A Text Feature Recognition Model for Examination Papers Based on Optical Character Recognition Integrated with YOLO

Jia-Yu Lai [1], Alisa Huang[1], Jia-Wen Liou [2] and Chi-Sheng Huang*[1][0000-0001-5763-1143]

[1] Department of Computer Science and Information Engineering, National Taichung University of Science and Technology, Taiwan
[2] Department of Business Administration, National Taichung University of Science and Technology, Taiwan
vcshuang@nutc.edu.com.tw

Abstract. With the growing demand for digital assessments, developing an automated grading system is decisive for improving grading efficiency and reducing human error. This study focuses on Optical Character Recognition (OCR) technology in exam papers, aiming to enhance the recognition accuracy for both handwritten and printed text. We constructed a Convolutional Neural Network (CNN) model using TensorFlow Keras, which extracts features through multiple convolutional layers combined with Rectified Linear Unit (ReLU) activations and reduces dimensionality using Average Pooling and Global Average Pooling. A Fully Connected Layer followed by a Softmax function is then employed to classify eight different character categories. To address the diversity in data sources and noisy environments, a data augmentation strategy was integrated during training, and the model was optimized using the Adaptive Moment Estimation (Adam) optimizer along with Sparse Categorical Crossentropy as the loss function. Experimental results demonstrate that the model achieves nearly 95% accuracy after approximately 10 training epochs, exhibiting strong generalization and stability, thereby providing a solid technical foundation for the OCR module within an automated grading system.

Keywords: Automated Grading System, Optical Character Recognition, Convolutional Neural Network, Data Augmentation, Educational Technology.

1 Introduction

With the rising demand for digital assessments, traditional paper-based exam grading is not only time-consuming but also prone to subjective biases, which can compromise both the fairness and accuracy of evaluations. Although previous studies have applied object detection techniques for segmenting exam paper regions, a large volume of handwritten and printed text still fails to achieve satisfactory recognition accuracy when processed using conventional Optical Character Recognition (OCR) methods. The variability in handwriting, inconsistent writing styles, and noise interference further impose higher requirements on the stability and precision of OCR systems.

This study focuses on enhancing the recognition performance of the OCR module in exam papers, aiming to improve the accuracy of both handwritten and printed character recognition. By leveraging deep learning techniques, a Convolutional Neural Network (CNN) model was designed and trained, and a data augmentation strategy was integrated to address the challenges posed by diverse data sources and noisy environments. Through in-depth optimization of the OCR component, this research aspires to provide a more robust and precise text recognition solution for automated grading systems, thereby reducing the subjective errors and labor costs associated with manual grading, and promoting the digital and automated transformation of educational assessments.

2 Related Work and Literature Review

2.1 Development of Exam Paper Image Recognition Grading Systems

With the promotion of digital assessment in the education sector, traditional manual grading is time-consuming and subjective. In recent years, research on automated scoring has attracted increasing attention. Object detection models (such as YOLO) [1] can accurately segment the layout of exam papers, while OCR extracts text. This technology has been successfully applied in invoice[2] and license plate recognition [3], but its application in educational contexts is still emerging. This study integrates both technologies to enhance scoring efficiency and consistency.

2.2 Applications of Data Preprocessing and Data Augmentation

In this study, we used the LabelImg tool [4] for precise annotation, dividing exam sheet images into key areas (questions, answer areas, handwriting, question numbers), and further categorizing handwritten data (A, B, C, D, E, F, true/false questions). In addition, data augmentation techniques such as rotation, flipping, cropping, and brightness adjustment [5] were employed to simulate diverse shooting conditions and enhance the model's generalization ability, laying a solid foundation for accurate automated grading.

2.3 Applications of Sorting and Matching Algorithms in Result Integration

This study proposes an automated exam paper image recognition process, focusing on the use of sorting and matching algorithms to integrate YOLOv7-detected blocks with CNN-based OCR outputs. The system first sorts and groups the detected blocks, then draws inspiration from the Hungarian algorithm used in DeepSort[6], combined with Euclidean distance, for matching to ensure precise correspondence between student answers and standard answers. Inspired by DeepSort's efficient matching strategy, we believe it is possible to further develop a customized matching algorithm tailored to the characteristics of exam paper images. This approach not only effectively reduces matching errors but also enhances the accuracy and stability of automated grading, laying the foundation for large-scale applications and detailed teaching analysis.

3 Methods and Experimental Results

This study aims to build an intelligent grading system using YOLOv7 and OCR technologies to address the time-consuming nature, subjective errors, and high labor costs associated with traditional paper-based grading. The system is designed to automatically recognize key regions within exam papers, and data augmentation together with transfer learning strategies are applied to enhance the overall accuracy and robustness of the model. The following sections describe the experimental design and results across several stages.

