

## Особливості використання нейронних мереж для автоматичного розпізнавання рукописного тексту

Вінницький національний технічний університет

### Анотація

У роботі розглянуто проблему автоматичного розпізнавання рукописного тексту. Запропоновано та реалізовано гібридну архітектуру, що поєднує згорткові нейронні мережі для виділення візуальних ознак, рекурентні нейронні мережі для аналізу послідовностей та метод конекціоністської часової класифікації для навчання без посимвольної розмітки. Проведено експериментальне дослідження на еталонному наборі даних IAM, досягнуто показник помилки символів на рівні 14,6%.

**Ключові слова:** автоматичне розпізнавання рукописного тексту; згорткові нейронні мережі; рекурентні нейронні мережі; конекціоністська часова класифікація; комп'ютерний зір; глибоке навчання.

### Abstract

The paper considers the problem of automatic handwritten text recognition. A hybrid architecture combining convolutional neural networks for visual feature extraction, recurrent neural networks for sequence analysis, and connectionist temporal classification for training without character-wise segmentation is proposed and implemented. Experimental research was conducted on the IAM benchmark dataset, achieving a Character Error Rate of 14.6%.

**Keywords:** automatic handwritten text recognition; convolutional neural networks; recurrent neural networks; connectionist temporal classification; computer vision; deep learning.

### Вступ

Автоматичне розпізнавання рукописного тексту залишається однією з найскладніших задач комп'ютерного зору через високу варіативність почерків, наявність шумів та деградацію історичних документів. Класичні методи на основі прихованих марковських моделей поступово витісняються підходами глибинного навчання, які дозволяють створювати «end-to-end» системи [1].

Актуальність теми зумовлена зростаючим попитом на цифровізацію архівів, обробку медичних записів та автоматизацію документообігу, де ручне перенесення даних є надмірно трудомістким. Метою роботи є дослідження ефективності гібридної архітектури нейронної мережі для розпізнавання англійських рукописних рядків без попередньої сегментації на символи, а також аналіз технічних факторів, що впливають на точність транскрипції.

### Результати дослідження

Для вирішення задачі було обрано архітектуру типу CRNN (Convolutional Recurrent Neural Network, згорткова рекурентна нейронна мережа), яка складається з трьох функціональних блоків [2].

1. Візуальний енкодер (CNN): 4-блокова згорткова мережа, що перетворює вхідне зображення розміром 64 на 800 пікселів у послідовність карт ознак. У кожному блоці застосовано функції активації ReLU та шари нормалізації для прискорення збіжності, а також операції «пулінгу» для зменшення розмірності. Використання згорток дозволяє системі бути стійкою до локальних деформацій символів.

2. Контекстний декодер (Bi-LSTM): два шари двонаправленої довгої короткострокової пам'яті (256 прихованих станів). Цей блок враховує контекст як зліва направо, так і справа наліво, що є критично важливим для розрізнення схожих за написанням символів.

3. Блок вирівнювання (СТС): використання конекціоністської часової класифікації дозволяє навчати мережу на парах «зображення – текст» без необхідності трудомісткої розмітки координат кожного символу, що вирішує фундаментальну проблему сегментації злитого рукописного тексту [3].

Загальна кількість параметрів реалізованої моделі становить близько 4,6 млн, що забезпечує баланс між виразною здатністю мережі та обчислювальною ефективністю.

Як еталонний набір даних використано IAM Handwriting Database, що містить зразки почерку понад 400 авторів [4]. Для підвищення узагальнюючої здатності моделі реалізовано конвеєр попередньої обробки даних: вирівнювання нахилу (deskew), нормалізація висоти рядка та аугментація. Аугментація

включала додавання гаусівського шуму, розмиття, зміну контрасту та еластичні деформації, що застосовувалися випадковим чином під час навчання.

Навчання проводилося протягом 50 епох із використанням оптимізатора Adam та розміром міні-пакету (batch size) 32 зразки. Час навчання на графічному апаратному прискорювачі склав близько 6 годин. За результатами експериментів, найкраща точність була досягнута на 44-й епосі. Фінальні метрики на тестовій вибірці (з використанням жадібного декодера) склали:

- частота помилок у розпізнаванні символів – Character Error Rate (CER): 15,24% (найкращий результат на валідації – 14,60%);
- частота помилок у розпізнаванні слів – Word Error Rate (WER): 36,82%;
- точність – асигурація (повне співпадіння рядків): 63,18%.

Аналіз помилок показав, що точність розпізнавання корелює з довжиною слів: для коротких слів (1–3 символи) точність сягає 84%, тоді як для слів довжиною понад 10 символів вона падає до 50% і нижче. Аналіз матриці плутанини виявив, що левова частка помилок припадає на візуально подібні пари символів, такі як «а»-«о», «г»-«п», «l»-«t», а також на пропуски вузьких символів (наприклад, «i») у швидкому почерку.

### Висновки

Розроблена інформаційна технологія на базі CNN та Bi-LSTM продемонструвала свою ефективність для задач офлайн-розпізнавання рукописного тексту. Застосування алгоритму CTC дозволило уникнути складної сегментації символів. Отриманий результат (CER ~15%) свідчать про придатність системи для індексації архівів. Водночас показник WER залишається високим, що вказує на необхідність вдосконалення мовної складової. Для повноцінної автоматичної оцифровки необхідно впровадження пост-обробки з використанням словників або N-грамних моделей. Перспективним напрямком подальших досліджень також є перехід до трансформерних архітектур та використання механізмів уваги, які здатні краще моделювати довгострокові залежності у тексті.

### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-Based Learning Applied to Document Recognition // Proceedings of the IEEE. 1998. Vol. 86, No. 11. P. 2278–2324.
2. Graves A., Liwicki M., Fernández S. et al. A novel connectionist system for unconstrained handwriting recognition // IEEE Trans. Pattern Anal. Mach. Intell. 2009. Vol. 31, № 5. P. 855–868.
3. Graves A., Fernández S., Gomez F. Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks // Proc. 23rd Int. Conf. on Machine Learning (ICML). 2006. P. 369–376.
4. Marti U.-V., Bunke H. The IAM-database: an English sentence database for off-line handwriting recognition // Int. J. Document Anal. Recognit. 2002. Vol. 5, No. 1. P. 39–46.

**Ищенко Гліб Андрійович** – студент групи 6KN-22б, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: [glebishshenko1000@gmail.com](mailto:glebishshenko1000@gmail.com)

**Хмельівський Юрій Сергійович** – асистент кафедри комп'ютерних наук, факультет інтелектуальних інформаційних технологій та автоматизації, Вінницький національний технічний університет, Вінниця, e-mail: [yu.khmelivskiy@vntu.edu.ua](mailto:yu.khmelivskiy@vntu.edu.ua)

**Ishchenko Hlib A.** – student of group 6KN-22b, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: [glebishshenko1000@gmail.com](mailto:glebishshenko1000@gmail.com)

**Khmelivskiy Yuriy S.** – Assistant Professor of the Department of Computer Sciences, Faculty of Intelligent Information Technologies and Automation, Vinnytsia National Technical University, Vinnytsia, e-mail: [yu.khmelivskiy@vntu.edu.ua](mailto:yu.khmelivskiy@vntu.edu.ua)