# **THÔNG TIN CHUNG CỦA NHÓM**

* Link YouTube video của báo cáo (tối đa 5 phút):   
  [*https://youtu.be/49S6kkh\_D4A*](https://youtu.be/49S6kkh_D4A)
* Link slides:   
  [*https://github.com/HoangSonDeveloper/CS2205/blob/026f9e26243674a1dfc9104be19bd3bdb2b554c5/S%C6%A1n%20Nguy%E1%BB%85n%20Ho%C3%A0ng%20-%20CS2205.SEP2025.DeCuong.FinalReport.Template.Slide.pdf*](https://github.com/HoangSonDeveloper/CS2205/blob/026f9e26243674a1dfc9104be19bd3bdb2b554c5/S%C6%A1n%20Nguy%E1%BB%85n%20Ho%C3%A0ng%20-%20CS2205.SEP2025.DeCuong.FinalReport.Template.Slide.pdf)

| * Họ và Tên: Nguyễn Hoàng Sơn * MSSV: 250101058 | * Lớp: CS2205.CH201 * Tự đánh giá (điểm tổng kết môn): 9/10 * Số buổi vắng: 0 * Link Github:   <https://github.com/HoangSonDeveloper/CS2205> |
| --- | --- |

# **ĐỀ CƯƠNG NGHIÊN CỨU**

| **TÊN ĐỀ TÀI (IN HOA)**  NGHIÊN CỨU VÀ ĐỀ XUẤT GIẢI PHÁP THÍCH NGHI MÔ HÌNH DCDM CHO BÀI TOÁN SIÊU PHÂN GIẢI VĂN BẢN TIẾNG VIỆT |
| --- |
| **TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)**  RESEARCH AND PROPOSAL OF DCDM ADAPTATION FOR VIETNAMESE SCENE TEXT SUPER-RESOLUTION |
| **TÓM TẮT**  Trong các hệ thống nhận dạng ký tự quang học (Optical Character Recognition – OCR), chất lượng hình ảnh văn bản đầu vào đóng vai trò quyết định đến độ chính xác của kết quả nhận dạng. Tuy nhiên, trong thực tế, hình ảnh văn bản thường có độ phân giải thấp, bị mờ hoặc nhiễu do điều kiện thu nhận không lý tưởng. Bài toán siêu phân giải văn bản trong cảnh (Scene Text Super-Resolution) vì vậy trở thành một hướng nghiên cứu quan trọng nhằm nâng cao hiệu quả của các hệ thống OCR.  Gần đây, mô hình DCDM (Diffusion-Conditioned Diffusion Model) đã cho thấy hiệu quả vượt trội trong việc phục hồi chi tiết văn bản nhờ kết hợp mô hình khuếch tán ảnh với thông tin ngữ nghĩa từ văn bản. Tuy nhiên, DCDM được thiết kế và huấn luyện chủ yếu cho tiếng Anh, vốn chỉ sử dụng bảng chữ cái Latinh không dấu, nên chưa phù hợp để xử lý văn bản tiếng Việt – một ngôn ngữ có hệ thống dấu thanh và nguyên âm biến thể phức tạp.  Nghiên cứu này tập trung phân tích và đề xuất giải pháp thích nghi mô hình DCDM cho bài toán siêu phân giải văn bản tiếng Việt. Cụ thể, đề tài xây dựng một quy trình tạo dữ liệu giả lập văn bản tiếng Việt với đầy đủ các ký tự có dấu, đồng thời đề xuất mở rộng và tinh chỉnh module mã hóa văn bản trong DCDM nhằm bảo toàn cấu trúc hình thái và ngữ nghĩa của ký tự tiếng Việt trong quá trình phục hồi ảnh. Kết quả kỳ vọng của nghiên cứu là nâng cao chất lượng ảnh văn bản sau siêu phân giải và cải thiện đáng kể độ chính xác của các hệ thống OCR khi áp dụng trong bối cảnh thực tế tại Việt Nam. |
| **GIỚI THIỆU**  Trong bối cảnh chuyển đổi số, nhu cầu trích xuất thông tin tự động từ hình ảnh ngày càng gia tăng trong các lĩnh vực như giao thông, tài chính, hành chính công và lưu trữ tài liệu. Các hệ thống nhận dạng ký tự quang học (Optical Character Recognition – OCR) đóng vai trò quan trọng trong quá trình này, tuy nhiên hiệu quả của chúng phụ thuộc lớn vào chất lượng hình ảnh văn bản đầu vào.  Trong thực tế, hình ảnh văn bản thường được thu nhận từ các thiết bị phổ thông như camera giám sát hoặc điện thoại di động, dẫn đến hiện tượng mờ, nhiễu và mất chi tiết do điều kiện ánh sáng không ổn định, rung lắc hoặc khoảng cách thu nhận lớn. Đối với văn bản trong cảnh (scene text), sự đa dạng về phông chữ, nền ảnh và biến dạng hình học càng làm gia tăng độ khó của bài toán. Vì vậy, siêu phân giải văn bản (Scene Text Super-Resolution) được xem là một bước tiền xử lý quan trọng nhằm cải thiện độ rõ nét của ký tự và nâng cao hiệu quả của các hệ thống OCR.  