

**ПОРІВНЯЛЬНИЙ АНАЛІЗ СИСТЕМИ ОПТИЧНОГО
РОЗПІЗНАВАННЯ ТЕКСТУ TESSERACT OCR З МОДЕЛЯМИ НА
БАЗІ ГЛИБОКИХ НЕЙРОННИХ МЕРЕЖ**

Абросімов Є.О., Зелений О.П.

e-mail: yehor.abrosimov@nure.ua, oleksandr.zeleniy@nure.ua

Харківський національний університет радіоелектроніки, каф. МСТ
м. Харків, Україна

Optical Character Recognition (OCR) systems have evolved from traditional approaches like TesseractOCR, which relies on classical image processing, to modern deep learning frameworks that integrate convolutional and recurrent neural networks. While TesseractOCR, a very popular system for solving tasks of Optical Character Recognition, it performs well on standardized text, but struggles with non-standard fonts, handwritten text, and low-quality images, whereas contemporary end-to-end neural models utilize data augmentation and language model post-processing to enhance accuracy. Transitioning from conventional OCR methods to deep learning – based solutions might significantly improve performance across diverse conditions, advancing document digitization and text recognition.

Оптичне розпізнавання тексту (OCR) є ключовою технологією для оцифрування документів і автоматизації обробки друкованого тексту. Однією з найвідоміших систем у цій сфері є TesseractOCR – класичне рішення, яке стало стандартом для розробників більш сучасних підходів [4]. Хоча TesseractOCR демонструє високу точність на друкованих документах і якісних зображеннях, його універсальність обмежена, особливо за наявності рукописних текстів або зображень низької якості [1, 2, 5].

Актуальність обраної теми дослідження полягає в тому що з розвитком цифрових технологій та необхідністю швидкої обробки великого обсягу текстової інформації OCR-системи стають все більш затребуваними. Сучасні потреби включають розпізнавання тексту у реальному часі, обробку багатомовних документів та адаптацію до неякісних зображень. Тому метою дослідження є аналіз сучасних підходів до OCR, їхньої ефективності та можливостей інтеграції класичних алгоритмів і нейронних мереж для підвищення точності розпізнавання.

TesseractOCR застосовує послідовність класичних алгоритмів: від попередньої обробки (бінаризація, усунення шумів) до сегментації й розпізнавання символів на основі шаблонів чи статистичних моделей [4]. Цей підхід працює ефективно для стандартизованих форматів, проте втрачає точність за складніших умов, як-от незвичні шрифти або низька контрастність. Сучасні системи повного циклу розпізнавання (наприклад, CRNN, TrOCR) об'єднують згорткові й рекурентні нейронні мережі, що дає змогу водночас отримувати ознаки із та обробляти послідовності символів

[2]. За рахунок урахування контексту й інтеграції мовних моделей знижується кількість помилок, спричинених шумом або візуально подібними символами [3]. Такі моделі краще адаптуються до різноманітних типів зображень, перевершуючи TesseractOCR за точністю та гнучкістю.

Нові технології, як-от UPOCR (Unified Pixel-Level OCR Interface), спрямовані на забезпечення гнучкішого підходу до розпізнавання тексту, працюючи безпосередньо на рівні пікселів [1]. Вони дозволяють підвищити стійкість системи до варіацій у форматі зображень та різних типів текстів, що робить OCR ще більш універсальним (рисунком 1). Основна ідея UPOCR (Unified Pixel-Level OCR Interface) полягає в тому, щоб створити універсальний підхід до розпізнавання тексту, який працює на рівні окремих пікселів. На відміну від традиційних OCR-систем, що спираються на сегментацію символів та аналіз контурів, UPOCR безпосередньо інтерпретує зображення як суцільну текстову структуру.

