

ІНФОРМАЦІЙНА СИСТЕМА РОЗПІЗНАВАННЯ МОВИ ЖЕСТІВ З ВИКОРИСТАННЯМ ШТУЧНОГО ІНТЕЛЕКТУ

Вінницький національний технічний університет

Анотація

Розроблено інформаційну систему розпізнавання мови жестів з використанням штучного інтелекту. Використано методи глибокого навчання для розпізнавання жестів української жестової мови в реальному часі. Зокрема, застосовано згорткові нейронні мережі для виявлення рук та їх ключових точок. Основна увага приділяється оптимізації архітектури моделі, підбору ефективних наборів даних та забезпеченню високої точності розпізнавання. Розроблена система дозволяє конвертувати жести у текст і озвучувати їх, що сприяє покращенню комунікації для людей з порушенням слуху.

Ключові слова: нейронні мережі, комп'ютерний зір, жестова мова, розпізнавання.

Abstracts

An information system for sign language recognition using artificial intelligence has been developed. Deep learning methods have been applied to recognize Ukrainian Sign Language gestures in real time. Specifically, convolutional neural networks (CNNs) are used to detect hands and identify their key points. The main focus is on optimizing the model architecture, selecting effective datasets, and ensuring high recognition accuracy. The developed system enables the conversion of gestures into text and their vocalization, thereby improving communication for people with hearing impairments.

Keywords: neural networks, computer vision, sign language, recognition.

Вступ

Штучний інтелект та нейронні мережі стали невід'ємною частиною сучасних технологій, дозволяючи автоматизувати складні процеси, аналізувати великі обсяги даних та вирішувати завдання, які раніше потребували людського втручання. Одним із ключових напрямів їх застосування є комп'ютерний зір, що дає змогу розпізнавати об'єкти, жести, міміку та текст, відкриваючи нові можливості для комунікації та взаємодії між людьми й машинами.

Метою роботи є розробка інформаційної системи для розпізнавання жестів української жестової мови в реальному часі. Система використовує методи глибокого навчання та комп'ютерного зору для ідентифікації рук і жестів, конвертує їх у текст і озвучує, що сприяє покращенню комунікації для людей із порушенням слуху.

Структура нейронних мереж

Штучний інтелект (ШІ) є однією з провідних галузей сучасної науки, що спрямована на створення алгоритмів та систем, здатних виконувати когнітивні завдання, подібні до людських. ШІ базується на використанні математичних моделей, машинного навчання та нейронних мереж, що дозволяють аналізувати великі обсяги даних, виявляти закономірності та приймати рішення [1].

Однією з ключових технологій у сфері штучного інтелекту є штучні нейронні мережі (ШНМ). Це обчислювальні моделі, натхненні біологічними нейронами людського мозку. Вони складаються з багатьох взаємопов'язаних шарів нейронів, які проходять процес навчання на основі великих наборів даних. Глибокі нейронні мережі (DNN) та згорткові нейронні мережі (CNN) є найбільш ефективними для аналізу зображень і відео, тоді як рекурентні нейронні мережі (RNN) використовуються для обробки послідовних даних [2].

У даній роботі застосовується згорткова нейронна мережа для ідентифікації жестів. Основний принцип роботи згорткової нейронної мережі (CNN) полягає у застосуванні фільтрів для виявлення важливих ознак на зображенні через операції згортки, активації та пулінгу, після чого використовуються повнозв'язні шари для класифікації зображень на основі цих ознак (рис. 1). Згорткові мережі зазвичай використовуються в задачах комп'ютерного зору, таких як розпізнавання об'єктів, класифікація зображень, виявлення аномалій, тощо. CNN допомагає виділяти основні ознаки, як-от контури і форми рук та пальців, що необхідно для точного розпізнавання жестів [3].

