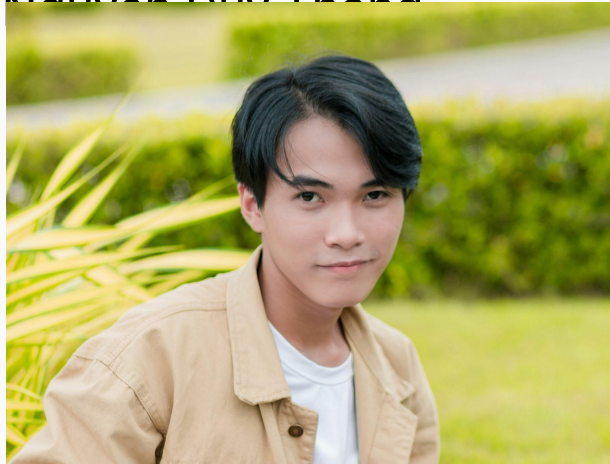


MÔ HÌNH NHẬN DIỆN KÝ TỰ QUANG HỌC
(OCR) SỬ DỤNG TRANSFORMER
CHO NHẬN DIỆN CHỮ VIẾT TAY TIẾNG VIỆT.

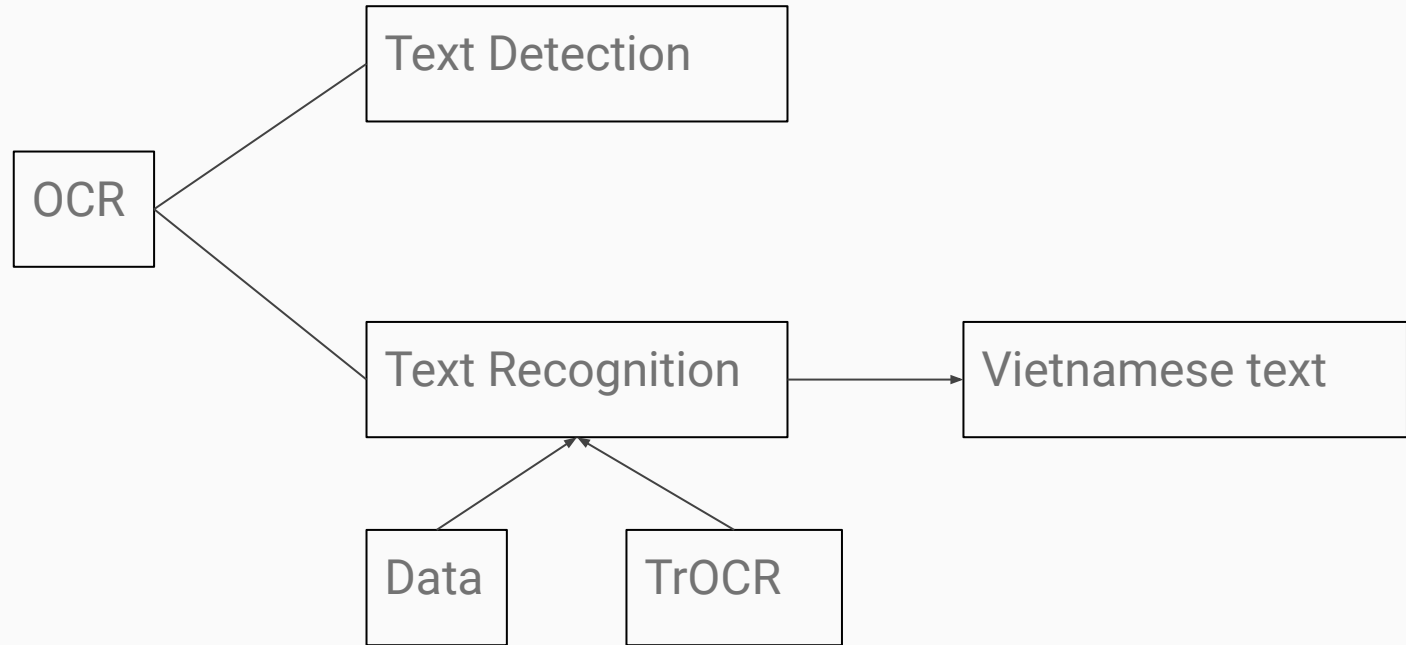
Nguyễn Duy Thắng - 22521333

Tóm tắt

- Lớp: CS519.021.KHTN
- Link Github: <https://github.com/DyThen-Kumo/CS519.021.KHTN>
- Link YouTube video:
<https://www.youtube.com/watch?v=9Ds2L92Kt6w>
- Ảnh + Họ và Tên: Nguyễn Duy Thắng



Giới thiệu



Mục tiêu

1. Cải thiện độ chính xác Word Accuracy của mô hình TrOCR trong tác vụ nhận diện hình ảnh chữ viết tay tiếng Việt.

$$\text{Word Accuracy} = \text{Words Correct} / (\text{Words Correct} + \text{Words Misspelled} + \text{Words Skipped} + \text{Words Added})$$

2. Đào tạo một mô hình đào tạo sẵn (pre-trained model) để dễ dàng tùy chỉnh cho nhiều dữ liệu hình ảnh chữ viết tay tiếng Việt với đa dạng kiểu chữ hơn.
3. Áp dụng mô hình cho một ứng dụng cụ thể: nhận diện chữ viết tay tiếng Việt từ một hình ảnh đầu vào trong ứng dụng Translate.

