

NÂNG CAO NHẬN THỨC VÙNG KÝ TỰ ĐỂ PHÁT HIỆN VĂN BẢN

Lê Minh Hoàng - 21522089

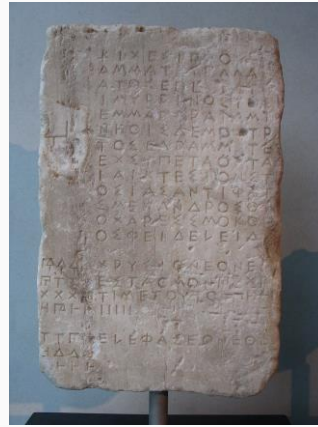
Tóm tắt

- Lớp: CS519.011
- Link Github của nhóm: <https://github.com/leminhhoang1008/CS519.011>
- Link YouTube video: <https://youtu.be/0t625H1Tf0I>



Lê Minh Hoàng

Giới thiệu



How?

to make it faster when running

to make the result better

to make it easier to use



Mục tiêu

- *Nâng cao độ chính xác*: Phát triển mô hình để nhận biết vùng chứa ký tự với độ chính xác cao, đặc biệt là đối với các ký tự có hình dạng và kích thước đa dạng.
- *Tăng cường hiểu biết vị trí*: Cải thiện khả năng của mô hình nhận thức về vị trí cụ thể của ký tự trong ảnh, bao gồm cả các trường hợp ký tự gần nhau, chồng lấn hoặc ký tự trên nền ảnh phức tạp.
- *Tối ưu hóa tốc độ xử lý*: Cải thiện hiệu suất của hệ thống phát hiện văn bản để đảm bảo khả năng xử lý nhanh chóng và hiệu quả, đặc biệt là đối với các hình ảnh có kích thước lớn và nền phức tạp.

Nội dung và Phương pháp

Nội dung

- Thực hiện nghiên cứu các phương pháp SSTD[1], CLRS[2], CRAFT[3] để tìm hiểu về cấu trúc, thuật toán của từng phương pháp
- Rút ra các kết luận về ưu, nhược điểm của từng phương pháp, từ đó tổng hợp các ưu điểm để xây dựng một mô hình mới
- Thực hiện huấn luyện trên bộ dữ liệu SynthText[4] và đánh giá trên các bộ dữ liệu phục vụ cho bài toán “Scene Text Detection”, rút ra kết luận và điều chỉnh

Nội dung và Phương pháp

Phương pháp:

- Hiểu rõ cơ sở lý thuyết, kiến trúc mô hình, và ưu điểm, nhược điểm của các phương pháp SSTD[1], CLRS[2], CRAFT[3]
- Dựa trên ưu điểm của mỗi phương pháp, thực hiện xây dựng lại-tối ưu hóa model CRAFT[3] ở các giai đoạn:
 - Ground Truth Label Generation
 - Weakly-Supervised Learning
 - Inference
- Tìm hiểu về các độ đo liên quan đến các phương pháp như F-measure, H-mean,... được sử dụng, lựa chọn phương pháp đánh giá chung tối ưu nhất cho bài toán
- Tiến hành tái huấn luyện trên tập dữ liệu SynthText[4] để thực hiện so sánh với kết quả ban đầu và rút ra các kết luận củng cố, hỗ trợ cho việc tiếp tục tối ưu hóa mô hình khi đạt kết quả mong muốn
- Tạo bộ dữ liệu ngẫu nhiên khoảng 10000 hình ảnh là hình ảnh có chứa kí tự, thực hiện đánh giá mô hình trên các độ đo Precision, Recall, F-measure
- Xây dựng ứng dụng web giúp auto labeling vùng trong hình ảnh có chứa kí tự để phục vụ cho các bài toán khác

Kết quả dự kiến

- Báo cáo các phương pháp tối ưu mà chúng tôi đã làm được đối với đề tài
- Xây dựng ứng dụng có khả năng tự động phát hiện các vùng kí tự, khoanh vùng bằng bounding box, cho phép tải xuống hình ảnh output để có thể sử dụng làm dữ liệu cho các bài toán khác

Tài liệu tham khảo

- [1] Pan He, Weilin Huang, Tong He, Qile Zhu, Yu Qiao, Xiaolin Li: Single Shot Text Detector with Regional Attention. ICCV, 2017
- [2] Pengyuan Lyu, Cong Yao, Wenhao Wu, Shuicheng Yan, Xiang Bai: Multi-Oriented Scene Text Detection via Corner Localization and Region Segmentation. CVPR, 2018
- [3] Youngmin Baek, Bado Lee, Dongyoon Han, Sangdoo Yun, Hwalsuk Lee: Character Region Awareness for Text Detection. CVPR, 2019
- [4] Ankush Gupta, Andrea Vedaldi, and Andrew Zisserman: Synthetic Data for Text Localisation in Natural Images. Dept. of Engineering Science, University of Oxford.