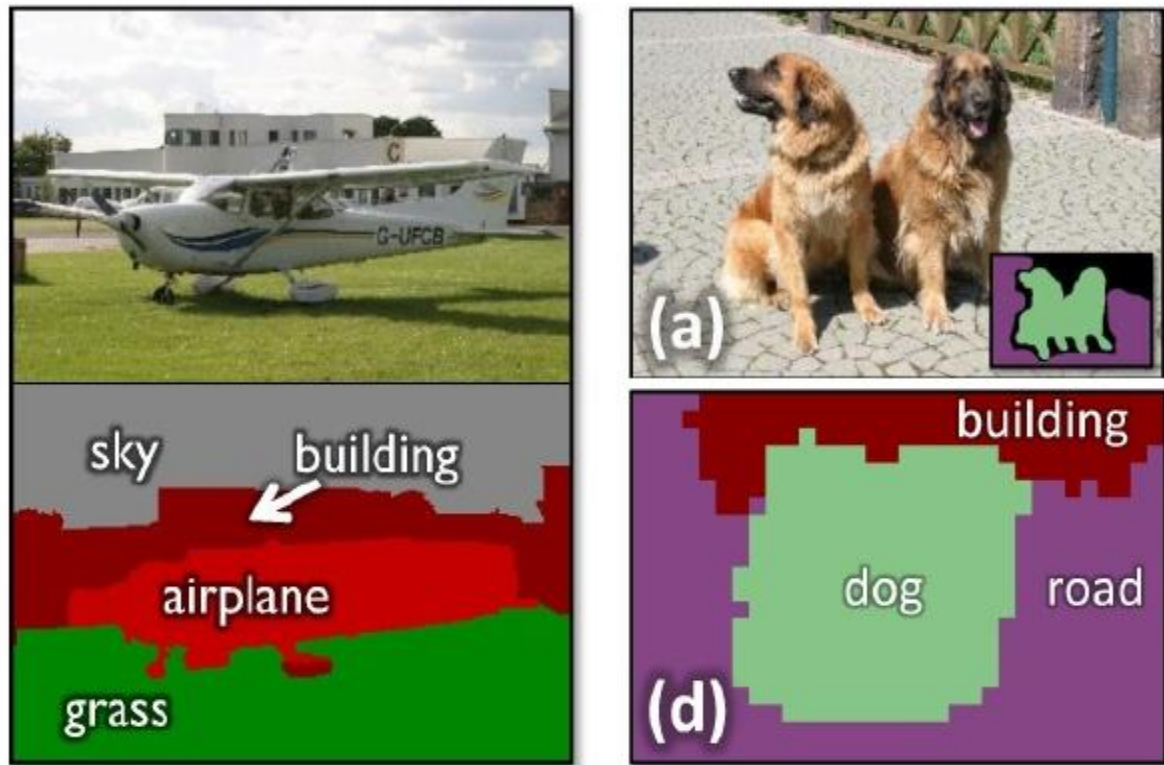


Рисунок 2.3 – Семантична сегментація зображень

Визначення такої моделі - дуже складна проблема через неможливості опису не тільки характеристик створюваної системи, її окремих компонентів, послідовності виконуваних робіт на кожному етапі проектування, одержуваних технічних рішень, але і опису функцій розробників, застосовуваних методологій, технологій, методів, інструментальних засобів. Тому найбільш прийнятним є другий спосіб, а саме категорно-функторні, що має на увазі проектування створюваної системи на "родинні" її системи і за властивостями проєкцій (відображень) встановлює внутрішню структуру елементів її життєвого циклу у вигляді категорій.

Наявність категорій елементів циклу дає можливість отримати математичну модель проектування системи у вигляді композиції функторів. фактично така математична модель описує процес створення складної інформаційної системи на макрорівні. Однак подальше уточнення моделі, що припускає визначення об'єктів у вигляді структурованих множин і морфізма між ними, з подальшою розробкою відповідних алгоритмів, дозволяє перейти від макропроектування до мікропроектування. Це в певній мірі дає можливість формалізувати зазначену проблему і знизити втрати по використуваних ресурсів, так як є можливість від концептуального проектування перейти до фізичного шляхом моделювання даного процесу.

На основі категорно-функторні моделі є можливість побудувати імітаційну модель процесу проектування, маючи на увазі завдання мети, функцій, обмежень і отримання параметрів створюваної системи у вигляді результатів від реалізованих функцій. Однак фізичне проектування буде ефективним при



застосуванні типових проектних рішень в рамках застосовуваних інформаційних технологій.

### 2.2.1 Геометричний і структурний підходи

Кожен раз, коли стикаються з незнайомими завданнями, з'являється природне бажання представити їх у вигляді деякої легко розуміється моделі, яка дозволяла б осмислити завдання в таких термінах, які легко відтворюються нашою уявою. А так як ми існуємо в просторі і в часі, найбільш зрозумілою для нас є просторово-часова інтерпретація задач.

Будь-яке зображення, яке виникає в результаті спостереження будь-якого об'єкта в процесі навчання або іспиту, можна представити у вигляді вектора, а значить і в вигляді точки деякого простору ознак. Якщо стверджується, що при показі зображень можливо однозначно віднести їх до одного з двох (або декількох) образів, то тим самим стверджується, що в деякому просторі існує дві (або декілька) області, які не мають спільних точок, і що зображення - точки з цих областей. Кожній такій області можна приписати найменування, т. Е. Дати назву, яка відповідає образу.

Якщо проінтерпретувати в термінах геометричної картини процес адаптації розпізнавання образів, обмежившись поки випадком розпізнавання тільки двох образів. Заздалегідь вважається відомим лише тільки те, що потрібно розділити дві області в деякому просторі і що показуються точки тільки з цих областей. Самі ці області заздалегідь не визначені. Немає будь-яких відомостей про розташування їх меж чи правил визначення приналежності точки до тієї чи іншої області.

В ході навчання пред'являються точки, випадково вибрані з цих областей, і повідомляється інформація про те, до якої області належать пред'являються точки. Ніякої додаткової інформації про ці області. Про розташування їх меж, в ході навчання не повідомляється. Мета навчання полягає або в побудові поверхні, яка розділяла б не тільки показані в процесі навчання точки, але і всі інші точки, що належать цим областям, або в побудові поверхонь, що обмежують ці області так, щоб в кожній з них знаходилися тільки точки одного образу. Інакше кажучи, мета навчання полягає в побудові таких функцій від векторів-зображень, які були б, наприклад, позитивні на всіх точках одного і негативні на всіх точках іншого способу. У зв'язку з тим, що області не мають спільних точок, завжди існує ціла безліч таких поділяють функцій, а в результаті навчання повинна бути побудована одна з них.

Якщо пред'являються зображення належать не двом, а більшій кількості образів, то завдання полягає в побудові по показаним в ході навчання точкам поверхні, що розділяє всі області, які відповідають цим образам, один від одного. Завдання це може бути вирішена, наприклад, шляхом побудови функції, що приймає над точками кожної з областей однакоє значення, а над точками з різних областей значення цієї функції повинно бути по-різному.

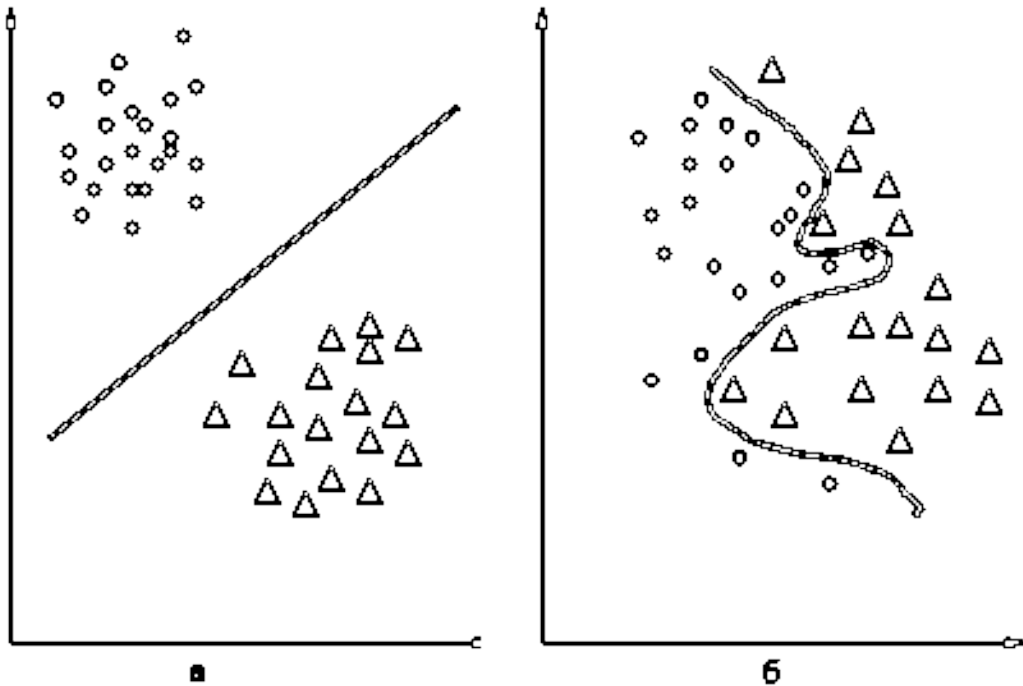


Рисунок 2.4 – Приклад з двома образами

На перший погляд здається, що знання всього лише деякої кількості точок з області недостатньо, щоб відокремити всю область. Дійсно, можна вказати незліченна кількість різних областей, які містять ці точки, і як би не була побудована за ним поверхню, що виділяє область, завжди можна вказати іншу область, яка перетинає поверхню і в той же час містить показання точки. Однак відомо, що завдання про наближення функції за інформацією про неї в обмеженій множині точок, істотно більш вузької, ніж вся безліч, на якому функція задана, є звичайною математичною завданням про апроксимації функцій. Зрозуміло, рішення таких задач вимагає введення певних обмежень на класі розглянутих функцій, а вибір цих обмежень залежить від характеру інформації, яку може додати учитель в процесі навчання.

Однією з таких підказок є гіпотеза про компактності образів. Інтуїтивно ясно, що апроксимація розділяє функції буде завданням тим легшою, ніж більш компактні і чим більше рознесені в просторі області, що підлягають розподілу.

Так, наприклад, в разі, показаному на рисунку 2.4(а), поділ свідомо більш просто, ніж в разі, показаному на рисунку 2.4(б). Дійсно, в разі, зображеному на рисунку 2.4(а), області можуть бути розділені площиною, і навіть при великих погрішності у визначенні розділяє функції вона все ж буде продовжувати розділяти області. У разі ж на рисунку 2.4(б), поділ здійснюється хитромудрою поверхнею і навіть незначні відхилення в її формі призводять до помилок поділу. Саме це інтуїтивне уявлення про порівняно легко поділених областях призвело до гіпотези компактності.

Поряд з геометричною інтерпретацією проблеми адаптації розпізнавання образів існує й інший підхід, який названий структурним, або лінгвістичним. Пояснимо лінгвістичний підхід на прикладі розпізнавання зорових зображень.

Спочатку виділяється набір вихідних понять - типових фрагментів, що зустрічаються на зображеннях, і характеристик взаємного розташування фрагментів - "зліва", "знизу", "всередині" та інші. Ці вихідні поняття утворюють словник, що дозволяє будувати різні логічні висловлювання, іноді звані припущеннями. Завдання полягає в тому, щоб з великої кількості висловлювань, які могли б бути побудовані з використанням цих понять, відібрати найбільш суттєві для даного конкретного випадку.

Далі, переглядаючи кінцеве і по можливості невелике число об'єктів з кожного образу, потрібно побудувати опис цих образів. Побудовані опису повинні бути настільки повними, щоб вирішити питання про те, до якого образу належить даний об'єкт. При реалізації лінгвістичного підходу виникають два завдання:

- завдання побудови вихідного словника;
- завдання побудови правил опису з елементів заданого словника.

В рамках лінгвістичної інтерпретації проводиться аналогія між структурою зображень і синтаксисом мови. Прагнення до цієї аналогії було викликано можливістю використовувати апарат математичної лінгвістики.

