

# TRÍCH XUẤT THÔNG TIN TỪ CHỨNG MINH THƯ SỬ DỤNG YOLO, VIETOCR VÀ CƠ CHẾ ATTENTION

*EXTRACTING INFORMATION FROM ID CARDS USING YOLO, OCR AND  
ATTENTION MECHANISM*

Nguyễn Phương Lan - 240101015

# Tóm tắt

- Lớp: CS2205.CH183
- Link Github của thành viên:  
<https://github.com/nguyenphuonglanuit/CS2205.CH183.git>
- Link YouTube video: <https://youtu.be/ytmr3-OnKVg>
- Họ và Tên: Nguyễn Phương Lan



# Giới thiệu

- Trích xuất thông tin từ hình ảnh (chứng minh thư, bằng lái xe,...) là vấn đề phổ biến, yêu cầu tốc độ, tiện lợi và độ chính xác cao. Xử lý ngôn ngữ tiếng Việt gặp khó khăn do đặc trưng dấu câu và ký tự phức tạp. Mô hình VietOCR kết hợp Yolo và tiền xử lý phù hợp và hiệu quả trong xử lý chữ viết tay và tài liệu tiếng Việt.
- Đầu vào mặt trước thẻ căn cước công dân không gắn chip hoặc mặt trước chứng minh nhân dân. Đầu ra là các thông tin văn bản được trích xuất từ hình ảnh, đảm bảo độ chính xác cao, xử lý tốt các đặc trưng ngôn ngữ tiếng Việt (bao gồm cả dấu và ký tự đặc biệt).



```
{
  "typeID" : "12so_front"
  "id" : "001097018379"
  "name" : "MAI THÀNH CÔNG"
  "gender" : "Nam"
  "birth_day" : "28/01/1997"
  "expire_date" : "28/01/2022"
  "nationality" : "Việt Nam"
  "origin_location" : "Nam Đồng, Đồng Đa, Hà Nội"
  "recent_location" : "TDP Trung 5, Xuân Đình, Bắc Từ Liêm, Hà Nội"
}
```

# Mục tiêu

- Tự động hóa quy trình xác thực để giảm thời gian xử lý hồ sơ, hạn chế sai sót nhập liệu và nâng cao hiệu quả làm việc.
- Hỗ trợ hệ thống xác thực tự động nhằm ứng dụng trong eKYC, kiểm tra an ninh, hỗ trợ ngân hàng, bảo hiểm và dịch vụ công. Đặc biệt là hệ thống còn thúc đẩy quy trình quản lý khách hàng.
- Cải thiện độ chính xác bằng độ nhận diện văn bản: Xử lý các trường hợp ảnh kém chất lượng, góc chụp lệch bằng cách sử dụng YOLO, OCR và Attention Mechanism để tăng hiệu suất nhận diện văn bản.
- Xây dựng hệ thống trích xuất thông tin thực tế bằng cách kết hợp Yolo V8 và VietOCR và sử dụng các độ đo đánh giá WER VÀ BLEU để nâng cao tính ứng dụng vào trong thực tế.

# Nội dung và Phương pháp

- Nội dung 1: Dataset, nhận diện đối tượng bằng yolo và tiền xử lý hình ảnh
  - Tập dataset: Mặt trước Chứng minh nhân dân 9 số và mặt trước Căn Cước Công Dân không gắn chip thu thập từ Internet.
  - Đầu tiên, sử dụng Yolo V8 để xác định vùng ảnh chứa các thông tin quan trọng như số ID, tên, ngày sinh. Đồng thời có thể xử lý được các trường hợp ảnh đặc biệt như ảnh bị nghiêng hoặc bị mất một phần
  - Tiếp theo, tiền xử lý để đảm bảo chất lượng đầu vào của các thuật toán: Điều chỉnh độ sáng, độ tương phản, và giảm nhiễu bằng các phương pháp như Gaussian Blur. Sử dụng các thuật toán xoay và cắt để căn chỉnh các vùng căn bản (xử lý góc chụp). Đồng thời, chuẩn hóa kích thước ảnh giúp hệ thống xử lý các ảnh từ nhiều nguồn khác nhau một cách nhất quán.

