THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo (tối đa 5 phút):
  https://youtu.be/ytmr3-OnKVg
- Link slides (dạng .pdf đặt trên Github của nhóm):
  (ví dụ: https://github.com/nguyenphuonglanuit/CS2205.CH183.git*)*
- *Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới*
- *Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in*
- *Lớp Cao học, mỗi nhóm một thành viên*

- Họ và Tên: Nguyễn Phương Lan
- MSSV: 240101015

- Lớp: CS2205.CH183
- Tự đánh giá (điểm tổng kết môn): 9.0/10
- Số buổi vắng: 0
- Số câu hỏi QT cá nhân: 8
- Link Github:
  https://github.com/nguyenphuonglanuit/CS2205.CH183.git

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

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

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

EXTRACTING INFORMATION FROM ID CARDS USING YOLO, OCR AND ATTETION MECHANISM

TÓM TẮT *(Tối đa 400 từ)*

Công nghệ nhận diện văn bản từ CMND/CCCD ứng dụng AI và Computer Vision, giúp trích xuất thông tin cá nhân tự động bằng Deep Learning và OCR, giảm sai sót và tăng tốc độ xử lý trong hành chính, tài chính và dịch vụ công. Trong ngân hàng, hệ thống eKYC quét và đối chiếu dữ liệu thay vì nhập thủ công, nâng cao trải nghiệm khách hàng và giảm rủi ro gian lận. Trong quản lý hành chính, số hóa dữ liệu giúp đơn giản hóa tra cứu hồ sơ và tăng hiệu suất làm việc.

Công nghệ này còn tăng cường bảo mật, sử dụng mô hình YOLO, SSD để phát hiện giấy tờ giả và tích hợp nhận diện khuôn mặt. Các nghiên cứu gần đây ứng dụng LSTM và Attention Mechanism giúp nâng cao hiệu suất, đảm bảo hoạt động ổn định ngay cả với hình ảnh chất lượng thấp.

Nhìn chung, nhận diện văn bản từ CMND/CCCD không chỉ tối ưu quy trình mà còn góp phần xây dựng nền tảng số hiện đại. Với sự phát triển của AI, công nghệ này hứa hẹn sẽ tiếp tục mở rộng và ứng dụng rộng rãi trong tương lai.

GIỚI THIỆU *(Tối đa 1 trang A4)*

Trích xuất thông tin từ hình ảnh là một lĩnh vực quan trọng trong xử lý ảnh và thị giác máy tính, đặc biệt là trong việc số hóa các tài liệu như thẻ căn cước, hóa đơn, hay giấy tờ tùy thân. Công nghệ này không chỉ giúp tự động hóa quy trình quản lý dữ liệu mà còn tăng tốc độ xử lý và giảm thiểu sai sót so với việc nhập liệu thủ công. Đặc biệt, nó đã trở thành một công cụ quan trọng trong các thủ tục hành chính, tài chính,

và dịch vụ công.

Hệ thống trích xuất thông tin từ hình ảnh yêu cầu ba yếu tố chính: tốc độ, độ chính xác và tính linh hoạt với các định dạng tài liệu khác nhau. Nhiều mô hình học sâu như mạng nơ-ron tích chập (CNN), mạng nơ-ron hồi tiếp (RNN), và các biến thể như LSTM hoặc GRU đã được áp dụng để nhận diện văn bản từ hình ảnh. Tuy nhiên, khi làm việc với các loại tài liệu phức tạp như giấy tờ tùy thân hoặc chữ viết tay, các mô hình này có thể gặp thách thức do sự đa dạng trong cách trình bày và các yếu tố gây nhiễu như góc nhìn, độ mờ của ảnh.

Các nghiên cứu chỉ ra rằng các mô hình tiên tiến như CTPN, Faster R-CNN, RetinaNet, và Tesseract OCR có khả năng cải thiện đáng kể độ chính xác của hệ thống. Đặc biệt, việc kết hợp giữa mô hình YOLO và OCR đã mang lại kết quả khả quan, giúp nâng cao cả về tốc độ và độ chính xác trong nhận diện ký tự. Tuy nhiên, các phương pháp truyền thống như phát hiện biên cạnh hay cắt viền giấy tờ vẫn gặp nhiều hạn chế khi áp dụng trong thực tế với dữ liệu chứa nhiều nhiễu.