Các phương pháp siêu phân giải truyền thống dựa trên mạng nơ-ron tích chập (CNN) hoặc mạng đối sinh (GAN) đã đạt được một số kết quả trên ảnh tự nhiên, nhưng thường bộc lộ hạn chế khi áp dụng cho ảnh chứa văn bản. Những mô hình này chủ yếu tối ưu chất lượng hình ảnh tổng thể mà chưa đảm bảo bảo toàn cấu trúc hình thái và ngữ nghĩa của ký tự, dẫn đến hiện tượng nhòe nét hoặc biến dạng chữ, làm suy giảm độ chính xác nhận dạng.  Gần đây, các mô hình khuếch tán (Diffusion Models) đã cho thấy hiệu quả vượt trội trong các bài toán sinh và phục hồi ảnh. Mô hình DCDM (Diffusion-Conditioned Diffusion Model) là một hướng tiếp cận tiêu biểu, kết hợp mô hình khuếch tán ảnh với thông tin ngữ nghĩa ở mức ký tự nhằm cải thiện chất lượng siêu phân giải cho ảnh văn bản. Tuy nhiên, DCDM được thiết kế và huấn luyện chủ yếu cho tiếng Anh – một ngôn ngữ không có dấu thanh.  Khi áp dụng cho tiếng Việt, mô hình gặp hạn chế do không nắm bắt được các nguyên âm biến thể và dấu thanh, trong khi đây là yếu tố mang tính quyết định về mặt ngữ nghĩa. Việc phục hồi sai hoặc thiếu dấu có thể làm thay đổi hoàn toàn nội dung văn bản. Xuất phát từ hạn chế này, đề tài tập trung nghiên cứu và đề xuất giải pháp thích nghi mô hình DCDM cho bài toán siêu phân giải văn bản tiếng Việt, nhằm nâng cao chất lượng ảnh văn bản và hiệu quả của các hệ thống OCR trong bối cảnh ứng dụng tại Việt Nam. |
| **MỤC TIÊU**   1. **Phân tích và đánh giá mô hình DCDM** trong bài toán siêu phân giải văn bản, làm rõ các hạn chế khi áp dụng mô hình này cho văn bản tiếng Việt có dấu thanh và nguyên âm biến thể. 2. **Xây dựng bộ dữ liệu văn bản tiếng Việt phục vụ huấn luyện và đánh giá**, đảm bảo bao phủ đầy đủ các ký tự và dấu thanh đặc trưng, nhằm hỗ trợ quá trình thích nghi mô hình siêu phân giải cho tiếng Việt. 3. **Đề xuất và thực nghiệm giải pháp thích nghi mô hình DCDM cho tiếng Việt**, thông qua việc mở rộng không gian biểu diễn ký tự và tinh chỉnh mô hình, hướng tới bảo toàn cấu trúc hình thái của ký tự và nâng cao hiệu quả nhận dạng văn bản. |
| **NỘI DUNG VÀ PHƯƠNG PHÁP**  **Xây dựng bộ dữ liệu văn bản tiếng Việt giả lập**  Do thiếu hụt các bộ dữ liệu chuẩn cho bài toán siêu phân giải văn bản tiếng Việt, nghiên cứu này đề xuất xây dựng một bộ dữ liệu giả lập nhằm phục vụ quá trình huấn luyện và đánh giá mô hình. Quy trình xây dựng dữ liệu được thiết kế để phản ánh sát các điều kiện thu nhận ảnh trong thực tế.  Cụ thể, văn bản tiếng Việt được sinh tự động từ các nguồn ngôn ngữ phong phú và được hiển thị bằng nhiều phông chữ Unicode phổ biến trên các nền ảnh đa dạng. Các kỹ thuật hạ chất lượng ảnh được áp dụng có kiểm soát, bao gồm làm mờ Gaussian, làm mờ chuyển động và giảm độ phân giải, nhằm tạo ra các cặp ảnh chất lượng cao và chất lượng thấp tương ứng. Bộ dữ liệu thu được đảm bảo bao phủ đầy đủ các ký tự đặc trưng của tiếng Việt, bao gồm nguyên âm có dấu và các dấu thanh, từ đó tạo nền tảng cho việc nghiên cứu khả năng phục hồi chi tiết văn bản của mô hình.  **Thích nghi mô hình DCDM cho văn bản tiếng Việt**  Trọng tâm của nghiên cứu là đề xuất giải pháp thích nghi mô hình DCDM cho bài toán siêu phân giải văn bản tiếng Việt. Trên cơ sở mô hình DCDM gốc, nghiên cứu tiến hành mở rộng không gian biểu diễn ký tự trong module khuếch tán văn bản tiềm ẩn nhằm đáp ứng số lượng và cấu trúc ký tự phong phú của tiếng Việt.  