

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА**

Факультет комп'ютерних наук та кібернетики
Кафедра теоретичної кібернетики

Кваліфікаційна робота
на здобуття ступеня бакалавра
за спеціальністю 122 Комп'ютерні науки

на тему:

**ВИЗНАЧЕННЯ ТА РОЗПІЗНАВАННЯ ЖЕСТІВ ДАКТИЛЬНОЇ
АБЕТКИ**

Виконав студент 4 курсу
Петренко Юрій Анатолійович _____

Науковий керівник:
доктор фізико-математичних наук
Крак Юрій Васильович _____

Засвідчую, що в цій роботі немає
запозичень з праць інших авторів без
відповідних посилань.

Студент _____

Роботу розглянуто й допущено
до захисту на засіданні кафедри
теоретичної кібернетики

«___» ___ 201_ р.,

протокол № _____

Завідувач кафедри

Ю.В. Крак _____

РЕФЕРАТ

Обсяг роботи 44 сторінки, 19 рисунків, 18 використаних джерел.
НЕЙРОМЕРЕЖА, ВИЗНАЧЕННЯ ЖЕСТІВ, НАВЧАННЯ НЕЙРОМЕРЕЖІ,
ДАКТИЛЬНА АБЕТКА, АРХІТЕКТУРА НЕЙРОМЕРЕЖІ.

Об'єкт розроблення програмного засобу – процес створення програми визначення жесту на фото.

Мета роботи полягає в:

1. Огляді загальних підходів теорії ідентифікації до розпізнавання елементів дактильної абетки.
2. Аналізі існуючих підходів до задач ідентифікації та розпізнавання елементів дактильної мови.
3. Аналізі та виборі найкращої архітектури нейронної мережі з підвищеною ефективністю розпізнавання та ідентифікації.
4. Розробці програми для визначення і розпізнавання жестів дактильної абетки на фото.

Інструменти, що використовуються:

- Середовище розробки – Google Colab.
- Мова Python.
- Модулі numpy, Tensorflow, keras.

Результати роботи: Обрано метод та джерело навчання нейромережі (готовий датасет жестів). Здійснено процес навчання нейромережі. Встановлено необхідні модулі та здійснена реалізація програми мовою Python. Отримані результати повністю відповідають темі дипломної роботи.

ЗМІСТ

РЕФЕРАТ	2
ЗМІСТ	3
СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАЧЕННЯ	5
ВСТУП	6
Оцінка сучасного стану об'єкта дослідження або розробки.	6
Актуальність роботи та підстави для її виконання.	6
Мета й завдання роботи.	8
Об'єкт і методи дослідження або розроблення.	8
Можливі сфери застосування.	8
РОЗДІЛ 1. АНАЛІЗ СИСТЕМ РОЗПІЗНАВАННЯ ЖЕСТІВ	10
1.1. Загальна інформація про жестову мову	10
1.2. Структура жестової мови	11
1.3. Структура систем розпізнавання жестів	11
РОЗДІЛ 2. АЛГОРИТМИ ТА ПІДХОДИ ДО РОЗПІЗНАВАННЯ ЖЕСТІВ ДАКТИЛЬНОЇ АБЕТКИ	13
2.1. Використання відеокамер	14
2.2. Використання рукавичок	16
2.3. Використання штучних нейронних мереж	17
РОЗДІЛ 3 НЕЙРОМЕРЕЖІ. ВИДИ ТА ЗАСТОСУВАННЯ	19
3.1. Історія виникнення нейромережі	19
3.2. Застосування нейромереж	20
3.3. Види нейромереж	21

	4
3.3.1. Нейромережа з прямим зв'язком	21
3.3.2. Згорткова нейромережа	21
3.3.3. Рекурентні нейромережі	23
Висновок	24
РОЗДІЛ 4. АНАЛІЗ АРХІТЕКТУР ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ	25
4.1.	264.2.
	264.3.
	284.4.
28РОЗДІЛ 5. НАВЧАННЯ НЕЙРОМЕРЕЖІ	
	30
5.1. Навчання з вчителем	30
5.2. Навчання без вчителя	31
5.3 Змішане навчання	31
5.4. Дані для навчання	32
5.4.1 Датасети	32
5.4.2 Веб-платформи зі статистикою	33
5.4.3 Парсинг сторонніх сайтів	33
РОЗДІЛ 6. РЕАЛІЗАЦІЯ ОБ'ЄКТА ДОСЛІДЖЕННЯ	35
6.1. Нейронна мережа	35
6.2. Необхідні засоби	35
6.3. Програмна реалізація	36
ВИСНОВКИ	40
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ	42

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАЧЕННЯ

API (Application Programming Interface) – інтерфейс програмування, інтерфейс створення додатків. Якщо говорити більш зрозумілою мовою, то API – це готовий код для спрощення роботи програміста.

Dataset – колекція однотипних даних, що застосовується в задачах машинної обробки даних.

Синапс – з'єднання, що використовується для відправки-отримання інформації між нейронами.

Бекенд (англ. Backend) – програмно-апаратна частина сервісу.

ML – сімейство строгих мов функціонального програмування з розвиненою параметрично поліморфною системою типів і па модулями.

РНМ – рекурентна нейронна мережа

ЗНМ – згорткова нейронна мережа

ШНМ – штучна нейронна мережа

ВСТУП

Оцінка сучасного стану об'єкта дослідження або розробки.

На даний момент технології розпізнавання жестів є затребуваними. Розробки в цій галузі досягли великих успіхів та використовуються у багатьох додатках та сервісах.

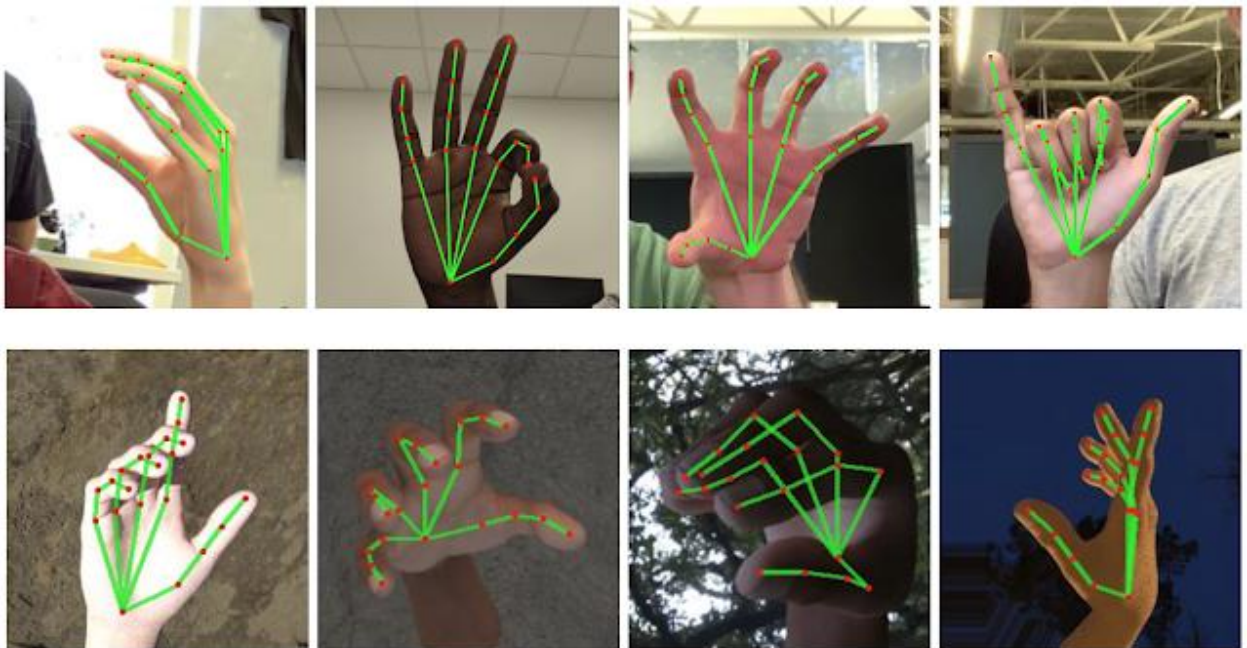
Наприклад, технологія Deep Sign дозволяє переводити звук в жестову мову та навпаки. Але через високу складність інтерпретації жестів у вільному спілкуванні, програма використовується у вузьких тематичних рамках («Прийом у лікаря», «Консультація в банку» і так далі). Це дозволило на початковому етапі контролювати дерево діалогів і попереджати контекстні помилки перекладу.

На сьогодні компанія Microsoft є лідером в технології розпізнавання жестів. Лінійка пристроїв розпізнавання руху для ігрових консолей Xbox 360 і Xbox One і персональних комп'ютерів Windows зосереджена навколо додаткового периферійного пристрою. Пристрій дозволяє користувачам управляти і взаємодіяти з їх ігровою консоллю або комп'ютером без ігрового контролера, через природний користувальницький інтерфейс за допомогою жестів.

Актуальність роботи та підстави для її виконання.

