Yolov7 모델을 이용한 영수증 항목 탐지

마흐무드 카셈, 모하메드 마흐무드, 강현수*
충북대학교 정보통신공학과

mahmoud.salah@chungbuk.ac.kr, mohamedabokhalil@chungbuk.ac.kr,

*hskang@cbnu.ac.kr

Receipts Items detection Using Yolov7 Model

Mahmoud Kasem, Mohamed Mahmoud, Kang Hyun Soo*

Chungbuk National University

Abstract

Recent strides in deep learning, computer vision, and natural language processing have fundamentally reshaped how machines handle unstructured data. Fields like Scanned Receipts OCR and Information Extraction (SROIE) represent the convergence of computer vision and natural language processing. SROIE is crucial for an end-to-end process, recognizing text in scanned images, like receipts, and structuring the extracted information. Its significance extends to various document intelligence applications, holding substantial potential for business. This paper introduces a deep learning approach for detecting receipt items based on YOLOv7 on the SROIE dataset, achieving a promising accuracy of 95.3%. This research not only showcases the potential of YOLOv7 in receipt item detection but also underscores the evolving landscape of unstructured data processing, offering valuable insights for future endeavors in this dynamic field.

Ⅰ. Introduction

Deep learning, a significant subset of machine learning, is widely adopted across various industries. In computer vision, it excels in object detection, image classification, and video analysis. In Natural Language Processing (NLP), deep learning is used for text classification, question answering[1], sentiment analysis, sentence similarity[2], machine translation, speech recognition, and table detection[3,4]. In face masking detection[5] The healthcare sector benefits from deep learning in disease diagnosis, treatment planning, drug discovery, and medical imaging analysis[6,7]. Robotics utilizes it for autonomous navigation, object recognition, and control. Additionally, deep learning is effective in handwritten recognition across languages[8]-[10]. In the Internet of Things (IoT), it aids in intrusion detection[11], while in finance, it enhances fraud detection, algorithmic trading, and risk management. The energy sector leverages deep learning for predictive maintenance, energy consumption forecasting, and customer profiling[12,13], highlighting its vast potential for future advancements.

The domain of Scanned Receipt OCR and Information Extraction(SROIE)[13] involves extracting and structuring text from scanned receipts for tasks like archiving and financial analytics. Deep learning, particularly YOLOv7, has advanced real-time object detection, making it valuable for receipt item detection. YOLO's grid-based approach predicts bounding boxes and class probabilities, enabling rapid and accurate detection.YOLOv7[14], with its enhanced performance, holds promise in improving document intelligence and streamlining business processes, especially in receipt OCR tasks.

Ⅱ. Methodology and Experimental Result

Dataset: The SROIE dataset includes 626 receipts for training and 347 for testing, each with four key text fields: company, address, date, and total. Predominantly featuring digits and English characters, the dataset has varied layouts and complex structures, as shown in Figure [1]. It includes text bounding boxes and corresponding transcripts, making it a valuable resource for text localization, recognition, and information extraction tasks.

Figure 1: Samples of the SROIE dataset

The YOLOv7 configuration for receipt item detection involves adapting the architecture to output bounding boxes and class probabilities specific to

receipt elements, with anchor boxes defined by typical item sizes and input dimensions matching the dataset resolution. Training parameters like batch size, learning rate, and optimization algorithms are optimized, and data augmentation strategies are used to diversify the training set. The training process includes preprocessing annotated receipt images, iterative training with dynamic learning rate adjustments, and applying regularization techniques to prevent overfitting, leveraging YOLOv7's real-time detection capabilities for accurate receipt item localization.

The model accurately identifies and delineates text elements within preprocessed images, demonstrating its capability to handle complex document structures. Evaluated on the SROIE Task 1 test dataset, the YOLOv7 model achieved an impressive 95.3% accuracy, highlighting its robustness in real-world document analysis. Table [1] shows that our model

outperforms others with superior precision, recall, and F1-score, and a competitive mean Average Precision (mAP), confirming its efficiency in receipt item detection.