Đối với tiếng Việt, một ngôn ngữ có hệ thống dấu phức tạp, thách thức lớn nhất là nhận diện chính xác ký tự và ghép từ. Để giải quyết vấn đề này, các nhà nghiên cứu đã kết hợp cơ chế Attention với mô hình OCR và YOLO, giúp hệ thống tập trung vào các vùng quan trọng trong hình ảnh, từ đó cải thiện khả năng nhận diện tiếng Việt. Sự kết hợp này đã cho thấy hiệu quả trong việc giảm thiểu lỗi và tăng độ chính xác.

Tóm lại, công nghệ nhận diện văn bản từ hình ảnh đóng vai trò then chốt trong việc tự động hóa và số hóa dữ liệu. Sự phát triển và ứng dụng các mô hình học sâu tiên tiến đã giúp cải thiện hiệu suất và mở rộng tiềm năng của công nghệ này trong thực tế.

MỤC TIÊU *(Viết trong vòng 3 mục tiêu)*

1. Tự động hóa quy trình xác thực:

Giảm thời gian xử lý, hạn chế sai sót, nâng cao hiệu suất, hỗ trợ eKYC, kiểm tra an ninh, ngân hàng, bảo hiểm và dịch vụ công.

2. Cải thiện độ chính xác trong 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.

3. Xây dựng hệ thống trích xuất thông tin thực tế:

Kết hợp Yolo V8 và VietOCR, đánh giá bằng WER (Word Error Rate) và BLEU (BiLingual Evaluation Understudy), đảm bảo hiệu suất cao và khả năng mở rộng trên dữ liệu thực tế.

NỘI DUNG VÀ PHƯƠNG PHÁP

1. Nội dung

Nghiên cứu OCR và ứng dụng Transformer trong nhận dạng chữ viết tay tiếng Việt, bao gồm: tổng quan hệ thống OCR, phân tích mô hình Transformer và ứng dụng VietOCR.

- Nghiên cứu các phương pháp OCR từ truyền thống đến deep learning, đánh giá hiệu quả, ưu điểm và hạn chế của từng phương pháp.
- Nghiên cứu cấu trúc Transformer, cơ chế self-attention và sự phối hợp giữa các thành phần để xử lý dữ liệu.
- Ứng dụng VietOCR, một mô hình Transformer, để nhận dạng chữ viết tay tiếng Việt, so sánh với OCR truyền thống và tối ưu hóa mô hình cho bài toán này.
- Các tiêu chí đánh giá hiệu suất: Nghiên cứu các phương pháp đánh giá độ chính xác của hệ thống OCR, bao gồm:
    - **Word Error Rate (WER):** Tiêu chí phổ biến để đánh giá lỗi trong nhận dạng văn bản.
    - **BLEU Score:** Chỉ số đo lường mức độ tương đồng giữa văn bản được nhận diện và văn bản gốc, được sử dụng trong lĩnh vực xử lý ngôn ngữ tự nhiên (NLP).

2. Phương pháp

- **Dữ liệu đầu vào**

Bộ dữ liệu sử dụng bao gồm hình ảnh mặt trước của Chứng minh nhân dân (CMND) 9 số và Căn cước công dân (CCCD) không gắn chip được thu thập từ Internet.

- **Xác định vùng thông tin quan trọng**

Sử dụng mô hình YOLO V8 để nhận diện và trích xuất các vùng quan trọng trên ảnh như:

- ❖ Số ID: Mã số CMND/CCCD.
- ❖ Họ tên: Họ và tên đầy đủ của người sở hữu.
- ❖ Ngày sinh: Ngày tháng năm sinh.
- ❖ Xử lý các trường hợp hình ảnh bị nghiêng, mất một phần dữ liệu hoặc có điều kiện ánh sáng kém.