В даний час проблема комунікації людей з обмеженими можливостями залишається досить актуальною. За даними Всесвітньої організації охорони здоров'я більше 360 мільйонів людей страждають на глухоту або порушення слуху. Згідно зі статистикою за 2020 рік, в Україні 4 мільйона (близько 10%) людей володіють різними порушеннями слуху. З них близько 400 тисяч (близько 3%) страждають тяжкими порушеннями.

Значущість даної проблеми посилюється також у зв'язку з необхідністю створення систем, які забезпечать кращі умови для спілкування та сприйняття інформації людям із вадами слуху в соціумі. У 2019 році на CVPR було анонсовано новий підхід до визначення і сприймання рук та реалізовано в Mediapipe¹ (згодом на основі цього Google представила мобільний застосунок). У цьому підході використовується швидкісне і високоточне визначення і відслідковування 21 ключової точки руки всього із одного кадру. Далі новий розпізнаний жест порівнюється з базою вже відомих жестів – від алфавіту і цифр на мові глухонімих до жестів, які символізують, наприклад, «рок» «світ», «ОК».



Дана розробка є актуальною, оскільки є дослідженням та засобом використання штучного інтелекту з широкими можливостями застосування,

¹ Mediapipe – кросплатформенна платформа з відкритим вихідним кодом для створення конвеєрів для обробки перцептивних даних різних модальностей, таких як відео і аудіо.

одним із яких є надання простих можливостей комунікації людям із вадами слуху та/або мовлення, які спілкуються переважно жестовою мовою, при чому така можливість є достатньо гнучкою.

Мета й завдання роботи.

Метою дипломної роботи є вирішення важливої науково-прикладної задачі розробки інформаційної технології визначення та розпізнавання елементів дактильної мови. Для цього необхідно вирішити наступні завдання:

1. Оглянути загальні підходи теорії ідентифікації до розпізнавання елементів дактильної абетки.
2. Проаналізувати існуючі підходи до задач ідентифікації та розпізнавання елементів дактильної мови.
3. Проаналізувати та обрати найкращу архітектуру нейронної мережі з підвищеною ефективністю розпізнавання та ідентифікації.
4. Розробити програму для визначення і розпізнавання жестів дактильної абетки на фото.

Об'єкт і методи дослідження або розроблення.

Об'єктом розроблення програмного засобу є процес створення програми визначення жестів на фото. Робиться фото або береться із будь-якого ресурсу, і, після виконання програми, отримується фото із визначеним жестом на ньому та виводиться надпис визначеної літери. Тобто нейронна мережа показує який жест ми їй «показали».

Можливі сфери застосування.

Беручи до уваги співвідношення кількості глухонімих людей, людей, що знають дактильну абетку, та людей, що не знайомі з нею, цей напрямок

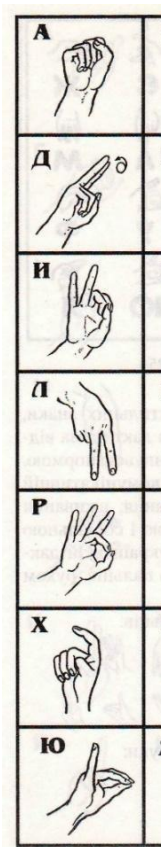
дає можливість розуміти, що говорять люди з вадами слуху й речового апарату. Наприклад, кожен з нас може встановити застосунок на телефон та в разі потреби розпізнати жести і дізнатися, що людина говорить.

Програми та застосунки можуть бути як вузькоспеціалізованими (прийом у лікаря), так і широкого спектру (звичайне спілкування). Технології розпізнавання жестів дактильної мови можуть використовуватися у різноманітних сферах, що стосуються людей с вадами слуху, тобто майже абсолютно у всіх.

РОЗДІЛ 1. АНАЛІЗ СИСТЕМ РОЗПІЗНАВАННЯ ЖЕСТІВ

1.1. Загальна інформація про жестову мову

В даний час у світі існує близько 135 мов жестів. Ми добре знаємо про розвиток мови жестів нікарагуанської мови на основі глухих. З самого початку в Нікарагуа не існувало універсальної мови жестів, але в 1977 році, коли була створена перша школа для глухих, вчені відразу ж почали документувати, як граматику та правила розвивалися протягом поколінь. Діти взаємодіють із попереднім поколінням, приймають і модифікують мову дорослих і навіть створюють правила, яких раніше не було. Майже всі мови жестів пройшли цей природній шлях створення мови. Для наочності на рис. 2 показана дактильна абетка української мови жестів.



1.2. Структура жестової мови

Жести можна розділити на чотири компоненти: конфігурацію руки, рух, локалізацію (місце виконання) і орієнтацію руки. Також враховується міміка, тобто: вираз обличчя і рух губ.

Буває і так, що поняття існує, але жест для нього ще не придуманий. Тоді використовується дактильна абетка – тобто жести, що відповідають буквам словесної мови. Іноді люди сприймають пальцевий алфавіт за жестову мову. Проте це різні речі: в той час, як пальцевим алфавітом передають звуки мови (буква за буквою), жестові знаки представляють цілі слова.

1.3. Структура систем розпізнавання жестів

Жест, в розумінні для реалізації різних інтерфейсів управління, – це форма невербального спілкування, при якій певна дія, рух тіла чи його частини, передає інформацію. При цьому, жест може містити в собі рух як однієї, так і декількох частин тіла та нести емоційне навантаження, яке робить його інформативним засобом спілкування і взаємодії.

Жести класифікуються за такими категоріями:

а) За типом:

- 1) Прості жести;
- 2) Комплексні жести;

б) За способом обробки:

- 1) Обробка після введення;

2) Обробка в процесі введення;

в) За тривалістю:

1) Статичні жести;

2) Динамічні жести.

Простими жестами є пальцеві жести (для розпізнавання таких жестів використовуються пристрої з сенсорними екранами) та жести руками (Kinect, МУО). Комплексні жести – це такі рухи, що задіюють декілька частин тіла чи складаються з декількох простих (Wii, Kinect).

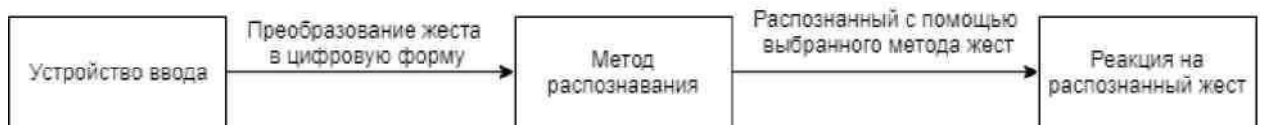
Класифікація жестів за шляхом обробки ділить їх на 2 категорії: жести, що оброблюються після введення та жести, що оброблюються в процесі введення. Приклад другого типу – збільшення / зменшення фотографії на сенсорних пристроях «Apple».

Тривалість жесту показує, чи змінюється він з часом. Тому жести можна розділити на статичні та динамічні. Статичні жести представляють одне зображення, тоді як динамічні жести представляють послідовність зображень.

Спрощена структура системи розпізнавання жестів включає в себе три блоки (див. рисунок 3):

- блок з пристроєм введення;
- блок з методом розпізнавання;
- блок реакції на розпізнавання.

Пристрій введення є початковою точкою входу системи. Він перетворює жест в цифрову форму та передає його в блок розпізнавання. Той, в свою чергу, порівнює отриману інформацію із шаблоном, що зберігається в базі даних, і в результаті повертає розпізнаний жест. У результаті блок реакції виконує функції відповідно до розпізнання жестів.



РОЗДІЛ 2. АЛГОРИТМИ ТА ПІДХОДИ ДО РОЗПІЗНАВАННЯ ЖЕСТІВ ДАКТИЛЬНОЇ АБЕТКИ

Дактильна мова – це спілкування за допомогою системи пальцевих знаків (дактилології, від грецьких: daktylos палець + logos вчення). Вона представляє собою такий алфавіт, в якому кожна буква «озвучується» пальцями рук. Слова показуються за допомогою відповідних комбінацій. Речення – за допомогою дотримання граматики звичайної розмовної мови.

На даний момент існують такі технології, за допомогою яких можна розпізнати жести, а саме використання:

- відеокамер;
- рукавичок з різноманітними датчиками та видами передавачів;

- технології Motion capture² ;
- камер стереозображення;
- 3D камер;
- штучних нейронних мереж.

Методи, в основі яких використовуються камери, можуть розпізнавати статичні та динамічні жести, а в деяких випадках їх точність розпізнавання перевищує 90%. Недоліком цих методів є головним чином те, що вони чутливі до змін освітлення. У різних роботах намагаються усунути цю проблему, використовуючи кольорові рукавички або створюючи однорідний фон, що робить взаємодію людини з комп'ютером незручною та неприродною.