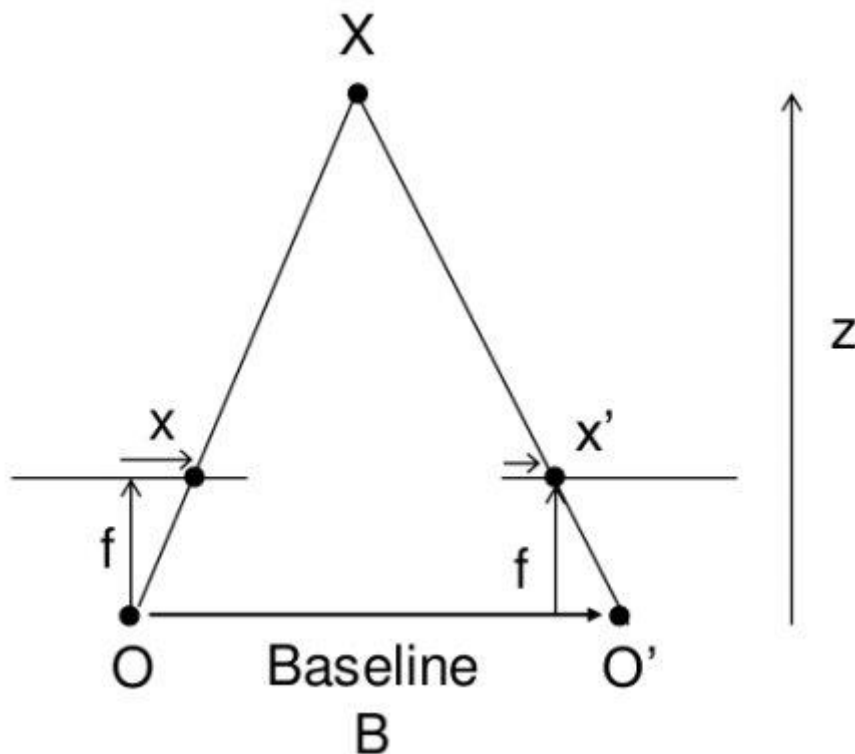


Рисунок 2.5 – Геометричний підхід

Методи за своєю природою є синтаксичними. Використання апарату математичної лінгвістики для опису структури зображень можна застосовувати

тільки після того, як проведена сегментація зображень на складові частини. Вироблені слова для опису типових фрагментів і методи їх пошуку. Після попередньої роботи, що забезпечує виділення слів, виникають власне лінгвістичні завдання, що складаються із завдань автоматичного граматичного розбору описів для розпізнавання зображень. При цьому проявляється самостійна область досліджень, яка вимагає не тільки знання основ математичної лінгвістики, а й оволодіння прийомами, які розроблені спеціально для лінгвістичної обробки зображень.

### 2.2.2 Гіпотеза компактності

Якщо припустити, що в процесі навчання простір ознак формується виходячи із задуманої класифікації, то тоді можна сподіватися, що завдання простору ознак саме по собі задає властивість, під дією якої образи в цьому просторі легко розділяються. Саме ці надії в міру розвитку робіт в області розпізнавання образів стимулювали появу гіпотези компактності, в якій мовиться: образам відповідають компактні безлічі в просторі ознак. Під компактним безліччю поки будемо розуміти якісь "згустки" точок в просторі зображень, припускаючи, що між цими згустками існують розділяють їх розрядження.

Однак цю гіпотезу не завжди вдавалося підтвердити експериментально, але, що найголовніше, ті завдання, в рамках яких гіпотеза компактності добре виконувалася як видно із рисунка 2.4(а), все без винятку знаходили просте рішення. І навпаки, ті завдання, для яких гіпотеза не підтверджувалася як видно із рисунка 2.4(б), або зовсім не вирішувалися, або вирішувалися з великими труднощами з залученням додаткових хитрувань. Цей факт змусив щонайменше засумніватися в справедливості гіпотези компактності, так як для спростування будь-якої гіпотези досить одного заперечує її прикладу. Разом з цим, виконання гіпотези всюди там, де вдавалося добре вирішити завдання навчання розпізнаванню образів, зберігало до цієї гіпотези інтерес. Сама гіпотеза компактності перетворилася в ознака можливості задовільного вирішення завдань розпізнавання.

Формулювання гіпотези компактності підводить впритул до поняття абстрактного образу. Якщо координати простору вибирати випадково, то і зображення в ньому будуть розподілені випадково. Вони будуть в деяких частинах простору розташовуватися більш щільно, ніж в інших. Назвемо деякий випадково обраний простір абстрактним зображенням. У цьому абстрактному просторі майже напевно будуть існувати компактні безлічі точок. Тому відповідно до гіпотезою компактності безлічі об'єктів, яким в абстрактному просторі відповідають компактні безлічі точок, розумно назвати абстрактними образами даного простору.

### 2.2.3 Адаптація та самонавчання

Всі картинки представлені, характеризують завдання навчання. У кожній з цих завдань задається кілька прикладів (навчальна послідовність) правильно вирішених завдань. Якби вдалося помітити якусь загальну властивість, яке залежить ні від природи образів, ні від їх зображень, а визначальне лише їх здатність до розділяємості, то поряд зі звичайною завданням адаптації розпізнавання, з використанням інформації про приналежність кожного об'єкта з навчальної послідовності того чи іншого образу можна було б поставити іншу класифікаційну завдання - так звану задачу навчання без учителя. Завдання такого роду на описовому рівні можна сформулювати наступним чином: системі одночасно або послідовно пред'являються об'єкти без будь-яких вказівок про їх належність до образам.

Вхідний пристрій системи відображає безліч об'єктів на безліч зображень і, використовуючи деякий закладене в неї заздалегідь властивість розділяємості образів, виробляє самостійну класифікацію цих об'єктів. Після такого процесу самонавчання система повинна придбати здатність до розпізнавання не тільки вже знайомих об'єктів (об'єктів з навчальної послідовності), але і тих, які раніше не пред'являлися. Процесом самонавчання деякої системи називається такий процес, в результаті якого ця система без підказки вчителя набуває здатності до вироблення однакових реакцій на зображення об'єктів одного і того ж образу і різних реакцій на зображення різних образів. Роль вчителя при цьому складається лише в підказці системі деякого об'єктивного властивості, однакового для всіх образів і визначає здатність до поділу безлічі об'єктів на образи.

Виявляється, таким об'єктивним властивістю є властивість компактності образів. Взаємне розташування точок в обраному просторі вже містить інформацію про те, як слід розділити безліч точок. Ця інформація і визначає то властивість розділяємості образів, що виявляється достатнім для самонавчання системи розпізнавання образів.

Більшість відомих алгоритмів самонавчання здатні виділяти тільки абстрактні образи, тощо. Компактні безлічі в заданих просторах. Різниця між ними полягає, по-видимому, в формалізації поняття компактності. Однак це не знижує, а іноді і підвищує цінність алгоритмів самонавчання, так як часто самі образи заздалегідь ніким не визначені, а завдання полягає в тому, щоб визначити, які підмножини зображень в заданому просторі є образи. Хорошим прикладом такої постановки завдання є соціологічні дослідження, коли по набору питань виділяються групи людей. В такому розумінні завдання алгоритми самонавчання генерують заздалегідь не відому інформацію про існування в заданому просторі образів, про які раніше ніхто не мав жодного уявлення.

Крім того, результат самонавчання характеризує придатність обраного простору для конкретного завдання навчання розпізнаванню. Якщо абстрактні образи, які виділяються в процесі самонавчання, збігаються з реальними, то

простір вибрано вдало. Чим сильніше абстрактні образи відрізняються від реальних, тим "незручно" вбрання простір для конкретного завдання.

Навчанням зазвичай називають процес вироблення в деякій системі тієї чи іншої реакції на групи зовнішніх ідентичних сигналів шляхом багаторазового впливу на систему зовнішньої коректування. Таку зовнішню коригування в навчанні прийнято називати "заохоченнями" і "покараннями". Механізм генерації цієї коригування практично повністю визначає алгоритм навчання. Самонавчання відрізняється від навчання тим, що тут додаткова інформація про вірність реакції системі не повідомляється.

Адаптація - це процес зміни параметрів і структури системи, а можливо, і керуючих впливів на основі поточної інформації з метою досягнення певного стану системи при початковій невизначеності і мінливих умовах роботи.

Навчання - це процес, в результаті якого система поступово набуває здатність відповідати потрібними реакціями на певні сукупності зовнішніх впливів, а адаптація - це підстроювання параметрів і структури системи з метою досягнення необхідної якості управління в умовах безперервних змін зовнішніх умов.

### 2.3 Розпізнавання за допомогою послідовної активації детекторів

Після формування і навчання детекторів в потоці даних, їх ієрархічна структура формує систему розпізнавання і може використовуватися для класифікації зображень. Розробка алгоритму розпізнавання зображень є ключовим етапом дисертаційного дослідження, оскільки алгоритм здійснює безпосередню процедуру класифікації зображень за категоріями з використанням попередньо навченої ієрархії детекторів.

Алгоритм розпізнавання складається з послідовної активації детекторів моделі, починаючи з першого рівня, на заданому зображенні. Нехай є шукана функція розпізнавання:  $\rightarrow$ , аргументом якої є зображення  $\in$ , представлені у вигляді вектора довжини, а значеннями функцій - безліч класів (категорій)  $\in$ , варійоване в залежності від поставленого завдання.  $\in$  підмножина пар аргументів із значень функції  $= \{(\cdot), \dots, (\cdot)\}$ . послідовно активується ієрархія детекторів, таким чином, реалізує функцію  $h: \rightarrow$ , яка апроксимує функцію на всій її області визначення, в тому числі в точках, які не включені в. Для обчислення значення  $h(\cdot)$  зображення надходить на вхід першого шару навченої системи, потім виконується послідовна

активація локальних детекторів на кожному з рівнів. вихідна значення ієрархії являє собою бінарне число, що визначає приналежність зображення до класу, при цьому вихідний рівень також проводить оцінку параметрів локалізації зображеного об'єкта, якщо значення активації дорівнює 1 тоді - зображення успішно розпізнано як належить класу.

У такому вигляді розроблена система розпізнавання використовується для рішення задачі унарною класифікації, коли безліч класів  $\in$  представлено двома елементами і одно  $\{0,1\}$ . Функція, таким чином, дорівнює 1 у випадках, коли

зображення, що служить її аргументом, містить об'єкт, належить цікавить класу, і 0 в іншому випадку. Для випадків, коли потрібно розпізнати зображення серед кількох можливих класів називається завданням мультикласової класифікації, проводиться навчання окремої ієрархії детекторів для кожного конкретного класу, і потім проводиться почергова перевірка зображення на позитивне відповідність кожної з них. В цьому випадку розглядається функція  $X'$ , певна на безлічі, значеннями якої є безліч класів  $Y'$ , таке що для обраного - того класу  $Y' = \{y_i, U_{j^*i} y_j\}$ .

Нехай дана навчена ієрархія детекторів  $M_c$  для деякого класу зображень (наприклад, людських осіб), або кілька ієрархій для завдання мультикласового розпізнавання, і зображення, яке необхідно розпізнати. Алгоритм розпізнавання складається з наступних кроків.

Крок 1. Виберемо ієрархію детекторів  $M_c$ .

Крок 2. Для 1-того рівня ієрархії, починаючи з першого, і для кожного детектора відповідного рівня  $D^{(s)}$  Складемо вектор вихідних даних  $Z^{(s)}$ .

Крок 3. Якщо  $= 0$ , то заповнимо вектор вихідних даних наступним чином: зробимо розрахунок значення функції ідентифікації детектора  $d_I^{(s)}$ . Для кожної ділянки зображення:

$$I(x \dots x + w, y \dots y + h) \quad (2.1)$$

де  $h$  - відповідає розмірам локального ділянки зображення,

$I$  - ідентифікатор детектора,

$d^{(s)}$  - значення впевненості,

і зробимо конкатенацію вектора  $()$  з вихідними значеннями детектора для тих ділянок:

$$d_I^{(s)} = 1 \text{ і } d_p^{(s)} > \quad (2.2)$$

де  $j$  - обраний поріг впевненості.

Крок 4. Якщо  $\neq 0$ , вектор вихідних даних отримаємо як:

$$Z_{(s)} = (D_{(s)}(Z_{(s-1)}) \parallel \dots \parallel D_{(s)}(Z_{(s-1)})) \parallel (0 \parallel \dots \parallel 0_{N_I}) \quad (2.3)$$

де  $Z$  - максимальне число можливих детекторів на рівні.

Крок 5. Якщо для рівня активації всіх детекторів цього рівня негативні (дорівнюють нулю), то зображення не належить до класу  $s$ .

Крок 6. Інакше, інкрементуємо значення і повторимо алгоритм з кроку 2.

Крок 7. Якщо активація останнього рівня позитивна, то зображення належить до класу. В іншому випадку повторимо алгоритм з кроку 1.

Використовуючи той факт, що низькорівневі локальні ділянки зображення які були взяті з достатнім масштабом, як правило, представляють собою краю, кордони і кути, слід зазначити має місце тенденцію до зростання різноманітності серед еквіваріантних детекторів на більш високих рівнях моделі. Експерименти показують, що кількість детекторів першого рівня при навчанні не перевищує  $\approx 10$ , при цьому ієрархії, навчені на об'єктах різних категорій, здатні розділяти між собою частину детекторів першого рівня, демонструючи ефект, що нагадує трансферне навчання або преднавчання без вчителя.

Для прикладів зображень об'єктів, розглянутих в даному дисертаційному дослідженні, таких як людські особи, число рівнів ієрархії, достатню для ефективного розпізнавання, не перевищує двох. В загальному випадку евристичне правило вибору кількості рівнів для моделі залежить від кількості ступенів свободи об'єкта, що цікавить, і не може бути однозначно встановлено по зображенню в силу впливу шуму і обмеженого простору спостереження.

Блок-схема алгоритму розпізнавання приведена на рисунку 2.6.



