

# Research on an Intelligent Book Recognition System Based on Computer Vision

Yanqing Zhong

Jingtang Library, Jiangmen, 529100, PR China

## Abstract

Accurate identification of books on shelves is important for intelligent library management. To improve efficiency, this paper proposes an intelligent book recognition system based on computer vision. Compared with the conventional LeNet-5 model, LeNet-5book improves the recognition accuracy from 95.71% to 98.35%. The results indicate that the proposed method is feasible, accurate, and suitable for smart shelf inspection in modern libraries.

## Keywords

Computer vision; library management; book recognition; Faster R-CNN; convolutional neural network.

## 1. Introduction

Recent progress in computer vision and deep learning has provided new solutions for automatic object detection and text recognition [1-4]. In library scenarios, book spines contain useful visual information, especially call numbers, which can be used to identify books and verify shelf order. However, practical recognition remains difficult because of uneven illumination, complex background, variable spine width, label deformation, and character diversity.

To solve these problems, this paper presents an intelligent book recognition system for shelf-level library management. The proposed method integrates image preprocessing, book spine detection, call number localization, character segmentation, and convolutional neural network recognition into a complete workflow. Faster R-CNN is adopted to improve the accuracy of call number region detection [3], while an improved LeNet-5 model is designed for character classification [1]. Experimental results demonstrate that the proposed system has good recognition performance and practical application value for intelligent library management.

## 2. System Framework

The overall framework of the proposed system is shown in Fig. 1 The workflow includes image acquisition, book spine detection, call number region localization, character segmentation, character recognition, and database storage. The system starts from the captured shelf image and ends with recognized call number information written into the library database.

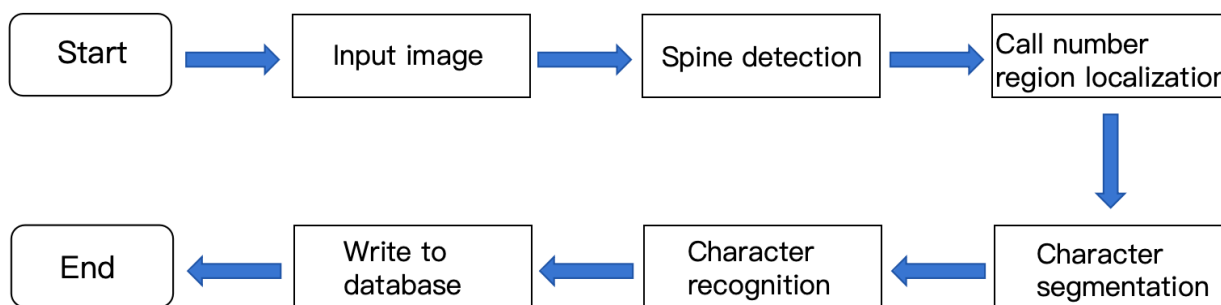


Fig. 1 Flowchart of the proposed book recognition system

As shown in the framework, the system is designed as a complete end-to-end recognition pipeline. Compared with manual inspection, this framework can significantly reduce repetitive labor and improve shelf management efficiency.

### 3. Methodology

#### 3.1. Book Spine Detection

Book spine detection is the first step in the proposed system. Since books on a shelf are generally arranged vertically, their boundaries can be extracted through edge and line features. In this study, edge detection is first applied to highlight the boundaries of adjacent books. A typical edge detection process includes grayscale transformation, Gaussian filtering, gradient calculation, non-maximum suppression, and double-threshold detection, as illustrated in Fig. 2.

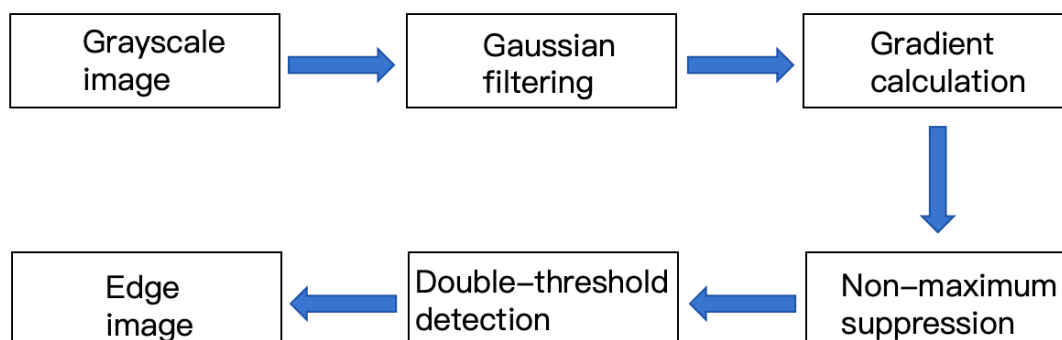


Fig. 2 Edge detection process

After edge enhancement, vertical line features are extracted to separate different book spines. This method is effective in regular shelf scenes and provides reliable input for subsequent call number analysis. An example of the original shelf image is shown in Fig. 3, and the corresponding edge detection result is shown in Fig. 4.

