

BÁO CÁO ĐỒ ÁN CUỐI KỲ

Môn học
CS519 - PHƯƠNG PHÁP LUẬN NGHIÊN CỨU
KHOA HỌC

Lớp học
CS519.O11

Giảng viên
PGS.TS. LÊ ĐÌNH DUY

Thời gian
09/2023 - 02/2024

THÔNG TIN NHÓM

Link Youtube:

<https://youtu.be/tDSVT1IK-mQ?si=FJAIXpUGU63C57IE>

Link slides: [CS519.011/Report.pptx at main · nghiatao30/CS519.011 \(github.com\)](#)

- Họ và Tên: Nguyễn Ích Tài
- MSSV: 18521362



- Lớp: CS519.O11
- Tự đánh giá (điểm tổng kết môn): 8/10
- Số buổi vắng: 5
- Số câu hỏi QT cá nhân: 2
- Số câu hỏi QT của cả nhóm: 6
- Link Github: <https://github.com/ichtai>
- Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:
 - o Lên ý tưởng nghiên cứu
 - o soạn ppt
 - o trưởng nhóm

- Họ và Tên: Nguyễn Quế Phong
- MSSV: 21520398



- Lớp: CS519.O11
- Tự đánh giá (điểm tổng kết môn): 7.5/10
- Số buổi vắng: 4
- Số câu hỏi QT cá nhân: 2
- Số câu hỏi QT của cả nhóm: 6
- Link Github: <https://github.com/NgQPhong/CS519>
- Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:
 - o Đóng góp ý tưởng
 - o Viết phần Poster
 - o Làm video YouTube

- Họ và Tên: Đường Minh Khang
- MSSV: 21520957



- Lớp: CS519.O11
- Tự đánh giá (điểm tổng kết môn): 8/10
- Số buổi vắng: 4
- Số câu hỏi QT cá nhân: 2
- Số câu hỏi QT của cả nhóm: 6
- Link Github: [nghiatao30 \(Đường Minh Khang\) \(github.com\)](https://github.com/nghiatao30)
- Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:
 - o Đóng góp ý tưởng
 - o Làm doc báo cáo
 - o Quay video YouTube

TÊN ĐỀ TÀI: NHẬN DẠNG CHỮ TIẾNG VIỆT Ở NGOẠI CẢNH

TÊN ĐỀ TÀI TIẾNG ANH: RECOGNITION OF VIETNAMESE TEXT IN OUTDOOR SCENES

Tóm Tắt:

Nhận dạng chữ viết thông qua hình ảnh là một bài toán khá phổ biến trong lĩnh vực máy học và học sâu không chỉ bởi độ đa dạng về ngôn ngữ trên toàn thế giới mà còn bởi sự thiết yếu của nó được phục vụ trong đời sống con người hàng ngày. Ở đây, mục tiêu đề tài chúng tôi chính là cho ra được một giải pháp có thể áp dụng để nhận dạng chữ cái Tiếng Việt trong những vấn đề thực tế khi mà bức ảnh hoặc dữ liệu đầu vào chứa nhiều nội dung gây nhiễu và không liên quan gây khó khăn và đặt ra thách thức cho việc xác định được thông tin chữ viết có trong đó. Tổng quan quá trình thực hiện gồm 2 giai đoạn. Thứ nhất, đề tài sẽ huấn luyện và sử dụng mô hình YOLOv5 qua tập dữ liệu Vintext và dữ liệu tự thu thập để tiến hành tìm và xác định vị trí của những vùng có chứa văn bản. Thứ hai đó chính là sử dụng mô hình VietOCR, là một mô hình được kết hợp giữa CNN và Transformer với tính tổng quát cực tốt và độ chính xác tương đối cao trong việc nhận dạng chữ viết tay, chữ đánh máy cho Tiếng Việt, để đưa vào những bức ảnh được cắt xén từ những vị trí có vùng chứa văn bản ở bước một và cho ra dự đoán nội dung tương ứng. Bên cạnh đó, chúng tôi có thử nghiệm với phương pháp kết hợp giữa Edge Detection và VietOCR để có thể dễ dàng so sánh kết quả với phương pháp chính cũng như nhìn nhận ra mặt hạn chế và xác định được hướng cải thiện và phát triển cho đề tài.