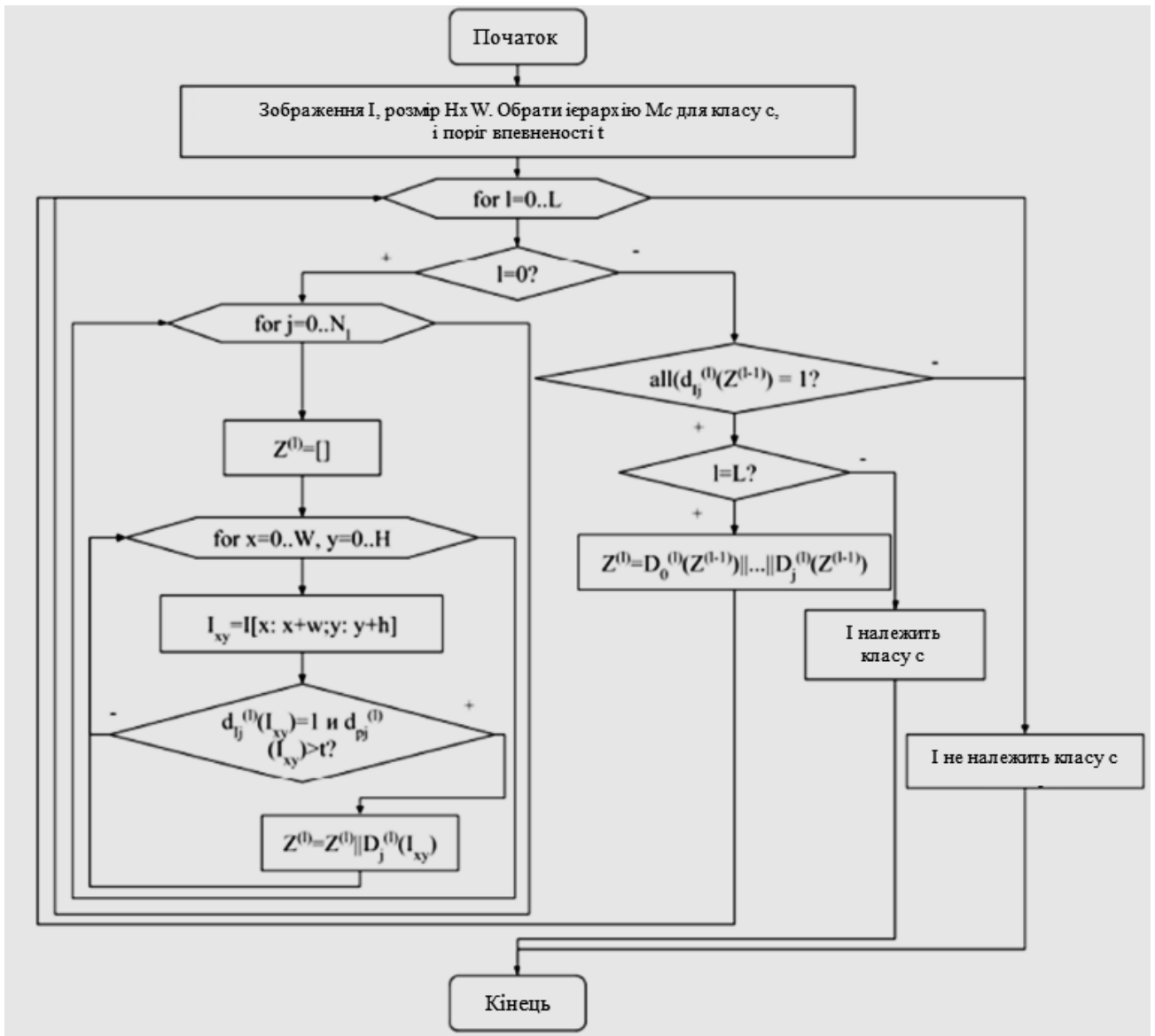


Рисунок 2.6 - Блок-схема алгоритму розпізнавання

Розроблений алгоритм розпізнавання відповідає вимогам поставленої завдання і дозволяє здійснювати розпізнавання зображених об'єктів з обчисленням параметрів інстанціювання в тривимірному просторі. Застосування ієрархічного підходу в реалізації алгоритму дозволяє здійснювати розпізнавання шляхом активацій обмеженого числа локальних ділянок зображення. Використання алгоритмом розпізнавання еківаріантних детекторів дозволяє алгоритму об'єднувати в одній категорії зображення, різні з інформатико-теоретичний точки зору (такі, як зображення людського обличчя в профіль і в фас), але відповідні при цьому однієї категорії приналежності.

Використання локальних детекторів дозволяє досягти стійкості до оклюзії (перекритті частини зображення стороннім фоном). Оскільки детектори моделі являють собою специфічні нейронні мережі, навчаються компактним репрезентаціям даних, окремий локальний детектор забезпечує стійкість до

ефектів шуму і розмиття в своїй області, дозволяючи алгоритму розпізнавання успішно справлятися з актуальними проблемами при розпізнаванні образів - нечіткістю вхідних даних і ефектами внутрікласовій варіативності.

#### 2.4 Використання вирішальною зворотного зв'язку при розпізнаванні

З огляду на присутність візуальних шумів на зображеннях і той факт, що найбільш чутливими до них є детектори першого рівня ієрархії, вірогідною представляється ситуація, при якій може бути допущено декілька помилок при активаціях детекторів першого рівня. Подібні помилки можуть зіграти значну роль при активації детекторів наступних рівнів, так як кожен високорівневий детектор приймає в якості вхідних параметрів конкатенацію вихідних параметрів детекторів попереднього рівня. У такій ситуації відсутність однієї або декількох активацій детекторів істотно спотворює вхідні дані і може стати причиною помилки активації високорівневого детектора, після чого ефект починає передаватися вгору по ієрархії і може служити причиною помилкової класифікації.

Так, попередні експерименти на зображеннях людських облич продемонстрували, що одного або двох помилково деактивованих детекторів досить, щоб порушити роботу детектора особи другого рівня, в ситуації, коли помилки відбуваються з високоінформативними локальними ознаками (такими, як ознаки очей).

Накопичення помилок в ієрархії при використанні алгоритму розпізнавання може бути істотно знижений за допомогою вирішальною зворотного зв'язку в ієрархії детекторів, що дозволяє зміцнювати або знизити впевненість активації детекторів попереднього рівня. Вирішальна зворотний зв'язок в ієрархії базується на наступних принципах.

Оскільки локальні ознаки об'єкта на зображенні відповідають елементам його просторової структури, передбачається що на зображенні вони повинні бути присутніми спільно.

У ситуації, коли частина локальних ознак відсутня (не розпізнає), високорівнева детектор виробляє передбачувану оцінку їх можливого розташування. Якщо оцінка збігається з розташуванням видимих локальних ознак, відсутні детектори попереднього активуються.

У ситуації, коли на зображенні присутні хибно-позитивні ознаки, аналогічним чином відбувається їх зіставлення з передбачуваної оцінкою вищестоящего детектора. Детектори, які не отримали підтвердження від верхнього рівня ієрархії, деактивуються.

Для реалізації зворотного зв'язку введемо додаткову функцію еквіваріантного детектора за аналогією з формулами 2.4 і 2.7 – функцію реконструкції. Розглянемо деякий детектор реконструкції. Розглянемо даний детектор  $D^{(s)}$  на рівні  $l$  ієрархії. Вхідними даними для цього детектора в загальному випадку є вектор  $Z^{(s-1)}$ , що представляє собою конкатенацію вихідних кортежів детекторів рівня  $l - 1$ , доповнених нульовими значеннями до

довжини  $4N_{s-1}$ . Розглянемо також вектор  $Z^{(s-1)}$ , що представляє собою результат впливу шуму на вектор  $Z^{(s-1)}$ , в якому деякі значення відсутні (дорівнюють нулю), що відповідає помилково-негативним видом помилок, або доповнені активації помилково-позитивних детекторів.

Тоді функція реконструкції  $d_R^{(s)}$  для  $i$ -того детектора  $l$ -того рівня проробляє реконструкцію оригінального вектора:

$$d(S) \underset{R_i}{(Zu(S-1))} = Z(S-1) \quad (2.4)$$

Слід зазначити, що функція реконструкції існує тільки для детекторів рівнів  $l > 0$ .

Реалізація функції реконструкції в загальному випадку залежить від апроксиматора, використовуюваного еквіваріантним детектором. Для розглянутих в даній роботі алгоритмів випадкового лісу і трансформує автоенкодера як показано в розділі 2.2 реалізація представляє собою, відповідно.

Для алгоритму випадкового лісу: додатковий елемент детектора, що представляє собою окремий випадковий ліс регресійних дерев, який навчається реконструювати вектор вхідних даних паралельно з іншими елементами детектора.

Для алгоритму розрідженого автоенкодера: останній (вихідний) шар нейронної мережі.

Підставою для введення функції реконструкції є той факт, що вектор вхідних даних  $Z^{(s-1)}$  є надлишковим за рахунок включення інформації про всіх локальних ознаках рівня  $l - 1$ . Так, для випадку з нульовим шумом очевидним є те, що знання параметрів інстанцірованія одного унікального ознаки досить для реконструкції інших ознак (в ситуації, коли тривимірний об'єкт не володіє додатковими ступенями свободи). На практиці використання навчаються апроксиматоров дозволяє здійснювати очищення вхідних даних від шуму за рахунок використання взаємної інформації про просторових поєднаннях локальних ознак.

Оскільки в загальному випадку число детекторів першого рівня для зображення не фіксоване і різниться в залежності від просторового положення об'єкта, для активації високорівневих детекторів відбираються детектори, що перевищили поріг впевненості при розпізнаванні. Розглянемо наступний приклад: в силу оклюзії (перекриття) частина ознак виявилася не розпізнає. Тоді трансформуючий автоенкодер детектора верхнього рівня здійснює реконструкцію вхідних даних, використовуючи тільки детектори, активовані з достатнім ступенем впевненості  $p > t$ , де  $t$  - емпірично обране порогове значення.

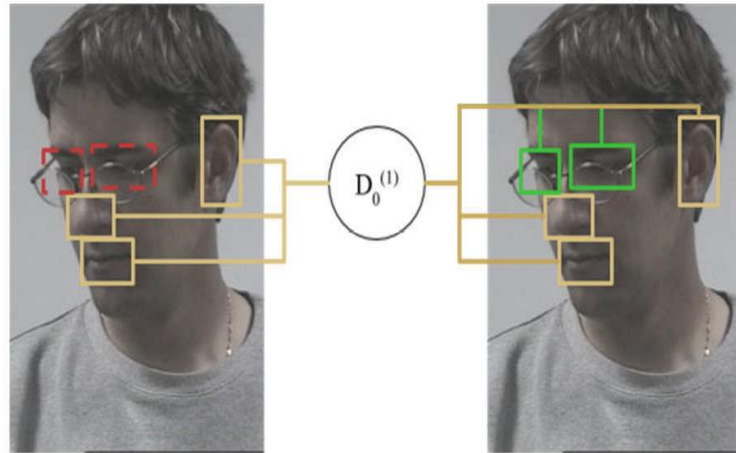


Рисунок 2.7 – Відновлення помилково негативні помилок детектором верхнього рівня за допомогою зворотного зв'язку

Нехай  $D^{(0)}, D^{(0)}, \dots, D^{(0)}$  – детектори, успішно активовані на першому рівні ієрархії. Кожному детектору відповідає скалярне значення впевненості активації  $p$ ; відберемо серед активованих детекторів таке підмножина  $0 \leq i \leq r$   $D^{(0)}, D^{(0)}, \dots, D^{(0)}$ , що для любого  $p, i \in 1..r, p > t$ , де  $t$  – 0-1 ріи порогове значення впевненості. Далі робиться припущення про те, що в силу впливу факторів візуальних шумів і варіативності об'єктів на першому рівні ієрархії може бути присутнім кілька помилково неактивованих детекторів  $D^{(0)}, D^{(0)}, \dots, D^{(0)}$ , тобто таких, для яких значення  $0 \leq h$   $h$  уверенности находится ниже порога положительной активации. Тогда входные данные высокоуровневого детектора  $D^{(1)}$  представлены следующим вектором:

$$Z^0 = (D^{(0)}(I_1) \parallel \dots \parallel D^{(0)}(I_r) \parallel 0 \parallel \dots \parallel 0) \quad (2.5)$$

де  $D^{(0)}$  – вектор вихідних значень  $i$ -того детектора,

$I_i$  –  $i$ -та область  $i$  зображення,

$0_j$  – нульовий вектор, відповідний пропущеному значенням  $j$ -того детектора, при цьому  $h$  вибирається таким чином, щоб сума  $h + r$  відповідала максимально можливого числа детекторів на першому рівні для навченої ієрархії,

$\parallel$  – операція конкатенації. Далі детектор другого рівня  $D^{(1)}$  здійснює операцію реконструкції.

Результати реконструкції є вектор:

$$d^{(1)} = (D'^{(0)} \parallel \dots \parallel D'^{(0)} \parallel Dd'^{(0)} \parallel \dots \parallel D'^{(0)}) \quad (2.6)$$

де  $D'^{(0)}$  – являє собою очікуване значення активації  $i$ -того детектора.

