International Journal of Science and Research Archive

eISSN: 2582-8185

Cross Ref DOI: 10.30574/ijsra

Journal homepage: https://ijsra.net/

(Review Article)

Intelligent document processing: AI-powered RPA for multilingual OCR of receipts

Spriha Deshpande * and Manoj Rajalbandi

Santa Clara, USA.

International Journal of Science and Research Archive, 2025, 14(01), 1164-1166

Publication history: Received on 10 December 2024; revised on 18 January 2025; accepted on 21 January 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.14.1.0163

Abstract

This paper explores the integration of Artificial Intelligence (AI) in Robotic Process Automation (RPA) for document processing, specifically focusing on Optical Character Recognition (OCR) of image receipts in multiple languages. By leveraging a hybrid Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) model, the system is capable of accurately extracting textual information from receipts. Furthermore, the model classifies these receipts into distinct categories, enabling efficient data management and analysis.

Keywords: RPA; CNN; RNN/LSTM; CTC; OCR

1. Introduction

With the increasing volume of documents and receipts in daily business transactions, manual processing has become time-consuming and error prone. Robotic Process Automation (RPA) combined with Artificial Intelligence (AI) offers an effective solution for automating document processing, particularly in extracting and classifying data from image-based receipts. This paper explores an AI-powered system utilizing a hybrid model of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to perform Optical Character Recognition (OCR) on receipts in multiple languages, including English, Spanish, French and so on. The proposed system not only enhances the accuracy of text extraction but also classifies receipts into different categories, improving overall efficiency and reducing human intervention in day-to-day activities.

1.1. Relevant formulas

- CNN Layer: The CNN extracts features from the input receipt image (1). This operation helps in extracting essential spatial features from the receipt image, such as text areas and important details.

$$Y_{i,j} = (X * W)_{i,j} + b \text{................ (1)}$$

X – Input Image
W – Convolution Filter (Kernel)
$*$ - Convolution
b – bias term
Y – Output of Convolution Layer

- RNN/LSTM Layer: After feature extraction using CNN, the LSTM/RNN layer captures the sequential dependencies in the text from the image. The LSTM network updates its state at each timestep t using the following formula (2). This process helps capture the sequential nature of the receipt's text (i.e., reading the receipt from left to right).

* Corresponding author: Deshpande Spriha

$$h_t = o_t \cdot \tanh C_t \ldots\ldots\ldots(2)$$

o_t – Output Gate
C_t – Cell State
tanh - hyperbolic tangent activation function

- CTC Loss: CTC loss is used for sequence-to-sequence tasks where the input and output sequences are not aligned. In OCR, the input is an image and output is the corresponding text.

$$\mathrm{L}_{CTC} = -\log(\textstyle\sum_{\mathrm{a}\in\mathrm{A}^L} \prod_{t=1}^{T} P(y_t = a_t)) \ldots\ldots\ldots(3)$$

- T - The number of timesteps (predicted sequence length).
- L - The number of characters in the true label sequence.
- a_t - The character at timestep t in a possible alignment
- $P(y_t{=}a_t)$ - The probability of the predicted character y_t matching the alignment character a_t.

2. Architecture

The flowchart in Figure 1 illustrates the overall pipeline for Optical Character Recognition (OCR) using a combination of deep learning components, including CNN, RNN, and the Connectionist Temporal Classification (CTC) loss function. The process begins with Image Acquisition, where input images of text are collected. These images undergo Image Preprocessing to enhance their quality, normalize their size, and convert them into a suitable format for feature extraction. This ensures the images are optimized for subsequent processing by the deep learning model. The preprocessed images are passed through a Convolutional Neural Network (CNN), which extracts meaningful spatial features, such as edges and textures, from the images. These features are then reshaped and fed into the next layer.

The extracted features are processed by RNN/LSTM Layers to capture temporal dependencies and sequence patterns, enabling the model to understand character sequences. The output from these layers is passed to a Dense Layer with SoftMax Activation, which classifies the probability distribution of each character in the sequence. Finally, the CTC Loss Function aligns the predicted sequence with the ground truth labels while handling variable-length inputs and outputs, making it ideal for OCR tasks. The decoded sequence is output as readable text. Feedback from the model allows for iterative improvements, ensuring accurate text recognition. This structured flow ensures a systematic and efficient approach to extracting textual information from images.