3.1 Data Preparation and Annotation

The dataset comprises open-source exam papers from various subjects at the middle and high school levels, provided in electronic files as well as photocopies, photographs, and scans of paper documents. Each exam paper was meticulously annotated using the Labellmg tool, with key regions—such as the question, answer area, handwritten responses, articles, diagrams, and item numbers—being delineated. Handwritten sections were further classified into eight categories (e.g., common options A, B, C, D, E, F and true/false symbols such as circles and crosses).

Initial experiments using unaugmented, directly captured images and electronic files yielded poor object localization due to insufficient samples and suboptimal data quality (e.g., noisy backgrounds and low contrast). To address these issues, various data augmentation techniques—including rotation, flipping, cropping, brightness adjustment, Gaussian blur (to simulate unfocused images), and sharpening (to reduce noise interference)—were applied to simulate extreme image conditions and increase data diversity, as illustrated in Fig. 1.

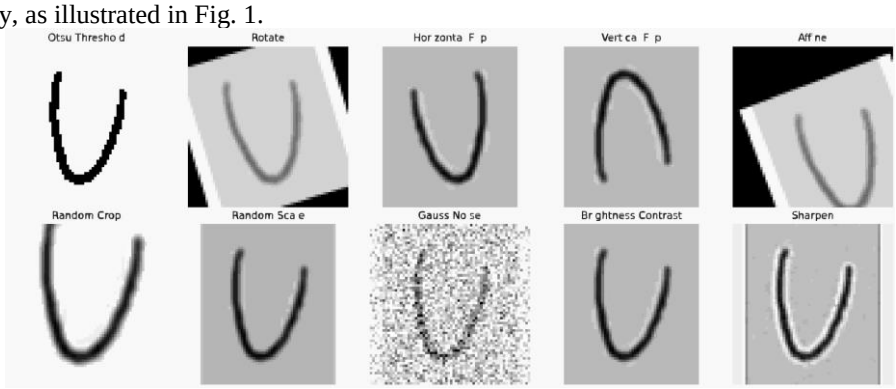

Fig. 1. Experimental data post-augmentation.

3.2 Model Training and Optimization Strategy

The core of the system is divided into two primary components: exam paper region detection and text recognition.

Region Detection. To accurately localize the question area, answer area, answer box, and item number within exam papers, we adopted the YOLOv7 model, which is renowned for its strong community support and balanced trade-off between computational efficiency and speed. During training, OpenCV-based image augmentation techniques were employed to simulate various lighting conditions, angles, and background interferences. Ultimately, we compiled a dataset comprising 1,458 augmented standard blank exam paper images (including both original and augmented versions), with a total of 32,250 annotated objects covering the categories of question, answer area, article, answer box, diagram, and item number. Training parameters were set to a batch size of 4 and 50 epochs, with the dataset split proportionally into training (80%), testing (15%), and validation (5%) sets.

Text Recognition. To recognize the text content extracted by the region detection module, we developed a CNN-OCR model using TensorFlow Keras. The model consists of multiple Conv2D layers with ReLU activations, coupled with average pooling and global average pooling layers for dimensionality reduction, and finally, a fully connected layer with a Softmax activation to classify the input into eight categories. During training, the model was optimized using the Adam optimizer and Sparse Categorical Crossentropy as the loss function, while image augmentation was simultaneously applied to bolster the model's robustness against variations in fonts and noise. For the handwritten responses, the original dataset of 2,976 samples was augmented to 23,806 samples, resulting in a total of 26,782 training samples. This stage was trained with 50 epochs and a batch size of 2, with an Early Stopping strategy implemented (terminating training if no improvement is observed over 5 consecutive epochs) to prevent overfitting. In addition to Gaussian blur—to simulate out-of-focus images—the sharpening filter was applied to reduce noise interference and enhance the capture of fine handwriting details.

3.3 Experimental Results and Analysis

The experiments were conducted in three stages, each addressing different data sources and challenges, and integrating sorting and matching algorithms to ensure accurate correspondence between exam regions.

Training and Testing on Exam Paper Data. Figure 2 compares YOLOv7's detection results on exam paper regions with and without data augmentation using Precision-Recall curves. With augmentation, the left chart shows superior performance across all categories (Question, Answer, Article, Answer box, Diagram, Item), with AP values of 0.969, 0.995, 0.980, 0.925, 0.993, and 0.966, and an mAP@0.5 of 0.971. The curve encloses a large, near-square area, indicating high accuracy and stability. Without augmentation, the right chart shows lower AP values (0.956, 0.979, 0.881, 0.987, 0.845) and an mAP@0.5 of 0.930, with a smaller, irregular curve and sharp Precision drop

after Recall exceeds 0.6, reflecting weaker performance. Data augmentation significantly boosts YOLOv7's detection, improving AP, convergence, and stability while addressing limited data issues.