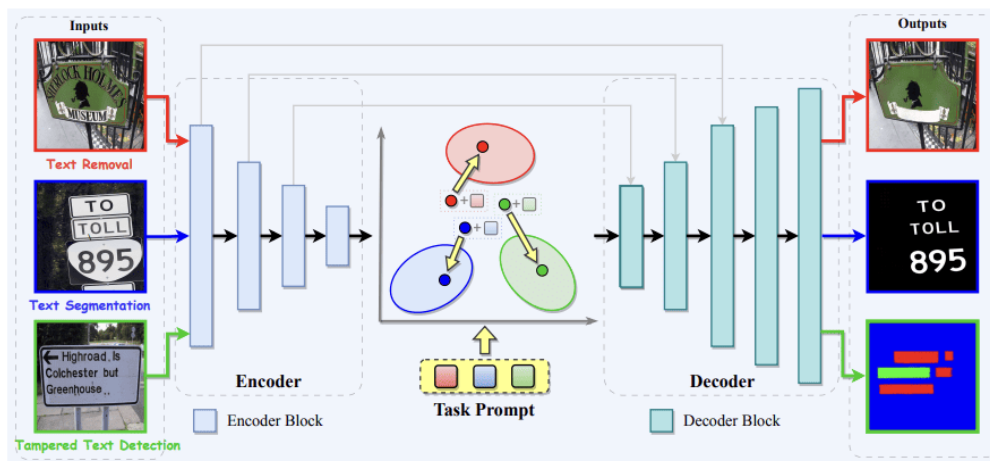


Рисунок 1 – Архітектура UPOCR

Крім того, інтелектуальні методи попередньої обробки, як data augmentation чи поліпшення якості зображень, підвищують стійкість систем до реальних викликів: змін освітлення, шуму та невисокої роздільної здатності [1, 2]. Поєднання класичних алгоритмів TesseractOCR із глибокими моделями створює гібридні рішення, що оптимізують вхідні дані для нейронних мереж, одночасно зберігаючи високу точність і масштабованість [3].

Актуальним напрямком є забезпечення роботи систем OCR у реальному часі на мобільних пристроях чи в середовищах з обмеженими ресурсами. Transfer learning і адаптивне навчання дають змогу переналаштовувати великі моделі на конкретні завдання, зберігаючи точність навіть за низькоякісних вхідних даних. Підхід self-supervised learning відкриває шляхи до використання неанотованих зображень, що надзвичайно важливо для подальшого прогресу OCR [2, 3].

Також важливою складовою OCR-систем є робота з багатомовними текстами. Нейромережеві підходи дозволяють моделювати особливості різних алфавітів та писемностей, забезпечуючи точніше розпізнавання складних символів та ієрогліфічних систем [3]. Однак такі моделі потребують значних обчислювальних ресурсів і великих наборів даних для навчання, що може бути обмеженням для малих компаній чи проєктів. Розвиток OCR також включає поліпшення точності у розпізнаванні рукописного тексту. Наприклад, PP-OCRv3, розроблений як легка OCR-система, дозволяє ефективно обробляти тексти низької якості, що робить її особливо корисною для мобільних пристроїв і вбудованих систем [2].

Отже, базова система TesseractOCR, хоча й залишається ефективною у певних випадках, має суттєві обмеження в універсальності та адаптивності. Сучасні методи глибокого навчання значно перевершують її за точністю, роботою з різноманітними форматами й багатомовністю. Перехід до новітніх нейронних моделей є ключовим для підвищення надійності та результативності OCR-технологій. Зокрема, використання CRNN, TrOCR та інших глибоких моделей дозволяє ефективно працювати з різними форматами тексту, включаючи рукописні документи та складні друковані тексти.

Варто зазначити, що майбутній розвиток OCR-технологій, ймовірно, зосередиться на покращенні ефективності для мобільних пристроїв, інтеграції self-supervised learning та використанні hybrid-методів, що поєднують класичні алгоритми з нейромережами. Це відкриє нові можливості для застосування OCR у сферах автоматизації документообігу, перекладу текстів та архівування історичних документів. Крім того, подальше вдосконалення методів розпізнавання рукописного тексту й адаптація OCR-систем до роботи у складних умовах забезпечать ще більшу універсальність та точність технології.

Список використаних джерел:

1. Peng, D., & Yang, Z. (2023). UPOCR: Towards Unified Pixel-Level OCR Interface. // ICML'24: Proceedings of the 41st International Conference on Machine Learning. Article No.: 1634, P. 40271-40294.
2. Li, C., Liu, W., & Guo, R. (2022). PP-OCRv3: More Attempts for the Improvement of Ultra Lightweight OCR System. ArXiv. <https://doi.org/10.48550/arXiv.2206.03001>
3. Dilmegani, C. (2025, лютий). State of OCR in 2025: Is it dead or a solved problem? AIMultiple. <https://research.aimultiple.com/ocr-technology/>
4. R. Smith. (2007). An Overview of the Tesseract OCR Engine. Ninth International Conference on Document Analysis and Recognition (ICDAR 2007), P. 629-633.
5. Абросімов Є.О. Дослідження сучасних методів аугментації текстових даних // 28-й Міжнародний молодіжний форум «Радіоелектроніка та молодь у XXI столітті». Зб. матеріалів форуму. Т. 6., Харків: ХНУРЕ. 2024. С. 35-36.