# **THÔNG TIN CHUNG CỦA BÁO CÁO**

* Link YouTube video của báo cáo (tối đa 5 phút):   
  *(ví dụ: https://www.youtube.com/watch?v=AWq7uw-36Ng)*
* Link slides (dạng .pdf đặt trên Github):   
  *(ví dụ: https://github.com/mynameuit/CS2205.APR2023/TenDeTai.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 Trường Khoa Nguyên * MSSV: 230201004   Không có mô tả ảnh. | * Lớp: CS2205.APR2023 * Tự đánh giá (điểm tổng kết môn): 8/10 * Số buổi vắng: 1 * Số câu hỏi QT cá nhân: * Link Github: <https://github.com/mynameuit/CS2205.APR2023/> * Link Youtube: |

# **ĐỀ CƯƠNG NGHIÊN CỨU**

|  |
| --- |
| **TÊN ĐỀ TÀI**  TẠO CÂU MÔ TẢ TIẾNG VIỆT MANG YẾU TỐ CẢM XÚC CHO ẢNH DỰA TRÊN TIẾP CẬN HỌC SÂU |
| **TÊN ĐỀ TÀI TIẾNG ANH**  VIETNAMESE PERSONALITY IMAGE CAPTIONING BASED ON DEEP LEARNING APPROACH |
| **TÓM TẮT**  Tự động sinh mô tả cho bức ảnh là bài toán tự động tạo một câu văn miêu tả các sự vật và hành động của chúng trong bức ảnh đầu vào. Các câu mô tả được ứng dụng trong việc hỗ trợ các mô hình máy học khác nhằm nâng cao tính hiệu quả trong việc xử lý các vấn đề mà các mô hình này hướng đến. Mặc dù đã có nhiều nghiên cứu về bài toán này, nhưng đa số đều tập trung vào việc áp dụng cho ngôn ngữ tiếng Anh. Điều này tạo ra một hạn chế lớn khi muốn ứng dụng các mô hình này vào các ngôn ngữ khác, như tiếng Việt. Hơn nữa, các câu mô tả thường thiếu đi tính cá nhân hóa, không thể phản ánh được ý kiến, cảm xúc hay phong cách của người mô tả. Trong bài viết này, chúng tôi đề xuất một mô hình tự động sinh câu mô tả tiếng Việt cho ảnh, sử dụng kiến trúc Encoder-Decoder và cơ chế Attention. Cũng như đề xuất thêm một tầng nhúng tính cách vào mô hình, nhằm tạo ra các câu mô tả mang theo yếu tố cảm xúc và cá nhân hóa, từ đó làm cho các mô tả trở nên sâu sắc và hấp dẫn hơn đối với người đọc. |
| **GIỚI THIỆU**  Việc sử dụng các hình ảnh để thể hiện nội dung của một chủ đề đang được sử dụng phổ rộng rãi không những trên các trang mạng xã hội, báo chí mà còn được sử dụng trong các văn bản tài liệu, sách giáo khoa. Bài toán tạo mô tả cho ảnh với mục tiêu là biểu diễn nội dung của bức ảnh thành dạng văn bản miêu tả đúng các sự vật và hành động trong bức ảnh đầu vào. Các câu mô tả này sẽ có thể được ứng dụng vào trong việc hỗ trợ các mô hình học máy khác nhằm nâng cao tính hiệu quả mà các mô hình học máy đó đang hướng tới, đồng thời việc sinh ra câu mô tả cho ảnh cũng có thể làm một phương tiện hỗ trợ những người khiếm thị trong việc đọc hiểu nội dung của một bức ảnh. Đã có nhiều nghiên cứu trong việc tự động sinh câu mô tả cho ảnh như: Retrieval based, Template based, ... Tuy nhiên những nghiên cứu này chỉ được áp dụng cho tiếng Anh, chưa có nghiên cứu nào được thực hiện nhằm giải quyết vấn đề cho tiếng Việt. Đề tài với phạm vi xây dựng mô hình tạo câu mô tả tiếng Việt mong muốn với một bức hình người dùng đưa vào, mô hình sẽ tạo ra câu mô tả tiếng Việt thể hiện lên được nội dung của bức ảnh.    Hình 1: Sơ đồ đề xuất sử dụng mô hình Show Attend an Tell kết hợp với sử dụng yếu tố cảm xúc để tạo mô tả ảnh  Input: một bức ảnh và cảm xúc lựa chọn  Output: câu mô tả tiếng Việt  Ví dụ:   |  |  | | --- | --- | | Input | Output | | Hình ảnh:    Cảm xúc: Ngạc nhiên | Khung cảnh thật đáng ngạc nhiên | |
