

**TĂNG CƯỜNG KHÔI PHỤC ẢNH
BẰNG TÍCH HỢP SÂU MÔ HÌNH
NGÔN NGỮ LỚN ĐA MÔ THỨC VÀO
KIẾN TRÚC ONERESTORE.**

Trần Tấn Đạt - 240101040

Tóm tắt

- Lớp: CS2205.FEB2025
- Link Github của nhóm:
<https://github.com/dattt19uit/CS2205.FEB2025>
- Link YouTube video: <https://youtu.be/nhREb3KO-hs>
- Trần Tấn Đạt



Giới thiệu

- Khôi phục ảnh được xem là bài toán then chốt trong Thị giác máy tính, ứng dụng rộng rãi (theo dõi đối tượng, y tế, bảo tồn di tích) [4, 5, 6].

World Scientific Connect

Journal of Mechanics in Medicine and Biology | Vol. 24, No. 09, 2440075 (2024)

MEDICAL IMAGE RESTORATION METHOD BASED ON NEIGHBORHOOD REGISTRATION AND ITS APPLICATION

YAXIONG YOU, XUNJIE SHI, ZHE LIU, RUI WANG✉, CHONGJI LI, and YANLONG YANG

SPRINGER NATURE Link

Image Restoration for Heritage Places Using GAN

Giới thiệu

- Restormer [1]: Hiệu quả với khử nhiễu, khử mờ, tăng cường ánh sáng yếu nhờ cơ chế attention → chưa xử lý tốt suy giảm đa dạng
- OneRestore [3]: Sử dụng cross-attention, tích hợp mô tả cảnh suy giảm bằng nhãn **tình**. → Chưa tận dụng ngữ nghĩa đa dạng.

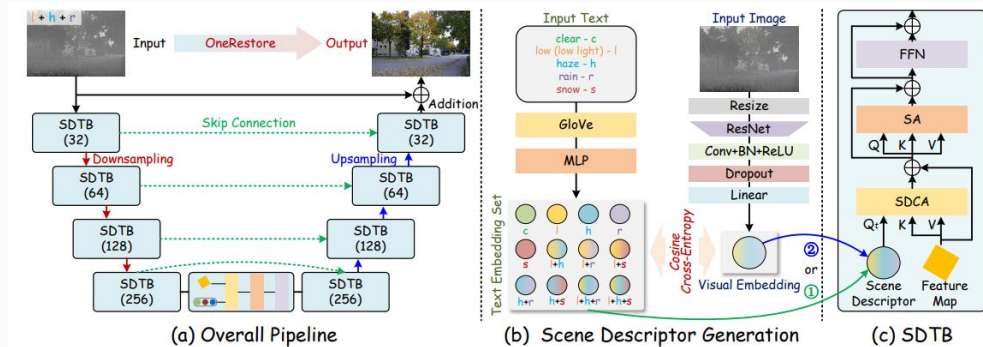
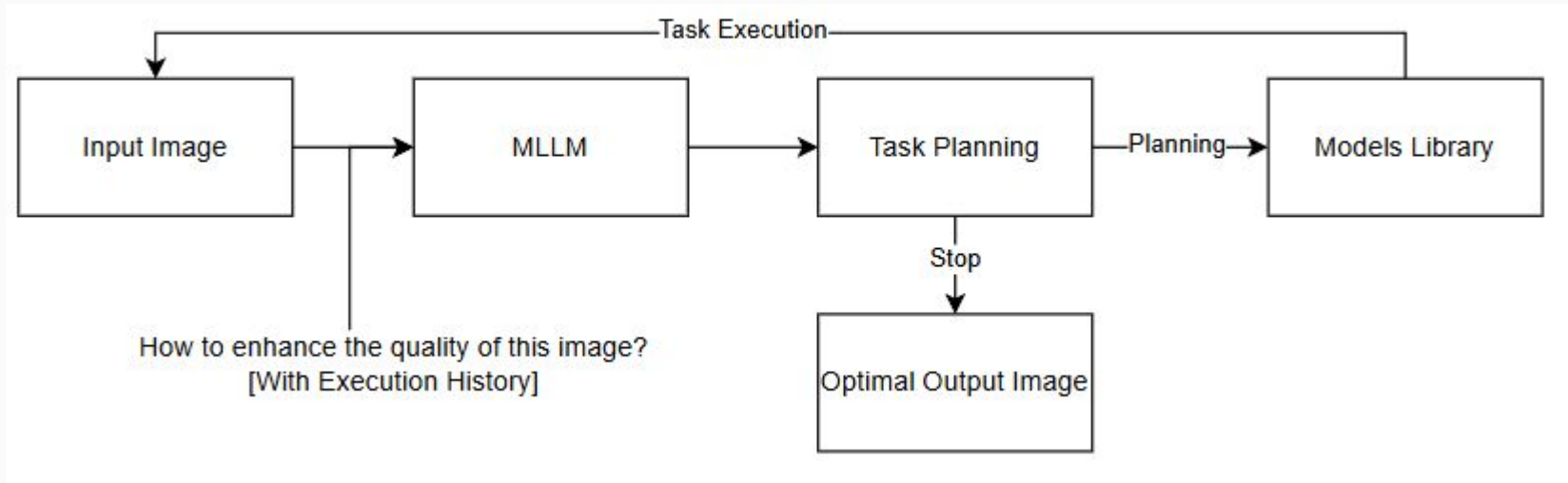


