

# МЕТОД БАГАТОПРОФІЛЬНОГО РОЗПІЗНАВАННЯ РУКОПИСНИХ ТЕКСТІВ OFFLINE

Вінницький національний технічний університет

## *Анотація*

*Досліджено основні проблеми та методи багатoproфільного offline-розпізнавання рукописного тексту. Запропоновано новий метод класифікації слів на базі згорткової нейронної мережі з визначеним словником.*

**Ключові слова:** рукописний текст, штучна нейронна мережа, ШНМ, згорткова, розпізнавання зображень, оптичне розпізнавання, OCR, класифікація.

## *Abstract*

*Researched main problems and methods of multi-profile offline handwritten text recognition. Proposed a new method for words classification based on convolutional neural network with determined dictionary.*

**Keywords:** handwritten, artificial neural network, ANN, convolutional, image recognition, optical recognition, OCR, classification.

## Вступ

Останнім часом стали популярними системи, що використовуються для розпізнавання рукописного тексту онлайн. Оскільки такі системи збирають іншу інформацію, окрім власне зображення тексту, точність роботи значно збільшується. Також система може пристосовуватися до почерку конкретної людини. При офлайн-розпізнаванні, коли обробляються статичні документи із почерком різних людей, це не виявляється можливим.

Саме тому проблема розпізнавання рукописного тексту офлайн є досить актуальною. Навіть у випадку рукодрукованого тексту, коли літери пишуться окремо одна від одної (без з'єднань), без зайвих артефактів (плями, неточності сканування, освітлення, фон і т.д.), точність розпізнавання зазвичай сягає 80 – 90%. Це досить низький показник, оскільки на кожному сторінку такого тексту приходиться кілька десятків помилок. Якщо ж казати про повноцінний рукописний текст, проблема не вирішена зовсім.

Розглянутий метод розпізнавання пропонує часткове вирішення питання розпізнавання рукописного тексту шляхом класифікації слів з обмеженого словника за допомогою згорткової нейронної мережі.

## Рукописний текст і методи його розпізнавання

Рукописний текст містить алфавіт букв, розділювальні знаки. Фундаментальна властивість тексту, що робить можливим спілкування, полягає в тому, що відмінності між різними символами значніші, ніж відмінності між різними написами того самого символу.

Почерк складається з тимчасової послідовності штрихів, тобто руху пера від дотику до паперу до його підняття. Символи рукописного тексту, як правило, формуються послідовно, один символ завершується перед початком наступного, а букви йдуть за певним просторовим порядком, наприклад, зліва направо. Але є і виключення – наприклад, в англійській мові крапки над «і» та пересічення на «t» як правило, затримуються. Спочатку пишеться основна частина слова, а потім слово завершується написанням крапок або запятих.

У всіх букв є статичні та динамічні властивості. Статичні відмінності можуть полягати, наприклад, у розмірі або формі символу. Відмінності в динаміці можуть полягати у кількості штрихів та їх порядку. Англійська мова може мати більше значимі зміни напрямку штриха, ніж китайська. В англійській мові також більше варіантів у присутності або відсутності повторних штрихів. Повторний штрих - це штрих у місці, де що-небудь вже написано, зазвичай він робиться, щоб уникнути піднімань пера ручки.

Найбільші проблеми у розпізнаванні рукописного тексту – це ті, що викликають труднощі у людей при спробі прочитати навіть власний почерк. По-перше, той факт, що більшість символів можуть

бути написані по-різному. Також рідко можна зустріти двох людей з однаковим почерком. Ця задача пов'язана з відмінністю шрифтів у класичній задачі розпізнавання тексту. Але на відміну від шрифтів, кожна буква в тексті однієї людини може мати інший стиль залежно від контексту, в якому здійснюється написання навколишніх букв, та багатьох інших факторів. Щоб впоратися з цією проблемою, безліч систем містять в собі компонент, який сам навчається отримуваному почерку, розрізняє користувачів, і застосовує ці дані при прийнятті рішення.

По-друге, кілька символів часто мають однаковий вигляд, або практично зовсім не відрізняються у почерку однієї людини. Більше того, деякі букви можуть бути написані необережно і об'єктивно бути схожими на абсолютно інші букви. Ця складність зазвичай долається шляхом додавання системи розпізнавання тексту та використання словника, в якому програма може знайти сумнівне слово, і таким чином можна уникнути неоднозначності.

Найбільш універсальним підходом до вирішення завдання розпізнавання рукописного тексту є нейромережевий. Основні переваги нейронних мереж полягають у можливості навчатися самостійно і автоматично на основі вибірки, бути продуктивними на зашумлених даних, можливості паралельної реалізації і можливості бути ефективними інструментами для обробки великих баз даних.

У цьому підході існує безліч різних методів. Найпопулярнішими можна назвати нечіткі нейронні мережі, мережу Хемінга, мережу Хопфілда, самоорганізаційну карту Кохонена та багато інших [1].

Основною перевагою застосування нейронних мереж та машинного навчання взагалі є алгоритми, що навчаються. Вони поділяються на дві значні групи: алгоритми, що навчаються «з учителем» та «без учителя» [2]. Відмінність полягає у наявності в першому випадку зазделегідь відомих правильних результатів, на яких система може навчитися узагальненню. У випадку навчання без учителя система націлена на виокремлення груп екземплярів на основі спільних признаков.

### Результати дослідження

Метод розпізнавання рукописного тексту, що розглядається у даній роботі, має наступні особливості:

- розпізнавання окремих слів незалежно від контексту;
- окремий модуль для розбиття на слова, заснований на спеціальних алгоритмах обробки зображень;
- зазделегідь відомий фіксований словник – можливе розпізнавання слів лише з цього словника;
- модуль для розпізнавання слів на растровому зображенні на базі згорткової нейронної мережі.

