

РОЗПІЗНАВАННЯ СИМВОЛІВ ТЕКСТОВИХ ДОКУМЕНТІВ ІЗ ВИКОРИСТАННЯМ НЕЙРОННОЇ МЕРЕЖІ

Вінницький національний технічний університет

Анотація

Запропоновано підхід по розпізнаванню текстових документів із використанням згорткової нейронної мережі.

Ключові слова: розпізнавання друкованих символів, згорткова нейронна мережа, навчання мережі.

Abstract

An approach to recognizing text documents using a convolutional neural network is proposed

Keywords: recognition of printed characters, convolutional neural network, network learning.

Вступ

Одним із напрямів розпізнавання образів є їх реалізація у системах оптичного розпізнавання текстів (Optical Character Recognition, OCR-системах), що являє собою чудовий приклад використання технологій штучного інтелекту. Система OCR реалізує автоматичне розпізнавання при допомозі спеціально розроблених програм зображень символів друкованого або ж рукописного тексту й переведення його в формат, який можна використати для подальшого оброблення редакторами текстів або текстовими процесорами та інше [1]. Але набагато складнішим стає це завдання при здійсненні автоматизованого розпізнавання друкованих текстів із дефектами, яке на даний час поки ще не має остаточного вирішення є однією складних задач розпізнавання даних. Розгляду одного із підходів по виділенню та розпізнавання текстових документів присвячений даний матеріал.

Розпізнавання текстових символів

На теперішній час сформувалася певна послідовність обробки сканованого текстового документу для виділення і розпізнавання тексту. Ця послідовність включає такі етапи: виділення фрагменту тексту, попередня обробка цієї частини документу, сегментація рядків, слів та у багатьох випадках і символів, а потім приступають до самої важливої роботи по обробці документів – власне розпізнавання із використанням класифікаторів [2].

Для розпізнавання символів у системах розпізнавання текстів використовуються такі класифікатори, як растровий, ознаковий та структурний [3]. Растровий класифікатор порівнює символ з набором еталонів, по черзі накладаючи зображення одне на одне. Еталонами виступають спеціально підготовлені зображення, кожне з них об'єднує в собі контури багатьох варіантів написання того або іншого символу. Ознаковий класифікатор висуває гіпотези, виходячи із ступеня схожості параметрів символу з еталонними значеннями. Оперує певними числовими ознаками, такими, наприклад, як довжина периметра, кількість чорних точок в різних областях або уздовж різних напрямів, тощо. Структурний класифікатор проводить структурний аналіз символу, розкладаючи останній на елементарні складові (відрізки, дуги, кола, точки) і формуючи точну схему аналізованого знаку. Потім отримана схема у вигляді структурного опису символів порівнюється з еталоном. Цей класифікатор працює повільніше растрового і ознакового, зате відрізняється високою точністю.

У останній час ці функції класифікаторів все частіше покладають на нейронні мережі, які у залежності від їх структури та налаштування поєднують у собі характеристики таких класифікаторів та дозволяють виконувати процес розпізнавання символів текстових документів [4, 5]. Тому для

розробки програмного продукту по розпізнаванню текстових документів було вирішено використати згорткову нейронну мережу. Такою мережею було вибрано згорткову нейронну мережу типу ImageNet, яка дозволяє формувати бажану конфігурацію із трьох типів шарів: згорткового шару, шару підвибірки та вихідного повнозв'язного шару нейронної мережі.

На основі такої послідовності обробки зображення із текстом у структурі програмного засобу введено модулі, що послідовно із отриманого зображення текстового документа виділяють сторінки, потім у них виділяють текстовий фрагмент і виконують сегментацію виділеного фрагменту із текстом на абзаци, рядки, слова і символи і потім виконують розпізнавання символів. У сформованій програмі процес розпізнавання символів покладається на нейронну мережу. Таку нейронну мережу необхідно попередньо налаштувати на розпізнавання символів тексту [6]. Частина текстового документа, у якій символи будуть не розпізнані, маємо виконати із залученням додаткових засобів. Також є ще один модуль для навчання згорткової нейронної мережі на виконання процедури розпізнавання друкованих символів та модуль виведення результатів розпізнавання символів.

