**CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM**

**Độc Lập - Tự Do - Hạnh Phúc**

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH

**TRƯỜNG ĐẠI HỌC**

**CÔNG NGHỆ THÔNG TIN**

ĐỀ CƯƠNG NGHIÊN CỨU

|  |
| --- |
| **TÊN ĐỀ TÀI: TỐI ƯU HÓA NHẬN THỨC VÙNG KÍ TỰ ĐỂ PHÁT HIỆN VĂN BẢN** |
| **TÊN TIẾNG ANH: OPTIMIZATION OF CHARACTER REGION AWARENESS FOR TEXT DETECTION** |
| **Thời gian thực hiện**: Từ ngày .../2024 đến ngày .../2024. |
| A person in a yellow shirt  Description automatically generated**Sinh viên thực hiện**:  Họ vè tên: Lê Minh Hoàng  MSSV: 21522089  Khoa: Khoa học Máy tính  Lớp: CS519.O11  Github: <https://github.com/leminhhoang1008>  Email: [21522089@gm.uit.edu.vn](mailto:21522089@gm.uit.edu.vn)  Sđt: 0981764872 |
| **Tóm tắt**:  Phương pháp phát hiện văn bản trong ảnh dựa trên mạng nơ-ron đã xuất hiện gần đây và cho kết quả hứa hẹn. Các phương pháp trước đây được huấn luyện với các hộp giới hạn ở cấp độ từ có giới hạn trong việc biểu diễn vùng văn bản có hình dạng bất kỳ. Tuy nhiên các phương pháp trước đó thường có độ chính xác không cao hoặc không thể phát hiện văn bản trên các hình ảnh có nền phức tạp, đi kèm theo tốc độ xử lý tương đối chậm. Đề tài tập trung vào việc nâng cao khả năng nhận biết vùng ký tự trong các hệ thống phát hiện văn bản. Ta có thể tập trung vào các phương pháp, thuật toán hoặc mô hình mới để cải thiện khả năng nhận diện và phân loại vùng chứa ký tự trong hình ảnh. Các kỹ thuật mới này có thể bao gồm sự tích hợp thông tin vùng ký tự để cải thiện độ chính xác và độ nhạy cảm của hệ thống phát hiện văn bản. Điều này có thể giúp ứng dụng thị giác máy tính và các hệ thống OCR (quét và nhận dạng ký tự) hiểu biết và xử lý tốt hơn các vùng chứa văn bản trong ảnh. |
| **Giới thiệu**:  Bài toán "Character Region Detection" là một lĩnh vực quan trọng trong thị giác máy tính và xử lý hình ảnh, nơi mà mục tiêu chính là phát hiện và định vị vùng chứa văn bản trong các hình ảnh. Đặc biệt, bài toán này đặt ra thách thức trong việc cải thiện khả năng nhận biết và hiểu biết về vị trí cụ thể của các ký tự trong một hình ảnh.  Trong ngữ cảnh của bài toán này, "Character Region Awareness" đề cập đến khả năng của mô hình hoặc thuật toán nhận thức và hiểu biết về vùng chứa ký tự trong hình ảnh một cách chính xác và đầy đủ. Thách thức chính là làm sao để mô hình có thể tự động nhận biết và chú ý đến các vùng chứa ký tự, bao gồm cả các kí tự có hình dạng, kích thước và vị trí đa dạng.  Trong đề tài này, tôi thực hiện nghiên cứu các phương pháp đã được công bố về “Phát hiện vùng kí tự” để tìm hiểu các thuật toán, mô hình,… từ đó tìm ra ưu/nhược điểm của các phương pháp, tổng hợp lại và tạo ra một phương pháp tối ưu hơn cho việc phát hiện vùng kí tự trong hình ảnh nhằm hướng tới việc tối ưu hóa các mô hình mạng nơ-ron, thuật toán phân loại, và phương pháp xử lý hình ảnh để cải thiện độ chính xác và hiệu suất của hệ thống phát hiện văn bản. Cụ thể:  **Input**: Hình ảnh bất kì có chứ các kí tự với đa dạng ngôn ngữ  **Output**: Các vùng chứa kí tự được thể hiện trên hình ảnh Top of Form |
| **Mục tiêu**:  Tìm hiểu các phương pháp gồm SSTD[1], CLRS[2], CRAFT[3] nhằm tìm ra các ưu điểm của từng phương pháp, từ đó tạo ra một thuật toán mới được:   1. *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. 2. *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. 3. *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**   1. **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 ưa 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  1. **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] * Tìm hiểu các phương pháp tối ưu mạng nơ-ron học sâu như LAMB (Layer-wise Adaptive Moments) hoặc RAdam (Rectified Adam). * Dựa trên ưu điểm của mỗi phương pháp, thực hiện xây dựng lại-tối ưu hóa model CRAFT[3] * 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 chứ 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 mong đợi**   * Báo cáo các phương pháp tối ưu mà chúng tôi đã làm được đối với đề tài * Ứng dựng 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 lafddm 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. |
| **Kế hoạch thực hiện:**  + Tuần 1-4: Tìm hiểu các phương pháp SSTD, CLRS, CRAFT  Kết quả dự tính:   * Hiểu cấu trúc, thuật toán của các phương pháp * Tài liệu độ đo F-measure * Tài liệu chi tiết về các đặc điểm và ưu/nhược điểm của 3 phương pháp   + Tuần 4-12: Xây dựng mô hình với các ưu điểm của 3 phương pháp dựa trên CRAFT và đánh giá mô hình  Kết quả dự tính:   * Bảng kết quả đánh giá của mô hình mới trên bộ dữ liệu tự thu thập   + Tuần 10-16: Xây dựng ứng dụng demo  Kết quả dự tính:   * Ứng dựng 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 lafddm dữ liệu cho các bài toán khác |