Для детекторів, чий активації не рівні нульового вектору, тобто для таких,

що  $i \leq r$ , очікується збіг реконструкції з актуальних значеннями активації, або  $D^{(0)} \approx D'^{(0)}$ . У випадку, якщо збіг не дотримано для  $i$  досить числа присутніх детекторів, робиться висновок про втрату критичної кількості інформації, необхідної трансформує автоенкодеру для реконструкції детекторів попереднього рівня - ситуація, яка може статися в разі, коли  $r < h$ , тобто велика частина детекторів ієрархії відсутній що може статися в разі недостатньо навченої ієрархії. Якщо присутні детектори реконструйовані коректно, то для реконструкцій  $D'^{(0)} \dots D'^{(0)}$ , відмінних від нуля, робиться висновок про існування  $r + 1$   $r + h$  помилково негативних помилок на попередньому рівні ієрархії, коректованих за допомогою зворотного зв'язку. Оскільки кожне вихідне значення детектора  $D'^{(0)}$  містить набір параметрів, що включають в себе відносні координати детектуємої ознаки і вектор параметрів інстанціювання, відновлені детектори можуть бути розміщені на попередньому рівні ієрархії з урахуванням їх позиції на зображенні.

Після корекції помилок за допомогою зворотного зв'язку і відновлення пропущених детекторів розраховується активація детектора другого рівня.

Аналогічна ситуація, але з використання негативного зворотного зв'язку, може спостерігатися в умовах хибно позитивних активацій детекторів які можуть виникнути в разі присутності незвичайного кількості візуально помітних структурних елементів на зображенні.

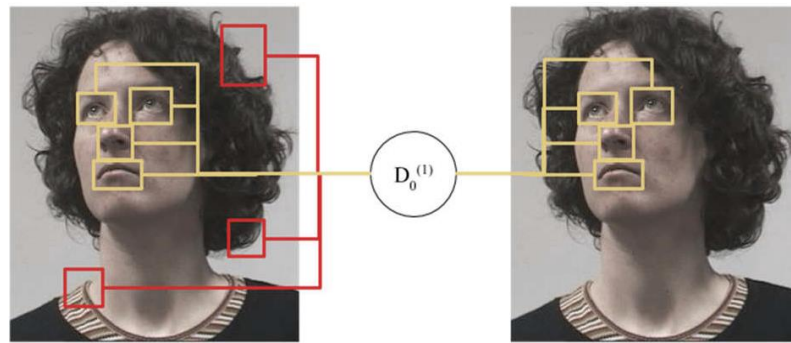


Рисунок 2.8 – Корекція хибно позитивних помилок детекторів верхнього рівня за допомогою зворотного зв'язку

У цьому випадку в якості вхідних значень детектора другого рівня присутній випадкові додаткові активації детекторів:

$$Z_{(s)} = (D_{(s)}(Z_{(s-1)}) \parallel \dots \parallel D_{(s)}(Z_{(s-1)})) \parallel (0 \parallel \dots \parallel 0_{N_1}) \quad (2.7)$$

При побудові реконструкції трансформує автоенкодера очікується, що реконструкція значень  $D^{(0)}$  ...  $D^{(0)}$  буде мати значення впевненості спрацьовування  $p_{r+1}$ , ...  $p_{r+z}$  що не перевищує порогу позитивної активації  $t$ . Таким чином, за допомогою негативного зворотного зв'язку зайві детектори придушуються.

В ході експериментів було встановлено, що високо рівневі детектори успішно класифікують і реконструюють вхідні дані з порогом шуму в 20%, що дозволяє коригувати помилки присутності детекторів при розпізнаванні в відповідному процентному відношенні.

## 2.5 Висновки з другого розділу

Проаналізувавши проблему розпізнавання зображень, можна зробити висновок що навчання розпізнавання образів є достатньо поширеною проблемою у сфері адаптації та розпізнавання зображень.

При вирішенні завдань управління методами розпізнавання образів замість терміна "зображення" застосовують термін "стан" та "ситуація". Популярні методи вирішення задач, коли треба уявити рішення у вигляді графічної моделі це:

- лінгвістичний метод;
- структурний.

При реалізації лінгвістичного підходу виникають два завдання:

- завдання побудови вихідного словника;
- завдання побудови правил опису з елементів заданого словника.

Перш ніж розпізнати зображення, необхідно його виділити або сегментувати. Сегментація як процедура пошуку однорідних, наприклад, по яскравості, кольору або текстурою, областей досить важка і до кінця не вирішена задача.

Для реалізації зазначеної проблеми використовуються два способи досліджень, дозволяють досягти бажаного результату. Перший передбачає декомпозицію проблеми, об'єкта на окремі фрагменти з визначенням не тільки фрагментів, але і зв'язків між ними. Другий пропонує семантичну сегментацію моделі зображення.

Для оцінки ефективності розробленої моделі та алгоритму розпізнавання пропонується використовувати статистичні методи оцінки ефективності розпізнавання, що дозволяють оцінити точність і повноту розпізнавання, визначити частку успішно розпізнаних зображень із загальної вибірки, а також оцінити точність визначення параметрів локалізації об'єкта.

Також були представлені і розглянуті гіпотеза компактності та методи оцінки ефективності розпізнавання зображень.

В ході проведення планування експерименту було визначено.

Для представленого методу і трьох альтернативних методів розпізнавання, було проведено комплексне тестування з використанням різних категорій тестових наборів даних, що включають в себе зображення під впливом шуму розмиття і оклюзії.

Запланована перевірка показала перевагу розробленого методу. Точність розпізнавання при використанні запропонованого методу зростає на 7-12%.

Розроблений метод більш стійкий до локального шуму: для зображень, підданих розмиття і оклюзії, падіння точності розпізнавання розробленого методу складає 3-5% проти 5-10% у альтернативних методів. Алгоритм розпізнавання використовує ієрархічний метод послідовної активації локальних детекторів, забезпечуючи стійкість алгоритму до різних категорій візуальних шумів, а також дозволяючи приймати рішення про розпізнавання на різних рівнях структури об'єкта. Алгоритм розпізнавання також проводить оцінку параметрів інстанціювання локальних ознак на кожному рівні ієрархії, дозволяючи не тільки класифікувати розглядається зображення з точки зору його належності до категорії розпізнавання, але і визначити параметри його розташування по відношенню до зорових сенсорів агента обробки інформації.

Точність розпізнавання розробленого методу при зміні орієнтації об'єкта на зображенні відчуває відхилення не більше ніж на 4-5%.

Розроблено алгоритм розпізнавання зображень з використанням представленої моделі репрезентації об'єктів.

### 3 ПРОВЕДЕННЯ ЕКСПЕРИМЕНТАЛЬНОГО ДОСЛІДЖЕННЯ

#### 3.1 Методи оцінки ефективності розпізнавання

Актуальним питанням при розпізнаванні зображень є оцінка ефективності роботи методу розпізнавання. Для отримання чисельного значення оцінки широко використовуються як загальні методи математичної статистики, так і специфічні показники, що застосовуються для оцінки алгоритмів машинного навчання.

Однією з найбільш простих метрик оцінки ефективності є процентна частка коректно розпізнаних зображень [6]. Коректним розпізнаванням вважається отримання на виході алгоритму класу, відповідного попередньо заданому класу. Для оцінки використовується вибірка, спроектована аналогічно навчальній вибірці, але містить зображення, до яких алгоритм не міг мати доступ в процесі навчання. Для цього, як правило, вихідна загальна вибірка поділяється на дві нерівні частини (Розмір тестової вибірки при цьому може відрізнятись, і складати 20-30% розміру загальної вибірки [5]). Тоді якщо функція  $c(x_i)$  являє собою функцію-індикатор коректності розпізнавання  $i$ -того зображення. Частка коректно розпізнаних зображень оцінюється наступним чином:

$$Q = \frac{\sum_{i=1}^n c(x_i)}{n} \quad (3.1)$$

де  $n$  - кількість зображень в тестовій вибірці,

1 - відповідає коректному розпізнаванню всіх зображень вибірки.

Таким чином, оцінка приймає значення на відрізку  $[0,1]$ .

Розглянутий показник є найбільш узагальненим і підходить для безлічі завдань розпізнавання з обмеженою кількістю класів. У завданнях, де кількість класів не фіксоване, і завдання розпізнавання являє собою задачу ідентифікації об'єкта певної категорії серед безлічі інших, потенційно необмежених категорій, замість неї застосовуються такі показники як точність і повнота оцінки. Їх використання дозволяє розрізнити помилково-позитивні (класифікатор прийняв позитивне рішення по зображенню, що не містить шуканого об'єкта) і помилково-негативні (класифікатор не пізнав об'єкт на зображенні, де він був присутній) помилки розпізнавання, або помилки першого і другого роду. Таким чином, точність оцінки в межах класу представляє собою частку зображень, 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 в тестовій вибірці. Інакше,

$$P = \frac{TP}{TP+FP}, R = \frac{TP}{TP+FN} \quad (3.2)$$



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

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

- кожен показник враховується індивідуально,
- для випадків, коли між показниками може спостерігатися спостерігається залежність, проводиться оцінка одного з показників при фіксації іншого (наприклад, оцінка точності при рівні повноти в 0.75);
- обидва показники можуть бути скомбіновані в один.

Серед прикладів агрегованих метрик для третьої методики зустрічаються такі показники як  $F$  - міра, коефіцієнт кореляції Меттьюз, регресивні коефіцієнти  $\Delta P$  і  $\Delta R$  [4]. Також, для деяких цілей розглядається зважене арифметичне середнє показника точності і значення, зворотного показником повноти, і навпаки. З точки зору теорії ймовірності ці показники можуть інтерпретуватися наступним чином: точність відповідає ймовірності того, що випадково обраний з безлічі позитивно упізнаних зображень дійсно розпізнано коректно, при цьому повнота є ймовірність того, що випадково обраний із загальної зображення буде коректно класифіковано алгоритмом (позитивно або негативно). Так, в залежності від програми завдання, до продуктивності методу розпізнавання можуть бути пред'явлені вимоги, що стосуються максимізації повноти (для випадків, коли певна кількість хибно-позитивних рішень допускається), так і збалансованого значення двох показників або  $F$  - міри, вираженої в такий спосіб:

$$F = 2 \frac{RP}{P+R} \quad (3.3)$$

де  $F$  - міри є гармонійне середнє значення показників точності і повноти [10].

Крім перерахованих показників, важливу роль в оцінці ефективності методів розпізнавання грають показники, що дозволяють оцінити метрики роботи алгоритмів, не пов'язаних безпосередньо з прийняттям класифікуючих рішень. Серед цих показників можна виділити розмір довірчого інтервалу для вибірки зображень, оцінка числа ітерацій і часу роботи алгоритму при навчанні, необхідних для досягнення збіжності, темп збіжності і т.д. [12].

Для оцінки розміру вибірки, що забезпечує статистично достовірні результати, використовується метод розрахунків довірчих інтервалів.

Для параметра  $\theta$  - числа зображень вибірки, що представляє собою випадкову величину  $X$  з рівнем довіри, таку, що  $X = (x_1, x_2, \dots, x_n)$  Довірчий 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:

$$\mathbb{P}(L \leq \theta \leq U) = p \quad (3.4)$$

де  $l$  - довірчого інтервалу,

$p$  - рівень довіри,

$U$  – довірча межа.

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

Досягнення збіжності при навчанні визначається як етап, коли

$\Delta j = j_i - j_{i-1} < a$ , тобто зміні функції ціни навчається моделі на черговому кроці оптимізації не перевищує деякого фіксованого мінімального значення, вибір якого може проводитися з урахуванням поставленого завдання.

Існує набір методик оптимізації, що дозволяють варіювати розміри цього порогового значення виходячи з темпів збіжності на конкретній ділянці області визначення функції, заданої навчається моделлю [7]. У деяких випадках, коли досягнення збіжності утруднено (наприклад, при оцінці метапараметрів моделі на етапі крос-валідації), число ітерацій навчання задається фіксованим значенням. Для сучасних алгоритмів оптимізації, таких як L-BFGS, minFunc і т.д., це значення, як правило, не перевищує двох тисяч ітерацій [9].

### 3.2 Проведення експерименту

У цьому розділі представлені експериментальні результати для розробленого в рамках дисертаційного дослідження методу розпізнавання зображень. Проведені експерименти діляться на три категорії.

Навчання трансформує автоенкодера окремо від моделі. Метою експериментів є визначення точності відновлення фрагментів зображення по компактним репрезентації автоенкодера, адекватність сформованого коду, його відповідність значень трансформацій. Крім цього, тестується ефективність роботи автоенкодера як функції-ідентифікатора. У завершальній частині проводиться зіставлення з іншими алгоритмами унарною класифікації та відновлення трансформації.

Розпізнавання зображень за допомогою повноцінної моделі. Об'єктами експерименту є показники помилки розпізнавання (в порівнянні з альтернативними алгоритмами), а також формуються моделлю просторові структури еквіваріантних детекторів.

Робота моделі на узагальнених відеофрагментів, що містять безліч об'єктів різних класів. Досліджується здатність моделі кластеризувати сцену, виявляючи окремі об'єкти і формуючи відповідні їм репрезентації.

