THÔNG TIN CHUNG CỦA NHÓM JQK

* Link YouTube video của báo cáo (tối đa 5 phút):   
  *...*
* Link slides (dạng .pdf đặt trên Github của nhóm):   
  *…*
* *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*

|  |  |
| --- | --- |
| * Họ và Tên: Ngô Thành Phát * MSSV: 19521994 | * Lớp: CS519.N11 * Tự đánh giá (điểm tổng kết môn): 9.5/10 * Số buổi vắng: 1 * Số câu hỏi QT cá nhân: 8 * Link Github:https://github.com/Pakm19 |

**ĐỀ CƯƠNG NGHIÊN CỨU**

|  |
| --- |
| **TÊN ĐỀ TÀI  (IN HOA)**  TĂNG CƯỜNG ĐỘ CHÍNH XÁC CỦA NHẬN DẠNG VĂN BẢN CẢNH BẰNG PHƯƠNG PHÁP HƯỚNG DẪN TỪ ĐIỂN |
| **TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)**  INCREASE THE ACCURACY OF SCENE TEXT RECOGNITION BY DICTIONARY-GUIDE METHOD |
| **TÓM TẮT *(Tối đa 400 từ)***  Bài toán Scene Text Recognition (Nhận dạng văn bản trên cảnh) là một trong những bài toán quan trọng trong lĩnh vực thị giác máy tính và xử lý ngôn ngữ tự nhiên, có ứng dụng rộng rãi trong nhiều lĩnh vực khác nhau. Bài toán này đặt ra thách thức trong việc nhận dạng và trích xuất văn bản từ các hình ảnh chứa văn bản được chụp từ môi trường thực tế. Các thách thức đối với bài toán này bao gồm độ phức tạp của các phông chữ, kích thước và hình dạng của các ký tự, độ tương phản và sự nhiễu của hình ảnh. Để giải quyết bài toán này, các phương pháp xử lý ảnh và học máy được sử dụng để phát hiện và trích xuất các đặc trưng của văn bản từ hình ảnh. Các thuật toán nhận dạng ký tự được áp dụng để chuyển các đặc trưng này thành văn bản có ý nghĩa. Các ứng dụng của bài toán này là rất đa dạng, từ nhận dạng biển số xe, địa chỉ, tên nhãn hiệu, đến nhận dạng văn bản trên các tài liệu cổ, hộ chiếu và giấy tờ tùy thân. Đầu tiên chúng tôi tiến hành nghiên cứu và phân tích các hướng tiếp cận truyền thống của bài toán Scene Text Recognition và từ đó đề xuất phương pháp mới kết hợp từ điển trong cả giai đoạn đào tạo và suy luận của hệ thống nhận dạng văn bản cảnh. Chúng tôi sử dụng từ điển để tạo danh sách các kết quả có thể xảy ra và tìm ra kết quả tương thích nhất với hình thức trực quan của văn bản. Bên cạnh đó, chúng tôi đóng góp Vintext, một bộ dữ liệu văn bản cảnh tiếng Việt. Bộ dữ liệu này sẽ là một tiêu chuẩn đầy thách thức để đo lường khả năng ứng dụng và độ mạnh các thuật toán Scene Text Recognition |
| **GIỚI THIỆU**  Phát hiện và nhận dạng văn bản cảnh là một vấn đề nghiên cứu quan trọng với nhiều ứng dụng, từ lập bản đồ và bản địa hóa đến điều hướng rô-bốt và nâng cao khả năng tiếp cận cho người khiếm thị. Tuy nhiên, nhiều trường hợp văn bản trong tự nhiên vốn đã mơ hồ do phong cách nghệ thuật, sự xuống cấp của thời tiết hoặc điều kiện ánh sáng bất lợi. Trong nhiều trường hợp, sự mơ hồ không thể được giải quyết mà không cần suy luận về ngôn ngữ của văn bản  Trên thực tế, một cách tiếp cận phổ biến để cải thiện hiệu suất của hệ thống nhận dạng văn bản cảnh là sử dụng từ điển và truyền đầu ra dự đoán dưới dạng một từ trong từ điển. Quy trình thông thường để xử lý hình ảnh đầu vào bao gồm: (1) phát hiện các thể hiện văn bản, (2) đối với mỗi thể hiện văn bản được phát hiện, tạo chuỗi ký tự có thể xảy ra nhất, dựa trên hình thức cục bộ của thể hiện văn bản mà không có mô hình ngôn ngữ, và (3) tìm từ trong từ điển có khoảng cách chỉnh sửa nhỏ nhất (còn gọi là khoảng cách Levenshtein [1]) đối với chuỗi ký tự được tạo và sử dụng từ này làm đầu ra nhận dạng cuối cùng.  