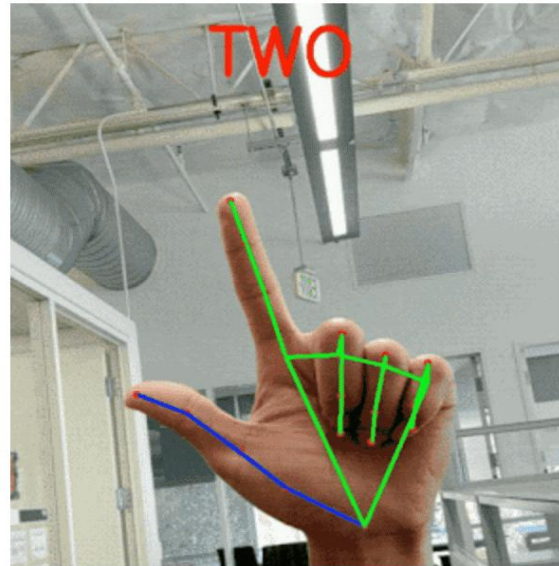
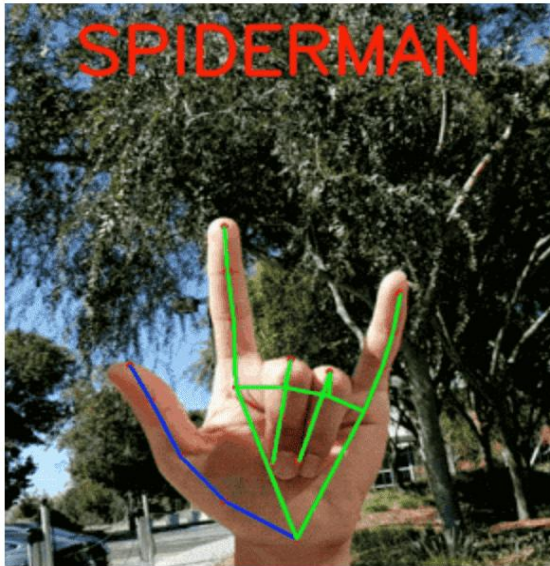
Методи на основі рукавичок дозволяють з високою якістю розпізнавати позиції ключових точок руки людини, але вимагають використання спеціальних рукавичок.

Методи, що беруть за основу використання тривимірних сенсорів дозволяють розпізнати ключові точки тіла людини в реальному часі. Крім того, ці методи є незалежними від зміни освітлення. Але, не зважаючи на це, на сьогоднішній день не існує готових рішень, які дозволяють з великою точністю розпізнавати конфігурації та динамічні жести руки людини.

2.1. Використання відеокамер

² Motion capture – технологія цифрового запису рухів, що використовується у розважальному, медичному, спортивному, анімаційному та кінематографічному програмному забезпеченні.

Для діагностики жестів необхідно зафіксувати положення руки і пальців. При цьому виникає багато схожих ситуацій, які можуть бути

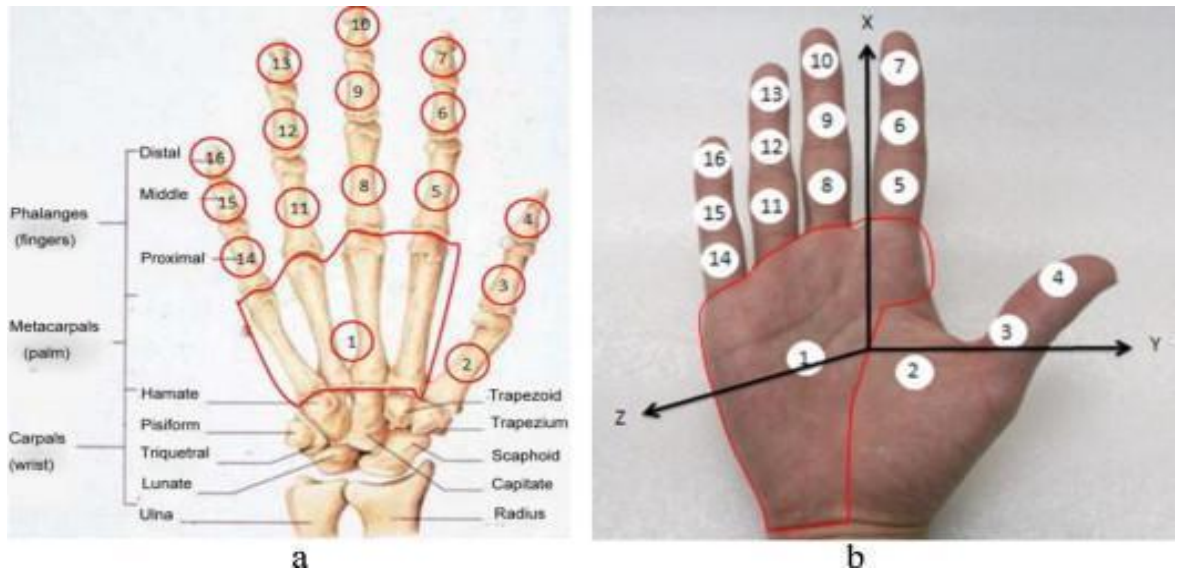


представлені як жест, але при цьому ним не являються.

При визначенні жестів, виникають питання, які потребують особливої уваги:

1. Період часу, коли певне положення руки вважається жестом. Тобто, мається на увазі час фіксації руки в статичному положенні для визначення жесту. Під час роботи з цією технологією людина може здійснювати мимовільні жестикуляції, що не несуть в собі ніякого сенсу та ніяк не пов'язані із взаємодією з пристроєм.
2. Вибір відеокамери. Відео та фото-камери мають ліміт швидкості сприйняття інформації (25-30 кадрів за секунду), що не дозволяє відслідковувати усі рухи руки. Адже окремі частини руки можуть рухатися із різною швидкістю. За такої ситуації при зйомці ми отримуємо змазаний жест, який важко розпізнати та інтерпретувати. Можна використовувати дорогі камери із високою кількістю кадрів за секунду, але система задумується для роботи зі звичайними середньостатистичними камерами.

3. Алгоритми обрахування. Висока складність алгоритмів та низька швидкодія елементів системи, що обраховує, не завжди дозволяє виконувати бажані дії за доли секунди та потребують певного часу. Що



залежить від розширення зображення, якості зйомки та інших факторів.

Для вирішення проблеми пошуку шкіри на зображенні пропонується метод сегментації зображення.

Системи розпізнавання дактильної та жестової інформації на основі відеоданих побудовані, у своїй більшості, на наступній послідовності кроків:

- 1) отримання вхідного зображення;
- 2) знаходження на зображенні обличчя і кистей рук;
- 3) виділення значимих характеристик (ознак);
- 4) зменшення кількості характеристик;
- 5) пошук у базі даних;
- 6) побудова розпізнаної послідовності (літери, слова).

Для того, щоб знайти обличчя та руки на зображенні, використовується метод пошуку області за кольором шкіри (модель шкіри Гауса або більш досконала модель суміші розподілу Гауса).

Методами, що використовуються для розпізнавання форми кисті, є:

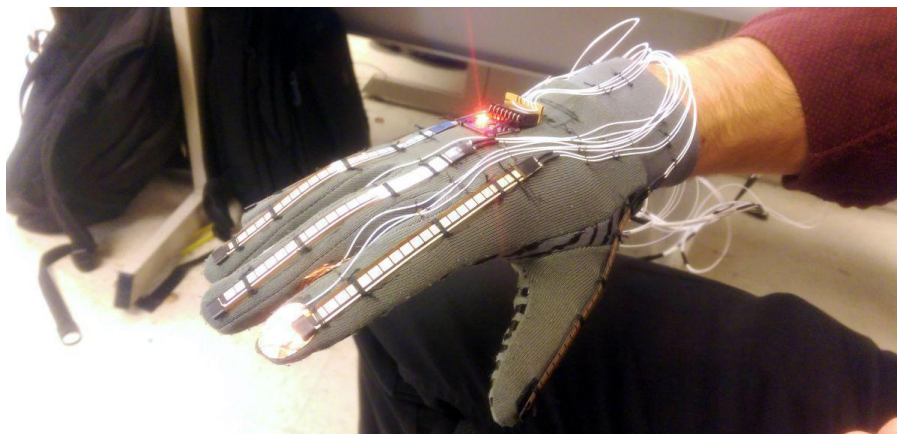
- 1) контур кисті та особливості контуру;
- 2) особливості, отримані в просторі основних компонентів;
- 3) особливості, отримані такими методами, як зона Хаара та інші.

2.2. Використання рукавичок

Пристрій «Language of Glove» було створено силами американських науковців. Функція цієї рукавички полягає в перетворенні англійської мови жестів в звичайний текст.

Система включає в себе пару рукавичок, у яких по довжині пальців протягнуті датчики. Причому такі, що мають здатність розтягуватися (адже пальці згинаються). Ці датчики фіксують рух руки та положення пальців, що позначають букви, слова, фрази.

Після цього рухи перетворюються в електричні сигнали та передаються на плату, що знаходиться на зап'ясті. Вона, в свою чергу, передає дані по безпроводній мережі на телефон, що переводить їх у слова і озвучує з приблизною швидкістю одне слово за секунду.



Також для вирішення задач класифікації жестів використовують кольорові рукавиці. Цей метод дозволяє за допомогою лише однієї відеокамери в реальному часі розпізнавати конфігурацію руки і відстежувати рухи долоні в тривимірному просторі.