Запропонований алгоритм розбиття зображення тексту на окремі зображення слів складається з наступних кроків:

- 1) розмити зображення (фільтром Гауса);
- 2) бінаризувати зображення;
- 3) видалити шуми;
- 4) згладити області з високою концентрацією темних пікселів (при цьому слова перетворюються на чорні «плями»);
- 5) знайти координати чорних областей та зберегти окремі зображення за цими координатами.

Даний підхід не є ідеальним і потребує подальшого допрацювання, але для базової реалізації методу розпізнавання він є достатнім. Також цілком можливо повністю його замінити на інший підхід – головне, отримати правильний результат на цьому етапі.

Далі кожне зображення слова окремо розпізнається класифікатором на базі згорткової нейронної мережі. Останнім часом згорткові нейронні мережі успішно застосовуються для розпізнавання зображень. За допомогою них було досягнуто точність понад 99.87% на базі даних MNIST, що наразі є найвищим значенням з усіх досягнутих [3]. На рисунку 1 наведено приклад архітектури такої мережі. Можливі інші варіації кількості та послідовності шарів, кількості нейронів у них тощо.

Вихідний шар нейронної мережі складається з такої кількості нейронів (у даному прикладі – 150), що відповідає кількості слів у словнику. Кожний вихідний нейрон відповідає одному зі слів, і завдяки використанню нормалізованої експоненційної функції (анг. “softmax”) значення на його виході позначає ймовірність знаходження даного слова на зображенні.

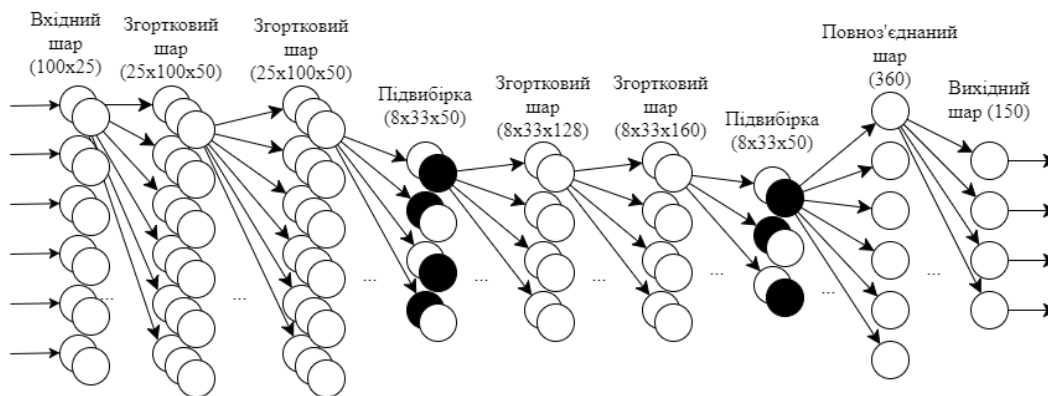


Рисунок 1 – Приклад архітектури згорткової нейронної мережі для задачі розпізнавання

Розмір словника можливо збільшувати, але з обмеженнями. Чим більше вихідних нейронів використовується, тим складніше тренувати модель і тим нижче точність розпізнавання. Таким чином, при збільшенні словника до кількох тисяч слів переваги даного методу нівелюються. Але він є цілком ефективним для застосування у вузькоспеціалізованій області для розпізнавання обмеженої кількості слів (до 2-4 тисяч) – наприклад, у медичинській сфері для розпізнавання лікарських рецептів.

### Висновки

Досліджено особливості та основні проблеми у розпізнаванні рукописного тексту offline. Хоча більшість існуючих методів розпізнавання полягають у розбитті тексту на слова, слів – на літери, та у розпізнаванні кожної літери (або складу) окремо, розглянутий метод пропонує розпізнавання слів цілком. Метод базується на згортковій нейронній мережі, що слугує для класифікації зображень слів за визначеним словником.

Основні переваги розглянутого методу полягають у простоті архітектури та відсутності потреби у значній кількості тренувальних зразків при збереженні високої точності розпізнавання. Очевидним недоліком є обмеженість словника кількома тисячами варіантів, що робить метод можливим до застосування лише у вузькоспеціалізованих галузях.

### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Zhang G. P. Neural networks for classification: a survey. / Zhang G.P. – IEEE Transactions on Systems, Man, and Cybernetics, 2000. – Part C: Applications and Reviews, 30(4). – p. 451–462.
2. Уоссермен Ф. Нейрокомпьютерная техника: Теория и практика: Пер. на русский язык / Ю. А. Зуев, В. А. Точенов. – 1992. — 184 с.
3. Multi-column deep neural networks for image classification. / Ciresan, Dan; Meier, Ueli; Schmidhuber, Jürgen. – IEEE Conference on Computer Vision and Pattern Recognition. – New York, NY: Institute of Electrical and Electronics Engineers (IEEE), 2012. – с. 3642–3649. ISBN 978-1-4673-1226-4.

**Гринько Олексій Олексійович** — студент групи ІКІ-16м, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: grinko.alexey@gmail.com

Науковий керівник: **Захарченко Сергій Михайлович** — канд. техн. наук, доцент кафедри обчислювальної техніки, Вінницький національний технічний університет, м. Вінниця

**Grynko Oleksii O.** — Department of Informational Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, email: grinko.alexey@gmail.com

Supervisor: **Zakharchenko Sergii M.** — Cand. Sc. (Eng.), Assistant Professor of Computer Technology, Vinnytsia National Technical University, Vinnytsia