Tuy nhiên, cách tiếp cận trên có ba vấn đề lớn. Đầu tiên, nhiều trường hợp văn bản là các từ nước ngoài hoặc từ tự tạo không có trong từ điển nên việc buộc đầu ra là một từ trong từ điển sẽ dẫn đến kết quả sai trong nhiều trường hợp. Thứ hai, không có vòng phản hồi trong quy trình xử lý chuyển tiếp nguồn cấp dữ liệu ở trên; ngôn ngữ trước đó không được sử dụng trong bước thứ hai để ghi điểm và tạo chuỗi ký tự có thể xảy ra nhất. Thứ ba, bản thân việc chỉnh sửa khoảng cách là không xác định và không hiệu quả trong nhiều trường hợp. Không rõ đầu ra là gì khi nhiều từ trong từ điển có cùng khoảng cách chỉnh sửa với chuỗi ký tự đầu ra trung gian. Hơn nữa, nhiều ngôn ngữ có các ký hiệu đặc biệt có vai trò khác với các ký tự chính của bảng chữ cái nên việc xử lý thống nhất các ký hiệu và ký tự trong khoảng cách soạn thảo là không phù hợp.  Theo kinh nghiệm, chúng tôi đánh giá phương pháp của mình trên một số bộ dữ liệu chuẩn bao gồm TotalText [4], ICDAR2013 [5] và ICDAR2015 [6] và nhận thấy rằng phương pháp sử dụng từ điển của chúng tôi mang lại lợi ích trong cả giai đoạn đào tạo và suy luận. Chúng tôi cũng chứng minh những lợi ích của phương pháp nhận dạng văn bản không phải tiếng Anh của chúng tôi. Đặc biệt, chúng tôi cho thấy cách tiếp cận của chúng tôi hoạt động tốt đối với tiếng Việt, một ngôn ngữ Nam Á dựa trên bảng chữ cái Latin có thêm các ký hiệu trọng âm (´, `, ?, . , ˜) và các ký tự phái sinh (ô, ê, â, a, ˘ ơ, ư). Là ngôn ngữ mẹ đẻ của 90 triệu người Việt Nam và 4,5 triệu người Việt di cư trên khắp thế giới, văn bản tiếng Việt xuất hiện trong nhiều hoạt cảnh nên việc phát hiện và nhận diện văn bản thực cảnh tiếng Việt là một bài toán quan trọng. Chữ viết tiếng Việt cũng tương tự như các chữ viết khác như tiếng Bồ Đào Nha, vì vậy một kỹ thuật học chuyển giao hiệu quả cho tiếng Việt cũng có thể áp dụng cho các ngôn ngữ khác. Để đạt được điều này, một đóng góp của bài báo của chúng tôi là giới thiệu bộ dữ liệu chú thích cho văn bản cảnh tiếng Việt, và các thử nghiệm của chúng tôi trên bộ dữ liệu này là một minh chứng có giá trị về lợi ích của phương pháp kết hợp ngôn ngữ được đề xuất. Tóm lại, những đóng góp của bài báo của chúng tôi là hai phần chính. Đầu tiên, chúng tôi đề xuất một cách tiếp cận mới để kết hợp mô hình ngôn ngữ vào nhận dạng văn bản cảnh. Thứ hai, chúng tôi giới thiệu bộ dữ liệu cho văn bản cảnh tiếng Việt với 2000 hình ảnh được chú thích đầy đủ và 56K thể hiện văn bản  + Input: một hình ảnh có chứa văn bản.    + Output: một chuỗi các từ tương ứng với các từ được phân đoạn trong hình ảnh. |