Bên cạnh đó, chiến lược học chuyển giao (transfer learning) được áp dụng bằng cách sử dụng các trọng số tiền huấn luyện của mô hình gốc để kế thừa khả năng phục hồi hình ảnh cơ bản, sau đó tinh chỉnh mô hình trên bộ dữ liệu văn bản tiếng Việt. Quá trình tinh chỉnh tập trung vào việc giúp mô hình học được mối quan hệ giữa cấu trúc hình ảnh và thông tin dấu thanh, qua đó hạn chế hiện tượng mất hoặc sai dấu trong ảnh văn bản được phục hồi.  **Phương pháp đánh giá**  Hiệu quả của mô hình sau khi thích nghi được đánh giá thông qua cả chỉ số định lượng về chất lượng ảnh và hiệu quả trên tác vụ nhận dạng văn bản. Các chỉ số phổ biến như PSNR (Peak Signal-to-Noise Ratio) và SSIM (Structural Similarity Index) được sử dụng để đánh giá mức độ phục hồi hình ảnh. Đồng thời, độ chính xác nhận diện ký tự của các hệ thống OCR tiếng Việt được dùng như một thước đo gián tiếp nhằm đánh giá mức độ bảo toàn thông tin ngữ nghĩa của văn bản sau khi siêu phân giải. |
| **KẾT QUẢ MONG ĐỢI**   * **Về dữ liệu:**  Xây dựng được một bộ dữ liệu văn bản tiếng Việt phục vụ bài toán siêu phân giải, bao phủ đầy đủ các ký tự có dấu và dấu thanh đặc trưng. Bộ dữ liệu này có thể được tái sử dụng cho các nghiên cứu liên quan đến xử lý ảnh văn bản và nhận dạng ký tự tiếng Việt trong tương lai. * **Về mô hình:**  Đề xuất và thực nghiệm thành công phương án thích nghi mô hình DCDM cho văn bản tiếng Việt, giúp cải thiện khả năng phục hồi chi tiết ký tự có dấu và hạn chế hiện tượng mất hoặc sai dấu so với mô hình gốc huấn luyện trên tiếng Anh. * **Về hiệu quả ứng dụng:**  Ảnh văn bản sau khi siêu phân giải bằng mô hình đề xuất đạt chất lượng cao hơn về mặt cấu trúc và tính dễ đọc, qua đó cải thiện hiệu quả của các hệ thống OCR tiếng Việt khi so sánh với việc sử dụng trực tiếp ảnh đầu vào chất lượng thấp. * **Về đóng góp khoa học:**  Nghiên cứu cung cấp phân tích thực nghiệm về ảnh hưởng của đặc trưng ngôn ngữ tiếng Việt đối với bài toán siêu phân giải văn bản dựa trên mô hình khuếch tán, đồng thời làm cơ sở cho các nghiên cứu tiếp theo về mở rộng mô hình xử lý ảnh văn bản đa ngôn ngữ. |
| **TÀI LIỆU THAM KHẢO**  [1] Chuhui Xue, Goon-Guek Han, Wenqing Zhang, et al. "DCDM: Diffusion-Conditioned Diffusion Model for Scene Text Editing and Super-Resolution". In Proceedings of the European Conference on Computer Vision (ECCV), 2024.  [2] Y. Yuan, Z. Zhang, and C. Li, et al. "Scene Text Image Super-Resolution with Text Style Embedding". In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, 2025.  [3] J. Wang, Y. Liu, and H. Hu, et al. "Efficient Scene Text Image Super-resolution with Semantic Guidance". arXiv preprint arXiv:2403.13330, 2024.  [4] L. Chen, X. Wang, and J. Li, et al. "Scene text image super-resolution with semantic-aware Interaction Network". Knowledge-Based Systems, vol. 285, 2025.  [5] T. H. Nguyen, K. V. Nguyen, and T. T. Nguyen, et al. "Vietnamese text recognition in scene images using deep learning with attention mechanism". Journal of Military Science and Technology, 2023.  [6] Q. L. Tran, H. A. Le, and T. M. Hoang, et al. "A Survey on Vietnamese Document Analysis and Recognition". arXiv preprint arXiv:2506.05061, 2025.  [7] M. T. Nguyen, et al. "VietOCR: A Tool for Vietnamese Optical Character Recognition". GitHub Repository: lynguyenminh/vietnamese-scenetext-detection-recognition, 2023. |