



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



Nguyễn Duy Thắng - 22521333

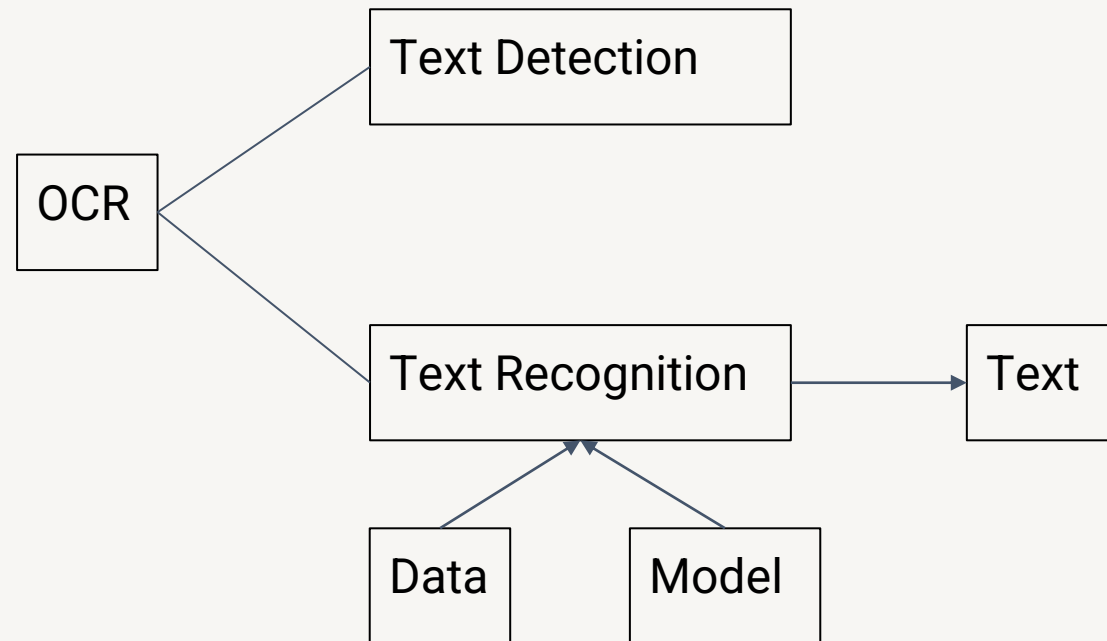
# ABOUT PROJECT

Lớp: CS519.O21.KHTN

Link Github:

<https://github.com/DyThen-Kumo/CS519.O21.KHTN>

# OCR SYSTEM



## OBSERVATIONS

Nhận diện chữ viết tay tiếng Việt gặp nhiều thách thức do tính đa dạng và phức tạp của các ký tự và dấu thanh. Các mô hình truyền thống thường kết hợp CNN để xử lý hình ảnh, RNN để tạo ra đầu ra và một mô hình ngôn ngữ để sửa lỗi chính tả, dẫn đến hệ thống phức tạp và khó tối ưu hóa. Kiến trúc Transformer, đã đạt nhiều thành công trong xử lý ngôn ngữ tự nhiên (NLP) và thị giác máy tính (CV), có thể giải quyết những hạn chế này.

## PROBLEMS

Tuy nhiên, các mô hình OCR dựa trên Transformer hiện nay thường được huấn luyện trên bộ dữ liệu tiếng Anh, giảm hiệu quả khi áp dụng cho tiếng Việt. Việc huấn luyện lại từ đầu rất tốn kém tài nguyên và thời gian. Liệu có cách nào để khắc phục điều đó?

## OBSERVATIONS + PROBLEMS

# INPUT - OUTPUT

**Input:** Một hoặc nhiều hình ảnh chữ viết tay tiếng Việt và dataset chữ viết tay tiếng Việt đã được gán nhãn.

**Output:** Văn bản được trích xuất từ hình ảnh đầu vào. Văn bản này có thể được lưu dưới dạng chuỗi ký tự hoặc dưới dạng tệp văn bản (.txt) hoặc các định dạng tài liệu khác (ví dụ: .docx, .pdf).



# COMPUTATIONAL THINKING

## DECOMPOSITION

Bài toán được chia thành 2 phần chính là **Thu thập dữ liệu** và **Nhận diện chữ viết tay tiếng Việt**.