Ця рукавичка складається з двадцяти сегментів, що забарвлені в десять різних кольорів. Використання невеликої кількості сегментів дозволяє розпізнати 20 кольорів певної точки зображення рукавички при різному освітленні. Варто зауважити, що спеціальне розміщення кольорових сегментів не дає можливість отримати ідентичні зображення при різних конфігураціях руки.

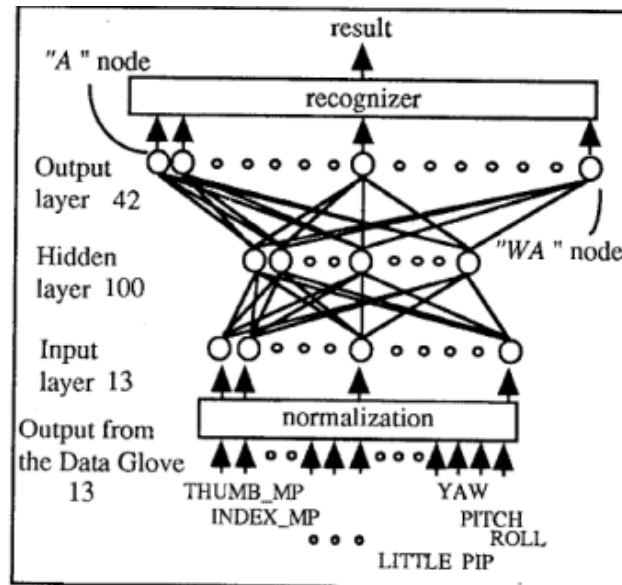


2.3. Використання штучних нейронних мереж

Штучні нейронні мережі (ШНМ) представляють собою математичну модель, що заснована на принципах біологічних нейронних мережах та їх функціонуванні. У завданнях на розпізнавання жестів, як правило, ШНМ використовують як інструмент машинного навчання. На вхід передаються характеристики жесту, а на виході отримується розпізнаний жест. ШНМ відрізняються між собою структурою і методом навчання.

Найпоширенішими видами ШНМ, що використовуються в задачах на розпізнавання, є мережі прямого поширення, згорткові нейронні мережі рекурентні нейронні мережі, самоорганізовані карти Кохонена і т.д.

На рис. 6 зображена рекурентна нейронна мережа, що використовується в Японії для допомоги у розпізнаванні людям з вадами слуху, статичних жестів дактильної абетки.

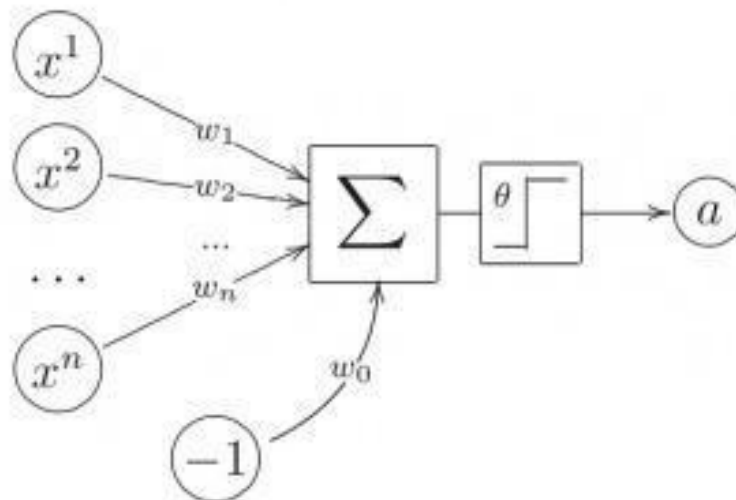


Вхідним пристроєм в роботі представлена рукавичка із вбудованими сенсорами, що повертає характеристики показаного жесту (10 точок і 3 кута). Вони нормалізуються та передаються на вхід нейронної мережі. Мережа на виході видає результат у вигляді масиву, що містить 42 елемента, які відповідають 42 жестам руки. Для зареєстрованих користувачів результати розпізнавання становлять – 71.4%, а для незареєстрованих – 47.8%.

РОЗДІЛ 3 НЕЙРОМЕРЕЖІ. ВИДИ ТА ЗАСТОСУВАННЯ

3.1. Історія виникнення нейромережі

Можливо, найдивніша властивість людського мозку – це здатність до самоаналізу. Природа думки становить інтерес для вчених в різних областях, від духовних основ існування до анатомії. Але ні висновки філософів, ні результати наукових експериментів не могли повністю вирішити цю проблему. Однак за останні кілька десятиліть нейробіологи та нейроанатоми доклали значних зусиль у цьому напрямку. Найбільш важливим проривом став механізм навчання нейронів. Це досягнення дозволило Маккаллоу і Піттс в 1943 році розробити основні принципи моделювання нейрона, які включали в себе механізм навчання і були названі формальним нейроном (рисунок 7).



Ц

я модель лягла в основу теорії логічних мереж та кінцевих автоматів. Крім того вона активно використовувалася психологами і нейрофізіологами для моделювання певних локальних процесів нервової діяльності. Беручи до уваги її дискретність, модель цілком узгоджується з комп'ютерною парадигмою та служить її «нейронним фундаментом».

Нехай n вхідних величин бінарних ознак (x_0, x_1, \dots, x_n) , що описують x .
Потрапляючи в нейрон імпульси складаються з вагами (w_0, w_1, \dots, w_n) . Потім на вхід подаються отримані дані і він виконує їх обробку за формулою:

$$\alpha(x) = \varphi \left(\sum_{j=1}^n w_j \cdot x^j - w_0 \right).$$

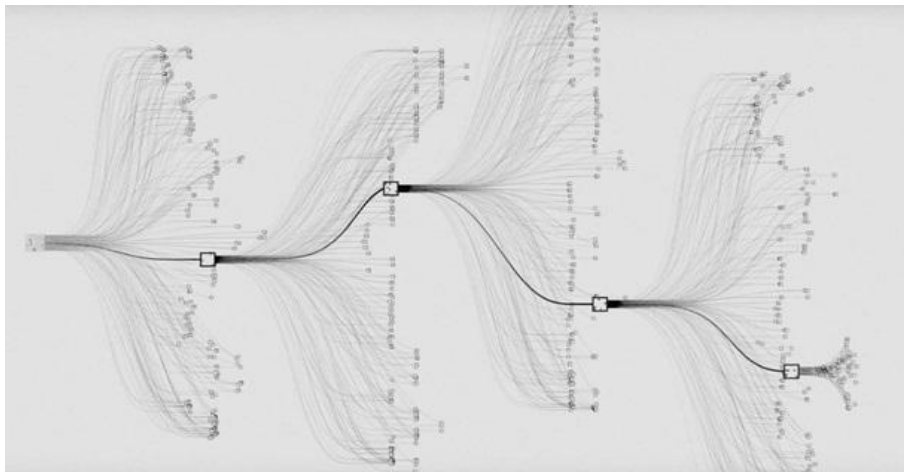
Якщо сумарний імпульс більший за поріг активації w_0 , то нейрон збуджується та на виході отримується 1, інакше 0.

Активаци́йна функція має наступний вигляд:

$$y = \begin{cases} 1 & \text{при } v \leq 0, \\ 0 & \text{при } v > 0. \end{cases}$$

3.2. Застосування нейромереж

Штучні нейромережі розробляються, щоб зрозуміти як працює мозок людини. Крім цього вони намагаються повторити його роботу. Нейромережа є системою, яка навчається і також працює за алгоритмами та на основі попереднього досвіду. Штучний нейрон – це спрощена модель природного.



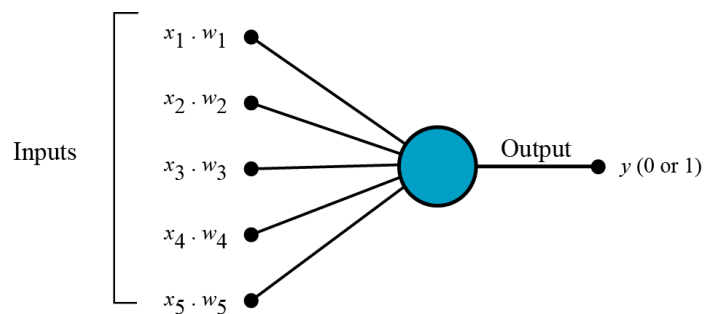
Простіше кажучи, нейронна мережа – це комп'ютерна програма, яка отримує та реагує на певну інформацію, а не виконує конкретні вказівки.

Нейромережі використовуються в різних галузях. Наприклад, у економіці – передбачають обмінний курс, ціни на нафту чи золото, банкрутство компаній. У робототехніці – роботи дозволяють оцінювати перешкоди, підтримувати рівновагу та керувати приборами.

3.3. Види нейромереж

3.3.1. Нейромережа з прямим зв'язком

Нейронна мережа прямого поширення – це штучна нейронна мережа, в якій зв'язки між вузлами не утворюють цикл. Повною протилежністю цього виду нейронної мережі є рекурентна нейронна мережа, в якій визначені шляхи повторюються циклічно. Модель з прямим зв'язком, як на рис.9, – найпростіша форма нейронної мережі, так як інформація обробляється лише в одному напрямку. Дані можуть проходити через декілька прихованих вузлів, проте вони завжди рухаються в одному напрямку і ніколи назад не повертаються.

