

УДК 004.8

Ярослав Корнага,
Микола Шиманський,
Андрій Барабаш

ЛОКАЛІЗАЦІЯ ТЕКСТУ НА ЗОБРАЖЕННЯХ ЗА ДОПОМОГОЮ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ

TEXT LOCALIZATION ON IMAGES USING CONVOLUTIONAL NEURAL NETWORKS

У статті розглядається питання локалізації тексту на зображеннях складних графічних сцен. Для локалізації застосовуються згорткова нейронна мережа із завчасно натренованою мережею для виділення ознак на основі моделі VGG-16. Для тренування використано набори зображень із ICDAR 2011 та ICDAR 2013.

Ключові слова: локалізація тексту, обробка зображень, машинний зір, нейронні мережі.

Рис.: 1. Табл.: 1. Бібл.: 3.

This paper deals with the problem of scene text localization. Convolutional neural network was used for this task. Pretrained model based on VGG-16 was used for feature extraction. Datasets from ICDAR and ICDAR 2013 were used for training.

Key words: text localization, image processing, computer vision, neural networks.

Fig.: 1. Table.: 1. Bibl.: 3.

Актуальність теми дослідження. Вирішення задачі локалізації тексту на зображеннях складних графічних сцен є необхідною умовою розпізнавання тексту на таких зображеннях, що може мати застосування в системах машинного зору.

Постановка проблеми. Текст на зображенні у повсякденному житті може мати різний колір та розмір шрифту, він може бути рукописним, що само собою означає, що написання може бути дуже варіативним по своїй формі. Колір фону та освітлення також будуть відрізнятися від зображення до зображення. Багатопарові нейронні мережі дозволяють встановлювати закономірності у множині різних можливих вхідних даних, що робить доцільним їх використання для такого роду задач.

Аналіз останніх досліджень і публікацій. За останні роки в області машинного зору було проведено ряд досліджень, які розглядають проблему локалізації тексту. У [1] автори пропонують метод виділення екстремальних регіонів, на основі класифікації яких проводиться пошук окремих символів. Після цього знайдені літери групуються повним перебором.

У [2] пропонується застосування неспеціалізованого генератора для формування регіонів-кандидатів, що можуть містити слова. Після цього застосовується конволюційна нейромережа для регресії текстових прямокутників.

Також у [3] описується метод виділення цілих текстових рядків з використанням їх симетричних характеристик.

Недосліджена частина проблеми. Дана робота присвячена вивченню та аналізу можливостей застосування багатошарових згорткових нейронних мереж для виділення як друкованого, так і рукописного тексту на зображеннях складних графічних сцен.

Постановка завдання. Завданням є тренування повністю згорткової нейронної мережі для виділення текстових областей на зображеннях складних графічних сцен.

Тренування нейромережі. В області машинного зору багатошарові згорткові нейромережі широко застосовуються для задач класифікації об'єктів на зображеннях. Перший шар згортки обчислює на основі оригінального зображення матрицю ознак – набір числових значень, що зазвичай характеризують відповідні їм області на зображенні. Кожен наступний згортковий шар обчислює додатковий набір ознак, на основі виходу попереднього шару. На даний момент існує багато різноманітних класифікаторів зображень на основі згорткових нейромереж, для яких доступні завчасно натреновані моделі. Успішність цих класифікаторів вказує на ефективність шарів нейромережі, що відповідають за виділення ознак. Для прискорення процесу навчання натреновані ваги з цих нейромереж можуть бути використані для ініціалізації ваг в моделі, що буде спеціалізована на класифікації областей зображення на текстові та нетекстові області. На рис. 1 зображена архітектура запропонованої нейронної мережі. Процес обробки зображення починається із виділення ознак на завчасно натренованих шарах, які коректуються на згортковому шарі Конв.1. На наступних шарах з Конв. 2 по Конв. 6 відбувається поступове зменшення розмірності матриці ознак за допомогою максимізаційного агрегування для можливості виділення текстових областей з різними розмірами шрифту: при обробці матриці ознак з великими розмірами більша вірогідність знайти дрібний текст і навпаки. У шарах текстових блоків проводиться обчислення координат кандидатів текстових областей. На останньому етапі обробки проводиться фільтрація масиву кандидатів для відсіювання занадто маловірогідних гіпотез та дублікатів.

Проведення експерименту. Для тренування і тестування системи було використано зображення із баз даних ICDAR 2011 та ICDAR 2013. Дані з цих вибірок використовуються у змаганнях з локалізації, сегментації та розпізнавання тексту на зображеннях. Нейронна мережа була натренована на зображеннях розміром 300×300 з використанням градієнтного спуску. Кількість зображень у тренувальній вибірці: 100000.

Таблиця 1.