| **MỤC TIÊU *(Viết trong vòng 3 mục tiêu)***   1. Nghiên cứu các hướng tiếp cận truyền thống của bài toán Scene Text Recognition 2. Đề xuất bộ dữ liệu tiếng Việt mới cho bài toán Scene Text Recognition . 3. Đề xuất phương pháp tăng cường độ chính xác cho bài toán Scene Text Recognition |
| **NỘI DUNG VÀ PHƯƠNG PHÁP**   1. **Mục tiêu 1**   Nội Dung   * Khảo sát các cách tiếp cận tiên tiến nhất trước đây cho bài toán Scence Text Recogniton. * Phân tích và đánh giá ưu nhược điểm của các của các công trình được công bố trước đây của bài toán   Phương Pháp   * Chúng tôi tìm kiếm các công trình nghiên cứu gần đây nhất đã được công bố trên các diễn dàng về Thị giác máy tính như CVPR, ICCV, ECCV,… để tìm kiếm các bài báo, luận văn, và tài liệu khác liên quan đến bài toán này. * Chúng tôi xác định các phương pháp tiếp cận tiên tiến nhất và có kết quả tốt nhất trong bài toán này. * Chúng tôi phân tích và đánh giá ưu nhược điểm của từng phương pháp, bao gồm hiệu suất, khả năng mở rộng, thời gian đào tạo và tính ứng dụng.  1. **Mục tiêu 2**   Nội Dung   * Xây dựng bộ dữ liệu bộ dữ liệu văn bản cảnh tiếng Việt lớn nhất từ trước đến nay * Phương Pháp * Chúng tôi xây dựng một đội ngũ riêng cho việc sưu tập dữ liệu với tiêu chí sưu tập các hình ảnh đại diện cho tập hợp đa dạng của các văn bản cảnh được bắt gặp trong cuộc sống hàng ngày ở Việt Nam và có những tiêu chuẩn, quy tắc cho công việc thu thập * Chúng tôi xây dựng một đội ngũ cho việc xử lý dữ liệu thô đầu vào có thể trở thành dữ liệu phục vụ cho bài toán Scence Text Recognition * Ví dụ về dữ liệu:      1. **Mục Tiêu 3** 2. Nội Dung   Đề xuất phương pháp tăng cường độ chính xác cho bài toán Scene Text Recognition Phương Pháp   * Trong các công trình trước đây, một từ điển đã được sử dụng để đảm bảo rằng chuỗi ký tự đầu ra là một từ hợp pháp từ từ điển và nó đã cải thiện độ chính xác rất nhiều. Hơn nữa, nếu một người có thể giảm kích thước của từ điển một cách chính xác (ví dụ: chỉ xem xét các từ xuất hiện trong tập dữ liệu), thì độ chính xác sẽ tăng hơn nữa. Tất cả những điều này là bằng chứng cho tầm quan trọng của từ điển, và từ điển được sử dụng như thế nào [9]. Tuy nhiên, việc sử dụng từ điển hiện tại dựa trên khoảng cách chỉnh sửa nhỏ nhất [1] là quá cơ bản * Chúng tôi cũng sử dụng từ điển trong giai đoạn đào tạo để đào tạo mô-đun nhận dạng để nhận ra trường hợp văn bản chính xác từ danh sách các ví dụ khó * Chúng tôi sử dụng kết hợp phương pháp của chúng tôi với các mô-đun tiên tiến nhất hiện nay, cụ thể là ABCNet [7] và MaskTextSpotterV3 [8] * Dưới đây là quy trình công nhận truyền thống (a) và quy trình đề xuất (b)     Hình 1b mô tả quy trình xử lý của quá trình nhận dạng giai đoạn của phương pháp của chúng tôi. Với một thể hiện văn bản được phát hiện x (được mô tả bằng hai đường cong Bezier [10]), một bản đồ đối tượng có kích thước cố định v sẽ được tính toán (sử dụng mô-đun căn chỉnh Bezier [10]). Từ v, chúng ta thu được đầu ra nhận dạng ban đầu yˆ. Sau đó, chúng tôi sẽ biên soạn một danh sách các từ ứng viên y1, . . . , yk, là những từ trong từ điển có khoảng cách chỉnh sửa nhỏ nhất đến y. Sau đó, chúng tôi sẽ tính điểm tương thích giữa từng từ ứng cử viên yi và bản đồ đặc trưng v và xuất ra từ có điểm tương thích cao nhất |