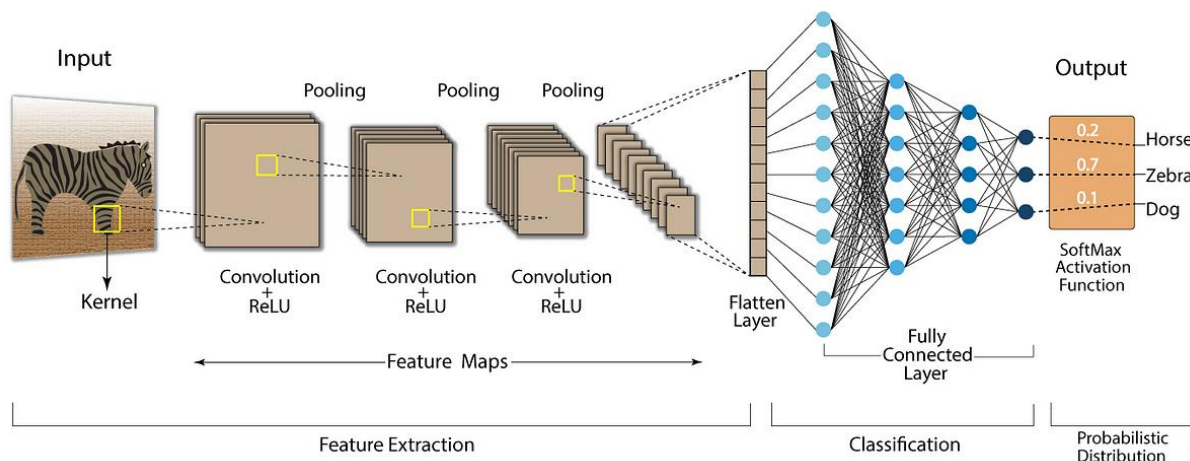


Рисунок 1 – Принцип роботи згорткової нейронної мережі

Нейромережа здатна розпізнавати унікальні характеристики. Для цього система знаходить характеристики базового рівня і на їх основі створює більш абстрактні припущення. При обробці картинки комп'ютер приймає на вхід набір пікселів.

Створення згортки допомагає знайти властивості зображення (лінії, кола, криві). Наприклад, у зображенні кімнати один із фільтрів шукає горизонтальні лінії, другий — вертикальні, а третій — похилі (рис. 2).

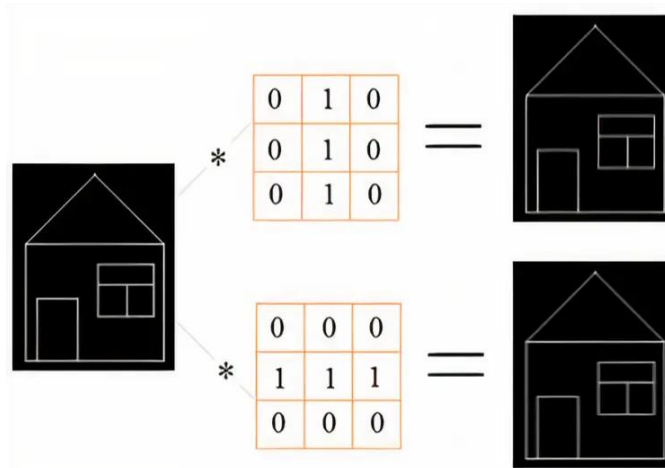


Рисунок 2 – Візуалізація роботи фільтрів

Щоб аналізувати обчислені ознаки на більшому масштабі, розмірність карт ознак скорочують за допомогою пулінгу. Пулінг-шар розбиває картинку, отриману від згорткового шару, на невеликі ділянки і залишає тільки піксель із найбільшим значенням. У результаті кількість пікселів зменшується в кілька разів, а нейрони наступного шару можуть виділяти більш загальні ознаки на зображенні (рис. 3).

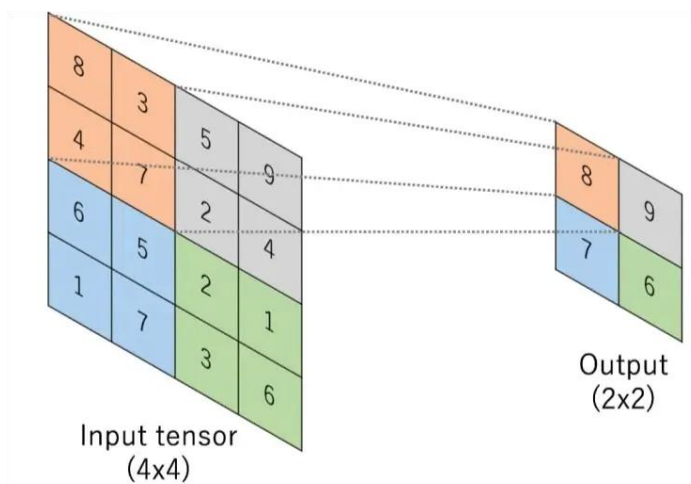


Рисунок 3 – Операція пулінгу

Останній повнозв'язний шар має видати результати нелінійної функції, з якою працювала нейромережа. Усі попередні шари потрібні для опрацювання зображення, тоді як повнозв'язний шар має виокремити ключові характеристики та виконати класифікацію — визначити, що на картинці, або знайти конкретний предмет [4].