Оскільки представлена модель вимагає для навчання обмежену вибірку з повно зв'язних відеофрагментів - тип даних, який не отримав поширення серед сучасних методів розпізнавання зображень - то навчання моделей для експерименту проводилося на даних, представлених автором. Навчальні вибірки включають в себе як відеофрагменти, отримані в результаті обробки згенерованих з використанням комп'ютерної графіки тривимірних моделей, так і фрагменти живої зйомки навколишнього світу, а також дані загальнодоступних камер спостереження.

Оцінка ефективності розпізнавання проводилася на базі відкритих баз даних LFW [40], Stanford Car Dataset [20], CVLAB Multi-View Car Dataset [24], Caltech Building Dataset [22]. Деякі моделі, що використовують альтернативні методи розпізнавання, надані бібліотекою Caffe [25]. Для тестування продуктивності трансформує автоенкодера використовувалася база даних HPID (Head Pose Image Database) [23], що містить маркіровані зображення людських облич в різних орієнтаціях щодо камери.

Експериментальна перевірка розробленого методу розпізнавання проведена відповідно до рекомендацій по навчанню і порівнянні моделей розпізнавання образів, прийнятими в області машинного навчання [54,5]. Для оцінки ефективності методів, які беруть участь в експерименті, використовувалися показники точності, повноти, і частки коректно розпізнання зображень.

Для оцінки вибірки застосовувався метод розрахунку довірчих інтервалів. Відповідно до стандартною практикою постановки експерименту з розпізнавання та рекомендаціями щодо вибору рівня довіри, значення рівня довіри вибрано рівним 0.95 [15]. Експериментальні дані продемонстрували, що для досягнення відповідного рівня досить вибірки близько 500 примірників, що для авторського методу відповідає короткому відеофрагменту, за часом не перевершує 15 с., Для якого частота зміни кадрів дорівнює 24 кадру в секунду, за умови використання методу пермутації.

Оскільки специфіка розробленого методу полягає у визначенні просторової локалізації об'єктів під різними кутами зору, то експериментальні вибірки для навчання і тестування містять набори відповідних зображень. Для оцінки здатності методу розпізнавання стійким чином обробляти об'єкти під впливом відповідних інваріантних перетворень, для кожного етапу експерименту результати обчислювалися за допомогою агрегації показників серії експериментальних перевірок. При цьому кожна серія складається з наступних елементів:

- навчання та тестування на загальній випадковою вибіркою;
- навчання і тестування на кластеризованій вибірці.

Кластеризація включає в себе розбивку вибірки по відеофрагментів, що містить зображення окремих об'єктів (наприклад, конкретних людських осіб), при цьому дані відеофрагментів навчання не перетинаються з тестовими.

Кількість кластеризації для кожного експерименту вибрано рівним чотирьом, відповідно до методики оцінки розпізнавання для варіативних об'єктів [11]. Таким чином, результат експерименту є агрегований показник по серії з п'яти стадій навчання і тестування з урахуванням розрахованої на базі складових показників стандартної помилки.

В рамках експерименту проводиться зіставлення результатів розробленого методу і альтернативних, що використовуються в даний час методів розпізнавання зображень. Вибрані альтернативи є відомі, широко застосовуються як у виробництві (метод Віюли - Джонса [14], нейронні мережі [13]), так і в академічних дослідженнях методи. При виборі альтернатив

використовувалися дані відкритих змагань по машинному навчання в області розпізнавання зображень, таких як ILSVRC [8]. Реалізації алгоритмів, що використовувалися в ході експерименту, представлені авторами відповідних методів, і отримані з відкритих джерел [11].

### 3.2.1 Обробка результатів роботи автоенкодера

В даному експерименті трансформує автоенкодер навчається на послідовності кадрів обертання людського обличчя. Виявлені на етапі оптичного трекінгу ділянки помітності відповідають елементам, які виділяються в складі людського обличчя і можуть бути згруповані за типами «очей», «ніс», «рід», «деталі зовнішнього контуру». На базі тестових відео фрагментів складена навчальна вибірка, що включає в себе фрагменти зображень кожного з локальних ознак.

Оскільки трансформуючий автоенкодер є мережею, що навчається репрезентаціям, для оцінки його ефективності використовувався показник регресійної точності (якості апроксимації зображень) - нормалізоване відхилення від абсолютного значення по модулю, розраховане наступним чином:

$$P_r(X, X') = 1 - \frac{X - X'}{X} \quad (3.5)$$

де  $X$  і  $X'$  відповідають відповідно до цільового зображення та апроксимації за допомогою репрезентаційної моделі [2]. Регресійна точність, таким чином, лежить в межах  $\{0,1\}$ . У наведених нижче таблицях, що містять експериментальні дані, цей показник представлений в процентному вираженні.

На базі отриманої вибірки, згрупованої попарно, проводиться навчання розрідженого трансформує автоенкодера. В експерименті брали участь чотири моделі автоенкодерів, зазначені в нижчеперелічених результатах як, і, і що відрізняються кількістю нейронів шару розпізнавання і породжує шару.

- А, 36 і 36 нейронів;
- В, 64 і 64 нейрона;
- С, 256 і 256 нейронів;
- D, 768 і 768 нейронів.

В рамках експерименту була виконана оцінка метопараметрів моделі з використанням методики кросвалідації. На рисунках 3.1 і 3.2 наведені графіки оптимізації функції ціни автоенкодера в залежності від варійованих метопараметрів розрідженості  $s$  і  $\beta$  (2.2, 2.3).

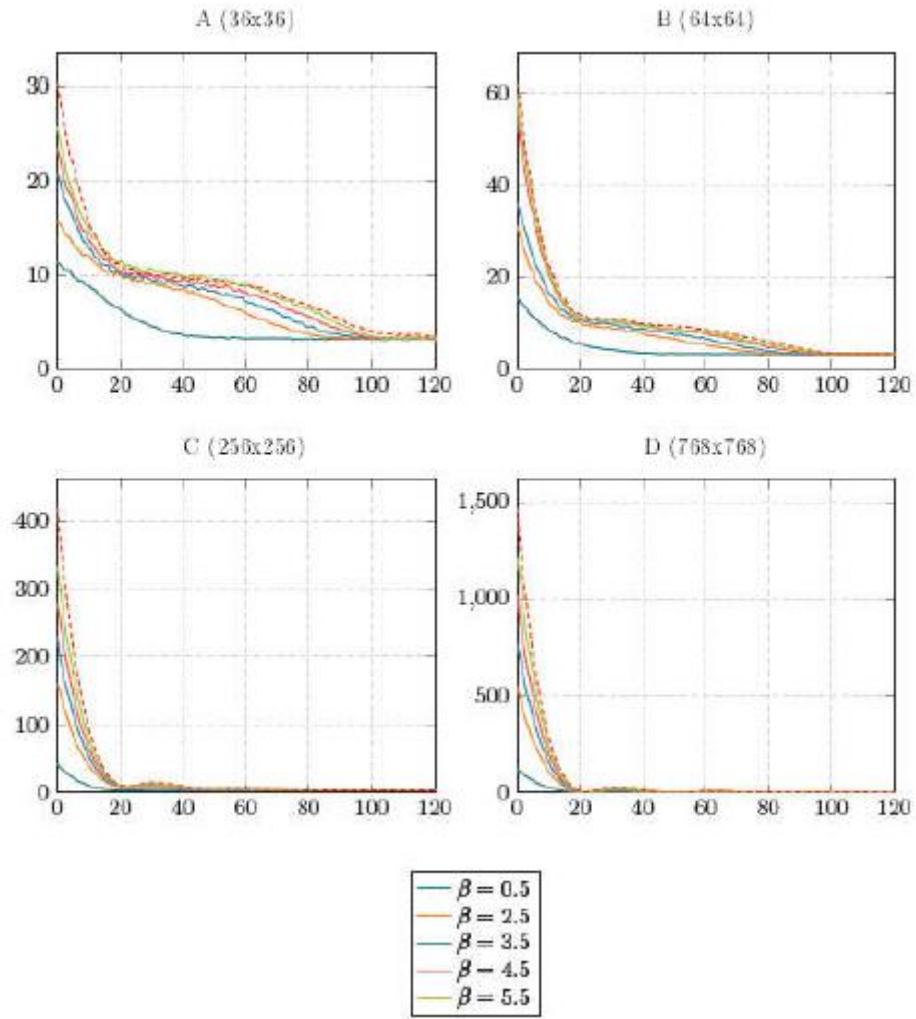


Рисунок 3.1 – Графіки навчання моделей в залежності від метапаметру  $\beta$

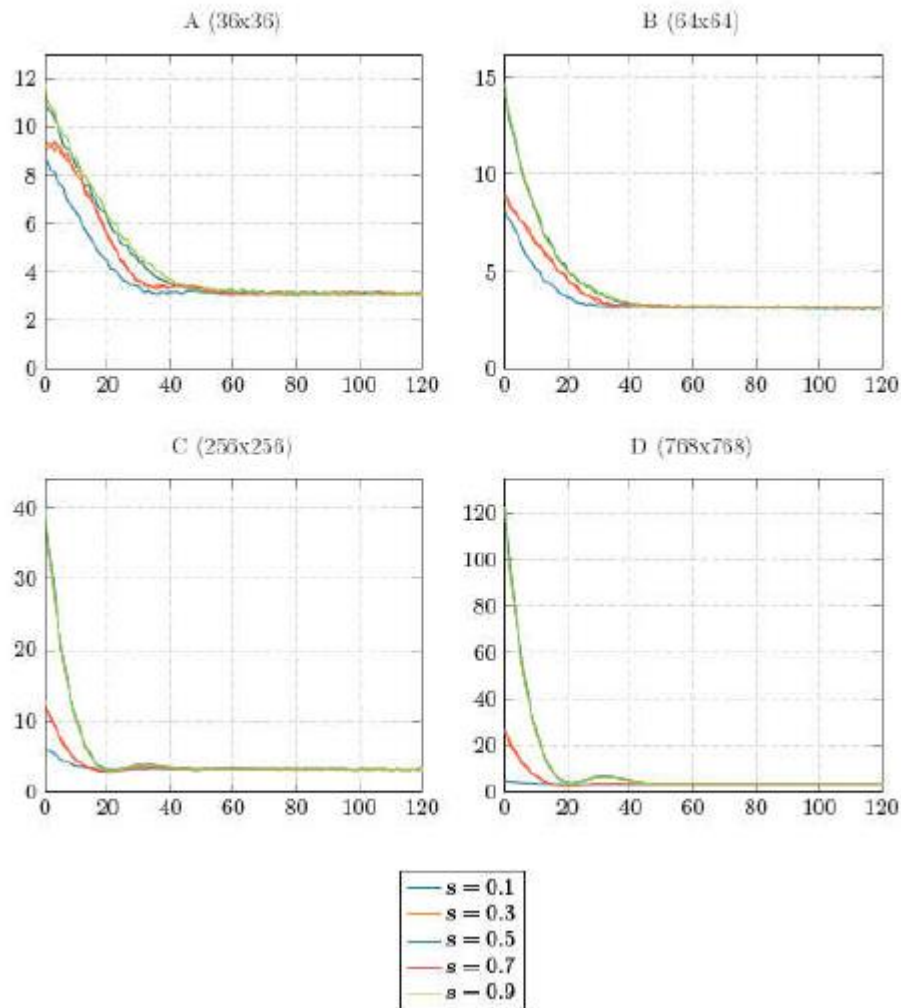


Рисунок 3.2 – Графіки навчання моделей в залежності від метапаметру  $s$

Вибір кількості нейронів проведено відповідно до методики навчання репрезентаційних нейронних мереж [9]. Трансформації представляють собою поворот камери по осях  $X$  і  $Z$ .

Результати реконструкції вхідних даних трансформуючим автоенкодером (по групах) наведені на рисунках. Кожен рисунок містить три рядки зображень - перший рядок відповідає вихідним зображенням, друга - результатами трансформації, третя - реконструкцій автоенкодера.

Візуальна схожість другий і третій рядки демонструє, що автоенкодер здатний відновлювати зображення на достатньому рівні для розпізнавання людиною. На рисунках 3.2-3.5 показані реконструкції трансформованих зображень.

