# Обґрунтування вибору теми/напрямку

Тема/напрямок: «Математичне та програмне забезпечення для розпізнавання та отримання тексту із зображень»

Дана тема була обрана, як тема бакалаврської роботи для напрацювання практичних навичок пов’язаних з:

1. Проектуванням та втіленням архітектури ПЗ, зокрема за допомогою сучасних інформаційних технологій та фреймворків.
2. Побудовою архітектури, відлагодженням, підбором гіперпараметрів нейронної мережі. Коректної побудови циклів навчання/тестування моделі.
3. Аналізом якості роботи мережі за допомогою обраних метрик.
4. Порівняльним аналізом роботи власноруч побудованої та відлагодженої нейронної мережі з отриманими раніше варіантами та більш досконалими, «state-of-the-art» моделями.

Оскільки в сучасному світі є потреба в автоматизації:

1. Ведення документообігу та, зокрема, сканування документів, з мінімальною похибкою.
2. Вилучення тексту написаного іноземною мовою в повсякденному житті (текст на вивісках, текст надрукований чи написаний на товарних пакунках)
3. Вилучення тексту в рамках наукової діяльності (транскрипція як древніх історичних текстів, написаних латиною, шумеро-аккадським клинописом, так і відносно сучасних, написаних за останні 100 років) для подальшого перекладу.
4. Вилучення тексту в рамках професійної діяльності (розпізнавання написаних власноруч викладачем оцінок/поміток на паперовій роздруківці перевірених тестів, сканування чеків, тощо)

Задача розпізнавання та вилучення як друкованого, так і рукописного тексту залишається актуальною. Особливого значення в рамках вирішення цієї задачі набуває досягнення високого рівня точності роботи моделей, оскільки ціна помилки може мати вирішальне значення. Помилки в розпізнаванні можуть призводити до втрати важливих даних, юридичних неточностей або навіть фінансових збитків, особливо в сферах документообігу, архівування та автоматизації бізнес-процесів. Саме тому сучасні системи оптичного розпізнавання символів (OCR) активно розвиваються, використовуючи методи глибокого навчання, що дозволяють підвищити точність і адаптивність алгоритмів до різних умов.

Зокрема, задача вилучення рукописного тексту видається більш складною, оскільки друкований текст – за рахунок стандартизації шрифтів – є більш легким для розпізнавання. На початкових етапах розвитку OCR-технологій для друкованого тексту використовували методи по символьного зіставлення та шаблонного порівняння, однак із розвитком машинного навчання стали застосовувати нейронні мережі, що суттєво покращило точність розпізнавання.

В свою чергу, рукописний текст може значно відрізнятися залежно від почерку кожної людини, змінюватися в межах одного документа, містити скорочення або нестандартні символи. Це створює додаткові виклики для систем розпізнавання, які мають враховувати контекст та особливості стилю написання, що представляє цікавість для аналізу.

Сучасні підходи до вилучення рукописного тексту використовують рекурентні та згорткові нейронні мережі, що дозволяє підвищити точність навіть у випадках нестандартного почерку або низької якості сканованого зображення. Вивчення цих підходів, зокрема застосування різних конфігурацій нейронних мереж на практиці для вирішення цієї задачі представляє інтерес для розгляду ще й з точки зору, як зміна гіперпараметрів та архітектури нейронної мережі вплине на її працездатність та якість роботи на різних тестувальних/валідаційних прикладах.

Постановка задачі/мета розробки і кінцевий результат

Постановка задачі:

1. Провести аналіз існуючих підходів до розпізнавання рукописного тексту, зокрема методів машинного навчання та нейронних мереж.
2. Дослідити математичні моделі попередньої обробки зображень, що покращують якість текстових даних (фільтрація шуму, бінаризація, сегментація тощо).
3. Розробити модель розпізнавання рукописного тексту на основі нейронної мережі, яка враховує особливості почерку та можливі варіації написання, зашумленість та низьку якість зображення.
4. Реалізувати програмне забезпечення, що забезпечує вилучення та розпізнавання слів із зображень.
5. Провести тестування та оцінити ефективність розробленого підходу на реальних даних, зокрема створених власноруч.
6. Провести порівняльний аналіз впливу різних архітектур/конфігурацій нейронних мереж на якість розпізнавання рукописного тексту.

Кінцевий результат:

* Розроблені моделі нейронної мережі CRNN (згорткової рекурентної нейронної мережі), здатні до розпізнавання та вилучення рукописних слів, написаних англійською мовою.
  + На вхід моделі подається зображення (фотографія) рукописного слова, наприклад *«example»*.
  + На виході модель видає слово в електронному вигляді, з кодуванням символів – «utf-8».
* Розроблене програмне забезпечення, обгортка для тренування та аналізу якості роботи нейронної мережі. Структура частини проєкту орієнтованої на це має бути організованою та легкою для розуміння й використання, оскільки модель не може бути досконалою, її архітектура буде постійно дороблюватися, а гіперпараметри змінюватися.

