Qeios

Peer Review

Review of: "Arabic-Nougat: Fine-Tuning Vision Transformers for Arabic OCR and Markdown Extraction"

Aiman Solyman[1]

1. University of Milan, Italy

The manuscript titled "*Arabic-Nougat: Fine-Tuning Vision Transformers for Arabic OCR and Markdown Extraction*" aims to propose an Arabic OCR model based on fine-tuning the Meta-Nougat. Although the research topic is interesting and relevant, the paper has several issues in terms of content and presentation; below are my comments:

General Comments

Academic Writing Issues: The paper requires improvements in academic tone and clarity. Grammar, coherence, and sentence structures need polishing.

Main Contribution: Fine-tuning Meta Nougat with synthetic data is not a strong contribution. Highlighting the dataset creation as the primary contribution would provide more value.

Introduction

1. Expand the introduction by providing a stronger foundation on prior work in Arabic OCR and Markdown parsing.
2. Clarify the dataset's legal and copyright status since it was collected from the Hindawi website.
3. Provide additional information about the Arabic books dataset, including their form, structure, and linguistic complexity (e.g., unique word count or vocabulary richness of the 1.1 billion Arabic tokens).
4. Contribution 4 in the contribution summary appears redundant and not novel. Either remove it or revise it to highlight a specific improvement.

Related Work

1. The related work section lacks citations for recent studies, especially those focused on Arabic OCR technologies.

2. A comparison table summarizing related methods and their performance metrics is recommended.

Methodology

1. The methodology is underdeveloped; a more detailed explanation of the architecture is required.

2. Figure 1 needs improvement. Annotate details of the subcomponents to make it more descriptive. Labels and arrows should clearly indicate the data flow and model transformations.

3. Include equations and mathematical representations for the model's components. Provide pseudocode for training steps to clarify implementation.

4. Provide statistics about the dataset, such as class distribution, text length, and types of formatting features present.

Empirical Evaluation

This section is weak; I would suggest deeply improving it or integrating it with the results section.

Results

Results are shallow, limited to a single table. Additional experiments should:

- Compare performance with other Arabic OCR models.

- Evaluate robustness across different text layouts, fonts, and noise levels.

- Include ablation studies to isolate the effects of tokenizer improvements and architectural enhancements.

Recommendations and Limitations

Merge recommendations and limitations into one section named Discussion and position it before the conclusion.

Conclusion

The conclusion should summarize key contributions more explicitly and emphasize the availability of datasets and tools as contributions to the broader Arabic NLP community. However, ending with a forward-looking statement would be better to encourage future research.

Declarations

Potential competing interests: No potential competing interests to declare.