Навчання згорткової мережі відбувається таким чином. Створюємо повноз'єднаний блок для моделі нейронної мережі, виконуємо ініціалізацію параметрів кількості вимірів для цього блоку. Формуємо функцію виклику екземпляра класу для повноз'єданого шару нейронної мережі, що об'єднує всі параметри в $N \times N$ матрицю. Використовуємо нелінійну функцію активації Gaussian Error Linear Units. Для цього повноз'єднаний шар, що з'єднується в кількість каналів вхідного зображення, переводиться у нове подання. Формуємо конволюційний шар для згортання і розділення зображення на фрагменти, виконуємо конкатенацію фрагментів зображення по осям висоти та ширини, здійснюємо ініціалізацію та проведення операцій налаштування із вхідним зображенням, виконуємо нормалізацію та усереднення. Формуємо вихідний результат у вигляді імовірнісного розподілу по класам. Перевіряємо функцію оновлення градієнту, що використовується для тренування нейронної мережі. Використовуємо функцію крос-ентропії для мультикласового розпізнавання. Для цього формуємо крос-ентропію з імовірнісними одиницями, що є необробленим вектором імовірностей конкретного вхідного зображення. Виконуємо оновлення вагів відповідно до розрахованого градієнту. Для отримання інформації з датасету для навчання моделі нейронної мережі вказуємо конфігурацію моделі, клас моделі та вказуємо кількість кроків тренування, що модель має пройти. Формуємо графік зміни швидкості тренування, чим вона менша тим більш прискіплива модель до змін. На початку тренування бажано мати якнайбільшу швидкість тренування, адже так модель зможе запам'ятати найбільш загальні патерни датасету, для покращення точності використовується менша швидкість тренування.

Виконуємо глобальну нормалізацію та ініціалізацію оптимізатора для знаходження глобального мінімуму цільової loss-функції. Здійснюємо ініціалізацію функції оновлення градієнтів. Нульовий крок визначає, що тренування проводиться з нуля. Здійснюємо операцію проведення тренування нейронної мережі, виконуємо розрахунок градієнтів та здійснюємо оновлення ваг нейронів відповідно до виконаних розрахунків. Підготовлена таким чином нейронна мережа використовується для розпізнавання символів текстового документа. Налаштування нейронної мережі виконується на підготовленій базі даних. У ролі такої бази даних виступає Letter Dataset, EMNIST Dataset [7] та сформований набір символів, які дозволили виконати навчання нейронної мережі для розпізнавання друкованих та у деякій мірі і рукописних символів.

Для тестування роботи сформованого програмного продукту були відібрані різні типи документів із чорним та кольоровим поданням тексту, із різними типами шрифтів та розмірів символів у них. Були створені різні зображення з друкованими та невеликими фрагментами рукописних символів. Виведення даних розпізнаного тексту виконується як звичайний текст, виділяючи тільки літери великі та прописні. Фрагмент рукописного тексту, написання якого максимально наближене до друкованого варіанту, також був розпізнаний. Це отримано завдяки навчанню нейронної мережі на значному наборі даних, що мають у томі числі варіанти написання рукописних символів у вигляді цифр. Результати тестування показали високу ефективність роботи створеного програмного додатку.

Висновки

Запропонований підхід може бути використаний у комп'ютерних системах виділення та розпізнавання друкованих символів за отриманим цифровим зображенням текстових документів.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Chaudhuri A. Optical character recognition systems for different languages with soft computing. / A. Chaudhuri, K. Mandaviya, P. Badelia, S. K. Ghosh // Springer, 2017. - 260 p. https://doi.org/10.1007/978-3-319-50252-6_1.
2. Оптичне розпізнавання символів [Електронний ресурс] – Режим доступу: https://ua.wikipedia.org/wiki/Оптичне_розпізнавання_символів.
3. Жихаревич В. В. Аналіз методів розпізнавання символів тексту / В. В. Жихаревич, С. Е. Остапов, І. В. Миронів // Радіоелектронні і комп'ютерні системи. 2016, № 5. – С. 137–142.
4. Субботін С. О. Нейронні мережі : теорія та практика: навч. посіб. / С. О. Субботін. – Житомир : Вид. О. О. Євенок, 2020. – 184 с..
5. Тимченко О. В. Нейромережеві методи розпізнавання зображень текстів /О. В. Тимченко, Б. М. Гавриш, Б. В. Дурняк // Поліграфія і видавнича справа, 2021, № 1 (81). – С.72-88.
6. Кушнір Н.О. Використання згорткових нейронних мереж у задачах розпізнавання та класифікації об'єктів зображень / Н.О. Кушнір, Т.М. Локтікова, А.В. Морозов, В.О. Юрченко // Технічна інженерія, №1 (89), 2022. – С. 93-100.
7. The EMNIST dataset. [Електронний ресурс]. – Режим доступу: <https://www.nist.gov/itl/iad/image-group/emnist-dataset>.

Дідич Олександр Дмитрович — студент групи 2КІ-22м факультету інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, e-mail: sasha.didych@outlook.com

Мартинюк Тетяна Борисівна — доктор технічних наук, професор кафедри обчислювальної техніки, Вінницький національний технічний університет, Вінниця.

Очкуров Микола Андрійович — старший викладач кафедри обчислювальної техніки, Вінницький національний технічний університет, м. Вінниця.

Didych Olexandr D., student of group 2KI-22m, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: sasha.didych@outlook.com

Martyniuk Tetiana B. — Dr. Sc. (Eng), Professor of the Chair of Computer Techniques, Vinnytsia National Technical University, Vinnytsia.

Ochkurov Mykola A. — Senior lecturer of the Computer Techniques Chair, Vinnytsia National Technical University, Vinnytsia.