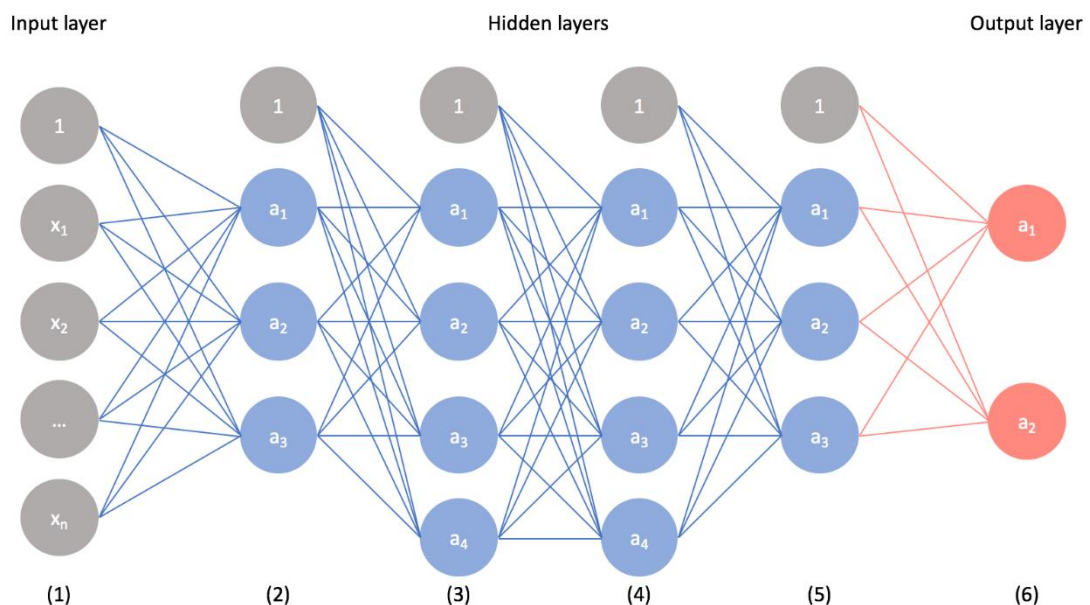


Нейронні мережі з прямим зв'язком застосовуються в галузях, які вимагають контрольованого навчання, наприклад, комп'ютерного зору. Крім цього, цей вид нейронної мережі найчастіше використовується в системах розпізнавання об'єктів та розпізнавання мови.

3.3.2. Згорткова нейромережа

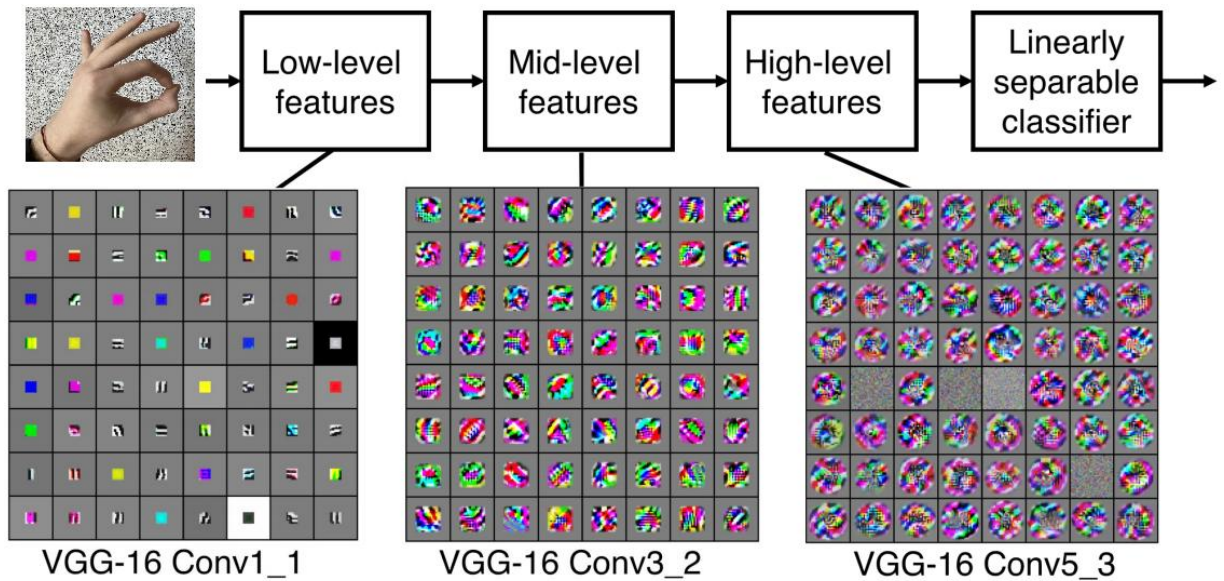
Згорткова нейронна мережа (ЗНМ) – це алгоритм глибокого навчання, який може приймати вхідне зображення, призначати важливість різних об'єктів та аспектів на зображенні й уміє розрізняти одне від іншого. У ЗНМ попередня обробка, набагато нижча в порівнянні з іншими алгоритмами класифікації. При достатній підготовці, ЗНМ можуть вивчати характеристики фільтрів, незважаючи на те, що в примітивних методах вони розроблені вручну.

Архітектура згорткових нейронних мереж аналогічна структурі зв'язку нейронів в мозку людини (рисунок 10). За натхнення було взято організацією зорової кори. Окремі нейрони реагують на подразники лише в обмеженій області зорового поля, відомому як рецептивне поле. Колекція таких полів накладається на всю зону зорового огляду.



Один згортковий шар можна сформувати, якщо до зображення застосовується багато фільтрів, як на рис. 11, і на виході отримуються різні карти активації. Щоб створити нейромережу, шари чергуються один за

одним, а між ними додаються функції активації (ReLU чи інші) та спеціальні pooling-шари, що зменшують розмір карт ознак.



У найперших шарах згорткові фільтри, як правило, співвідносяться з низькорівневими ознаками зображення – наприклад з краями. У середині знаходяться складніші особливості, такі як кути і кола. А в фінальних шарах фільтри більше нагадують певні специфічні ознаки, які можна інтерпретувати більш широко.

3.3.3. Рекурентні нейромережі

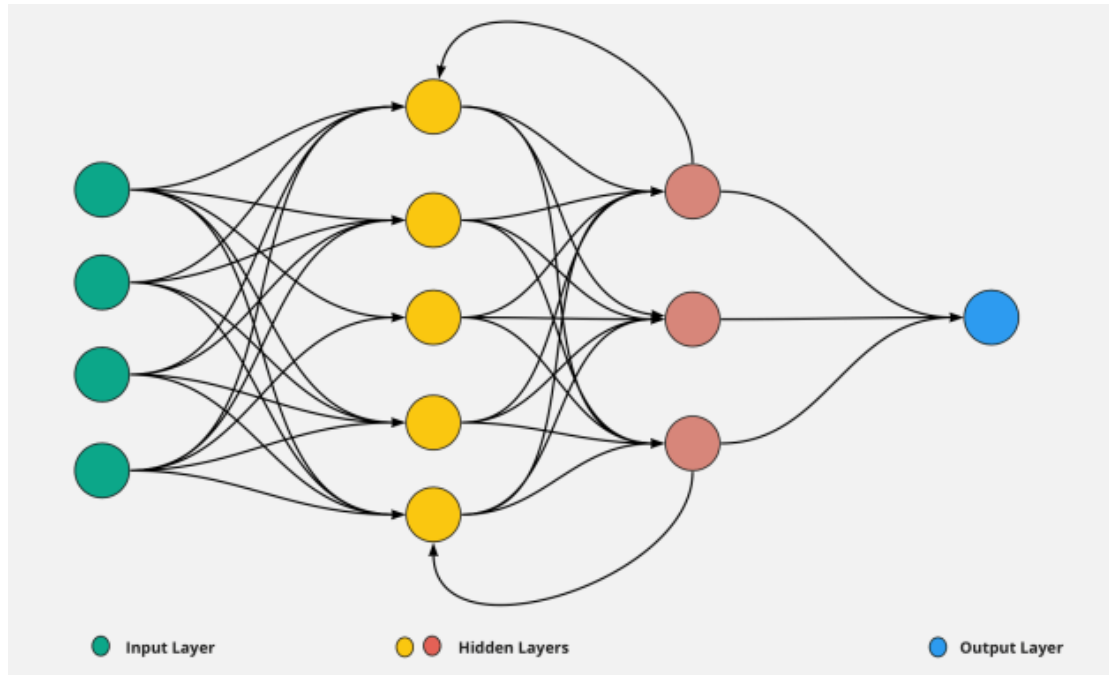
Рекурентна нейронна мережа (РНМ) – це тип такої штучної нейронної мережі, що використовує послідовні дані чи дані часових рядів. Вона використовується для вирішення порядкових або часових проблем, наприклад, переклад мови, розпізнавання мови та субтитри до зображень. Ці алгоритми включені в наступні популярні програми – Siri, голосовий пошук та Google Translate.

