Application of Deep Learning in Khmer Invoice Document Image Analysis: A Comparative Study of CNN, CNN-RNN Hybrid, and Transformer Models

Sidavid Sin

Master of Science in Management of Information Systems, Paragon International University, Phnom Penh, Cambodia

*Corresponding author's e-mail: ssin3@paragoniu.edu.kh
doi: https://doi.org/10.21467/proceedings.174.9

Abstract

This paper explores the application of deep learning (DL) techniques to the document image analysis of Khmer invoices, a niche but increasingly important area within the field of optical character recognition (OCR). Given the unique challenges posed by the Khmer script and the complexity of invoice layouts, this study reviews three pioneering DL approaches to understand their strengths and limitations in addressing these challenges. The first approach employs Convolutional Neural Networks (CNNs) for their superior text recognition capabilities, particularly adept at deciphering the intricate patterns of Khmer characters. The second method integrates CNNs with Recurrent Neural Networks (RNNs), leveraging the latter's proficiency in contextual and sequential data analysis to enhance layout understanding and text extraction. The third strategy introduces Transformer-based models, capitalizing on their self-attention mechanisms to foster a deeper comprehension of document context and relationships. Through a comparative analysis, this paper delineates the advantages of each model, such as the CNN's accuracy in character recognition, the CNN-RNN hybrid's effective layout and text analysis, and the Transformer's comprehensive document understanding. Conversely, it also discusses their disadvantages, including issues related to computational demands, training data requirements, and adaptability to diverse invoice formats. Concluding with a synthesis of findings, the study proposes a hybrid model that combines the robust feature extraction of CNNs with the contextual awareness of Transformers as a promising solution for Khmer invoice processing. This approach aims to mitigate the identified limitations and suggests directions for future research, emphasizing the need for lightweight architectures and the development of comprehensive benchmark datasets.

Keywords: *DeepLearning, Document ImageAnalysis, Optical Character Recognition (OCR)*

1 Introduction

The advent of Deep Learning (DL) has revolutionized many fields of computer science, with Document Image Analysis (DIA) being no exception. Traditional methods for DIA have relied heavily on hand-crafted features and classical machine learning algorithms. However, these approaches often fall short when dealing with the complexity and variability of document layouts and scripts, particularly for languages with intricate characters like Khmer. [1] [2] Recent research has shifted towards leveraging DL for its ability to learn hierarchical features, showing significant improvements in both accuracy and efficiency of document analysis. The Khmer language, with its complex script and unique numeral system, presents specific challenges for DIA, such as character segmentation, recognition, and contextual analysis of the textual content within invoices. This paper reviews three innovative DL methods tailored to address these challenges, comparing their approaches and outcomes to identify best practices for Khmer invoice processing.

2 Model Advantages Comparative

The first model, based on Convolutional Neural Networks (CNNs), focuses on text recognition, specifically designed to capture the intricate features of the Khmer script. CNNs are renowned for their ability to learn spatial hierarchies of features, making them exceptionally suited for recognizing the complex patterns in Khmer characters. The primary advantage of this approach lies in its robust feature extraction capabilities, significantly reducing the error rate in character recognition compared to traditional OCR methods. [1] The authors of this model argue that it outperforms existing approaches by adapting to the nuances of the Khmer script without requiring extensive pre-processing.

In contrast, the second model integrates CNNs with Recurrent Neural Networks (RNNs) to address both layout analysis and text recognition challenges. By leveraging the sequential nature of text through RNNs, particularly Long Short-Term Memory (LSTM) units, this hybrid model excels at understanding the context and flow within texts, which is crucial when processing invoices with variable information placement. This approach not only

Proceedings DOI: 10.21467/proceedings.174; Series: AIJR Proceedings; ISSN: 2582-3922; ISBN: 978-81-984081-6-7

recognizes characters with high accuracy but also effectively maps the structural layout of documents, distinguishing between different types of content, such as headers, item lists, and totals. The synergy between CNN and RNN components enables this model to demonstrate superior performance in holistic document understanding compared to standalone models.

The third approach adopts a Transformer-based architecture, leveraging self-attention mechanisms to understand the complex relationships between different document elements [3]. Transformers have shown remarkable success in various NLP tasks, and their application to DIA introduces the ability to process entire invoices in an end-to-end manner without needing separate stages for text detection and recognition. This model's advantage lies in its comprehensive understanding of document context, allowing it to accurately extract and classify information from various invoice sections. The authors of this method assert that it reduces the need for extensive data pre-processing and manual feature engineering, offering a more scalable solution for invoice processing.

3 Model Disadvantages Comparative

Despite its strengths, the CNN-based model's main limitation is its rigidity in handling non-standard layouts. While it excels in character recognition, its performance declines when faced with invoices that deviate from the trained formats, highlighting a lack of flexibility in adapting to new or unseen document layouts. The CNN-RNN hybrid, while powerful, faces significant computational demands for both training and inference. This poses a challenge, particularly for real-time applications or devices with limited processing power. Additionally, the complexity of this model requires a large amount of labeled data for training, making it less practical for scenarios where annotated documents are scarce. The Transformer-based model, although strong in its contextual understanding, relies heavily on large datasets for effective training. The scarcity of comprehensive, annotated Khmer invoice datasets can hinder the model's performance, potentially requiring substantial effort in dataset creation and curation.

4 Conclusion

The reviewed papers collectively underscore the potential of DL in transforming Khmer invoice processing through advanced character recognition, layout analysis, and context understanding. Each method presents unique advantages: CNNs for their precision in character recognition, CNN-RNN hybrids for their dual capacity to analyze text and layout, and Transformers for their holistic understanding of document context. However, the challenges identified—such as the need for extensive computational resources, large annotated datasets, and flexibility in processing diverse invoice formats—suggest that no single approach is without limitations. Considering the specific context of Khmer invoice analysis, a hybrid approach that combines the robust feature extraction of CNNs with the contextual comprehension of Transformers appears most promising. This recommendation is contingent on further research to optimize model efficiency and address the scarcity of annotated training data. Future work should focus on developing lightweight model architectures and semi-supervised learning techniques to alleviate the dependency on large datasets. Additionally, creating a comprehensive benchmark dataset for Khmer invoices could significantly advance the field, enabling more effective comparisons of different DL approaches.

5 Declarations

5.1 Competing Interests

The author declares no conflict of interest regarding the publication of this paper.

5.2 Publisher's Note

AIJR remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

How to Cite

Sidavid Sin (2025). Application of Deep Learning in Khmer Invoice Document Image Analysis: A Comparative Study of CNN, CNN-RNN Hybrid, and Transformer Models. *AIJR Proceedings*, 50-51. https://doi.org/10.21467/proceedings.174.9

References

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, 2015.
[2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, vol. 1. MIT Press, 2016.
[3] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, pp. 6000-6010, 2017.