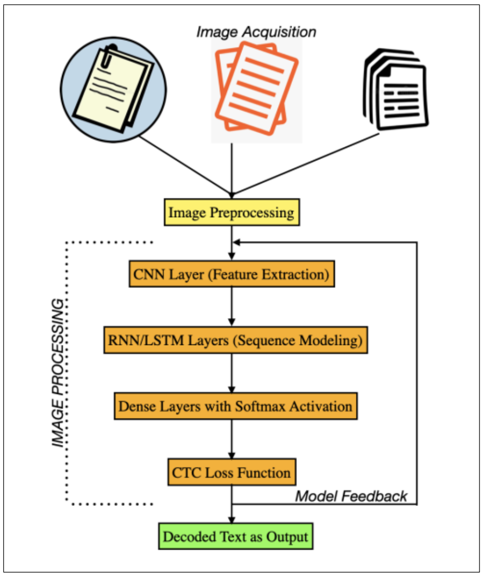

Figure 1 Flowchart of the OCR Pipeline with CNN, RNN and CTC Loss

3. Conclusion

This paper presents an AI-powered Robotic Process Automation (RPA) system that integrates Optical Character Recognition (OCR) with a hybrid CNN-RNN model and Connectionist Temporal Classification (CTC) loss for efficient multilingual receipt processing. The proposed approach enhances text extraction accuracy and facilitates automated classification of receipts into distinct categories, streamlining document management and reducing human intervention. By leveraging advanced deep learning techniques, the system demonstrates the ability to handle variability in receipt formats, languages, and image quality, making it suitable for real-world business applications. Future work can focus on incorporating advanced attention mechanisms, improving scalability, and exploring the integration of Natural Language Processing (NLP) for semantic analysis of extracted data. The proposed solution represents a significant step forward in automating and optimizing document processing workflows across industries.

Compliance with ethical standards

Acknowledgments

I would like to express my gratitude to my mentor Sahil Nyati, Dr Pavan Kumar Gautam for providing me resources to extend my research in the domain of automation. I also thank my organization NVIDIA for giving me the exposure to a world which is influenced by AI.

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] R. R and K. R. Anne, "CNN-RNN Hybrid Model to Classify a Local Language Slangs using Spectral features," *2024 International Conference on Inventive Computation Technologies (ICICT)*, Lalitpur, Nepal, 2024, pp. 600-607, doi: 10.1109/ICICT60155.2024.10544381.

[2] Z. Dong, R. Zhang and X. Shao, "A CNN-RNN Hybrid Model with 2D Wavelet Transform Layer for Image Classification," *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, Portland, OR, USA, 2019, pp. 1050-1056, doi: 10.1109/ICTAI.2019.00147.

[3] S. Lyu and J. Liu, "Hybrid Framework of Convolution and Recurrent Neural Networks for Text Classification," *2020 IEEE International Conference on Knowledge Graph (ICKG)*, Nanjing, China, 2020, pp. 313-320, doi: 10.1109/ICBK50248.2020.00052.

[4] J. Zhang, Y. Li, J. Tian and T. Li, "LSTM-CNN Hybrid Model for Text Classification," *2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, Chongqing, China, 2018, pp. 1675-1680, doi: 10.1109/IAEAC.2018.8577620.

[5] X. She and D. Zhang, "Text Classification Based on Hybrid CNN-LSTM Hybrid Model," *2018 11th International Symposium on Computational Intelligence and Design (ISCID)*, Hangzhou, China, 2018, pp. 185-189, doi: 10.1109/ISCID.2018.10144.

[6] X. Shi, T. Wang, L. Wang, H. Liu and N. Yan, "Hybrid Convolutional Recurrent Neural Networks Outperform CNN and RNN in Task-state EEG Detection for Parkinson's Disease," *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Lanzhou, China, 2019, pp. 939-944, doi: 10.1109/APSIPAASC47483.2019.9023190.