

**NGHIÊN CỨU VÀ ĐỀ XUẤT GIẢI PHÁP THÍCH NGHỊ
MÔ HÌNH DCDM
CHO BÀI TOÁN SIÊU PHÂN GIẢI VĂN BẢN TIẾNG VIỆT**

Nguyễn Hoàng Sơn - 250101058

Tóm tắt

- Lớp: CS2205.CH201
- Link Github của nhóm: <https://github.com/HoangSonDeveloper/CS2205>
- Link YouTube video: https://youtu.be/49S6kkh_D4A

Giới thiệu

- OCR được sử dụng rộng rãi trong hành chính, giao thông, tài chính
- Ảnh văn bản thực tế thường bị mờ, nhiễu, độ phân giải thấp
- Siêu phân giải văn bản giúp cải thiện chất lượng đầu vào cho OCR
- Các mô hình SR truyền thống chưa xử lý tốt chi tiết ký tự nhỏ

Giới thiệu

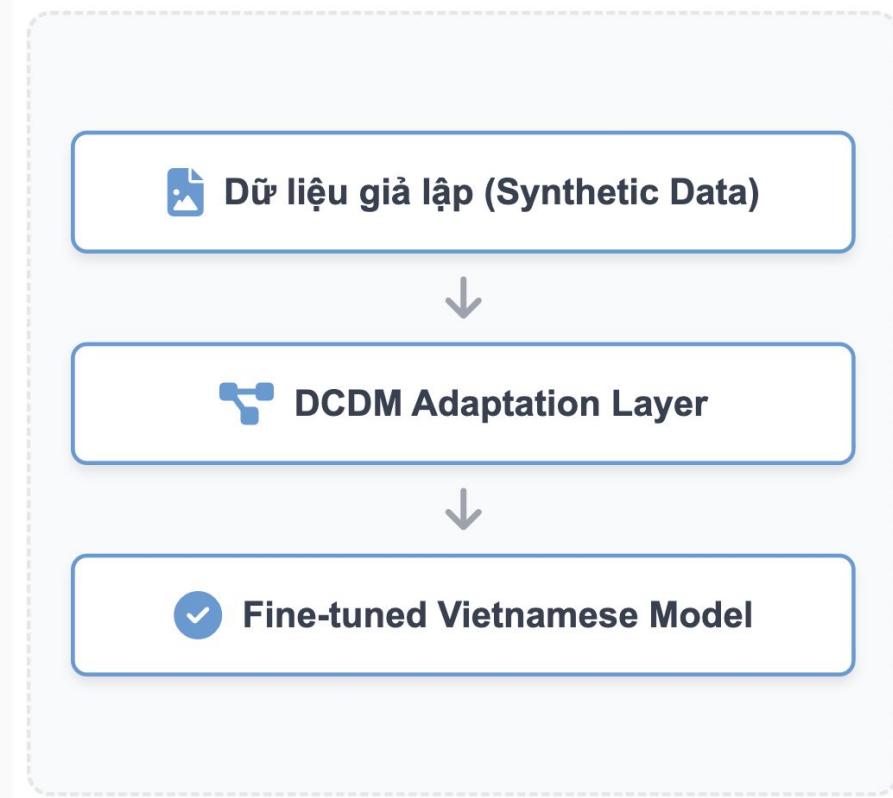
- DCDM (ECCV 2024) là mô hình SR văn bản dựa trên Diffusion
- Khai thác thông tin ngữ nghĩa ở mức ký tự
- Mô hình được tối ưu chủ yếu cho tiếng Anh
- Chưa phù hợp với tiếng Việt có dấu thanh và nguyên âm biến thể

Mục tiêu

- Phân tích hạn chế của mô hình DCDM khi áp dụng cho tiếng Việt
- Xây dựng dữ liệu phục vụ bài toán siêu phân giải văn bản tiếng Việt
- Đề xuất giải pháp thích nghi mô hình DCDM cho văn bản có dấu

Nội dung và phương pháp

- Xây dựng bộ dữ liệu văn bản tiếng Việt giả lập (đa dạng phông chữ, nền ảnh, mức nhiễu)
- Mở rộng không gian biểu diễn ký tự trong mô hình để hỗ trợ dấu tiếng Việt
- Fine-tune mô hình DCDM trên tập dữ liệu tiếng Việt đã chuẩn bị



Đánh giá

- Đánh giá chất lượng ảnh (Image Quality):
 - PSNR (Peak Signal-to-Noise Ratio)
 - SSIM (Structural Similarity Index)
- Đánh giá hiệu dụng (Usability):
 - Đo lường độ chính xác nhận dạng của OCR tiếng Việt trên ảnh phục hồi
- Kịch bản so sánh (Comparison):

Ảnh gốc (LR)

vs.

DCDM gốc

vs.

DCDM thích nghi

Kết quả dự kiến

- Phục hồi tốt hơn các ký tự có dấu thanh.
- Giảm hiện tượng mất hoặc sai dấu.
- Cải thiện hiệu quả của hệ thống OCR tiếng Việt.
- Đóng góp bộ dữ liệu phục vụ nghiên cứu sau này.

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/vietnamesescenetext-detection-recognition, 2023.