| **KẾT QUẢ MONG ĐỢI**   * Khi kết quả thành công (kết quả thực nghiệm cao hơn các mô-đun tiên tiến nhất hiện tại, cụ thể là H-mean score) sẽ mở ra một hướng tiếp cận mới cho toán Sence Text Recogniton * Bộ dữ liệu VinText sẽ là bộ dữ liệu tiêu chuẩn cho bài toán Scene Text Recognition ở Việt Nam * Triển khai vào ứng dụng thực tế để cải thiện hiệu suất như tự động ghi lại thông tin chi tiết của container xuất nhập khẩu (mã, loại, kích cỡ,…) từ tàu vào bờ và từ bờ đến. |
| **TÀI LIỆU THAM KHẢO *(Định dạng DBLP)***  [1]. V.I.Levenshtein. Binary codes capable of correcting insertions and reversals. 1966  [2]. Chee Kheng Chng and Chee Seng Chan. Total-text: A comprehensive dataset for scene text detection and recognition. In IAPR International Conference on Document Analysis and Recognition (ICDAR), 2017.  [3]. Dimosthenis Karatzas, F. Shafait, S. Uchida, M. Iwamura, L. G. I. Bigorda, Sergi Robles Mestre, J. M. Romeu, D. F. Mota, Jon Almazán, and L. D. L. Heras. Icdar 2013 robust reading competition. 2013 12th International Conference on Document Analysis and Recognition, pages 1484–1493,  [4]. Chee Kheng Chng and Chee Seng Chan. Total-text: A comprehensive dataset for scene text detection and recognition. In IAPR International Conference on Document Analysis and Recognition (ICDAR), 2017  [5] Dimosthenis Karatzas, F. Shafait, S. Uchida, M. Iwamura, L. G. I. Bigorda, Sergi Robles Mestre, J. M. Romeu, D. F. Mota, Jon Almazán, and L. D. L. Heras. Icdar 2013 robust reading competition. 2013 12th International Conference on Document Analysis and Recognition, pages 1484–1493, 2013.  [6] Dimosthenis Karatzas, L. G. I. Bigorda, A. Nicolaou, S. Ghosh, Andrew D. Bagdanov, M. Iwamura, Jiri Matas, Lukas Neumann, V. Chandrasekhar, S. Lu, F. Shafait, S. Uchida, and Ernest Valveny. ICDAR2015 competition on robust reading. In International Conference on Document Analysis and Recognition (ICDAR), 2015.  [7]Y. Liu, Hao Chen, Chunhua Shen, Tong He, Lian-Wen Jin, and L. Wang. ABCNet: Real-time scene text spotting with adaptive bezier-curve network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020.  [8] Minghui Liao, Guan Pang, J. Huang, Tal Hassner, and X. Bai. Mask textspotter v3: Segmentation proposal network for robust scene text spotting. ArXiv, 2020  [9] X. Yang, Kaihua Tang, Hanwang Zhang, and J. Cai. Autoencoding scene graphs for image captioning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019  [10] Y. Liu, Hao Chen, Chunhua Shen, Tong He, Lian-Wen Jin, and L. Wang. ABCNet: Real-time scene text spotting with adaptive bezier-curve network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020. |