Рекурентні нейромережі набагато складніші, ніж мережі прямого зв'язку і можуть виконувати складніші завдання, ніж базове розпізнавання зображень. Осмислення та генерація природної мови передбачає набагато складнішу обробку, ніж розпізнавання зображень, яку можуть виконувати повторювані нейронні мережі завдяки своїй архітектурі. У той час як у

нейронних мережах прямого зв'язку з'єднання ведуть лише від одного нейрона до нейронів у наступних шарах без будь-яких зворотних зв'язків, а рекурентні нейронні мережі дозволяють з'єднанням повертатися до нейронів у тому ж шарі, що забезпечує ширший спектр операцій (рисунок 12).

Висновок

Для задач розпізнавання образів та класифікації найбільш ефективним



типом мереж є згорткові нейронні мережі, що є прогресивним напрямком розвитку. Крім перевірених структур мереж, постійно з'являються нові види загорткових мереж, що демонструють кращу точність чи швидкість роботи.

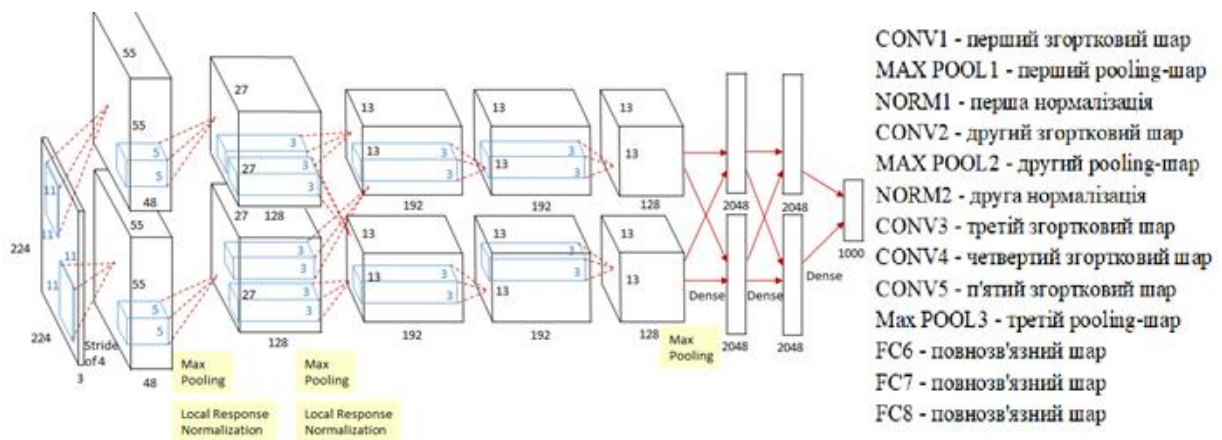
В порівнянні з повнозв'язними мережами, в яких кожен нейрон пов'язаний з усіма нейронами наступного шару, згорткові мережі мають певні особливості структури. Вони дозволяють обробляти великі масиви даних, наприклад зображення або звук.

Згорткові шари виконують операцію згортки пересуваючи фільтр по зображенню з невеликим кроком. Саме така структура дозволяє зменшити кількість параметрів у порівнянні з повнозв'язною мережею, оскільки нейрони поєднуються не з усіма нейронами наступного шару, а тільки з тими, які використовувались під час виконання згортки.

РОЗДІЛ 4. АНАЛІЗ АРХІТЕКТУР ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ

4.1. Архітектура AlexNet

Архітектура нейромережі AlexNet складається з 5 згорткових шарів (рисунок 13). Між ними знаходяться pooling – шари і шари нормалізації. А



завершують нейромережу 3 повнозв'язних шари.

Параметри нейромережі:

1. Це перша архітектура, в якій застосовується нелінійність ReLU^3 .
2. Для нормалізації використовуються окремі шари.
3. На етапі попередньої обробки використовується велике доповнення даних.
4. При Dropout – 50% відключається половина нейронів.

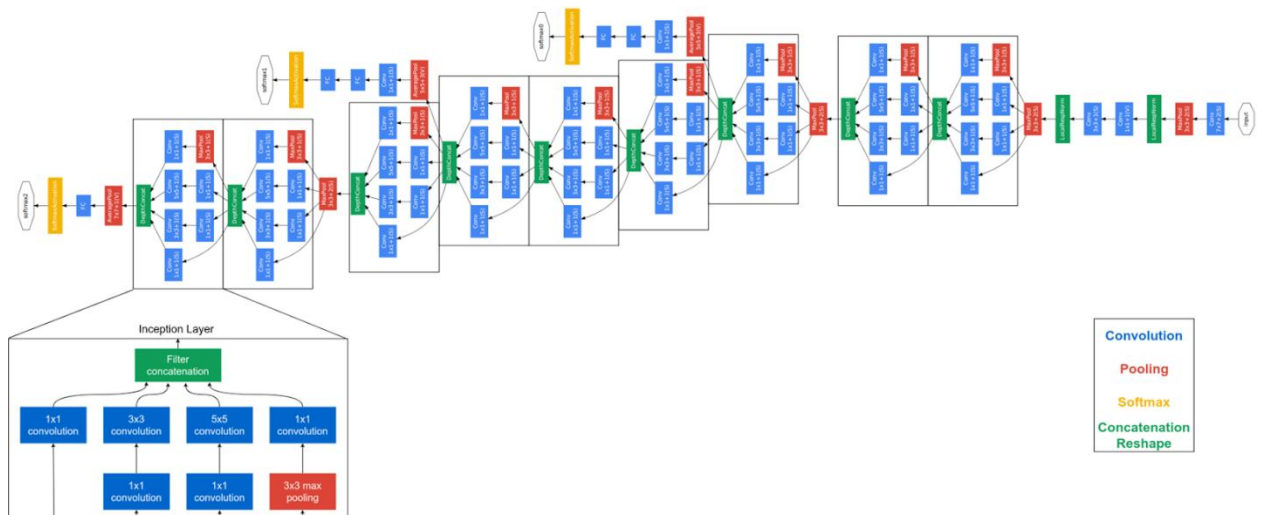
4.2. Архітектура VGGNet

Головна ідея VGG-архітектур – використання більшого числа шарів з фільтрами меншого розміру (рисунок 14). Існують версії VGG-16 і VGG-19 з

³ ReLU – функція активації.

4.3. Архітектура GoogleNet

GoogleNet – більш глибока архітектура, що містить 22 шарів (рисунок 15). Для більшої ефективності обчислення Google придумала спеціальний модуль Inception. Його архітектура складається із багатьох



модулів, що йдуть один за одним.

Також GoogleNet не має повнозв'язних шарів. Крім цього вона містить всього п'ять мільйонів параметрів – в 12 разів менше, ніж у AlexNet.

У GoogleNet є маленька підмережа – Stem Network. Вона складається з трьох згорткових шарів і містить два pooling-шари, та розташовується на початку архітектури.

4.4. Архітектура ResNet

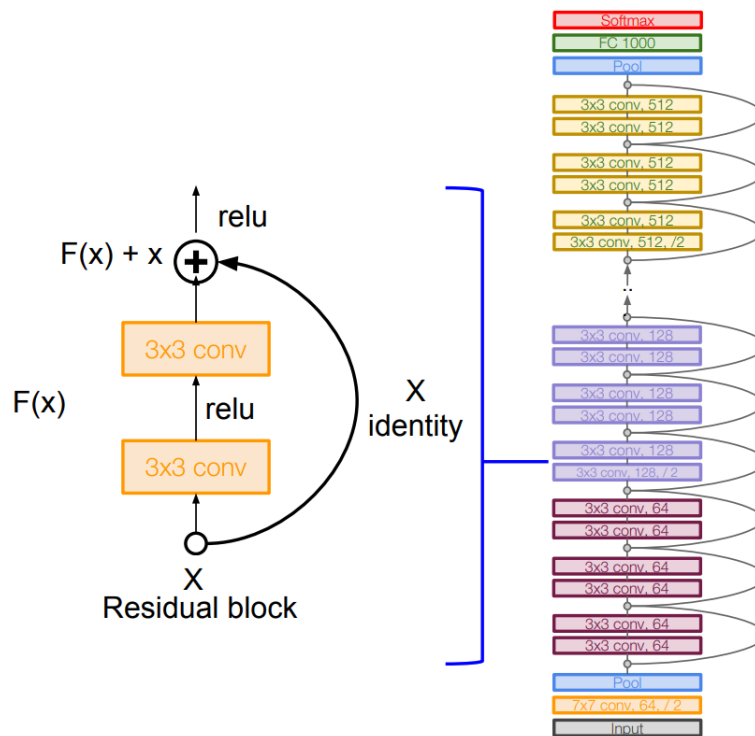


Рисунок 16 - Архітектура ResNet

Нейромережа ResNet складається з великого стека однакових залишкових блоків, кожен з яких має два згорткових шара розмірністю 3×3 , як показано на рис. 16. Періодично число фільтрів збільшується в два рази і розмірність зменшується з кроком 2. На початку архітектури знаходиться додатковий згортковий шар.

Крім цього у цієї нейромережі немає повнозв'язних шарів в кінці – використовується тільки один шар з вихідними класами. Якщо необхідно збільшити кількість шарів, то використовуються такі ж додаткові шари, як і в GoogleNet для зменшення розмірності зображення.