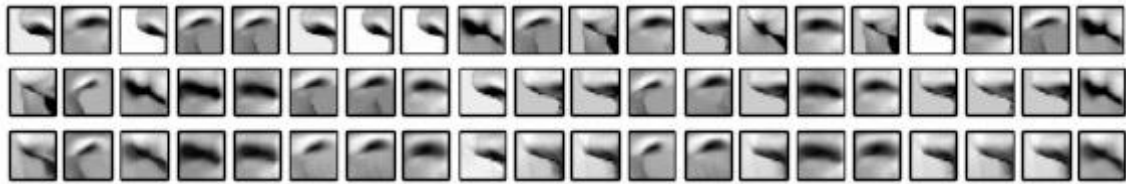


Рисунок 3.3 – Риси обличчя: ніс

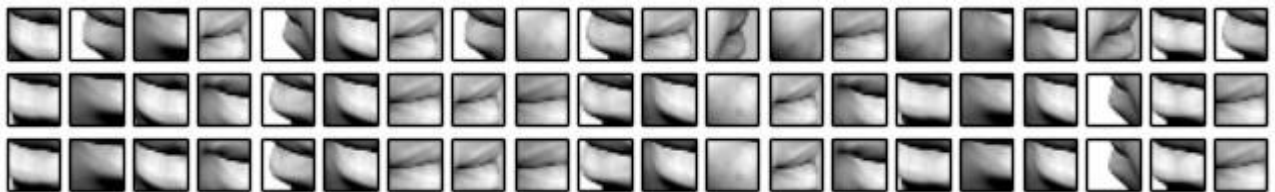


Рисунок 3.4 – Риси обличчя: рід

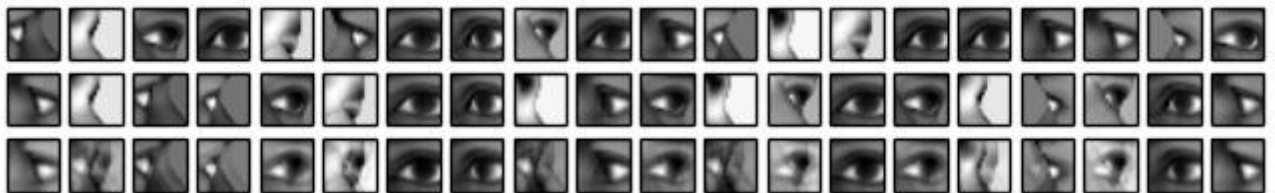


Рисунок 3.5 – Риси обличчя: око

Точність реконструкції оцінюється як середньоквадратичне відхилення цільового значення від реконструкції автоенкодера. У таблиці 3.1 наведені результати для різних категорій ділянок помітності, витягнутих з відеофрагментів (складання вибірки проводилося в автоматичному режимі, методом трекінгу; категорії промарковані назвами для спрощення і зручності аналізу):

Таблиця 3.1 - Показники точності реконструкцій автоенкодера, %

	A (36x36)	B (64x64)	C (256x256)	D (768x768)
Риси обличчя: очі	79 ± 4	95 ± 3	96 ± 3	96 ± 3
Риси обличчя: ніс	82 ± 3	92 ± 3	93 ± 3	94 ± 2
Риси обличчя: рід	84 ± 3	97 ± 4	97 ± 3	97 ± 3
Риси обличчя: вуха	72 ± 3	91 ± 3	93 ± 4	93 ± 3
Риси обличчя: деталі контуру	92 ± 4	96 ± 3	97 ± 4	98 ± 3



## Продовження таблиці 3.1

Автомобілі: колеса	72 ± 4	92 ± 4	94 ± 3	95 ± 4
Автомобілі: фари	74 ± 3	91 ± 3	92 ± 2	92 ± 2
Автомобілі: лобове скло	80 ± 3	94 ± 4	95 ± 2	95 ± 3
Автомобілі: заднє скло	82 ± 4	92 ± 2	93 ± 3	93 ± 3
Автомобілі: бампер	79 ± 4	90 ± 3	92 ± 4	93 ± 4
Автомобілі: деталі контуру	78 ± 2	89 ± 4	90 ± 3	91 ± 3
Будинки: вікна	73 ± 3	90 ± 4	91 ± 4	91 ± 3
Будинки: двері	71 ± 3	90 ± 4	91 ± 4	91 ± 4
Будинки: кути	77 ± 4	94 ± 2	95 ± 3	96 ± 4

З отриманих даних можна зробити висновок про оптимальність моделі - збільшення числа нейронів веде до незначного підвищення точності реконструкції, при цьому збільшуючи обчислювальну навантаження. Також більше число нейронів вимагає більшого числа ітерацій навчання для досягнення збіжності.

Отримані результати узгоджуються з результатами, отриманими іншими дослідниками при навчанні моделей автоенкодерів [16]. Слід зазначити відмінності в показниках точності для окремих категорій зображень - так, статистично більш прості елементи (кути, деталі контурів) піддаються реконструкції успішніше, ніж складні (окремі риси обличчя, наприклад, очі). Складність в даному випадку може бути виражена за допомогою агрегованої інформаційної ентропії зображень, що надходять на вхід автоенкодера.

Для використання трансформує автоенкодера в моделі представлення зображень необхідно відповідність критеріям еквіваріантного детектора. Таким чином, точність реконструкції автоенкодера є корисний, але допоміжний показник. Основним показником є здатність ефективно виявляти відповідні ділянки на зображенні. Для оцінки роботи трансформує автоенкодера в режимі еквіваріантного детектора використовувалася вибірка НРІД. Додатково проводився розрахунок показника узгодженості, визначеного як ймовірність того, що при отриманні позитивного рішення в тому кадрі детектор винесе позитивне рішення в  $j + 1$ -м кадрі, або  $P(D_I(I_{j+1}) = 1 | D_I(I_j) = 1)$ .

Таблиця 3.2: Показники точності, повноти і узгодженості трансформує автоенкодера на вибірці HPID, %

	P	R	C
Риси обличчя: очі	96 ± 3	95 ± 3	97 ± 3
Риси обличчя: ніс	98 ± 4	96 ± 4	96 ± 4
Риси обличчя: ріт	97 ± 4	96 ± 4	98 ± 4
Риси обличчя: вуха	98 ± 2	94 ± 3	97 ± 2
Риси обличчя: деталі контуру	99 ± 2	95 ± 3	96 ± 3

Результати активацій трансформує автоенкодера на зображеннях вибірки HPID наведені на малюнку в додатку А. Обмежувальними прямокутниками відзначені області активності детекторів, при цьому різні детектори виділені кольором.

Найбільш проблемними випадками для роботи еквіваріантного детектора є аномальні ситуації, не зустрілися в відеофрагменті - наприклад, оклюзія шуканого ділянки зображення сторонніми предметами (так, детектор демонструє зростання помилки до 10% при обробці зображень осіб, на яких присутні окуляри). Для зображень, семантично відповідних відеофрагменту, що використовувався для навчання, детектор демонструє гнучкість в умовах обмежене вибірки - так, 2-3 відеофрагментів, отриманих з використанням комп'ютерної графіки, виявилось досить для розпізнавання вибірки HPID з присутністю 15 осіб, що відрізняються підлогою, расою та особливостями міміки.

Отримані в результаті експерименту показники демонструють здатність трансформує автоенкодера виконувати роль детектора характерних ділянок зображення з дотриманням умови еквіваріантності. На рисунку 3.4 наведені дані візуалізації, що відповідають значенням ваг нейронів розпізнає і породжує шару автоенкодера.

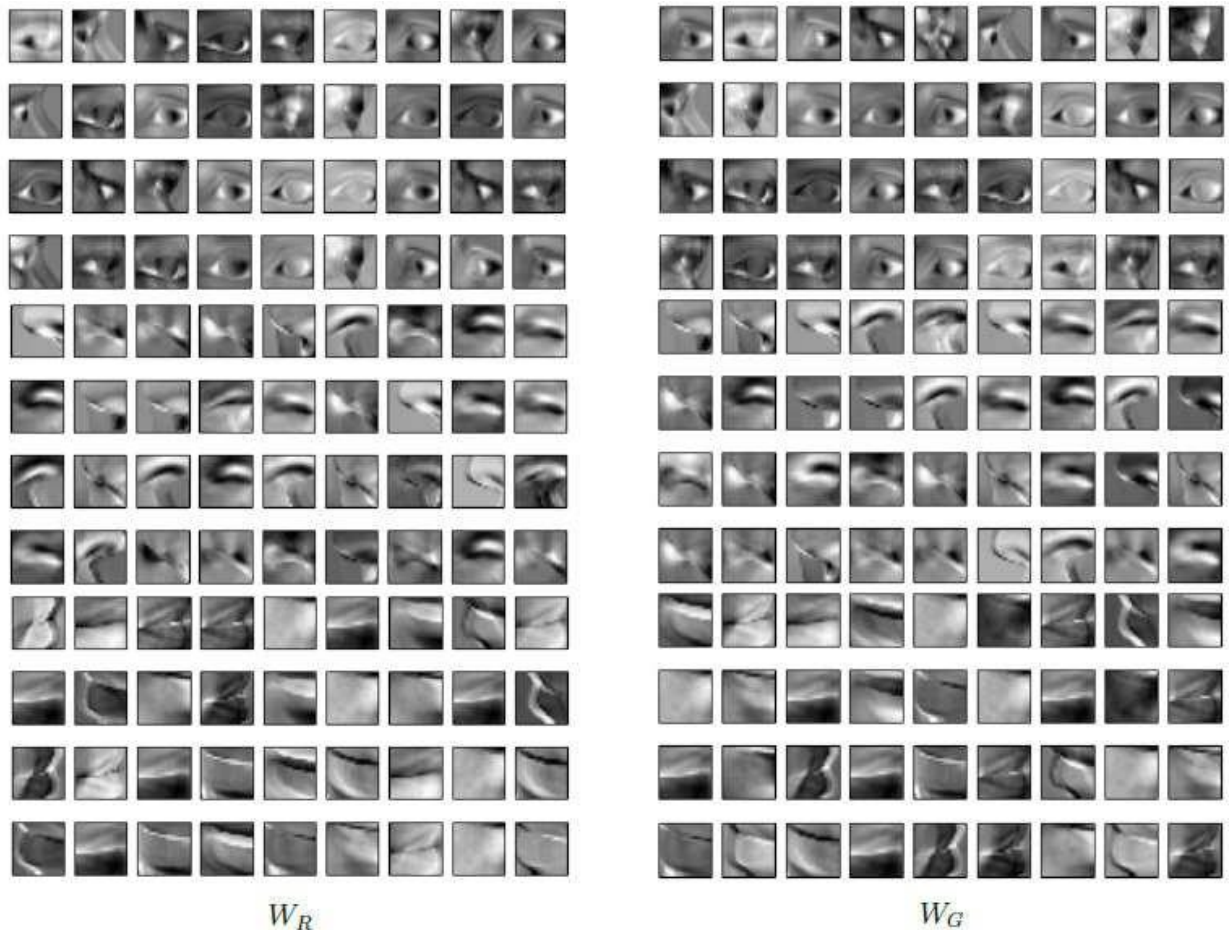


Рисунок 3.6 – Візуалізація ваг нейронів шару розпізнавання і породжує шару

Отримані ваги демонструють тенденцію до формування розрідженого коду. Серед ваг окремого автоенкодера практично не зустрічаються дублікати, при цьому характерна наявність декількох (3-5) «змішаних нейронів», візуально відповідних комбінації (накладення) різних ознак. За рахунок відповідних нейронів автоенкодер здатний реконструювати результати впливу трансформацій, відсутніх в його навчальній вибірці. Так, якщо вибірка містить два результату повороту об'єкта на кут і відповідно, то навчений автоенкодер здатний передбачити проміжний результат трансформації  $Y_1 < Y < Y_2$ .

Оскільки в якості одного з основних переваг даної моделі позиціонується еківаріантність - здатність до ідентифікації об'єктів на зображенні в різних орієнтаціях - то експериментальні вибірки згруповані таким чином: розглядається деяка вихідна позиція об'єкта з координатами обертання (кутів Ейлера)  $(0,0,0)$ , при цьому в групу, що характеризується значеннями  $(Y_{min}, Y_{max})$ , входять зображення об'єктів, які зазнали трансформації обертання ( $\cdot$ ), такий, що для будь-якого  $y \in (Y_i, Y_j, Y_k)$  вірно нерівність  $Y_{min} < Y < Y_{max}$ .

Тестування розпізнавання осіб проводилося шляхом зіставлення методом Віюлі-Джонса [12], класифікатора SVM в поєднанні з обчисленням гістограми орієнтованих градієнтів [3] і згорткової мережі, навченої на вибірці ImageNet. Навчені моделі були надані бібліотеками Caffe [33] і OpenCV [19].

Метод Віоли-Джонса має на увазі наявність окремої стадії навчання для кожної орієнтації. В рамках експерименту використовувалася преднавчана модель, яка продемонструвала зіставні результати для фронтальної і профільної орієнтації осіб, але вкрай низькі результати в проміжних станах. При використанні класичного методу розпізнавання (SVM) в поєднанні з методами зниження розмірності і підвищення інваріантності (гістограма орієнтованих градієнтів) для класифікатора характерно зниження точності в міру включення до вибірки зображень різних орієнтацій. Така поведінка пов'язана з тим, що модель в процесі навчання намагається виробити ознаки, однаково підходять для всіх зображень у вибірці, в результаті отримуючи рівномірно розподілені невисокі значення точності.

Серед порівнянних методів авторська модель поступається тільки глибокої згорткової мережі, здатної до навчання різних (що не змішуються) локальних ознак для різних орієнтацій об'єкта.

Деякі приклади результатів розпізнавання осіб наведені в рисунках. Слід зазначити, що на відміну від методу Віоли-Джонса, авторський метод дозволяє виділити комплексну структуру голови людини, не обмежуючись ділянкою, що містить очі і рот. Такий ефект є наслідком використання ознак деталей контуру обличчя.