Giới Thiệu:

Trước sự bùng nổ của dữ liệu hình ảnh trên Internet và sự phổ biến của các thiết bị di động với khả năng chụp ảnh ngày càng cao, nhu cầu tự động hóa việc nhận diện và phân tích chữ viết trên ảnh đã trở nên ngày càng quan trọng. Bối cảnh nghiên cứu này đặt ra những thách thức đáng kể khi cần xác định và hiểu nội dung văn bản trong các hình ảnh, đặc biệt là khi chúng chứa những yếu tố như góc chụp, ánh sáng không đồng đều, hoặc background quá phức tạp. Vấn đề chính của đề tài là phát triển các phương pháp và thuật toán hiệu quả để nhận diện chữ viết trong các điều kiện góc ngách và đa dạng này. Việc thực hiện nghiên cứu này là quan trọng vì nó mang lại nhiều ứng dụng thực tế như: Tự động nhận diện biển số xe trong các hình ảnh giao thông đến việc quét và chuyển đổi thông tin từ các tài liệu giấy, khả năng phát hiện chữ viết thông qua ảnh mở ra một loạt các ứng dụng trong cuộc sống hàng ngày. Đồng thời, nó cũng là một bước quan trọng trong việc phát triển công nghệ hỗ trợ người khuyết tật và giúp cải thiện hiệu suất của các hệ thống xử lý ảnh tự động. Hơn hết, đây là đề tài có thể được ứng dụng, cải thiện và phát triển cho người Việt Nam chúng ta.

Trong đề tài lần này, từ một bức ảnh là dữ liệu đầu vào chúng tôi đề xuất một giải pháp bao gồm sự kết hợp từ 2 mô hình YOLOV5 và VietOCR. Trong đó, mô hình YOLOV5 với độ chính xác thực nghiệm và tốc độ xử lý đáng kinh ngạc của nó chính là một trong những sự lựa chọn tốt nhất cho bài toán nhận diện đối tượng và được sử dụng ở đây chính là xác định các vùng ảnh có chứa ký tự từ ảnh đầu vào. Bên cạnh đó, mô hình nhận dạng chữ cái tiếng Việt VietOCR cũng được sử dụng bởi độ thân thiện và dễ cài đặt đi kèm với độ chính xác tin cậy sẽ cung cấp cho ta một giải pháp tổng quan cho toàn cảnh bài toán với mục tiêu áp dụng được vào nhiều vấn đề thực tiễn, dễ sử dụng và kết quả sai số không quá cao.

Input: bức ảnh

Output: nội dung dự đoán tương ứng từ hình ảnh.

Mục tiêu:

- Xây dựng một mô hình tổng quan, dễ cài đặt với độ chính xác cao.
- Áp dụng được mô hình với nhiều vấn đề thực tiễn.
- Nghiên cứu, thử nghiệm và đánh giá mô hình để đưa ra định hướng phát triển và hoàn thiện các vấn đề hạn chế.

Nội dung và phương pháp:

• Nội dung:

- o Thử nghiệm phương pháp nhận dạng vùng ảnh có chứa chữ viết với Edge Detection
- o Tiến hành thử nghiệm phương pháp nhận dạng vùng ảnh có chứa chữ viết với YOLOv5
- o Huấn luyện mô hình VietOCR với tập dữ liệu Vintext bản Original với định dạng vị trí 4 đỉnh và nhãn.
- o Đánh giá kết quả sau khi đưa từng phương pháp qua mô hình VietOCR đã được huấn luyện.
- o So sánh kết quả và đúc kết ra mặt hạn chế cần cải thiện.

• Phương Pháp:

1. Edge Detection

- Từ ảnh input đầu vào, sử dụng thuật toán Sobel Edge Detection và các hàm. Có sẵn trong thư viện OpenCV để tìm vị trí chứa văn bản.
- Cắt ảnh các vị trí đã tìm được.
- Sử dụng phép Morphology Closing để lấp những khoảng trống với kích thước kernel là 5x5.
- Dùng hàm contour để tìm các đường bao quanh các biên cạnh tìm thấy
- được sau khi sử dụng Closing.

- Cắt các vùng được tìm thấy bởi contour và đưa ảnh đã cắt vào mô hình VietOCR để dự đoán

2. YOLOv5

- Từ ảnh input đầu vào, đưa ảnh qua mô hình YOLOv5 để tìm xác định vị trí của vùng chứa văn bản.
- Tiến hành cắt ảnh từ các vị trí tìm thấy qua mô hình YOLOv5.