Fig. 3: Architecture of our OneRestore. (a) Overall pipeline, where 32, 64, 128, and 256 represent the number of channels. (b) Scene descriptor generation, where scene descriptors are fed into each (c) Scene Descriptor-guided Transformer Block (SDTB) by manual text embeddings or automatic extraction based on visual attributes.

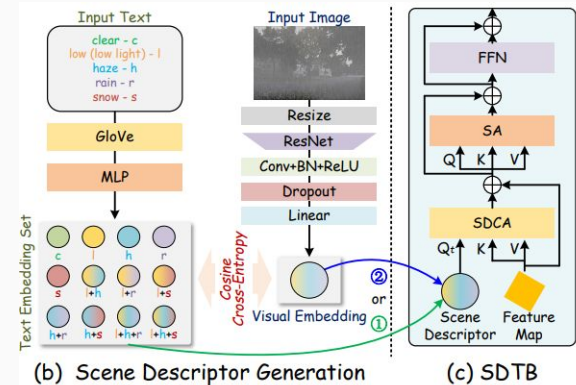
Giới thiệu

- RestoreAgent [2] tận dụng MLLM, hiệu suất cao nhưng điều phối bên ngoài, cả hai mô hình độc lập → Chi phí tính toán cao, hiệu suất không ổn định.



Giới thiệu

- Động lực nghiên cứu: **"Làm thế nào để cả 2 mô hình hỗ trợ lẫn nhau, giảm chi phí tính toán và nâng cao hiệu suất khôi phục ảnh?"**
- Giải pháp: Tận dụng OneRestore làm baseline, nghiên cứu cross-attention để tích hợp embedding động từ MLLM, tạo ảnh khôi phục chất lượng cao thay vì tận dụng embedding tĩnh từ OneRestore gốc.



Mục tiêu

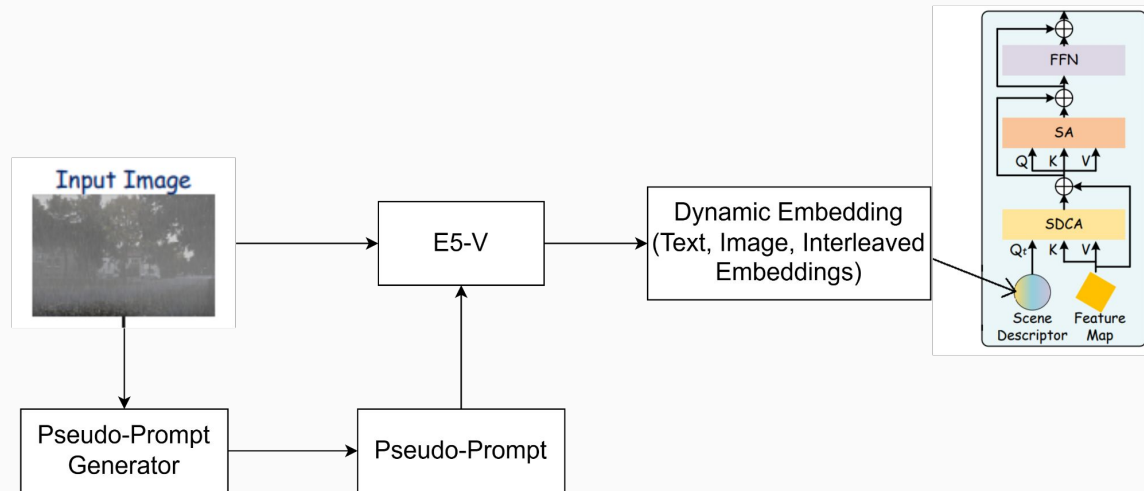
- Chuẩn bị tập dữ liệu thử nghiệm gồm ảnh suy giảm đơn và ảnh suy giảm phức tạp để phục vụ huấn luyện, kiểm định và đánh giá mô hình.
- Tích hợp E5-V [7] vào pipeline OneRestore bằng cross-attention để tạo embedding động.
- Huấn luyện lại (pretrain) mô hình tích hợp trên tập LSDIR [8], tinh chỉnh (finetune) trên tập validation và đánh giá đầu ra trên tập CDD-11 thông qua các chỉ số PSNR, SSIM.