## ABSTRACTION

Về thu thập dữ liệu, đây là công việc khá đơn giản nhưng cũng vất vả.

Về nhận diện, có thể khái quát hoá thành bài toán nhận diện chữ viết tay bằng Transformer nói chung.

## PATTERN RECOGNITION

Bài toán nhận diện chữ viết tay dựa trên Transformer đã được đào tạo trước nhưng chỉ trên dữ liệu tiếng Anh.

## ALGORITHM

Finetuning mô hình TrOCR được đào tạo trên dữ liệu tiếng Anh và đào tạo lại trên dữ liệu tiếng Việt.



## SOLUTION

# FINETUNING

Tinh chỉnh (finetuning) một mô hình đã được đào tạo trên ngôn ngữ có bảng chữ cái Latinh để đào tạo thêm về tiếng Việt vì mô hình đã học được một số nét chữ cơ bản.

# 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 + 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 + PHƯƠNG PHÁP

## PHƯƠNG PHÁP

Thu Thập và Chuẩn Bị Dữ Liệu.

Mô Hình Huấn Luyện Ban Đầu.

FineTuning với Dữ Liệu Chữ Viết Tay Tiếng Việt.

Đánh Giá và Tinh Chỉnh.

Xây Dựng Ứng Dụng được tích hợp Pre-Trained Model.

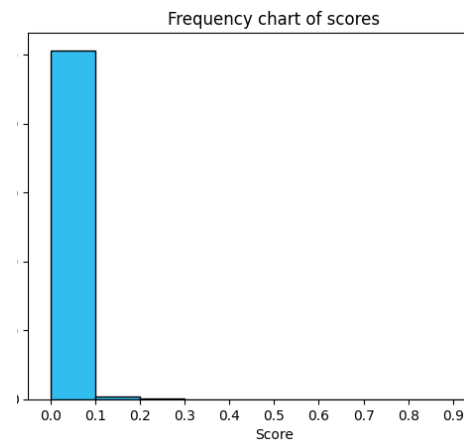
## KẾT QUẢ MONG ĐỢI

**Độ chính xác của mô hình:** Độ chính xác mong đợi dựa theo Word Accuracy trung bình là lớn hơn 80%.

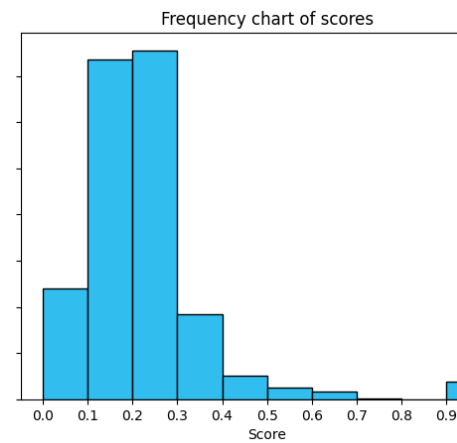
**Pre-trained Model:** Mô hình đào tạo trước sẽ có kích thước ít hơn 1.5GB để dễ dàng cài đặt vô ứng dụng.

**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.

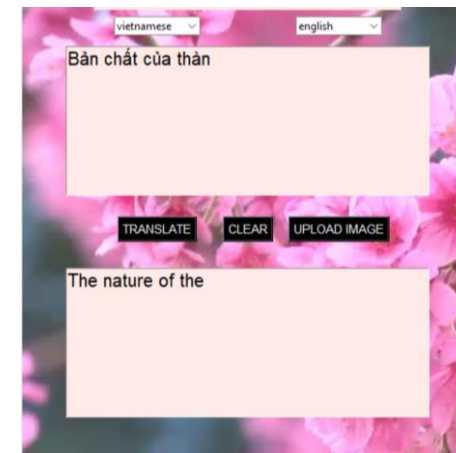
# KẾT QUẢ



Trước khi finetuning



Sau khi Finetuning



Ứng dụng trong dịch máy

# REFERENCE

[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)

A series of thin, light brown lines forming an abstract, overlapping geometric pattern on the left side of the slide. The lines intersect to create various polygonal shapes, some of which are filled with a very light brown color.

# THANK YOU