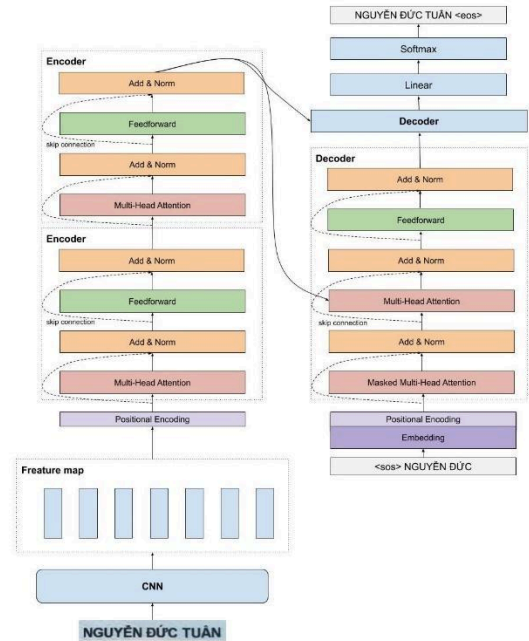


- Bộ dữ liệu sử dụng là bản Converted dataset Vintext với định dạng đầu vào là COCO format.
- Để tiến hành huấn luyện với mô hình YOLOv5 với định dạng đầu vào là hình chữ nhật, nhóm đã sử dụng một công cụ hỗ trợ cho việc chuyển đổi dữ liệu từ định dạng COCO sang định dạng của YOLOv5 cho việc huấn luyện mô hình là Roboflow.
- Tiến hành huấn luyện với mô hình YOLOv5 với epochs dự đoán khoảng 100.



3. VietOCR

- Để cho ra được kết quả cao, nhóm tiến hành pre-trained VietOCR trên bộ dữ liệu Vintext bản Original với định dạng vị trí 4 đỉnh và nhãn. Tiến hành cắt ảnh từ các vị trí tìm thấy qua mô hình YOLOv5.
- Tiến hành cắt ảnh với vị trí 4 đỉnh tương ứng với nhãn của chúng.
- Đưa dữ liệu đã được xử lý và huấn luyện trên mô hình pre-trained VGG19 – Transformer (vgg_transformer) với 5000 iters.



Kế hoạch thực hiện:

Tuần 1 tới tuần 3: Tìm hiểu YOLOV5, thuật toán Sobel Edge Detection, mô hình VietOCR và các mô hình phù hợp khác.

Kết quả dự kiến:

- Tập tài liệu chi tiết về YOLOV5, Sobel Edge Detection và VietOCR

Tuần 2 tới tuần 6: Thu thập dữ liệu có sẵn Vintext và tập dữ liệu tự thu thập thêm từ các biển báo, các cửa hiệu, những bức ảnh có bao hàm văn bản tiếng việt phạm vi ước tính xung quanh Thủ Đức.

Kết quả dự kiến:

- Tập dữ liệu có sẵn Vintext gồm khoảng 2000 hình ảnh
- Tập dữ liệu tự thu thập gồm khoảng thêm 1000 hình ảnh từ các góc chụp, điều kiện và phong nền khác nhau.

Tuần 6 tới tuần 12: Xây dựng và huấn luyện mô hình YOLOV5, Sobel Edge Detection và VietOCR. Ghi nhận kết quả vào bảng và so sánh 2 phương pháp YOLOV5 và Sobel Edge Detection khi kết hợp với VietOCR. Tìm ra mặt hạn chế và cần cải thiện, đồng thời nghiên cứu thêm các phương pháp khác nếu nó thích hợp hơn với mục tiêu đề án đề ra.

Tuần 12 tới tuần 17: Xây dựng chương trình demo đơn giản cho đề tài.

Kết quả mong đợi:

- Dự kiến phương pháp YOLOV5 + VietORC sẽ mang lại kết quả khả quan hơn trên tập dữ liệu được sử dụng.
- Tạo ra một mô hình nhận diện chữ cái tiếng việt có thể sử dụng được trên nhiều bức ảnh thực tế với độ sai số không đáng kể.
- Cải thiện được những mặt hạn chế, nghiên cứu thêm nhiều phương pháp khác nhằm vận dụng và hoàn thiện mô hình nhưng vẫn giữ được mục tiêu đặt ra ban đầu.

Tài liệu tham khảo:

- [1] Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., Kwon, Y., Michael, K., ... & Jain, M. (2022). ultralytics/yolov5: v7. 0-yolov5 sota realtime instance segmentation. *Zenodo*.
- [2] Jocher, G., Stoken, A., Borovec, J., Chaurasia, A., Changyu, L., Hogan, A., ... & Ingham, F. (2021). ultralytics/yolov5: v5. 0-YOLOv5-P6 1280 models, AWS, Supervise. ly and YouTube integrations. *Zenodo*.
- [3] Huu, Phat Nguyen, Thanh Tran Ngoc, and Quang Tran Minh. "Proposing Vietnamese Text Recognition Algorithm Combining CRAFT and VietOCR." *2022 11th International Conference on Control, Automation and Information Sciences (ICCAIS)*. IEEE, 2022.
- [4] *Quoc Pham Viet - Nhận Dạng Tiếng Việt Sử Dụng Mô Hình Transformer và AttentionOCR— Data Scientist at Overspace (pbcquoc.github.io)*
- [5] https://docs.opencv.org/3.4/d2/d2c/tutorial_sobel_derivatives.html
- [6] <https://viblo.asia/p/part1-edge-detection-voi-opencv-L4x5xLVB5BM>