OPTICAL CHARACTER RECOGNITION (OCR) MODEL USING TRANSFORMER FOR VIETNAMESE HANDWRITING RECOGNITION.

Nguyễn Duy Thắng [1]

[1] Trường ĐH Công Nghệ Thông Tin (UIT)

What ?

We introduce a Pre-trained TrOCR for the Vietnamese handwriting recognition task, in which we have:

- Improve the Word Accuracy of the TrOCR model in the task of recognizing images of Vietnamese handwriting.
- Train a pre-trained model to easily customize more Vietnamese handwriting image data with more diverse fonts.
- Apply the model to a specific application: recognize Vietnamese handwriting from an input image in the Translate application.

Why ?

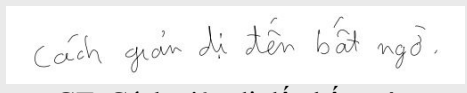
Optical Character Recognition (OCR) is an important technology for converting text from images into digital writing, playing a key role in many applications such as document digitization, handwriting recognition, and text extraction. information from images.

However, OCR still faces many challenges, especially when working with complex languages like Vietnamese, requiring OCR models to have sophisticated and accurate processing capabilities.

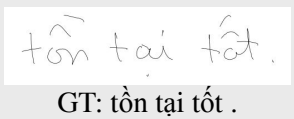
Overview

Collect Data → Fine-tune Pre-trained TrOCR → Evaluate → Application

GT: Bản chất của thành công



GT: Cách giản dị đến bất ngờ .



GT: tồn tại tốt .

Figure 1. Sample.

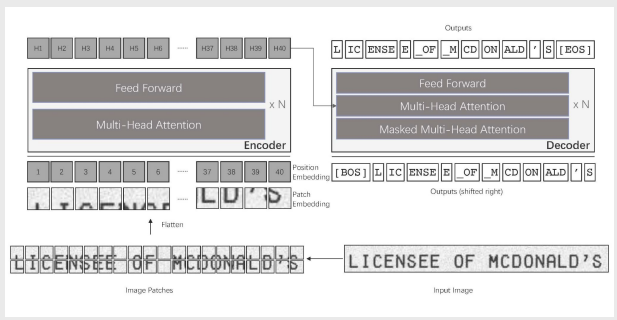


Figure 2. TrOCR pipeline.

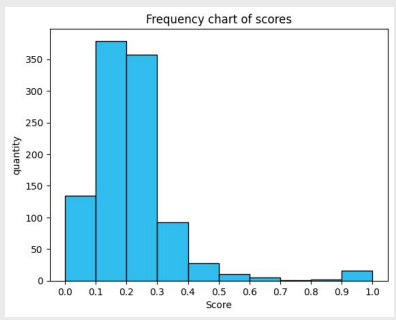


Figure 3. Testing results.

Figure 4. Application

Description

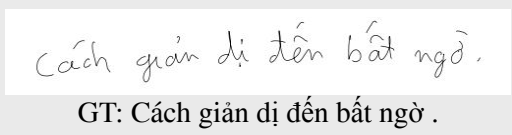
1. Collect Data

- Labeled Vietnamese handwriting data is collected and prepared from Kaggle. This data is then divided into training, validation, and test sets for number: 5097 - 1161 - 1024.
- Take data labels from the above three sets and put them into a single .csv file for convenience in training and evaluation.

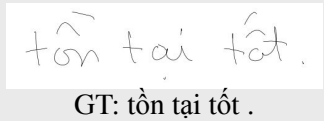
2. Fine-tune Pre-trained TrOCR

- We use TrOCR-Base-IAM, trained model on IAM Dataset, to learn basic features of English handwriting images, was selected to proceed..
- The pre-trained model will be fine-tuned on the Vietnamese handwriting dataset above to adapt with Vietnamese.

GT: Bản chất của thành công



GT: Cách giản dị đến bất ngờ .



GT: tồn tại tốt .

Figure 1. Sample.

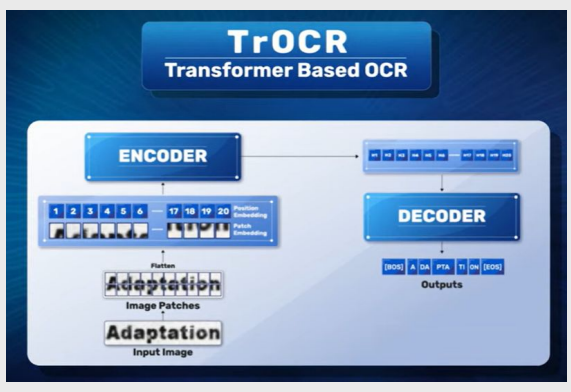


Figure 5. TrOCR architecture.

3. Evaluate

- The model is evaluated on a validation set to check its performance according to Word Accuracy.
- However, because of limited time, the model can only be trained for 5 epochs. The evaluation results on the test set are as figure 3:

4. Application

- Above is the orc integrated application: Figure 4.
- Youtube and Github for demo and more:
  - https://www.youtube.com/watch?v=9Ds2L92Kt6w
  - https://github.com/DyThen-Kumo/CS519.O21.KHTN

NII

Nguyễn Duy Thắng – Trường Đại học Công Nghệ Thông Tin UIT
TEL : 096**621 Email : 22521333@gm.uit.edu.vn**