Nội dung và Phương pháp

Nội dung

- Nghiên cứu các hệ thống OCR hiện nay và cách chúng hoạt động.
- Nghiên cứu về cấu trúc Transformer và quá trình xử lý của nó.
- Nghiên cứu về TrOCR, mô hình OCR theo Transformer đã được đào tạo sẵn cho các tác vụ nhận dạng chữ viết tay.
- Thu thập dữ liệu chữ viết tay tiếng Việt đã được gán nhãn để có bộ dataset đủ lớn để có thể finetuning mà không bị overfitting.
- Huấn luyện mô hình TrOCR base trên bộ dataset đã thu thập, đánh giá kết quả và lưu lại pre-trained model.
- Xây dựng ứng dụng cho pre-trained model đã có.

Nội dung và Phương pháp

Phương pháp

- **Thu Thập và Chuẩn Bị Dữ Liệu:** Dữ liệu chữ viết tay tiếng Việt đã được gán nhãn được thu thập và chuẩn bị từ Internet. Dữ liệu này sau đó được chia thành các tập train, validation, test.
- **Mô Hình Huấn Luyện Ban Đầu:** Pre-trained TrOCR (TrOCR-Base-IAM), được huấn luyện trên một tập dữ liệu lớn và đa dạng (IAM Dataset) để học các đặc trưng cơ bản của hình ảnh chữ viết tay tiếng Anh, được lựa chọn để tiến hành finetuning.
- **FineTuning với Dữ Liệu Chữ Viết Tay Tiếng Việt:** Pre-trained model sẽ được finetuning trên tập dữ liệu chữ viết tay tiếng Việt. Quá trình này bao gồm việc điều chỉnh các tham số mô hình để phù hợp với các đặc điểm ngữ cảnh và hình ảnh của tiếng Việt.
- **Đánh Giá và Tinh Chỉnh:** Mô hình được đánh giá trên tập validation để kiểm tra hiệu suất. Dựa trên các kết quả này, mô hình sẽ được tinh chỉnh thêm để đạt được độ chính xác cao nhất rồi lưu lại làm pre-trained model.
- **Xây Dựng Ứng Dụng được tích hợp Pre-Trained Model:** Xây dựng một ứng dụng dịch máy đơn giản, cho phép upload một bức ảnh chữ viết tay tiếng Việt và nhận diện nội dung bức ảnh làm input cho phân dịch nghĩa sang ngôn ngữ khác.

Kết quả dự kiến

- **Độ chính xác của mô hình:** Độ chính xác mong đợi dựa theo Word Accuracy trung bình là lớn hơn 80%.
- **Pre-trained Model:** Mô hình đào tạo trước sẽ có kích thước ít hơn 1.5GB để dễ dàng cài đặt vào ứng dụng.
- **Ứng dụng được tích hợp Pre-trained Model:** Ứng dụng có thể cho phép người dùng upload hình ảnh chữ viết tay tiếng Việt và nhận diện chính xác.

Tài liệu tham khảo

- [1]. Minghao Li, Tengchao Lv, Jingye Chen, Lei Cui, Yijuan Lu, Dinei A. F. Florêncio, Cha Zhang, Zhoujun Li, Furu Wei: TrOCR: Transformer-Based Optical Character Recognition with Pre-trained Models. AAAI 2023: 13094-13102
- [2]. Jan Kohút, Michal Hradis: Fine Tuning Is a Surprisingly Effective Domain Adaptation Baseline in Handwriting Recognition. CoRR abs/2302.06308 (2023)
- [3]. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin: Attention is All You Need. NIPS 2017: 5998-6008
- [4]. Michael Jungo, Lars Vögtlin, Atefeh Fakhari, Nathan Wegmann, Rolf Ingold, Andreas Fischer, Anna Scius-Bertrand: Impact of Ground Truth Quality on Handwriting Recognition. CoRR abs/2312.09037 (2023)
- [5]. Vittorio Pippi, Silvia Cascianelli, Christopher Kermorvant, Rita Cucchiara: How to Choose Pretrained Handwriting Recognition Models for Single Writer Fine-Tuning. CoRR abs/2305.02593 (2023)
- [6]. Mst. Shapna Akter, Hossain Shahriar, Alfredo Cuzzocrea, Nova Ahmed, Carson K. Leung: Handwritten Word Recognition using Deep Learning Approach: A Novel Way of Generating Handwritten Words. CoRR abs/2303.07514 (2023)