HỢP NHẤT PHÁT HIỆN VÀ PHÂN TÍCH BỐ CỤC VĂN BẢN TRÊN BỘ DỮ LIỆU BẢNG HIỆU TIẾNG VIỆT PHÂN CẤP

GVHD: PGS. TS Lê Đình Duy

22520004 - Trần Như Cẩm Nguyên
22520361 - Trần Thị Cẩm Giang

Tóm tắt

- Lớp: CS519.O21.KHTN
- Link Github: https://github.com/Yangchann/CS519.O21.KHTN
- Link YouTube video: https://youtu.be/C8iep3TsBYY

Trần Như Cẩm Nguyên
22520004

Trần Thị Cẩm Giang
22520361

Giới thiệu

- Phát hiện văn bản trong ảnh cảnh (scene text detection) và phân tích bố cục tài liệu (document layout analysis) từ lâu được coi là hai nhiệm vụ riêng biệt trong thị giác máy tính.
  - **Scene text detection** là quá trình nhận diện và định vị văn bản xuất hiện trong các hình ảnh chụp từ môi trường thực tế [1].
  - **Document layout analysis** là quá trình xác định và phân loại các thành phần khác nhau của một tài liệu, chẳng hạn như tiêu đề, đoạn văn, hình ảnh, bảng biểu và chú thích [2].

➢ Hợp nhất phát hiện và phân tích bố cục văn bản trong hình ảnh đã tạo ra một nhánh nghiên cứu quan trọng và thu hút nhiều sự quan tâm của giới nghiên cứu.

Giới thiệu

- **Tiếng Việt** là một ngôn ngữ phức tạp với hệ thống dấu câu phong phú, tạo ra thách thức lớn trong việc nhận diện chính xác các ký tự.
- **Bảng hiệu (signboard)** là một dạng thông tin có cấu trúc đặc biệt, chứa đựng nhiều thông tin quan trọng, hữu ích cho các ứng dụng thực tế như:
  - Thu thập và phân tích dữ liệu trong bảng hiệu để nghiên cứu thị trường.
  - Tự động cập nhật tên đường, cửa hàng và các địa điểm quan trọng khác lên bản đồ số, ...

➢ Phát hiện và phân tích bố cục văn bản trên bảng hiệu tiếng Việt có nhiều thách thức và ứng dụng trong thực tế.

Giới thiệu

- **INPUT:** Ảnh chứa một bảng hiệu.
- **OUTPUT:** Danh sách các đoạn văn bản phát hiện được trên bảng hiệu. Mỗi đoạn có nhãn để phân loại các thành phần trên bảng hiệu (tên cửa hàng, địa chỉ…) và danh sách các dòng văn bản. Mỗi dòng gồm danh sách các từ với tọa độ đỉnh bounding box và nội dung từ.

Hình 1. Minh họa về kết quả của bài toán phát hiện và phân tích bố cục văn bản trên bảng hiệu tiếng Việt

Mục tiêu

- Xây dựng bộ dữ liệu bảng hiệu tiếng Việt phân cấp - **Hierarchical Vietnamese Signboard**, bộ dữ liệu được chú thích theo cấu trúc phân cấp (hierarchical annotations) để phục vụ cho quá trình huấn luyện và đánh giá.
- **Nghiên cứu hợp nhất hai nhiệm vụ** Scene Text Detection và Document Layout Analysis và đề xuất phương pháp mới.
- **Đánh giá** hiệu xuất của một số phương pháp SOTA và phương pháp do nhóm đề xuất trên bộ dữ liệu Hierarchical Vietnamese Signboard.

Nội dung và Phương pháp

- **Nội dung 1: Tìm hiểu tổng quan đề tài**
  - Phương pháp: tìm hiểu tổng quan các phương pháp để giải quyết từng nhiệm vụ scene text detection và document layout analysis.
- **Nội dung 2: Nghiên cứu các phương pháp hợp nhất hai nhiệm vụ phát hiện và phân tích bố cục văn bản xuất hiện trên ảnh (scene text)**
  - Phương pháp: tìm hiểu các phương pháp SOTA hiện có như Unified Detector, Upstage KR,. . . từ các công trình đã được công bố trên các top conference.
- **Nội dung 3: Nghiên cứu và đề xuất phương pháp mới**
  - Phương pháp:
    - Nghiên cứu kiến trúc của Transformer và Interactive Attention.
    - Xây dựng model để giải quyết bài toán.

Nội dung và Phương pháp

- **Nội dung 4: Xây dựng bộ dữ liệu Hierarchical Vietnamese Signboard**
  - Phương pháp:
    - Tạo guideline hướng dẫn để thu thập dữ liệu và lên kế hoạch label.
    - Sử dụng điện thoại, máy ảnh để chụp bảng hiệu của các cửa hàng. Sau đó chọn lọc những bảng hiệu đạt yêu cầu và bắt đầu xử lý, gán nhãn.

- **Nội dung 5: Chạy thực nghiệm và đánh giá trên bộ dữ liệu Hierarchical Vietnamese Signboard**
  - Phương pháp:
    - Tiến hành chạy thực nghiệm các phương pháp hiện có và phương pháp do chúng tôi đề xuất trên bộ dữ liệu Hierarchical Vietnamese Signboard.
    - Thống kê và đánh giá các phương pháp.

Kết quả dự kiến

- Bộ dữ liệu **Hierarchical Vietnamese Signboard**.
- **Phương pháp mới** để hợp nhất hai nhiệm vụ phát hiện và phân tích bố cục văn bản xuất hiện trong ảnh (scene text), cụ thể là trên bảng hiệu tiếng Việt.
- **Kết quả đánh giá** giữa phương pháp đề xuất và các phương pháp SOTA hiện có.

Tài liệu tham khảo

[1]. Shangbang Long, Xin He, Cong Yao: Scene Text Detection and Recognition: The Deep Learning Era. Int. J. Comput. Vis. 129(1): 161-184 (2021)

[2]. Jilin Wang, Michael Krumdick, Baojia Tong, Hamima Halim, Maxim Sokolov, Vadym Barda, Delphine Vendryes, Chris Tanner: A Graphical Approach to Document Layout Analysis. ICDAR (5) 2023: 53-69

[3]. Shangbang Long, Siyang Qin, Dmitry Panteleev, Alessandro Bissacco, Yasuhisa Fujii, Michalis Raptis: Towards End-to-End Unified Scene Text Detection and Layout Analysis. CVPR 2022: 1039-1049

[4]. Maoyuan Ye, Jing Zhang, Juhua Liu, Chenyu Liu, Baocai Yin, Cong Liu, Bo Du, Dacheng Tao: Hi-SAM: Marrying Segment Anything Model for Hierarchical Text Segmentation. CoRR abs/2401.17904 (2024)

[5]. Shangbang Long, Siyang Qin, Yasuhisa Fujii, Alessandro Bissacco, Michalis Raptis: Hierarchical Text Spotter for Joint Text Spotting and Layout Analysis. WACV 2024: 892-902

[6]. Shangbang Long, Siyang Qin, Dmitry Panteleev, Alessandro Bissacco, Yasuhisa Fujii, Michalis Raptis: ICDAR 2023 Competition on Hierarchical Text Detection and Recognition. ICDAR (2) 2023: 483-497

[7]. Xingyu Wan, Chengquan Zhang, Pengyuan Lyu, Sen Fan, Zihan Ni, Kun Yao, Errui Ding, Jingdong Wang: Towards Unified Multi-granularity Text Detection with Interactive Attention. CoRR abs/2405.19765 (2024)