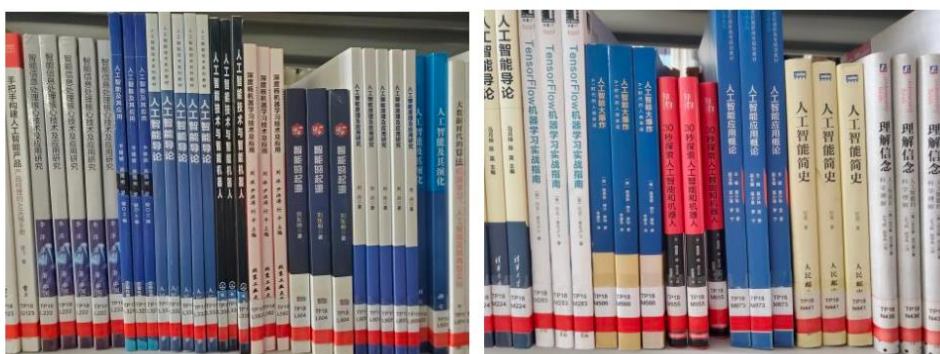


Fig. 3 Original image of books captured from the shelf



Fig. 4 Input image of book spines and the corresponding edge detection result

### 3.2. Call Number Region Localization

After book spines are extracted, the next task is to determine the call number area on each spine. Because the position, size, and appearance of call number labels differ from book to book, traditional handcrafted rules are often unstable. Therefore, this paper uses region-based deep learning detection methods.

R-CNN first generates candidate regions, extracts CNN features from these regions, and then classifies them [5]. Based on this idea, Faster R-CNN introduces a region proposal network and significantly improves detection efficiency and accuracy [3]. In this work, Faster R-CNN is adopted to localize the call number region on book spines.

The localization result of call number extraction is shown in Fig. 5. The detected regions can be accurately cropped and used for character segmentation.

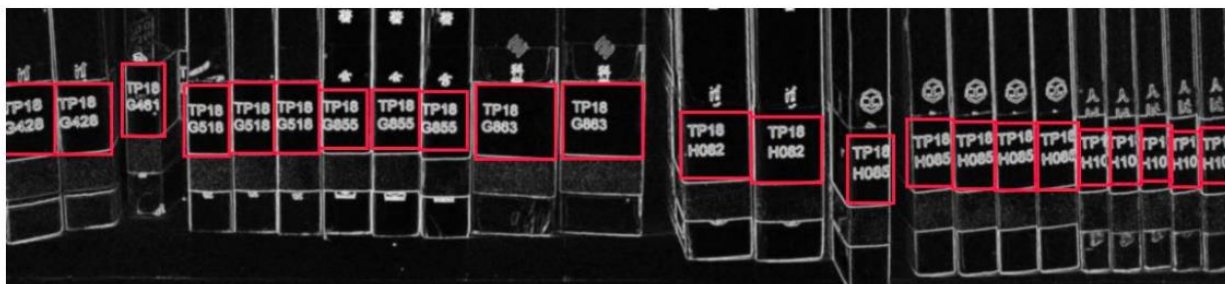


Fig. 5 Localization and extraction of call number regions

### 3.3. Character Segmentation

After the call number region is extracted, character segmentation is performed. Since a call number usually contains letters and digits, it is necessary to divide the target region into separate character images before recognition. In this paper, binarization and connected component analysis are adopted to segment individual characters.

An example of the segmentation result is shown in Fig. 6. Each segmented character is normalized and then sent to the recognition network for classification. Proper character segmentation is essential because segmentation quality directly affects recognition accuracy.

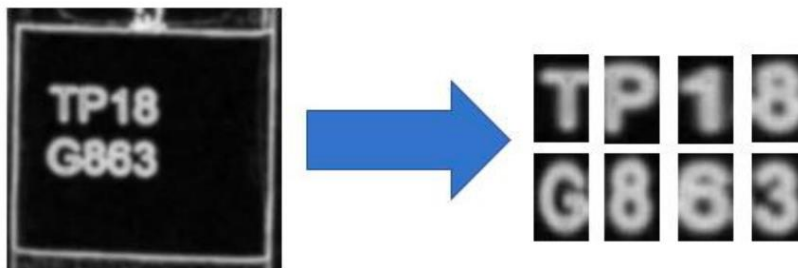


Fig. 6 Character segmentation of the call number

### 3.4. Character Recognition Model

For character recognition, this study proposes an improved convolutional neural network based on LeNet-5. The original LeNet-5 is a classical CNN structure for handwritten and printed character recognition [1]. To better fit the features of library call number characters, the network is optimized and named LeNet-5book. The improved model enhances feature extraction capability and classification performance for mixed alphanumeric call numbers. Compared with the conventional LeNet-5, the improved model is more suitable for library book call number recognition tasks in complex shelf environments.

## 4. Experimental Results and Discussion

To evaluate the effectiveness of the proposed method, experiments were carried out on collected shelf images. The system completed the entire workflow, including image input, spine detection, call number localization, character segmentation, character recognition, and database writing. The implementation interface of the developed intelligent book recognition system is shown in Fig. 7.



Fig. 7 Intelligent book recognition system

The test results of the improved model and the baseline model are listed in Table 1. The LeNet-5book model correctly recognized 658 samples with an average accuracy of 98.35%, while the conventional LeNet-5 correctly recognized 635 samples with an average accuracy of 95.71%. Although the recognition time of LeNet-5book is slightly longer, its accuracy improvement is significant.

Table 1. Comparison of test results between the improved model and the conventional model

Model	Number of Correct Recognitions	Accuracy (%)	Recognition Time (s)
LeNet-5book	658	98.35	2.87
LeNet-5	635	95.71	2.62

According to the results, the improved LeNet-5book model increases accuracy by 2.64 percentage points. This indicates that the optimized network is more effective for alphanumeric call number recognition.

## 5. Conclusion

This paper proposes an intelligent book recognition system based on computer vision. The method integrates book spine detection, call number localization, character segmentation, and character recognition into a unified workflow. Experimental results show that the proposed method can effectively complete the process from shelf image acquisition to database storage. The recognition accuracy of LeNet-5book reaches 98.35%, outperforming the conventional LeNet-5 model. Therefore, the proposed system is feasible and valuable for smart library shelf inspection.

## References

- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149.
- Girshick, R. (2015). Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1440–1448). Santiago, Chile.
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 580–587). Columbus, OH, USA.