Design and implementation of an optical character recognition system and invoice information extraction using Computer Vision combined with Large Language Model

Le Bich Thao

Abstract: The research aims to automate the financial data entry process by combining OCR and LLM technology. The author uses the VietOCR model based on the Transformer architecture (VGG Seq2Seq) to recognize Vietnamese text from invoice images, overcoming the limitations of traditional methods when dealing with noisy images. Subsequently, the LLM plays a role in semantic post-processing, correcting spelling errors and converting raw data into JSON structure. Experimental results show that the fine-tuned VietOCR model achieves a character accuracy of 97.55% and a word error rate (WER) reduced to 3.26%, effectively meeting the digital transformation needs of businesses.

Keywords: OCR, VietOCR, LLM, Information extraction, Electronic invoice

References

[1] T. Do et al., "Key information extraction and recognition from rich text images," Vietnam J. Comput. Sci., vol. 11, no. 4, pp. 569–594, 2024.

[2] A. C. Tran, L. T. Ho, and H. T. Nguyen, "Information extraction from invoices by using a graph convolutional neural network: a case study of Vietnamese stores," IEIE Trans. Smart Process. & Comput., vol. 11, no. 5, pp. 316–323, 2022.

[3] F. Loukil et al., "LLM-centric pipeline for information extraction from invoices," in Proc. Int. Conf. Foundation and Large Language Models (FLLM2024), 2024.

[4] J. Miano, Compressed Image File Formats: Jpeg, Png, Gif, Xbm, Bmp. Addison-Wesley Professional, 1999.

[5] F. Zhuang et al., "A comprehensive survey on transfer learning," Proc. IEEE, vol. 109, no. 1, pp. 43–76, 2020.

[6] L. N. Smith, "Cyclical learning rates for training neural networks," in Proc. 2017 IEEE Winter Conf. Applications of Computer Vision (WACV), 2017.

[7] H. S. Oluwatosin, "Client-server model," IOSR J. Comput. Eng., vol. 16, no. 1, pp. 67–71, 2014.