Процес навчання моделі має бути оптимізованим, для мінімізації часових та матеріальних витрат.

* Розроблене програмне забезпечення – REST API та веб-застосунок з користувацьким інтерфейсом для легкої роботи користувача з моделлю. ПЗ має передбачати функцію завантаження зображення з рукописним текстом (словом) через форму, та отриманням транскрипції зображеного слова в відповідному текстовому полі.

# Огляд та якісний аналіз існуючих рішень/подібних реалізацій

Системи оптичного розпізнавання символів (OCR) є важливим компонентом автоматизації процесів обробки текстових даних. Вони використовуються для перетворення друкованого та рукописного тексту на цифровий формат, що дозволяє зберігати, редагувати та аналізувати інформацію. Розвиток OCR-технологій відбувався в кілька етапів, від простих алгоритмів шаблонного порівняння до складних моделей глибокого навчання.

**Традиційні методи OCR**

Перші системи OCR базувалися на технологіях символьного розпізнавання, використовуючи шаблонне порівняння символів із заздалегідь визначеними еталонними зразками. Основні особливості таких систем:

* Висока залежність від типу шрифту.
* Потреба у великій кількості зразків для навчання.
* Відносно низька адаптивність до змін у тексті та зображенні.

Популярні програмні продукти, такі як **Tesseract OCR** (розроблений HP, зараз підтримується Google), використовували алгоритми сегментації зображення, виділення контурів символів та їх порівняння з вбудованими шаблонами.

**Сучасні методи OCR на основі нейронних мереж**

Сучасні підходи OCR широко застосовують машинне навчання, особливо згорткові нейронні мережі (CNN) та рекурентні нейронні мережі (RNN), а також комбінації CNN + LSTM. Завдяки цим технологіям OCR-системи можуть розпізнавати текст із високою точністю навіть у складних умовах, таких як:

* Погана якість зображення.
* Варіативність шрифтів та рукописного тексту.
* Присутність шуму або артефактів на зображенні.

**Приклади сучасних рішень:**

1. **Tesseract OCR** (Google) – підтримує нейронні мережі для розпізнавання друкованого тексту, але його ефективність для рукописного тексту залишається обмеженою.
2. **EasyOCR** – бібліотека на основі PyTorch, яка підтримує розпізнавання як друкованого, так і рукописного тексту з використанням CRNN (Convolutional Recurrent Neural Network).
3. **Google Cloud Vision OCR** та **Microsoft Azure OCR** – хмарні сервіси, які пропонують потужні алгоритми розпізнавання тексту, проте є комерційними рішеннями.
4. **PaddleOCR** – гнучка та ефективна бібліотека, яка використовує глибоке навчання для розпізнавання друкованого та рукописного тексту на багатьох мовах.

**Розпізнавання рукописного тексту**

Рукописний текст має значно більшу варіативність у порівнянні з друкованим, що створює складнощі для традиційних OCR-систем. Найбільш ефективні підходи до розпізнавання рукописного тексту включають використання гібридних нейронних архітектур:

* **CRNN (Convolutional Recurrent Neural Network)** – поєднання згорткових шарів для обробки зображення та рекурентних шарів для аналізу послідовностей символів.
* **Transformers (Vision Transformers, ViT)** – новітні моделі, які використовують механізм уваги (attention mechanism) для розпізнавання структурних особливостей тексту.
* **Seq2Seq (Sequence-to-Sequence)** – використовується в поєднанні з CTC-loss (Connectionist Temporal Classification) для транскрибування тексту без потреби у попередній сегментації.

**Комерційні та Open-Source рішення**

Вивести на презентацію

|  |  |  |  |
| --- | --- | --- | --- |
| **Назва OCR-системи** | **Тип (Open-Source/Комерційний)** | **Підтримка рукописного тексту** | **Використані технології** |
| Tesseract OCR | Open-Source | Обмежена | CNN + LSTM |
| EasyOCR | Open-Source | Так | CRNN |
| PaddleOCR | Open-Source | Так | CNN + RNN + Attention |
| Google Cloud Vision OCR | Комерційний | Так | Глибинне навчання |
| Microsoft Azure OCR | Комерційний | Так | CNN + Transformers |

Сучасні OCR-рішення забезпечують високу точність розпізнавання друкованого тексту, але розпізнавання рукописного тексту залишається складним завданням через його нестандартність. Використання CRNN, Seq2Seq, CTC-loss та Transformer-моделей значно покращує результати, що дозволяє автоматизувати завдання обробки рукописного тексту у різних сферах. Подальші дослідження спрямовані на підвищення точності, адаптивності моделей до різних почерків та оптимізацію швидкості їх роботи.

Сформувати ще одну таблицю з порівнянням різних архітектур нейронних мереж CRNN та їх перформансу з бумажок, які вже є і подивитися можливо інші бумажки. Якісний аналіз комерційних та опен сорс рішень.