# Nội dung và Phương pháp

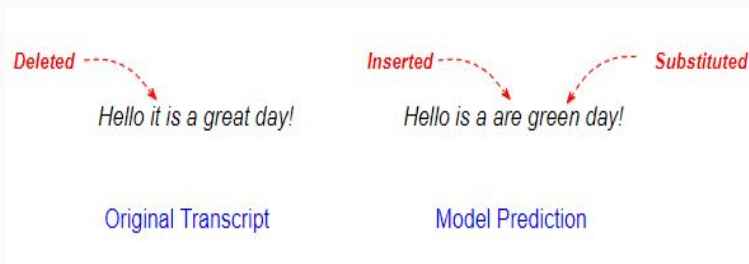
- Nội dung 2: Nhận diện văn bản bằng OCR, trích xuất và phân loại thông tin
  - Sau khi Yolo V8 xác định các vùng cần thiết, thuật toán OCR Tesseract sẽ trích xuất văn bản từ các vùng này và kết hợp với các mạng LSTM hoặc Attention Mechanism để cải thiện độ chính xác, đồng thời sẽ xử các văn bản phức tạp như chữ in mờ hoặc bị méo.
  - Các văn bản sau khi trích xuất sẽ được phân tích và gán nhãn vào các trường hợp tương ứng như: Từ khóa nhận diện các trường thông tin (giới tính, nơi cấp), xác định dạng số ID và ngày tháng bằng regex.

# Nội dung và Phương pháp

- Nội dung 3: Cải thiện kết quả bằng phương pháp nâng cao
  - Attention Mechanism: Tăng khả năng nhận diện chính xác các chi tiết nhỏ.
  - Học sâu: Kết hợp với các mạng CNN hoặc Transformer để xử lý ảnh phức tạp.
  - Hệ thống sẽ hoàn chỉnh và trả về đối tượng JSON chứa đầy đủ thông tin từ CMND/CCCD.

# Nội dung và Phương pháp

- Nội dung 4: Các thang đo được sử dụng
  - Độ đo đánh giá Word Error Rate (WER)



$$\begin{aligned}\text{Word Error Rate} &= \frac{\text{Inserted} + \text{Deleted} + \text{Substituted}}{\text{Total words in transcript}} \\ &= \frac{1 + 1 + 1}{6} \\ &= 0.5\end{aligned}$$

- Độ đo đánh giá BLEU

$$p_n = \frac{\sum_{C \in \{\text{Candidates}\}} \sum_{(n\text{gram} \in C)} \text{Count}_{\text{clip}}(n\text{gram})}{\sum_{C' \in \{\text{Candidates}\}} \sum_{(n\text{gram}' \in C')} \text{Count}(n\text{gram}')}$$

$$\log \text{BLEU} = \min\left(1 - \frac{r}{c}, 0\right) + \sum_{n=1}^N w_n \log p_n$$



# Kết quả dự kiến

**Độ chính xác cao:** Nhận dạng chữ viết tay và chữ in trên CMND/CCCD với độ chính xác  $>90\%$ . Giảm sai sót nhập liệu, nâng cao hiệu quả xử lý dữ liệu.

**Tốc độ xử lý nhanh:** Xử lý ảnh  $<1$  giây, phù hợp với ứng dụng cần xử lý dữ liệu lớn.

**Khả năng thích ứng:** Hoạt động tốt trên ảnh mờ, ánh sáng yếu, góc chụp lệch, đảm bảo nhận diện ổn định và chính xác.

**Tích hợp dễ dàng:** Xuất dữ liệu dưới dạng JSON, thuận tiện cho việc tích hợp hệ thống. Hỗ trợ truyền dữ liệu linh hoạt giữa các nền tảng.

# Tài liệu tham khảo

- [1]. **Redmon, J., Divvala, S., Girshick, R., & Farhadi, A.** "You Only Look Once: Unified, Real-Time Object Detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779-788. DOI: 10.1109/CVPR.2016.91.
- [2]. **Liu, W., Anguelov, D., Erhan, D., Szegedy, C., & Reed, S.** "SSD: Single Shot Multibox Detector," *European Conference on Computer Vision (ECCV)*, 2016, pp. 21-37. DOI: 10.1007/978-3-319-46448-0\_2.
- [3]. **Mishra, A., & Agrawal, A.** "A Survey on Text Detection and Recognition from Image," *International Journal of Computer Applications*, vol. 182, no. 5, pp. 19-24, 2019. DOI: 10.5120/ijca2019919077.
- [4]. **Xu, Y., & Yang, H.** "Text Detection and Recognition in Natural Images with Yolo and Tesseract OCR," *Proceedings of the International Conference on Computer Vision and Image Processing (CVIP)*, 2019, pp. 1-5. DOI: 10.1109/CVIP.2019.00001.
- [5]. **Lee, J., & Park, S.** "Real-Time ID Card Recognition Using YOLO and OCR with Edge Devices," *IEEE Access*, vol. 8, pp. 181204-181213, 2020. DOI: 10.1109/ACCESS.2020.3024593.