Крім ідентифікації ділянок зображення, вторинної функцією еквіваріантного детектора є оцінка параметрів інстанціонування або оцінка позиції об'єкта. Для експериментальної перевірки точності цієї оцінки автоенкодер працює в обмеженому режимі, при цьому значення нейронів шару інстанціонування с вважаються вихідними значеннями мережі в цілому. Для зіставлення результатів використовувалися такі методи як класичний алгоритм POSIT [31], і навчаються з учителем регресорів - випадковий ліс [30], SVM. Результати представлені в таблиці 3.3.

Виявляється, таким об'єктивним властивістю є властивість компактності образів. Взаємне розташування точок в обраному просторі вже містить інформацію про те, як слід розділити безліч точок. Ця інформація і визначає то властивість розділяємості образів, що виявляється достатнім для самонавчання системи розпізнавання образів.

Більшість відомих алгоритмів самонавчання здатні виділяти тільки абстрактні образи, т. Е. Компактні безлічі в заданих просторах. Різниця між ними полягає, по-видимому, в формалізації поняття компактності. Однак це не знижує, а іноді і підвищує цінність алгоритмів самонавчання, так як часто самі образи заздалегідь ніким не визначені, а завдання полягає в тому, щоб визначити, які підмножини зображень в заданому просторі є образи. Хорошим прикладом такої постановки завдання є соціологічні дослідження, коли по набору питань виділяються групи людей. В такому розумінні завдання алгоритми самонавчання генерують заздалегідь не відому інформацію про існування в заданому просторі образів, про які раніше ніхто не мав жодного уявлення.

Таблиця 3.3 - Точність оцінки позиції за допомогою трансформує автоенкодера, %

	POSIT	Випадковий ліс	SVM	Авторський метод
Риси обличчя: очі	75 ± 3	83 ± 5	82 ± 3	89 ± 3
Риси обличчя: ніс	73 ± 3	86 ± 2	83 ± 3	88 ± 3
Риси обличчя: ріт	82 ± 4	84 ± 4	81 ± 4	90 ± 4
Риси обличчя: вуха	76 ± 4	88 ± 4	81 ± 3	89 ± 2
Риси обличчя: деталі контуру	70 ± 2	86 ± 3	83 ± 4	86 ± 4
Розмивання: очі	73 ± 3	81 ± 3	80 ± 3	89 ± 4
Розмивання: ніс	68 ± 3	82 ± 3	79 ± 4	86 ± 3
Розмивання: рот	75 ± 3	85 ± 3	81 ± 4	88 ± 4
Розмивання: вуха	79 ± 2	83 ± 2	79 ± 2	86 ± 3
Розмивання: деталі контуру	68 ± 4	84 ± 4	82 ± 3	88 ± 3
Оклюзія: очі	68 ± 4	80 ± 3	81 ± 3	90 ± 5
Оклюзія: ніс	72 ± 2	84 ± 5	82 ± 4	86 ± 2
Оклюзія: ріт	67 ± 5	84 ± 2	82 ± 4	87 ± 3
Оклюзія: вуха	68 ± 3	82 ± 4	83 ± 2	89 ± 4
Оклюзія: деталі контуру	69 ± 4	83 ± 4	80 ± 2	87 ± 4

Трансформуючий автоенкодер демонструє істотно більш якісну оцінку позиції об'єктів. Істотну перевагу над класичними методами комп'ютерного зору пояснюється тим, що алгоритм POSIT вимагає для оцінки позиції наявності маркерів, які можуть бути спроектовані на об'єкт за допомогою методів епіполлярної геометрії. Так, оцінка позиції людського обличчя, як правило, проводиться тільки для обмеженого простору трансформацій, в умовах, коли очі і рот знаходяться в полі зору камери. Для зображень, незначно

відхиляються від фронтальних, ефективність трансформує автоенкодера порівнянна з альтернативними методами.

### 3.2.2 Обробка результатів розпізнавання зображень

Було проведено дослідження ефективності найпростішої дворівневої моделі як видно із рисунка 3.1 на трьох категоріях зображень людських облич: звичайні фотографії осіб, фотографії зі штучним зашумлення за допомогою оклюзії і розмиття. Ефективність моделі розраховувалася за допомогою показників точності і повноти і порівнювалася з аналогічними показниками альтернативних методів розпізнавання. Як альтернатив для зіставлення вибиралися поширені алгоритми розпізнавання, що використовуються як академічному середовищі, так і для вирішення прикладних завдань.

Звернемо увагу, що в обох випадках - і при розпізнаванні посвідчень особи, і при розпізнаванні осіб - ми маємо справу з надзвичайно чутливими даними, які всім учасникам процесу краще б ретельно оберігати від чужих очей.

З боку клієнтів забезпечити безпеку своїх персональних даних досить просто: ключове правило - не завантажувати зображення власного документа і власного обличчя при реєстрації на невідомі сторінки і в невідомих сервісах. З боку замовників ця відповідальність полягає у виборі підрядників при розробці програм і їх інтеграції у власні інформаційні системи, забезпечення безпеки зберігання та передачі отриманих даних користувача.

До системних інтеграторів абсолютною вимогою є уникати побудови конструкції, при якій розпізнавання йде на сторонніх неконтрольованих серверах, а дані передаються в незахищеному вигляді по нестійким або слабо захищених каналах.

Те ж саме вимога пред'являється і до тестування: процедура розпізнавання документів не повинна допускати передачу на сторонні ресурси, які знаходяться поза межами контролю сторони, яка отримує призначені для користувача дані.

Найбезпечнішим рішенням сьогодні є така схема взаємодії, при якій розпізнавання документів проводиться абсолютно автономно на кінцевому пристрої користувача без збереження і передачі образів документів на сторонні сервіси. Така схема мінімізує ризики витоку і не дає шахраям вкрати образи документів, продати або використати їх для різних протиправних дій.

Таким чином виходить, що в разі грамотної побудови системи автоматизації взаємодії з клієнтами їх віддаленої верифікації, розпізнавання осіб і розпізнавання документів виступають не як антагоністи або заміщають один одного процеси, але як необхідне доповнення, завдяки якому функціонал додатків розкривається найбільш повно.

Оскільки в якості одного з основних переваг даної моделі позиціонується еквіваріантність - здатність до ідентифікації об'єктів на зображенні в різних орієнтаціях - то експериментальні вибірки згруповані таким чином:

розглядається деяка вихідна позиція об'єкта з координатами обертання (кутів Ейлера)  $(0,0,0)$ , при цьому в групу, що характеризується значеннями  $(Y_{min}, Y_{max})$ , входять зображення об'єктів, які зазнали трансформації обертання  $(\cdot, \cdot)$ , такий, що для будь-якого  $y \in (Y_i, Y_j, Y_k)$  вірно нерівність  $Y_{min} < Y < Y_{max}$ .

Тестування розпізнавання осіб проводилося шляхом зіставлення методом Віюлі-Джонса [12], класифікатора SVM в поєднанні з обчисленням гістограми орієнтованих градієнтів [32] і згорткової мережі, навченої на вибірці ImageNet. Навчені моделі були надані бібліотеками Caffe [11] і OpenCV [19].

Метод Віюлі-Джонса має на увазі наявність окремої стадії навчання для кожної орієнтації. В рамках експерименту використовувалася передобучення модель, яка продемонструвала зіставні результати для фронтальної і профільної орієнтації осіб, але вкрай низькі результати в проміжних станах. При використанні класичного методу розпізнавання (SVM) в поєднанні з методами зниження розмірності і підвищення інваріантності (гістограма орієнтованих градієнтів) для класифікатора характерно зниження точності в міру включення до вибірки зображень різних орієнтацій. Така поведінка пов'язана з тим, що модель в процесі навчання намагається виробити ознаки, однаково підходять для всіх зображень у вибірці, в результаті отримуючи рівномірно розподілені невисокі значення точності.

Звернемо увагу, що ми говоримо виключно про активну технології розпізнавання осіб, яка проводиться на підставі добровільного "пред'явлення" власного обличчя користувача в камеру скануючого пристрою. У нашому випадку - це кінцеве призначення для користувача пристрій - смартфон або веб-камера.

При використанні пасивного розпізнавання осіб, що застосовується на вулицях і в громадських місцях, користувач часто не знає, в який момент його обличчя було розпізнано і з якою метою це було зроблено. Так, безумовно, декларується, що пасивне розпізнавання осіб необхідно для забезпечення громадської безпеки та правопорядку. Але одночасно ця процедура в явному вигляді є обмеженням свободи.

Для нас цілком очевидно, що будь-яка технологія, яка розробляється сьогодні, в умовах прискореного технологічного розвитку, повинна застосовуватися виключно за згодою людини, служити для розвитку і підтримки його прав і свобод. Особливо, коли мова йде про збір, накопичення, систематизації і передачі персональних даних. Рішення про те, яким чином здійснювати свою ідентифікацію, повинен приймати користувач, причому маючи можливість усвідомленого вибору і застосування альтернативних методів.

Активне розпізнавання особи, як і розпізнавання документів, є суто добровільними актами, які в умовах посилення вимог безпеки до віддаленої ідентифікації та верифікації, є кроком до реалізації прав і свобод. Зверніть увагу, що полегшуючи дистанційний доступ до своїх послуг в умовах ізоляції, компанії роблять його в прямому сенсі безпечніше, демонструючи турботу про

клієнта, його часу, здоров'я і комфорт. При цьому для кінцевого користувача зберігається свобода дії і право вибору.

Серед порівнянних методів авторська модель поступається тільки глибокої згорткової мережі, здатної до навчання різних (що не змішуються) локальних ознак для різних орієнтацій об'єкта.

Деякі приклади результатів розпізнавання осіб наведені в рисунках. Слід зазначити, що на відміну від методу Віоли-Джонса, авторський метод дозволяє виділити комплексну структуру голови людини, не обмежуючись ділянкою, що містить очі і рот. Такий ефект є наслідком використання ознак деталей контуру обличчя.

Порівняння результатів розпізнавання зображень осіб, підданих впливу шуму розмиття проводилося за допомогою методів SVM + HOG, глибокої згорткової мережі і алгоритму випадкового лісу [33]. Вибірка проводилася на базі CVLAB Dataset і включала в себе об'єкти, подані з різних кутів. Шум розмиття є розмиття по Гауса зі значенням  $a = 0.5 \dots 2.5$ .

Таблиця 3.5 - Показники точності розпізнавання розмитих осіб для груп різних орієнтацій, %

	Випадковий ліс	SVM+HOG	CaffeNet	Авторський метод
(0°, 15°)	75 ± 4	72 ± 4	85 ± 2	91 ± 4
(15°, 30°)	73 ± 4	74 ± 3	81 ± 3	88 ± 3
(30°, 45°)	75 ± 2	76 ± 4	80 ± 4	88 ± 3
(45°, 60°)	72 ± 3	74 ± 5	82 ± 4	89 ± 4
(60°, 90°)	71 ± 3	73 ± 4	83 ± 4	91 ± 2
(90°, 120°)	73 ± 3	72 ± 4	83 ± 5	87 ± 4

При впливі ефектів розмиття дисперсія результатів по відношенню до орієнтації об'єкта знижується. Експериментальні показники узгоджуються з результатами по розпізнаванню «чистих» осіб.

Як завершальній групи експерименту використовувалися зображення людських облич під впливом шуму оклюзії (часткового перекриття).

Шум оклюзії згенерований за допомогою випадкового розміщення на зображеннях осіб непрозорих геометричних фігур. При генерації шуму оклюзії параметри розміщення фігур підібрані таким чином, щоб залишати відкритою мінімум дві третини площі зображення.



Таблиця 3.6 - Показники точності розпізнавання розмитих осіб для груп різних орієнтацій, %

	Випадковий ліс	SVM+HOG	CaffeNet	Авторський метод
(0°, 15°)	72 ± 4	70 ± 2	84 ± 3	90 ± 5
(15°, 30°)	69 ± 3	66 ± 4	82 ± 3	89 ± 4
(30°, 45°)	70 ± 3	68 ± 4	83 ± 2	88 ± 3
(45°, 60°)	71 ± 4	70 ± 4	81 ± 4	89 ± 4
(60°, 90°)	70 ± 2	67 ± 3	84 ± 5	88 ± 3
(90°, 120°)	72 ± 3	67 ± 3	85 ± 4	90 ± 3

Як демонструють результати, оклюзія є істотною перешкодою для алгоритмів розпізнавання, що використовують компактні цілісні уявлення, таких як випадковий ліс і SVM + HOG - для деяких груп орієнтацій спостерігається падіння точності розпізнавання до 10%. Методи, що використовують локальні ознаки, представлені мережею CaffeNet і авторським методом, менш чутливі до часткового перекриття локальних областей зображення. У таких ситуаціях продуктивність представленого методу наближається до показників основного конкурента - згорткової мережі.

### 3.2.3 Аналіз результатів

В даному експерименті перевірялася здатність алгоритму за допомогою багаторівневої моделі аналізувати сцени, що містять безліч об'єктів, і автономно (без наявності вчителя) класифікувати їх між собою. Як вибірки використовувалися відкриті дані, що містять відеозаписи камер спостереження за дорожнім трафіком. Оцінка продуктивності алгоритму проводилася в такий спосіб: експериментатором підраховувалася кількість об'єктів в кадрі  $Q$ , аналогічним чином вручну вироблялося розбиття по групах  $G$ , потім ці значення зіставлялися зі значеннями  $Q'$  і  $G'$ , знайденими алгоритмом.