| **MỤC TIÊU**   * Nghiên cứu ứng dụng của mô hình Show Attend and Tell trong việc tạo mô tả cho hình ảnh. Đồng thời, tìm hiểu cách tích hợp lớp nhúng tính cách vào mô hình, nhằm tạo ra các mô tả cá nhân hóa và phong phú hơn. * Xây dựng bộ dữ liệu tiếng Việt cho bài toán tạo mô tả hình ảnh dựa trên phương pháp xây dựng tập dữ liệu về cá nhân từ Facebook. Đây là một phương pháp hiệu quả để tạo ra bộ dữ liệu đa dạng và phong phú, kết hợp với các nguồn dữ liệu MSCOCO và FLICKR30K. * Xây dựng ứng dụng tích hợp chung với mô hình với đầu vào là bức ảnh và cảm xúc người dùng chọn, đầu ra là một câu mô tả phù hợp với bức ảnh và cảm xúc đầu vô. |
| **PHẠM VI**   * Áp dụng cho tiếng Việt. * Chỉ sinh câu với 6 cảm xúc đã được xác định trước: Tức giận, hạnh phúc, ghê tởm, buồn bã, sợ hãi,  ngạc nhiên. * Áp dụng cho các bức ảnh có người. |
| **NỘI DUNG VÀ PHƯƠNG PHÁP**  **Nội dung thực hiện:**   * Tìm hiểu mục tiêu và yêu cầu bài toán, xác định mục tiêu cốt lõi và các ràng buộc. * Nghiên cứu và lựa chọn bài báo phù hợp, chọn lọc các bài báo giải quyết vấn đề tương tự, các phương pháp, kỹ thuật, độ đo. * Xây dựng bộ dữ liệu mô tả ảnh có cảm xúc tiếng Việt dựa vào các bộ dữ liệu phổ biến dùng cho bài toán mô tả ảnh MSCOCO, Flickr30k và bộ dữ liệu personality captions của Facebook. * Sử dụng mô hình Show Attend and Tell, và ứng dụng thêm layer để embed các yếu tố cảm xúc cho mô hình, train mô hình trên bộ dữ liệu đã xây dựng. * Đánh giá mô hình thông qua các độ đo như BLEU, METEOR, CIDEr, … * Xây dựng một ứng dụng demo để người dùng có thể thử nghiệm mô hình.   **Phương pháp thực hiện:**   * Tìm hiểu kiến trúc của mô hình Show Attend and Tell, nắm được cách xử lý dự liệu dạng ảnh và văn bản của bài toán và các thông số điều chỉnh để huấn luyện mô hình. * Dựa vào mô hình Show Attend and Tell thiết kế thêm layer để embed các yếu tố cảm xúc trong quá trình huấn luyện ảnh. * Tự xây dựng bộ dữ liệu personality captions, thông qua việc tham khảo cách làm của Facebook về cách phân loại bộ dữ liệu, sử dụng các API hiện có hỗ trợ quá trình chuyển đổi tiếng Anh sang tiếng Việt. * Tìm hiểu về các độ đo để đánh giá mô hình. * Huấn luyện mô hình với bộ dữ liệu xây dựng được và đánh giá kết quả mô hình. * Xây dựng, triển khai ứng dụng tích hợp mô hình để tăng khả năng tương tác. |
| **KẾT QUẢ MONG ĐỢI**   * Mô hình chạy thành công trên tập dữ liệu đã xây dựng, với câu mô tả tạo ra phù hợp với nội dung của bức hình và cảm xúc đầu vào, * Xây dựng bộ dữ liệu tiếng Việt đa dạng và phong phú, phản ánh được sự đa dạng của các đối tượng và hành động trong ảnh, cũng như tính cảm xúc và cá nhân hóa của người mô tả. * Triển khai thành công một ứng dụng tích hợp mô hình để người dùng có thể tương tác, sử dụng mô hình. |
| **TÀI LIỆU THAM KHẢO** *(Định dạng DBLP)*  [1]. Shuang Bai and Shan An. “A Survey on Automatic Image Caption Generation”. in: Neurocomputing 311 (may 2018). doi: 10.1016/j.neucom. 2018.05.080  [2]. Nicholas Becherer andothers. “Improving optimization of convolutional neural networks through parameter fine-tuning”. in: Neural Computing and Applications (2017), pages 1–11.  [3]. Oriol Vinyals andothers. “Show and Tell: A Neural Image Caption Generator”. in: CoRR abs/1411.4555 (2014). arXiv: 1411.4555. url: http: //arxiv.org/abs/1411.4555.  [4] ] K. Shuster, S. Humeau, H. Hu, A. Bordes, and J. Weston, “Engaging image captioning via personality,” in CVPR, 2019 |