- **Tiền xử lý ảnh**

Để tối ưu hóa chất lượng đầu vào cho mô hình OCR, kỹ thuật tiền xử lý sau được áp dụng:

- ❖ Điều chỉnh độ sáng, độ tương phản: Đảm bảo hình ảnh có độ rõ ràng tốt nhất.
- ❖ Giảm nhiễu: Sử dụng Gaussian Blur để loại bỏ nhiễu
- ❖ Xoay và cắt ảnh: Căn chỉnh góc chụp để chuẩn hóa kích thước ảnh

- **Nhận diện văn bản**

Sau khi xác định vùng chứa văn bản bằng YOLO V8, áp dụng hệ thống OCR để trích xuất nội dung, trong đó:

- ❖ Tesseract OCR: Sử dụng như một công cụ cơ bản để trích xuất văn bản từ ảnh.
- ❖ LSTM, Attention Mechanism: Kết hợp để nâng cao độ chính xác, đặc biệt đối với các văn bản bị mờ hoặc biến dạng.
- ❖ Cải tiến nhận dạng với Attention Mechanism: Cơ chế tự chú ý để tăng độ chính xác trong nhận diện các chi tiết nhỏ trên văn bản.

- **Phân loại và gán nhãn dữ liệu**

Sau khi trích xuất nội dung, hệ thống phân loại và gán nhãn cho dữ liệu để đảm bảo tính chính xác và hoàn thiện thông tin:

- ❖ Nhận diện trường thông tin: Hệ thống phân loại các đoạn văn bản trích xuất được vào các trường như số ID, họ tên, ngày sinh, giới tính, nơi cấp.
- ❖ Kiểm tra định dạng: Sử dụng biểu thức chính quy (Regex) để xác minh tính hợp lệ của số ID và định dạng ngày tháng.

- **Ứng dụng học sâu vào nhận diện văn bản**

Kết hợp CNN hoặc Transformer để trích xuất văn bản chính xác hơn từ ảnh phức tạp, kể cả trong điều kiện ánh sáng kém hoặc chất lượng thấp.

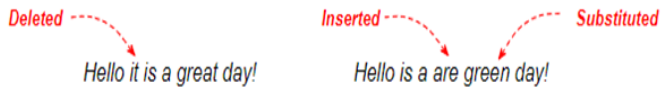
- **Định dạng đầu ra**

Hệ thống xuất kết quả nhận dạng dưới dạng tệp JSON, trong đó chứa đầy đủ thông tin trích xuất từ CMND/CCCD:

```
{
    "id": "123456789",
    "name": "Nguyễn Văn A",
    "dob": "01/01/1990",
    "gender": "Nam",
    "place_of_issue": "Hà Nội"
}
```

Định dạng giúp đảm bảo dữ liệu được tích hợp vào các hệ thống quản lý thông tin thực tế.

Độ đo đánh giá Word Error Rate (WER)



Original Transcript Model Prediction

$$\text{Word Error Rate} = \frac{\text{Inserted + Deleted + Substituted}}{\text{Total words in transcript}}$$

$$= \frac{1 + 1 + 1}{6}$$

$$= 0.5$$

Độ đo đánh giá BLEU

$$p_n = \frac{\sum_{C \in \{Candidates\}} \sum_{(ngram \in C)} Count_{clip}(ngram)}{\sum_{C' \in \{Candidates\}} \sum_{(ngram' \in C')} Count(ngram')}$$

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

KẾT QUẢ MONG ĐỢI

1. Độ chính xác cao

Hệ thống OCR cần đạt độ chính xác trên 90% khi nhận dạng chữ trên CMND/CCCD, giảm sai sót nhập liệu và nâng cao hiệu quả xử lý dữ liệu.

2. Tốc độ xử lý nhanh

Hệ thống OCR phải xử lý ảnh dưới 1 giây, đảm bảo hiệu suất cao và khả dụng trong môi trường thực tế.

3. Khả năng thích ứng

Hệ thống phải nhận diện chính xác trên ảnh chất lượng kém, như mờ, ánh sáng yếu, tương phản thấp hoặc bị méo, đảm bảo ổn định và đáng tin cậy.

4. Tích hợp dễ dàng

Kết quả nhận diện xuất ra dưới dạng JSON, giúp tích hợp linh hoạt vào các hệ thống, đơn giản hóa triển khai và nâng cao khả năng mở rộng.

5. Cải thiện mô hình qua thời gian

Hệ thống cần học hỏi từ dữ liệu mới, cập nhật mô hình liên tục để nâng cao độ chính xác và hiệu suất nhận diện.

TÀI LIỆU THAM KHẢO (Định dạng DBLP)

[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.

[6] Xie, E., & Tang, X. "Efficient Text Detection and Recognition in Natural Scenes Using YOLO and LSTM Networks," Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 1799-1808. DOI: 10.1109/ICCV.2017.00208.