Структура інформаційної системи

Роботу інформаційної системи можна представити у вигляді блок-схеми (рис. 4).



Рисунок 4 – Структура інформаційної системи розпізнавання жестів

Методика збору та підготовки даних

Для навчання нейронної мережі розпізнаванню жестів необхідно використовувати комбінований підхід, що включає зображення та відео. Для розпізнавання літер та окремих фраз доцільно застосовувати зображення жестів з різними ракурсами та умовами освітлення. Це забезпечить більш точний кінцевий результат.

Для розпізнавання речень слід використовувати відеоматеріали. Вони дозволять зафіксувати динаміку рухів, відобразити контекст та забезпечать розпізнавання динамічних жестів, що є критично важливим для розуміння складних фраз.

Збір даних включає пошук та завантаження картинок та відеороликів з інтернету, де люди виконують різні жести. Потрібно забезпечити різноманітність даних, використовуючи різні джерела, з різними умовами освітлення та фоном.

Далі дані необхідно конвертувати у формат, придатний для навчання нейронної мережі, наприклад, окремі кадри у форматі PNG або JPG. Розмір кадрів потрібно привести до уніфікованого розміру, нормалізувати яскравість та контрастність, видалити зайві елементи.

Для збільшення різноманітності навчальної вибірки потрібно використовувати методи аугментації даних, такі як повороти, масштабування, зміна освітлення, додавання шуму, зміна швидкості відтворення відео.

Висновки

Запропонована інформаційна система дозволить створити надійний та якісний набір даних для розпізнавання жестів. Основою роботи є комбінований підхід, що поєднує використання зображень статичних жестів та відеоматеріалів з інтернету для динамічних рухів. Додаткові методи аугментації даних, такі як повороти, масштабування, зміна освітлення, додавання шуму та зміна швидкості відтворення відео, значно розширять набір даних, підвищать стійкість моделі до варіацій та покращать її здатність до узагальнення. Це забезпечить необхідну різноманітність даних, що є запорукою ефективного навчання нейронної мережі.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Що таке штучний інтелект простими словами? [Електронний ресурс] – Режим доступу до ресурсу: <https://biznestrendy.com.ua/shcho-take-shtuchnyy-intelekt-prostymy-slovamy/>
2. Що таке нейронна мережа, як вона працює та які завдання може вирішувати [Електронний ресурс] – Режим доступу до ресурсу: <https://mc.today/uk/shho-take-nejronna-merezha/>
3. Використовуємо CNN для обробки зображень. Частина перша [Електронний ресурс] – Режим доступу до ресурсу: <https://dou.ua/forums/topic/48368>
4. Max-Pooling, Combining Channels using 1×1 convolutions, Receptive Field calculation [Електронний ресурс] – Режим доступу до ресурсу: <https://anilbhatt1.tech.blog/2020/02/19/max-pooling-1x1-convolutions-receptive-field-calculation/>
5. Book Chapter. Order Forecasting System for Vehicles Based on Previous Statistics Requests Kyrylashchuk S., Horodetska O., Voitsekhovska O., Zakharchenko S. Lecture Notes on Data Engineering and Communications Technologies, 2024, 219, pp. 116-131.
6. Порівняльний аналіз архітектур для виявлення об'єктів на зображеннях / І. О. Бабійчук, О. С. Городецька // Матеріали LIII науково-технічної конференції факультету інформаційних технологій та комп'ютерної інженерії. Вінниця 2024 р. 3 с. [Електронний ресурс]. Режим доступу: <https://conferences.vntu.edu.ua/index.php/all-fitki/all-fitki-2024/paper/view/20857/17309>.
7. Мартинюк, Т., Войцеховська, О., Городецька, О. і Рижков, А. 2024. Модуль інтеграції вебзастосунків із штучним інтелектом. Інформаційні технології та комп'ютерна інженерія. 59, 1 (Трав 2024), 5–12. DOI:<https://doi.org/10.31649/1999-9941-2024-59-1-5-12>

Городецька Оксана Степанівна – к.т.н., доцент кафедри обчислювальної техніки, Вінницький національний технічний університет, Вінниця, e-mail: horodecka.os@gmail.com

Шпачинська Анастасія Степанівна – студентка групи ІКІ-21б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: n.shpachynska@gmail.com

Horodetska Oksana Stepanivna – Ph.D., Associate Professor of the Computer Engineering Department, Vinnytsia National Technical University, Vinnytsia, e-mail: horodecka.os@gmail.com

Shpachynska Anastasiia Stepanivna – student of group ІКІ-21b, faculty of information technologies and computer engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: n.shpachynska@gmail.com