**BÁO CÁO ĐỒ ÁN CUỐI KỲ**

**Môn học**

**CS519 - PHƯƠNG PHÁP LUẬN NGHIÊN CỨU KHOA HỌC**

**Lớp học**

**CS519.M11 - CS519.M11.KHCL**

**Giảng viên**

**PGS.TS. LÊ ĐÌNH DUY  
Thời gian**

**09/2021 - 12/2021**

*----- Trang này cố tình để trống -----*

# **THÔNG TIN CHUNG CỦA NHÓM**

* Link YouTube video của báo cáo (tối đa 5 phút):   
  [*https://youtu.be/gSecIaQq-iQ*](https://youtu.be/gSecIaQq-iQ)
* Link slides (dạng .pdf đặt trên Github của nhóm):   
  <https://github.com/18520285/CS519_M11/blob/main/B%C3%A1o_C%C3%A1o_Cu%E1%BB%91i_K%E1%BB%B3_HAD.pdf>
* *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: Nguyễn Lê Hoàng Hùng * MSSV: 18520285 | * Lớp: CS519.M11 * Tự đánh giá (điểm tổng kết môn): 7.5/10 * Số buổi vắng: 1 * Số câu hỏi QT cá nhân: 12 * Số câu hỏi QT của cả nhóm: 3 * Link Github:   <https://github.com/18520285/CS519_M11>   * Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:   + Viết phần Đối tượng, Nội dung, Phương Pháp.   + Đóng góp phần Giới thiệu.   + Làm slide thuyết trình và video YouTube. |
| * Họ và Tên: Đinh Hoàng Linh Đan * MSSV: 19521309 | * Lớp: CS519.M11 * Tự đánh giá (điểm tổng kết môn): 8/10 * Số buổi vắng: 0 * Số câu hỏi QT cá nhân: 12 * Số câu hỏi QT của cả nhóm: 3 * Link Github: [18520285/CS519\_M11 (github.com)](https://github.com/18520285/CS519_M11) * Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:   + Viết phần Giới thiệu, Mục tiêu, Phạm vi   + Đóng góp vào phần Tóm tắt, Kế hoạch thực hiện.   + Làm Poster |
| * Họ và Tên: Trần Nguyễn Quỳnh Anh * MSSV: 19521217 | * Lớp: CS519.M11 * Tự đánh giá (điểm tổng kết môn): 8/10 * Số buổi vắng: 0 * Số câu hỏi QT cá nhân: 12 * Số câu hỏi QT của cả nhóm: 3 * Link Github: [18520285/CS519\_M11 (github.com)](https://github.com/18520285/CS519_M11) * Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:   + Viết phần Kết quả mong đợi, Kế hoạch thực hiện, Tài liệu tham khảo   + Đóng góp vào phần Tóm tắt   + Đóng góp ý tưởng phần Phương pháp   + Làm poster |

