Improving OCR outputs through Output Fusion

December 2, 2024

Duration: 4-6 months

Host Organization: Laboratory of Advanced Technology and Intelligent Systems, ENISo Sousse, Tunisia

Contact: hamza.gbada@gmail.com
Description:

Optical Character Recognition (OCR) is a important task in information. However, the OCR perfromance can vary depending on image quality, font styles, or the language of the text. Currently, many open-source OCR engine each of them give better performance under a point-wise conditions. This internship aims to develop approaches for merging the outputs of multiple OCR systems to improve the quality of the results. The work will focus on:

- A comparative analysis of the performance of various OCR systems.

- Designing and implement a method for fusion (e.g., based on majority voting, probabilistic models, or machine learning techniques).

- Evaluating the performance of the fused system on diverse datasets.

Requirements:

- Strong programming skills, preferably in Python.

- Familiarity with image processing techniques, OCR systems and deep learning

- Experience with machine learning techniques, data handling and evaluation

- Ability to work independently and in a team

- Problem-solving and analytical skills

- Familiarity with version control systems (Git)

References

1. Zhou, J., Yang, C., Zhu, Y., & Zhang, Y. (2024). Cross-region feature fusion with geometrical relationship for OCR-based image captioning. *State Key Laboratory of Media Convergence and Communication, the Communication University of China*, 100024, Beijing, China. Received 21 September 2023, Revised 11 February 2024, Accepted 14 July 2024, Available online 17 July 2024, Version of Record 24 July 2024.

2. Wilczok, E., & Lellmann, W. (2004). Adaptive Combination of Commercial OCR Systems. In A. Dengel, M. Junker, & A. Weisbecker (Eds.), *Reading and Learning* (Lecture Notes in Computer Science, vol. 2956, pp. 111–118). Springer, Berlin, Heidelberg.

   https://doi.org/10.1007/978-3-540-24642-8_8

3. Semkovych, V., & Shymanskyi, V. (2022). Combining OCR methods to improve handwritten text recognition with low system technical requirements. In The International Symposium on Computer Science, Digital Economy and Intelligent Systems* (pp. 693-702). Cham: Springer Nature Switzerland.

4. Kuang, Z., Sun, H., Li, Z., Yue, X., Lin, T. H., Chen, J., & Lin, D. (2021). MMOCR: A comprehensive toolbox for text detection, recognition, and understanding. In *Proceedings of the 29th ACM International Conference on Multimedia* (pp. 3791-3794).

5. Hsu, C. J., Chen, Y. C., Liao, F. T., Ho, P. C., Wang, Y. H., Hsu, P. C., & Shiu, D. S. (2024). Let's Fuse Step by Step: A Generative Fusion Decoding Algorithm with LLMs for Multi-modal Text Recognition. *arXiv preprint* arXiv:2405.14259.