Таблиця 3.7 – Оцінка результатів алгоритму композитних сцен

	Об'єктів виявлено, % від фактичного значення	Груп об'єктів виявлено, % від фактичного значення
Дорожній трафік 1	70	100
Дорожній трафік 2	72.5	100
Дорожній трафік 3	75	75
Дорожній трафік 4	74	75

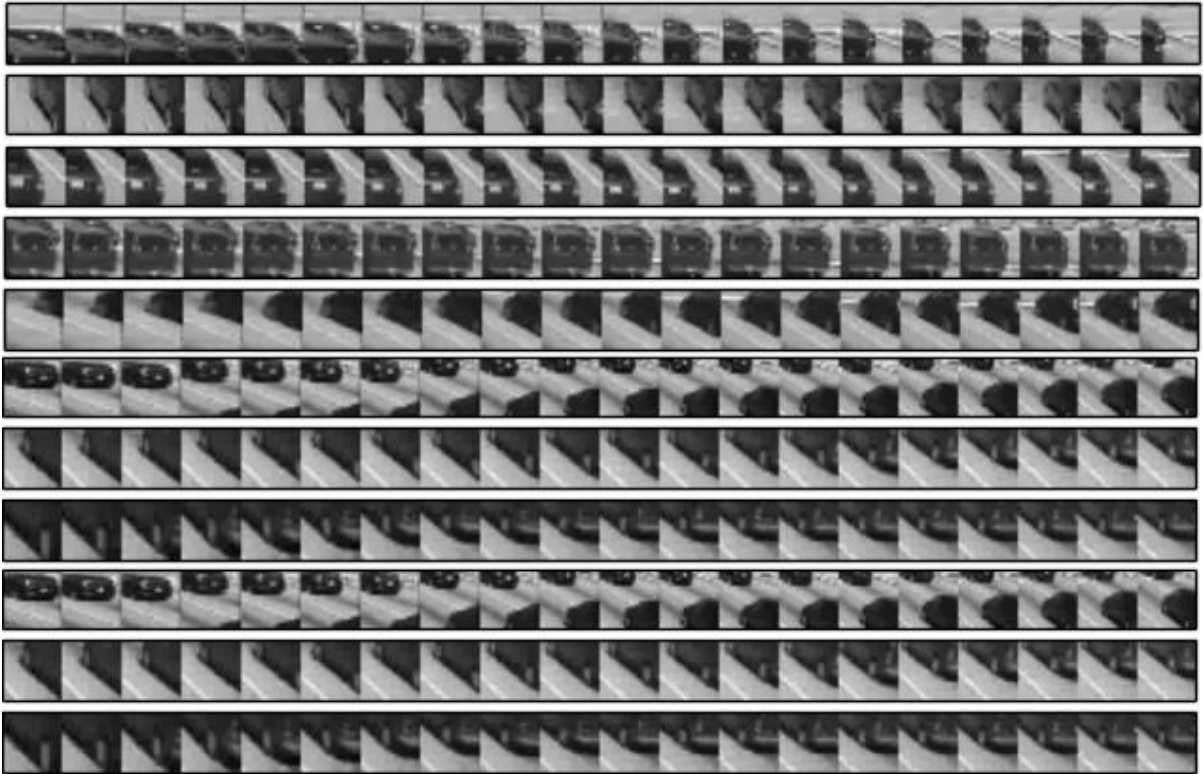


Рисунок 3.8 - Послідовності ділянок локальних фрагментів, виявлених при аналізі даних дорожніх камер

Слід зазначити, що на відміну від ситуації з контрольованим трекінгом і рухом камери навколо об'єкта, алгоритм в режимі аналізу композитних сцен зі сторонніх відеофрагментів позбавлений інформації про фактичний рух об'єкта навколо камери. Так, алгоритм не здатний без додаткової інформації визначити, що рух віддаленого від камери автомобіля являє собою переміщення по площині, не паралельній площині сенсора камери. У загальному випадку в такій ситуації алгоритм знаходиться в умовах невизначеності і здатний сформулювати тільки обмежене уявлення об'єкта за допомогою моделі еквіваріантних детекторів.

Існує можливість, використовуючи методи проективної геометрії, ввести деякі припущення в розрахунок оптичного трекінгу - так, для розглянутих експериментальних випадків стадія оптичного трекінгу була доповнена умовою жорсткості (фіксованою форми) об'єктів в кадрі. При цьому трекер інтерпретував зменшення розмірів об'єкта на зображенні як видалення від камери, і на підставі відносного зміни площі об'єкта  $\Delta S_r$  і показників оптичного потоку обчислював напрямок його руху.

### 3.3 Висновки до третього розділу

В даному розділі проведено експериментальне дослідження та отримані результати та дані до кожного методу та сценарію, на основі яких були створені порівняльні таблиці, графіки та фотографії.

За результатами проведених експериментів можна зробити наступні висновки.

Для оцінки ефективності розробленої системи розпізнавання зображень було проведено комп'ютерний експеримент з використанням різних категорій тестових наборів даних, що включають в себе зображення під впливом шуму розмиття і оклюзії, в ході якого було проведено порівняльна оцінка розробленої системи з альтернативними відомими методами розпізнавання.

Було проведено оцінку і аналіз результатів.

Експериментальна перевірка показала перевагу розробленої системи. Точність розпізнавання при використанні запропонованого методу зростає на 3-5%.

Розроблена система розпізнавання більш стійка до локального шуму: для зображень, підданих розмиття і оклюзії, падіння точності розпізнавання розробленої системи становить 3-6% проти 5-10% у альтернативних відомих методів.

Точність розпізнавання з використанням розробленої системи при зміні орієнтації об'єкта на зображенні відчуває відхилення не більше ніж на 4-5%.

## ВИСНОВКИ

В ході теоретичних і експериментальних досліджень, виконаних в дипломній роботі, були отримані наступні результати.

В роботі виконано аналіз сучасних методів розпізнавання зображень, створена класифікація типів існуючих підходів до розпізнавання. Визначено проблема розпізнавання тривимірних об'єктів під дією інваріантних просторових перетворень.

Проведено дослідження проаналізованих методів, отримані залежності швидкості навчання і точності розпізнавання від категорій розпізнаються зображень і метапараметрів поділи. Показана ефективність розробленого методу в зіставленні з альтернативними, застосовуються в сучасних умовах методами розпізнавання зображень, при вирішенні задачі розпізнавання зображення з проведенням просторової локалізації.

Результати експериментальних досліджень, наведені в роботі, становлять практичний інтерес при проектуванні систем управління та обробки інформації в області комп'ютерного зору і розпізнавання зображень, для тих завдань, де існує необхідність визначення просторових параметрів зображених об'єктів.

## ПЕРЕЛІК ПОСИЛАНЬ

1. Ben Krose, Valter van de Smagt. Introduction to neural networks. М.:2011 – 374 с.
2. Р. Каллан, Введение в нейронные сети. 2001, (третье издание). [Текст] — С. 159. — ISBN 5-7015-0361-X.
3. Саймон Хайкин, Нейронные сети полный курс. М.:Техносфера, 2012. - 448 с.
4. Gupta Jin Homma, Statical and Dynamical neural networks. М.:Техносфера, 2016. - 117 с.
5. Дж.Ту, Р.Гонсалес. Принципы распознавания образов. М., Мир, 1978.
6. Гонсалес Р., Вудс Р. Цифровая обработка изображений. М.:Техносфера, 2005. - 1072 с.
7. Каплан Р. Головні концепції нейронних мереж - The Essence of Neural Networks First Edition. — (перше видання). — «Вильямс», 2001. — С. 288. — ISBN 5-8459-0210-X.
8. S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” arXiv preprint arXiv:1502.03167, Feb. 2015. [Online]. Available: <https://arxiv.org/abs/1502.03167>
9. A. G. Howard, “Some improvements on deep convolutional neural network based image classification,” arXiv preprint arXiv:1312.5402, Dec. 2013. [Online]. Available: <https://arxiv.org/abs/1312.5402>.
10. S. J. Nowlan and G. E. Hinton, “Simplifying neural networks by soft weight-sharing,” Neural computation, vol. 4, no. 4, pp. 473–493, 1992. [Online]. Available: <https://www.cs.toronto.edu/~hinton/absps/sunspots.pdf>
11. Кудрявцев Л. Д. Математичний аналіз для нейронних мереж. — 3-е видання. — М.: Высшая школа, 2008.
12. Т. Xiao, J. Zhang et al., “Error-driven incremental learning in deep convolutional neural network for large-scale image classification,” in International Conference on Multimedia, no. 22. ACM, 2014, pp. 177–186.

13. J. Ortigosa-Hernández, I. Inza, and J. A. Lozano, “Towards competitive classifiers for unbalanced classification problems: A study on the performance scores,” arXiv preprint arXiv:1608.08984, Aug. 2016. [Online]. Available: <https://arxiv.org/abs/1608.08984>
14. F. Chollet, “Keras,” <https://github.com/fchollet/keras>, 2015
15. M. Abadi, A. Agarwal et al., “Tensorflow: Large-scale machine learning on heterogeneous distributed systems,” arXiv preprint arXiv:1603.04467, Mar. 2016. [Online]. Available: <https://arxiv.org/abs/1603.04467>
16. G. Huang, Z. Liu, and K. Q. Weinberger, “Densely connected convolutional networks,” arXiv preprint arXiv:1608.06993, Aug. 2016. [Online]. Available: <https://arxiv.org/abs/1608.06993v1>
17. P. Sermanet and Y. LeCun, “Traffic sign recognition with multiscale convolutional networks,” in International Joint Conference on Neural Networks (IJCNN), Jul. 2011, pp. 2809–2813. [Online]. Available: <http://ieeexplore.ieee.org/document/6033589/>
18. M. Thoma, “The HASYv2 dataset,” arXiv preprint arXiv:1701.08380, Jan. 2017. [Online]. Available: <https://arxiv.org/abs/1701.08380>.
19. S. Chetlur, C. Woolley et al., “cuDNN: Efficient primitives for deep learning,” arXiv preprint arXiv:1410.0759, Oct. 2014. [Online]. Available: <https://arxiv.org/abs/1410.0759>
20. N. McLaughlin, J. M. D. Rincon, and P. Miller, “Dataaugmentation for reducing dataset bias in person re-identification,” in International Conference on Advanced Video and Signal Based Surveillance (AVSS), no. 12, Aug. 2015, pp. 1–6. [Online]. Available: <http://ieeexplore.ieee.org/abstract/document/7301739/>.
21. Тарасенко-Клятченко О.В., Буц В.В. стаття «Організація багатоступового методу навчання згорткової нейронної мережі». Журнал «Інтернаука» випуск 6 (березень 2018). [Інтернет-ресурс].
22. “Прогнозування в умовах ринку”, (третє видання). [Текст] / Л.П. Владимірова. М.:2012.

23. Grady, L. Random walks for image segmentation / L. Grady // Pattern Analysis and Machine Intelligence, IEEE Transactions on. — 2006. — no. 28.11. — Pp. 1768–1783.
24. Дослідження по технологічному передбаченню. [Інтернет-ресурс] URL: <http://ictt.by>.
25. Кузьмин М. Основи Elasticsearch. [Інтернет-ресурс] / Інформаційна система habrahabr.ru – Режим доступу: <https://habrahabr.ru/post/280488/>.
26. Білуха М. Т. Методологія наукових досліджень: Підручник. — К.: АБУ, 2004. — 514 с.
27. Hinton, G.E. Reducing the dimensionality of data with neural networks / G.E. Hinton, R.R. Salakhutdinov // Science. — 2006. — no. 313(5786). — Pp. 504–507.
28. Hinton, G.E. Transforming auto-encoders / G.E. Hinton, A. Krizhevsky, S. D. Wang // Artificial Neural Networks and Machine Learning–ICANN 2011. — 2014. — Pp. 44–51.
29. ImageNet Large Scale Visual Recognition Challenge [Електронний ресурс] / Olga Russakovsky, Jia Deng, Hao Su et al. — arXiv, 2010. — <http://www.image-net.org/challenges/LSVRC>.
30. Rublee, E. ORB: an efficient alternative to SIFT or SURF / E. Rublee // Computer Vision (ICCV), 2011 IEEE International Conference. — 2011. — Pp. 2564–2571.
31. Fei-Fei, L. Combining randomization and discrimination for fine-grained image categorization / L. Fei-Fei, B. Yao, A. Khosla // Computer Vision and Pattern Recognition. — 2011. — Pp. 1577–1584.
32. Fei-Fei, L. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories / L. Fei-Fei, R. Fergus, P. Perona // Computer Vision and Image Understanding. — 2007. — no. 106(1). — Pp. 59–70.
33. Fei-Fei, L. Learning generative visual models from few training examples:

An incremental bayesian approach tested on 101 object categories / L. Fei-Fei, R. Fergus, P. Perona // *Computer Vision and Image Understanding*. — 2007. — no. 106(1). — Pp. 59–70.