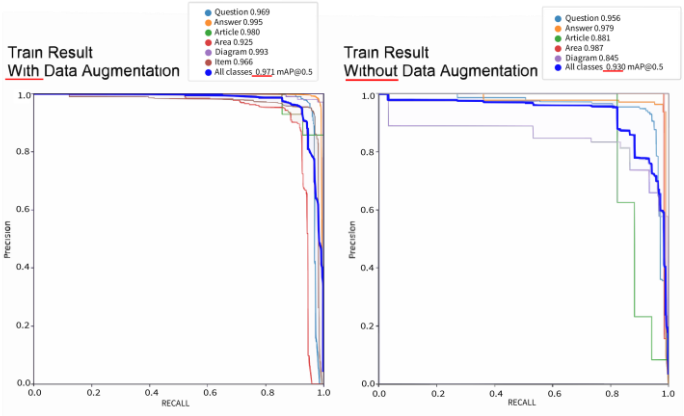

Fig. 2. YOLOv7 block detection results with and without data augmentation.

Specialized Optimization for Handwritten Data. For handwritten text recognition, the CNN-OCR model trained on the augmented dataset (totaling 26,782 samples) achieved markedly improved accuracy in detecting handwritten regions, reducing instances of misclassifying handwritten text as images. Validation results on known data reached 100% accuracy, and overfitting was not observed. However, when applied to unfamiliar data, the system exhibited high prediction confidence despite an error rate of approximately 50%, indicating that extreme cases (e.g., highly messy handwriting, mixed symbols, translucent paper, or residual eraser marks) still present challenges. Further optimization of training strategies and expansion of data diversity are needed to enhance robustness, as illustrated in Fig. 3.

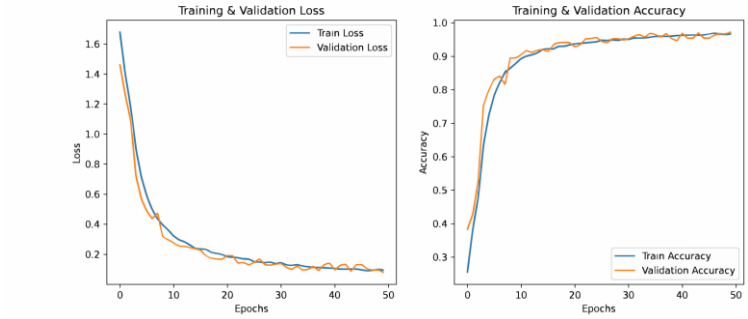

Fig. 3. TensorFlow Keras OCR training result.

Bounding Box Conversion and Q&A Matching Mechanism. Upon completing YOLOv7 detection, the output format is *<class name>*, *<x center>*, *<y center>*,

<width>, *<height>*. Based on those datasets, the object's vector can be computed. Subsequently, the system calculates the Euclidean distance between the detected question and item regions to establish a preliminary pairing. Based on the degree of overlap or minimum distance, items are then matched to the corresponding answer areas. Finally, the OCR module processes these matched regions to output a structured result in the form {item: answer}. Given that each exam paper typically contains no more than 1,000 bounding boxes, a nearest-neighbor matching strategy ($O(N^2)$) was adopted for its balance between efficiency and accuracy; for larger-scale applications, the Hungarian Algorithm or more complex matching techniques could be considered to achieve a globally optimal solution.

4 Conclusion and Future Work

In summary, this study successfully developed an automated exam paper image recognition system that integrates YOLOv7 and OCR technologies. By employing data augmentation and transfer learning strategies, the system has achieved significant improvements in both region detection and text recognition, thereby reducing the time cost and subjective errors associated with manual grading. Future work will focus on further optimizing the handwritten recognition module—by incorporating a more diverse dataset and exploring advanced OCR models—and investigating integration with digital archiving systems (e.g., national digital archives) to extend the application of this technology in large-scale automated grading and detailed educational analysis.

References

1. WongKinYiu/yolov7, https://github.com/WongKinYiu/yolov7, last accessed 2023/11/25.
2. Yao, X., Sun, H., Li, S., Lu, W.: Invoice Detection and Recognition System Based on Deep Learning. Comput. Intell. Neurosci. 2022, Article 8032726 (2022). https://doi.org/10.1155/2022/8032726
3. Sugiyono, A.Y., Adrio, K., Tanuwijaya, K., Suryaningrum, K.M.: Extracting Information from Vehicle Registration Plate using OCR Tesseract. Computer Science Department, Bina Nusantara University, Jakarta, Indonesia (2023).
4. HumanSignal/labelImg, https://github.com/HumanSignal/labelImg, last accessed 2023/09/23.
5. Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., Shen, F.: Image Data Augmentation for Deep Learning: A Survey. arXiv preprint arXiv:2204.08610 [cs.CV] (2022). https://doi.org/10.48550/arXiv.2204.08610
6. Wojke, N., Bewley, A., Paulus, D., Simple Online and Realtime Tracking with a Deep Association Metric. arXiv preprint arXiv:1703.07402 [cs.CV] (2017). https://arxiv.org/abs/1703.07402
7. TensorFlow CNN tutorial, https://www.tensorflow.org/tutorials/images/cnn, last accessed 2024/08/16.