Особливості нейромережі:

1. Вихідний сигнал повертається назад і подається як вхідний.
2. Вперше представлена концепція пропуску з'єднання.
3. Цю архітектуру можна розглядати як систему паралельних і одночасно послідовних модулів. Тобто у багатьох модулях inout-сигнал приходить

паралельно, а вихідні сигнали кожного модуля мають послідовне з'єднання.

4. ResNet, як правило, оперує блоками невеликої глибини в 20-30 шарів, що працюють паралельно, а не проганяє послідовно по всій довжині мережі.
5. Не використовується Dropout.
6. Після кожного згорткового шару використовується пакетна нормалізація.

РОЗДІЛ 5. НАВЧАННЯ НЕЙРОМЕРЕЖІ

Після того, як мережа була структурована для конкретного додатка, ця мережа готова до навчання. Для запуску цього процесу випадковим чином вибираються початкові ваги. Потім починається тренування або навчання.

Навчання нейромережі починається з підготовки матеріалів – розмічених зображень об'єктів, які потрібно буде розпізнавати. З їхньою допомогою ми навчаємо нейронну мережу розпізнавати, що зображено на фото. Алгоритм навчання означає процедуру, в якій використовуються правила навчання для налаштування вагів. Існують три парадигми навчання:

1. З вчителем.
2. Без вчителя.
3. Змішана.

Переважає більшість мереж використовують навчання з учителем. Навчання без вчителя використовується для виконання деяких початкових характеристик вхідних даних.

5.1. Навчання з вчителем

На етапі навчання правильний клас для кожного запису відомий (це називається контрольованим навчанням), і тому вихідним вузлам можуть бути присвоєні «правильні» значення – «1» для вузла, відповідного правильному класу, і «0». Таким чином, можна порівняти розраховані мережею значення для вихідних вузлів з цими «правильними» значеннями і обчислити помилку для кожного вузла (Правило «Дельта»). Ці терміни помилок потім використовуються для коригування ваг в прихованих шарах, щоб наступного разу вихідні значення були ближче до «правильних» значень.

Тобто навчання з учителем передбачає, що для кожного вхідного вектора X існує цільовий вектор YT , що необхідно виходом. Разом вони називаються навчальною парою. Звичайна мережа вчиться для деякої

кількості таких навчальних пар. У ході навчання відбувається зчитування вхідного вектора X , обчислюється вихід мережі Y та порівнюється з відповідним цільовим вектором YT . За допомогою зворотного зв'язку в мережу подається різниця $D \sim YT - Y$ і змінюються ваги W відповідно до алгоритму, який прагне мінімізувати помилку ϵ .

Зчитування векторів навчальної множини і налаштування ваг виконується доти, поки загальна помилка не досягне устанавленого низького рівня для всієї навчальної множини. Тобто, ваги налаштовуються так, щоб мережа робила відповіді якомога ближче до відомих правильних відповідей.

5.2. Навчання без вчителя

Навчання без вчителя являє собою набагато правдоподібнішу модель навчання в біологічній системі. Вона не потребує цільового вектора для виходів і не вимагає порівняння з правильними відповідями. Навчальна множина складається тільки з вхідних векторів. Алгоритм навчання налагоджує ваги таким чином, щоб виходили узгоджені вихідні вектори, тобто щоб виставляння близьких вхідних векторів давало однакові виходи.

Навчання без вчителя не вимагає знання правильних відповідей на кожен приклад навчальної вибірки. Тобто процес навчання виділяє статистичні властивості навчальної множини та групує в класи схожі між собою вектори. У цьому випадку розкривається внутрішня структура даних або кореляції між прикладами в системі даних, що дозволяє розподілити їх по категоріях.

5.3 Змішане навчання

У змішаному навчанні одна частина вагів визначається за допомогою навчання з вчителем, в той час як друга визначається за допомогою самонавчання. Теорія навчання розглядає три фундаментальні властивості, що пов'язані з навчанням за прикладами: ємність, складність зразків і обчислювальна складність. Під ємністю мається на увазі, скільки прикладів

може запам'ятати мережа та які функції й межі ухвалення рішень можуть бути сформовані на ній. Складність зразків визначає число навчальних прикладів, необхідних для досягнення здатності мережі до узагальнення. «Перенавчання» мережі може викликати дуже мала кількість прикладів. Тобто коли нейромережа добре функціонує на прикладах навчальної вибірки, але погано — на тестових прикладах, що відносяться до такого ж статистичного розподілу. Відомі чотири основних типи правил навчання:

- корекція за помилкою;
- навчання Больцмана;
- правило Хебба;
- навчання методом змагання.

5.4. Дані для навчання

Ефективність моделей машинного навчання (MachineLearning, ML) безпосередньо залежить від якості вихідних даних, потрібно багато часу на підготовку до моделювання: очищення, нормалізацію і генерацію змінних. При цьому в початковій вибірці можуть запросто бути відсутні дані, потрібні для формування предикторів і цільових змінних. Таким чином, виникає питання отримання даних. Звідки брати ці дані?

1. Використання готових датасетів;
2. Робота зі веб-платформами, які надають статистику;
3. Використання інформації зі сторонніх сайтів.

5.4.1 Датасети

Датасет – вибірка даних, набір даних, зазвичай в форматі «безліч з множини ознак» → «деякі значення» (якими можуть бути, наприклад, ціни на житло, або порядковий номер набору даних деяких класів), де X – безліч ознак, а y – ті самі деякі значення. Визначати,

наприклад, правильні індекси для безлічі класів – завдання класифікації, а шукати цільові значення (такі як ціна, або відстані до об'єктів) – завдання ранжування.

Структуровані вибірки даних з різних тематик можна завантажити самостійно з наступних джерел:

- Конкурсні платформи і змагання з DataScience і ML.
- Інтернет-портали і спільноти по DS і ML.
- Відкриті дані – інформація машино-зчитуваного формату, розміщена в Інтернеті і доступна для загального вільного використання та подальшого передруку без обмежень авторського права та інших способів контролю.

З часом різноманіття датасетів досягло тієї відмітки, що зараз можна використовувати будь-які з них. Починаючи від набору даних смертей з битв «Гри престолів» та закінчуючи даними біткоїна за всі роки існування.

5.4.2 Веб-платформи зі статистикою

Використовуючи веб-платформи зі статистикою, можна, наприклад, виявити частоту захворювань на COVID-19 по регіонах або аналізувати завантаженість доріг за допомогою GoogleMaps, щоб будувати оптимальну логістику по доставці товарів або розвезення пасажирів.

Для отримання статистичних даних, в тому числі в режимі реального часу, можна скористатися програмними інтерфейсами API, наданими сторонніми сервісами. Наприклад, API-рішення від Google, API різних сервісів Яндекс, трейдингові системи, які отримують інформацію з фондових, валютних і криптовалютних бірж для автоматизованої торгівлі цінними паперами. При цьому доступ до набору даних, як правило, в JSON-форматі, реалізується через URL-адресу. Подальший розбір вмісту JSON-файлу доведеться робити самостійно, за допомогою власних скриптів.

5.4.3 Парсинг сторонніх сайтів

Парсинг – це процес аналізу вмісту веб-сторінок за допомогою роботів-парсерів - спеціальних програм або скриптів. Найвідоміші парсери в мережі – це роботи пошукових систем, які аналізують сайти, зберігають результати аналізу в своїх базах, щоб видати при пошуку найбільш релевантні і актуальні документи. Можна написати свій власний веб-парсер на Python або PHP для автоматичного збору даних із зовнішніх ресурсів за заданими характеристиками (зображення, ключові слова і т.д.). Але найпростіше знайти вже готові рішення, бібліотеки які існують на всіх популярних мовах програмування.

РОЗДІЛ 6. РЕАЛІЗАЦІЯ ОБ'ЄКТА ДОСЛІДЖЕННЯ

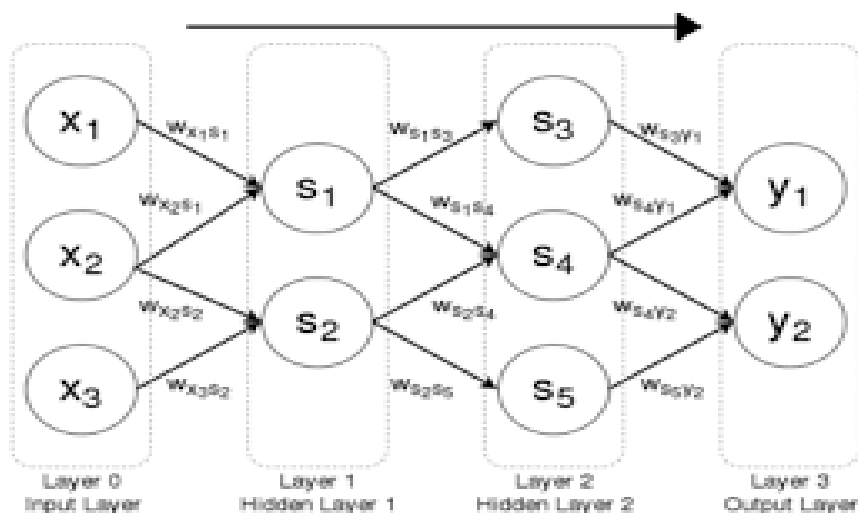
Беручи до уваги слабкі обчислювальні можливості ноутбуку, було вирішено виконувати реалізацію дослідження наступним способом.