Method	Precision	Recall	F1-Score	mAP
W Yu[15]	-	-	96.1	-
Zhang[16]	-	-	96.18	-
Our Model	96.6	97.2	96.79	95.3

Figures [2] illustrate the YOLOv7 model's effectiveness in text localization on the SROIE dataset.

Figure 2: Samples of the Text localization Output

Ⅲ.Conclusion

In conclusion, the YOLOv7 model significantly advances receipt item detection in the SROIE dataset, demonstrating high accuracy (95.3%) in text localization. Its real-time detection capabilities make it ideal for automating financial, accounting, and taxation processes. Key pre-processing and augmentation stages enhance robustness, addressing document skew and dataset diversity. The YOLOv7-based approach shows great potential for improving document intelligence and streamlining workflows across industries.

ACKNOWLEDGMENT

This work was supported by the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program (IITP-2024-2020-0-01462, 50%) supervised by the IITP, and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant 2023R1A2C1006944.

REFERENCES

[1] Abdallah, Abdelrahman, et al. "Arabicaqa: A comprehensive dataset for arabic question answering." arXiv preprint arXiv:2403.17848 (2024).

[2] Abdallah, Abdelrahman, Bhawna Piryani, and Adam Jatowt. "Exploring the state of the art in legal QA systems." Journal of Big Data 10.1 (2023): 127.

[3] Kasem, Mahmoud, et al. "Deep learning for table detection and structure recognition: A survey." ACM Computing Surveys (2022).

[4] Prasad, Devashish, et al. "CascadeTabNet: An approach for end to end table detection and structure recognition from image-based documents." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020.

[5] Mahmoud, Mohamed, Mahmoud SalahEldin Kasem, and Hyun-Soo Kang. "A Comprehensive Survey of Masked Faces: Recognition, Detection, and Unmasking." arXiv preprint arXiv:2405.05900 (2024).

[5] Nie, Liqiang, et al. "Disease inference from health-related questions via sparse deep learning." IEEE Transactions on knowledge and Data Engineering 27.8 (2015): 2107-2119.

[7] Mahmoud, Mohamed, and Hyun-Soo Kang. "Ganmasker: a two-stage generative adversarial network for high-quality face mask removal." Sensors 23.16 (2023): 7094.

[8] Toiganbayeva, Nazgul, et al. "Kohtd: Kazakh offline handwritten text dataset." Signal Processing: Image Communication 108 (2022): 116827.

[9] Kasem, Mahmoud SalahEldin, Mohamed Mahmoud, and Hyun-Soo Kang. "Advancements and Challenges in Arabic Optical Character Recognition: A Comprehensive Survey." arXiv preprint arXiv:2312.11812 (2023).

[10] Nurseitov, Daniyar, et al. "Handwritten Kazakh and Russian (HKR) database for text recognition." Multimedia Tools and Applications 80.21 (2021): 33075-33097.

[11] Mahmoud, Mohamed, et al. "Ae-lstm: Autoencoder with lstm-based intrusion detection in iot." 2022 International Telecommunications Conference (ITC-Egypt). IEEE, 2022.

[12] Hamada, Mohamed A., et al. "Neural network estimation model to optimize timing and schedule of software projects." 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST). IEEE, 2021.

[13] Kasem, Mahmoud SalahEldin, Mohamed Hamada, and Islam Taj-Eddin. "Customer profiling, segmentation, and sales prediction using AI in direct marketing." Neural Computing and Applications 36.9 (2024): 4995-5005.

[13] Huang, Zheng, et al. "Icdar2019 competition on scanned receipt ocr and information extraction." 2019 International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2019.

[14] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[15] Yu, Wenwen, et al. "PICK: processing key information extraction from documents using improved graph learning-convolutional networks." 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021.

[16] Zhang, Peng, et al. "TRIE: end-to-end text reading and information extraction for document understanding." Proceedings of the 28th ACM International Conference on Multimedia. 2020.