Показники якості роботи системи на тестових вибірках

	ICDAR 2011	ICDAR 2013
Точність	0,85	0,86
Відгук	0,8	0,81

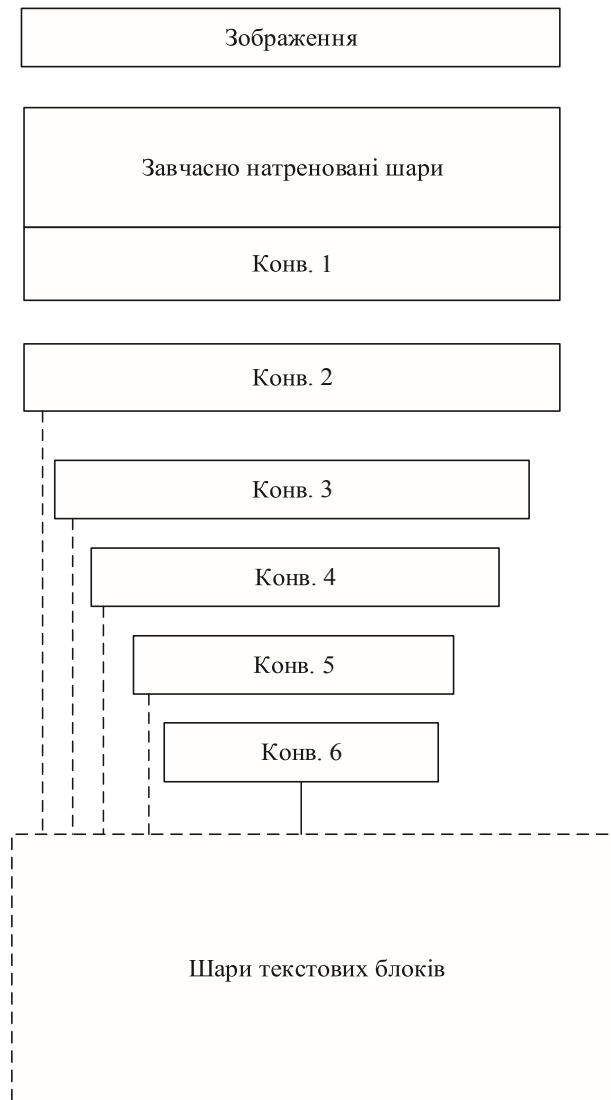


Рис. 1. Архітектура неймережі.

Штрихованими лініями позначено використання у шарах текстових блоків ядра згортки розміром 1×5 , а суцільною – 1×1

Висновки. Запропонована модель дозволяє забезпечити локалізацію тексту на зображеннях складних графічних сцен. В поєднанні з існуючими методами розпізнавання тексту можна отримати повну систему розпізнавання тексту, яка дозволяє виділяти із зображення текст і перетворювати його у формат, придатний для машинної обробки. Модель може бути використана для локалізації як друкованого, так і рукописного тексту.

Список використаних джерел

1. Neumann, L., and Matas, J. 2012. Real-time scene text localization and recognition. In Proc. CVPR, 3538–3545.
2. Jaderberg, M.; Simonyan, K.; Vedaldi, A.; and Zisserman, A. 2016. Reading text in the wild with convolutional neural networks. IJCV 116(1):1–20.
3. Zhang, Z.; Shen, W.; Yao, C.; and Bai, X. 2015. Symmetrybased text line detection in natural scenes. In Proc. CVPR, 2558–2567.

ДОВІДКА ПРО АВТОРІВ

Корнага Ярослав Ігорович - кандидат технічних наук, доцент кафедри технічної кібернетики Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського».

Kornaga Yaroslav - associate Professor of the Department of Technical Cybernetics, PhD of the National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute».

E-mail:slovyan_k@ukr.net

Шиманський Микола - студент кафедри Технічної Кібернетики Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського».

Shymanskyi Mykola - student of the Department of Technical Cybernetics of the National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute».

Барабаш Андрій - студент кафедри Технічної Кібернетики Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського».

Barabash Andrii - student of the Department of Technical Cybernetics of the National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute».

Yaroslav Kornaga, Mykola Shymanskyi, Andrii Barabash

TEXT LOCALIZATION ON IMAGES USING CONVOLUTIONAL NEURAL NETWORKS

Target setting. Scene text recognition can have many use cases in computer vision systems and text localization is a major part of this problem

Problem statement. Absence of effective method for text localization in natural scene images.

Actual scientific researches and issues analysis. There are different studies that cover various approaches to text localization task such as usage of maximally stable extremal regions, symmetric features and others.

Uninvestigated parts of general matters defining. This paper is dedicated to study and analysis of possibilities of deep convolutional neural networks for printed and handwritten text localization alike.

The research objective. The task at hand is to create and train a neural network model that can be used for effective text localization on natural scene images.

The statement of basic materials. This paper describes an architecture of a fully convolutional neural network that is proposed for the text localization task. The benefits of using pretrained models as feature extractors were described. Experimental results were presented.

Conclusions. Proposed model provides an effective text localization for natural scene images. It can be used with text recognition systems for text extraction from images and its further processing. This model can be trained for both printed and handwritten text localization.

Key words: text localization, image processing, computer vision, neural networks.