6.1. Нейронна мережа

Для розв'язання поставлених задач використаємо згорткову нейронну мережу, як найкращий загальний варіант розв'язання поставленої задачі.

Для навчання і перевірки мережі було зібрано вхідні дані із зображеннями жестів української дактильної абетки у кількості в середньому близько двохсот зображень на клас із різним фоном, масштабом, поворотом, нахилом та освітленням у форматі RGB і розміром 100×100 пікселів. Для навчання мережі необхідно сформувати спеціальну структуру, яку використовує Tensorflow.

Структура нейронної мережі виглядає наступним чином – в ній чергуються згорткові слої та слої підвибірки і є повнозв'язна частина нейронної мережі для класифікації.



6.2. Необхідні засоби

Для реалізації дослідження було обрано мову програмування Python.

Особливості Python:

- Широкий спектр доступних сторонніх бібліотек та модулів. Підтримувані ресурси дозволяють сильно розширити можливості мови, отримуючи доступ до майже будь-яких засобів, від веб фреймфорків до розрахунку ядерних реакторів.
- Динамічна і неявна типізація. Python не потребує попереднього визначення типу об'єкту та допустимих операцій на етапі написання коду – все це визначається динамічно під час виконання.
- Інтерпретована мова програмування. Код цієї мови не компілюється у виконуваний модуль, а натомість обробляється інтерпретатором під час виконання. І тому програму можна запустити де завгодно, де наявний інтерпретатор.

Також, беручи до уваги, слабкі обчислювальні потужності комп'ютера, було вирішено використовувати хмарний сервіс на основі Jupyter Notebook. Перевага цього сервісу в тому, що він дає доступ до швидких TPU⁴ і GPU⁵.

6.3. Програмна реалізація

Для реалізації програми були виконані наступні кроки:

1. Перш за все, встановлюються необхідні бібліотеки (модулі) – Tensorflow, Keras та numpy та імпортовано необхідні методи з них.
2. Далі записується розмір міні вибірки та розмір картинок.

⁴ TPU – тензорний процесор, що призначений для прискорення розрахунків штучного інтелекту. Розроблений Google спеціально для машинного навчання нейронних мереж.

⁵ GPU – графічний процесор.

3. Наступним кроком створюється Tensorflow датасет для тренування нашої нейромережі із каталогу з зображеннями та задається `validation_split` як 0.1 щоб 10% наших даних використовувались для перевірки, `seed` – початкове значення випадкових чисел ставлю як 42.
4. Аналогічно до попереднього кроку створюється перевірочний набір даних. Виходить як раз 10% від усіх зображень.
5. Потім задається список з назвами класів. Вони створюються автоматично по назвах каталогів (папок).
6. Створюється третій набір даних – для тестування. Всі вони будуть використовуватися для тестування.
7. Щоб збільшити швидкість навчання, налаштовується продуктивність Tensorflow датасетів. Викликається метод `prefetch` – він виконує попередню загрузку мінівбірок до того, як вони знадобляться. За допомогою AUTOTUNE виконується автоматичне налаштування.
8. Тепер створюється архітектура нейронної мережі. У ній чергуються згорткові слої і слої підвибірки. `Conv2D(16,(5,5))` – 16 карти ознак, розмір ядра згортки 5×5 . `Input_shape = (100, 100, 3)` означає що у нас 3 канали RGB, розмір 100×100 . Слой підвибірки `MaxPooling` означає, що із квадрата 2×2 вибирається максимальне значення. `Dropout` використовується для запобігання перенавчання нейромережі. Тобто, коли на вхід подається новий об'єкт, випадково виключається певна кількість нейронів з вірогідністю 0.2 (тобто 20%). Починаючи з `model.add(Flatten())` починається повнозв'язна частина класифікатора. `Flatten()` – виконує перетворення із двомірного вигляду в плоский. Повнозв'язний слой `Dense(1024,..)`, де 1024 означає кількість нейронів. Нарешті вихідний слой `model.add(Dense(33, ..))`. 33 – кількість нейронів, по кількості класів у наборі даних.
9. Компілюється модель та виконується навчання нейромережі. У метод `fit` передається набір даних для навчання. В якості перевірочного

набору вказується `validation_dataset`, які містять як і зображення з перевірконого набору даних, так і правильні відповіді для них, кількість епох 8. Після навчання доля правильних відповідей складає 0,756.

10. Оцінюється якість навчання нейромережі на тестових даних. Evaluate використовується для оцінки моделі з використанням даних валідації. І доля правильних відповідей склала більше 79%. Потім модель зберігається.

11. Після цього на вхід подається заздалегідь підготована картинка із жестом на ній. Виконуємо обробку зображення і за допомогою метода `prediction=model.predict(x)` виконується передбачення жесту на фото.

12.



Після виконання всіх кроків було отримано наступні результати:

Буква: С

13. І ще один приклад з іншим жестом. Як можна спостерігати, програма
правильно
визначила жести на
фото.

ВИСНОВКИ

Прискорений розвиток технологій став причиною значного зростання сфер діяльності, що пов'язані з вирішенням питань для людей з обмеженими можливостями. Особливо актуальними є задачі, що стосуються людей з вадами слуху.

Під час кваліфікаційної роботи було досліджено багато різноманітних питань. Зокрема було проаналізовано існуючі підходи до задач ідентифікації та розпізнавання елементів дактильної мови. Детально оглянуто кожен зі способів та обрано найзручніший і найбільш можливий підхід у даній ситуації.

Крім того, були оглянуті види нейромереж та встановлено, що для задач розпізнавання образів та класифікації найбільш ефективним типом мереж є згорткові нейронні мережі. Щоб обрати архітектуру мережі, було проаналізовано декілька найбільш популярних прикладів.

Щоб нейронна мережа могла коректно вирішувати поставлене завдання, її необхідно «прогнати» на певній кількості вхідних даних. Для досягнення мети кваліфікаційної роботи було зібрано набір даних, що містить 300 000 зображень 32 букв українського алфавіту. Дані склалися з навчального, перевірного та тестового наборів даних. В результаті встановлення необхідних компонентів та написання програми, були отримані результати, що були очікуваними. Спочатку відбувалося навчання нейромережі на вхідних даних, потім – тестування отриманої нейромережі для визначення точності роботи. Після отримання задовільних результатів тестування, завантажувалася картинка з жестом і в результаті роботи програми спостерігався вивід в консоль букви, що позначає жест на фото.

Програму було протестовано в різних освітленнях та на різних фокусних відстанях. Чим далі знаходиться рука із жестом від об'єктиву камери, тим нижча точність визначення. Схожа ситуація траплялася зі

знаходженням жесту на поверхнях, подібних до кольору руки. В цьому випадку декілька разів нейромережа допускала помилку.

Результати, що були отримані під час виконання кваліфікаційної роботи далекі від сучасних досягнень та стандартів, але дослідження в області розпізнавання жестів необхідно продовжувати з метою розкриття повноцінного потенціалу цієї галузі.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. <https://dev.by/news/google-sozdala-algoritm-raspoznavaniya-zhestov-rabotayushii-cherez-kameru-smartfona>
2. <https://ai.googleblog.com/2019/08/on-device-real-time-hand-tracking-with.html>
3. <https://worldvision.com.ua/articles/kak-rabotaet-tehnologiya-raspoznavaniya-zhestov>
4. <https://vc.ru/tribuna/38876-deep-sign-universalnaya-tehnologiya-raspoznavaniya-zhestov>
5. https://uk.wikipedia.org/wiki/Захоплення_руху
6. https://ela.kpi.ua/bitstream/123456789/32015/1/Konkina_magistr.pdf
7. <https://newsroom.ucla.edu/releases/glove-translates-sign-language-to-speech>
8. <https://deeptai.org/machine-learning-glossary-and-terms/feed-forward-neural-network>
9. <https://dzone.com/articles/six-types-of-neural-networks-you-need-to-know-about>
10. <https://www.ibm.com/cloud/learn/recurrent-neural-networks>
11. <https://www.reg.ru/blog/stehnfordskij-kurs-lekciya-9-arhitektury-cnn/>
12. <https://habr.com/ru/company/nix/blog/430524/>
13. «Штучні нейронні мережі. Обчислення» М.А. Новотарський, Б.Б. Нестеренко
14. <https://habr.com/ru/post/373245/>
15. «Dactyl Alphabet Modeling and Recognition Using Cross Platform Software» Ю.В. Крак, С.С. Кондратюк
16. «Система розпізнавання жестів на основі нейромережових технологій» А.В. Чуйков, А.М. Вульфін
17. <https://www.bbntimes.com/technology/exploring-the-3-types-of-artificial-neural-networks>

18. «Технологія розпізнавання елементів дактильно-жестової мови» Ю.В.
Крак, Д.В. Шкільнюк