# **ĐỀ CƯƠNG NGHIÊN CỨU**

|  |
| --- |
| **TÊN ĐỀ TÀI (IN HOA)**  NHẬN DIỆN VĂN BẢN TRONG HÌNH ẢNH KẾT HỢP TỪ ĐIỂN |
| **TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)**  DICTIONARY - GUIDED SCENE TEXT RECOGNITION |
| **TÓM TẮT** *(Tối đa 400 từ)*  Trong đề tài này, chúng tôi giải quyết những vấn đề thuộc lĩnh vực Computer Vision (Thị giác máy tính). Thứ nhất, chúng tôi nghiên cứu phương pháp hiện có của bài toán **Scene Text Recognition** (Nhận diện văn bản trong hình ảnh), từ đó giới thiệu một cách tiếp cận mới là **Dictionary-Guided Scene Text Recognition** (Nhận diện văn bản trong hình ảnh kết hợp từ điển) - một phương pháp kết hợp từ điển trong cả quá trình training (huấn luyện) và inference (suy luận) của hệ thống nhận diện văn bản trong hình ảnh. Thứ hai, chúng tôi giới thiệu bộ dữ liệu **VinText** - bộ dữ liệu dành cho nhận diện văn bản tiếng Việt trong hình ảnh. Đây là một bộ dữ liệu có tính thách thức vì tiếng Việt rất đa dạng, đặc biệt có dấu thanh - là một thành phần đóng vai trò quan trọng về nghĩa của từ trong tiếng Việt. Cuối cùng, chúng tôi xây dựng một chương trình ứng dụng để tăng tính ứng dụng thực tế cho đề tài. |
| **GIỚI THIỆU** *(Tối đa 1 trang A4)*   * Nhận dạng văn bản trong hình ảnh là một vấn đề cần thiết và giúp ích rất nhiều trong cuộc sống của chúng ta. Bài toán nhận dạng văn bản trong hình ảnh được ứng dụng rộng rãi trong nhiều lĩnh vực như bản đồ, điều hướng Robot, tự động hóa công nghiệp, tăng khả năng tiếp cận mọi thứ cho người khiếm thị, … * Tuy nhiên vẫn có những thách thức đối với lĩnh vực này. Văn bản trong hình ảnh đa dạng, không nhất quán về phông chữ, màu sắc, tỷ lệ ; văn bản được thể hiện theo phong cách nghệ thuật; do thời tiết làm văn bản xuống cấp hay do điều kiện ánh sáng bất lợi… có thể dẫn đến lỗi và kết quả của quá trình phát hiện và nhận dạng văn bản trong hình ảnh không được chính xác. * Trong đề tài này, chúng tôi sử dụng từ điển vào quá trình training (huấn luyện) dữ liệu và inference (suy luận) thay vì chỉ sử dụng từ điển trong quá trình inference (suy luận) như những phương pháp đã có. Điều này cho phép output không bắt buộc là từ có trong từ điển, mà từ điển sẽ được sử dụng để tạo ra danh sách các kết quả có thể xảy ra phù hợp với bối cảnh của input và tìm ra kết quả phù hợp nhất đối với văn bản trong hình ảnh.   ***Input***: Những hình ảnh có chứa văn bản (Tiếng Anh và Tiếng Việt)  ***Output***: Văn bản được nhận diện từ hình ảnh  .  *Hình 1* |
| **MỤC TIÊU** *(Viết trong vòng 3 mục tiêu)*   1. Nghiên cứu phương pháp thông thường hiện có của bài toán Scene Text Recognition và cải thiện độ chính xác và tốc độ xử lý. Cụ thể chúng tôi nghiên cứu cách tiếp cận mới đó là **Dictionary-Guided Scene Text Recognition**. 2. Tạo ra 1 bộ dữ liệu tiếng Việt cho bài toán Scene Text Recognition. 3. Xây dựng chương trình ứng dụng cho cách tiếp cận **Dictionary-Guided Scene Text Recognition**. |
| **NỘI DUNG VÀ PHƯƠNG PHÁP**   1. **Nội dung:**  * Nghiên cứu quá trình giải quyết bài toán **Scene Text Recognition** thông thường. * Đề xuất phương pháp mới cho bài toán **Scene Text Recognition** bằng cách sử dụng từ điển vào giai đoạn training (huấn luyện) và inference (suy luận). * Nghiên cứu thuật toán **ABCNet** [1] và **MaskTextSpotterV3** [2] và ứng dụng vào cả phương pháp thông thường và phương pháp cải tiến ở trên. * Xây dựng bộ dữ liệu tiếng Việt - **VinText** gồm 2000 ảnh chứa 56.084 văn bản tự nhiên được chú thích đầy đủ. * Huấn luyện mô hình **ABCNet** [1] và **MaskTextSpotterV3** [2] sử dụng bộ dữ liệu **TotalText** [3], **ICDAR2013** [4], **ICDAR2015** [5]và **VinText** bằng phương pháp thông thường và phương pháp mới được cải tiến. Sau đó so sánh và đánh giá kết quả giữa các mô hình. * Xây dựng chương trình ứng dụng cho cách tiếp cận cải tiến.  1. **Phương pháp:**  * Hệ thống nhận diện văn bản trong hình ảnh của chúng tôi gồm 2 giai đoạn: *detection* (phát hiện) và *recognition* (nhận diện). Chúng tôi tập trung chính vào phần *recognition* (nhận diện). * Ở giai đoạn *detection* (phát hiện), chúng tôi sử dụng **Bezier detection** và alignment module từ **ABCNet** [1].   *Hình 2*   * Hình ***2b*** mô tả các bước xử lý trong giai đoạn *recognition* (nhận diện) của phương pháp chúng tôi đề xuất. Từ văn bản ***x*** được giới hạn bởi 2 đường cong Bezier [1], một feature map (mạng đặc trưng) ***v*** được tính toán. Từ ***v***, chúng tôi có được output ban đầu, sau đó tạo ra 1 danh sách các từ . . . có edit distance nhỏ nhất (Levenshtein distance[6]) so với ***y***. Sau cùng, chúng tôi tính độ tương thích giữa mỗi từ với feature map ***v***, output cuối cùng sẽ là từ có độ tương thích cao nhất. * Tạo một bộ dữ liệu tiếng Việt mới - **VinText** bằng cách lấy ảnh trên mạng và được chụp lại bởi những người thu thập dữ liệu, nó đại diện cho tất cả những văn bản tự nhiên có thể được bắt gặp trong đời sống ở Việt Nam. **VinText** có tổng cộng 2000 tấm ảnh trong đó có 764 tấm tải từ Internet và 1236 tấm được chụp bởi những người thu thập dữ liệu. * Huấn luyện các mô hình **ABCNet** [1] và **MaskTextSpotter** [2] chạy trên các bộ dữ liệu **TotalText** [3], **ICDAR2013** [4], **ICDAR2015** [5] và bộ dữ liệu mới là **VinText** bằng cả phương pháp thông thường và phương pháp cải tiến. So sánh và đánh giá kết quả dựa trên độ đo **F1 score**. * Xây dựng một chương trình ứng dụng, cho phép người dùng chụp ảnh trực tiếp trên điện thoại, sau đó hình ảnh được đưa vào chương trình xử lý. |
| **KẾT QUẢ MONG ĐỢI**   * Phương pháp cải tiến **Dictionary-guided Scene Text Recognition** sẽ cải thiện được những hạn chế mà những cách tiếp cận hiện có gặp phải trong bài toán **Scene Text Recognition**. * Bộ dữ liệu **VinText** sẽ được ứng dụng và đóng góp vào việc phát triển nghiên cứu bài toán **Scene Text Recognition**. * Tạo chương trình ứng dụng thực tế cho cách tiếp cận cải tiến được đề xuất trên. Cho phép người dùng nhận diện văn bản trong hình ảnh trên điện thoại. Người dùng chụp ảnh trực tiếp và ảnh sẽ được đưa vào chương trình xử lý, sau đó văn bản đã được nhận diện sẽ hiển thị trên màn hình điện thoại. |
| **KẾ HOẠCH THỰC HIỆN:**  - **Tuần 1 đến tuần 5**: Tìm hiểu cách tiếp cận hiện có của bài toán Scene Text Recognition. Tìm hiểu về các bộ dữ liệu đã có sẵn: TotalText [3], ICDAR2013 [4], ICDAR2015 [5]. Từ đó nghiên cứu phương pháp mới là **Dictionary-guided Scene Text Recognition** . Bắt đầu thu thập dữ liệu cho bộ dữ liệu mới – **VinText**   * *Kết quả dự kiến:*   + Tài liệu về phương pháp tiếp cận bài toán **Scene Text Recognition** hiện có.  + Bộ dữ liệu TotalText [3], ICDAR2013 [4], ICDAR2015 [5] và tài liệu liên quan  + Mô hình cải tiến **Dictionary-guided Scene Text Recognition**  + Bộ dữ liệu **VinText** chưa hoàn chỉnh.  -  **Tuần 6 đến tuần 10**: Huấn luyện mô hình Scene Text Recognitionthông thường và Dictionary-guided Scene Text Recognition trên tập dữ liệu TotalText [3], ICDAR2013 [4], ICDAR2015 [5]. Tiếp tục thu thập dữ liệu cho bộ dữ liệu VinText.   * *Kết quả dự kiến:*   + Bảng kết quả của quá trình huấn luyện trên 3 tập dữ liệu.  + So sánh độ chính xác của hai phương pháp trên từng tập dữ liệu.  + Hoàn thiện bộ dữ liệu VinText.  -  **Tuần 11 đến tuần 13**: Huấn luyện mô hình Scene Text Recognition thông thường và Dictionary-guided Scene Text Recognition trên tập dữ liệu VinText.   * *Kết quả dự kiến:*   + Bảng kết quả của quá trình huấn luyện trên tập dữ liệuVinText.  + So sánh độ chính xác của 2 phương pháp.  -  **Tuần 14 đến tuần 16**: Xây dựng chương trình ứng dụng cho phương pháp Dictionary-guided Scene Text Recognition.   * *Kết quả dự kiến:*   + Chương trình ứng dụng |
| **TÀI LIỆU THAM KHẢO** *(Định dạng DBLP)*  [1] 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.  [2] Minghui Liao, Guan Pang, J. Huang, Tal Hassner, and X. Bai. Mask textspotter v3: Segmentation proposal network for robust scene text spotting. ArXiv, 2020.  [3] 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.  [4] 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.  [5] 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. [6] V. I. Levenshtein. Binary codes capable of correcting insertions and reversals. 1966. 1, 2, 3 |