Nội dung và Phương pháp

- Chuẩn bị dữ liệu:
 - LSDIR [8]: 100 ảnh train, 50 ảnh validation (suy giảm đơn).
 - CDD-11 [3]: 150 ảnh test (suy giảm phức tạp).
 - Chuẩn hóa ảnh: 256x256 (dùng PIL/OpenCV).
- Phân tích và tích hợp:
 - Phân tích OneRestore: Tập trung vào Scene Descriptor.
 - Tích hợp E5-V: Thay Scene Descriptor bằng Dynamic Embeddings từ MLLM.

Nội dung và Phương pháp

- Sơ đồ pipeline đề xuất: Input Image \rightarrow Pseudo-Prompt Generator \rightarrow Pseudo-Prompt \rightarrow (Input Image + Pseudo-Prompt) \rightarrow E5-V \rightarrow Dynamic Embeddings \rightarrow Scene Descriptor \rightarrow SDTB \rightarrow Output Image.



Nội dung và Phương pháp

- Huấn luyện và đánh giá:
 - Huấn luyện trên LSDIR, tinh chỉnh trên validation.
 - Đánh giá trên CDD-11: PSNR > 30, SSIM > 0.9, cải thiện 20% so với OneRestore.
- Ứng dụng demo bằng web app (Streamlit): Tải ảnh suy giảm, khôi phục, tính PSNR/SSIM.

Kết quả dự kiến

- Đánh giá trên tập CDD-11, mô hình đạt PSNR > 30, SSIM > 0.9, cải thiện 20% chất lượng khôi phục so với OneRestore gốc (theo PSNR trung bình).
- Xử lý hiệu quả ảnh suy giảm phức tạp (nhiều, mờ, vết mưa), nhờ tích hợp Dynamic Embeddings từ E5-V.
- Triển khai web app (Streamlit) để người dùng tải ảnh suy giảm, khôi phục, và tính PSNR/SSIM với ground-truth (nếu cung cấp).

Tài liệu tham khảo

- [1]. Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang: Restormer: Efficient Transformer for High-Resolution Image Restoration. CVPR 2022: 5718-5729
- [2]. Haoyu Chen, Wenbo Li, Jinjin Gu, Jingjing Ren, Sixiang Chen, Tian Ye, Renjing Pei, Kaiwen Zhou, Fenglong Song, Lei Zhu: RestoreAgent: Autonomous Image Restoration Agent via Multimodal Large Language Models. NeurIPS 2024
- [3]. Yu Guo, Yuan Gao, Yuxu Lu, Huilin Zhu, Ryan Wen Liu, Shengfeng He: OneRestore: A Universal Restoration Framework for Composite Degradation. ECCV (19) 2024: 255-272
- [4]. Zhiwen Yang, Hui Zhang, Dan Zhao, Bingzheng Wei, Yan Xu: Restore-RWKV: Efficient and Effective Medical Image Restoration with RWKV. CoRR abs/2407.11087 (2024)

Tài liệu tham khảo

- [5]. Yiran Li: Applications of Diffusion Model Image Restoration in the Field of Heritage Restoration: Overview and Outlook. CAIBDA 2023: 864-876
- [6]. Tianyang Xu, Yifan Pan, Zenhua Feng, Xuefeng Zhu, Chunyang Cheng, Xiao-Jun Wu, Josef Kittler: Learning Feature Restoration Transformer for Robust Dehazing Visual Object Tracking. Int. J. Comput. Vis. 132(12): 6021-6038 (2024)
- [7]. Ting Jiang, Minghui Song, Zihan Zhang, Haizhen Huang, Weiwei Deng, Feng Sun, Qi Zhang, Deqing Wang, Fuzhen Zhuang: E5-V: Universal Embeddings with Multimodal Large Language Models. CoRR abs/2407.12580 (2024)
- [8]. Yawei Li, Kai Zhang, Jingyun Liang, Jiezhong Cao, Ce Liu, Rui Gong, Yulun Zhang, Hao Tang, Yun Liu, Denis Demandolx, Rakesh Ranjan, Radu Timofte, Luc Van Gool: LSDIR: A Large Scale Dataset for Image Restoration. CVPR Workshops 2023: 1775-1787