

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM  
KHOA CÔNG NGHỆ THÔNG TIN

TÔ MÀU ẢNH ĐỘ XÁM  
DỰA TRÊN MÔ HÌNH KHUẾCH TÁN

Phạm Thái Huy - MSSV: 21120081

Tiêu Ân Tuấn - MSSV: 21120161

**Giảng viên hướng dẫn:**  
PGS.TS. Lý Quốc Ngọc và ThS. Đỗ Thị Thanh Hà

TP. Hồ Chí Minh, Ngày 16 tháng 12 năm 2025



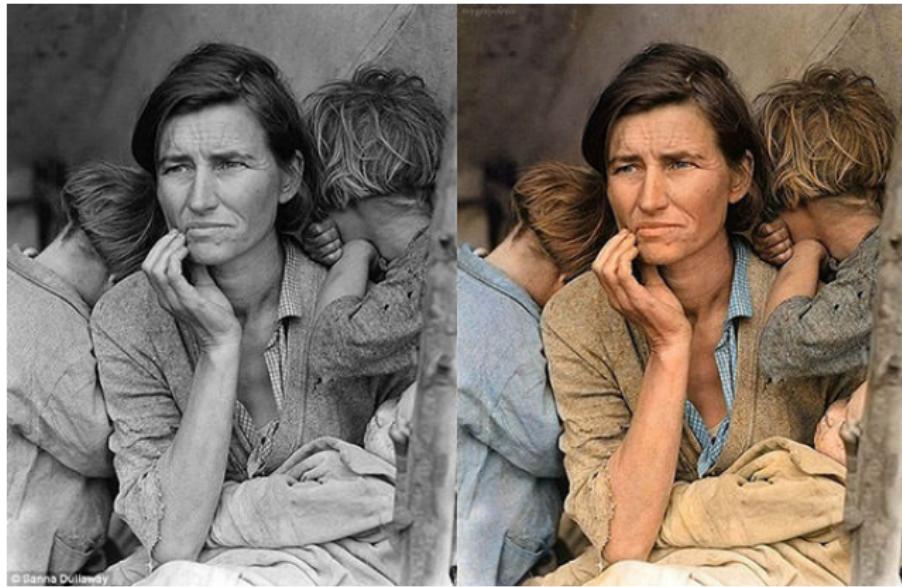
# Mục Lục

- ① Giới thiệu
- ② Các công trình nghiên cứu liên quan
- ③ Mô hình đề xuất
- ④ Thực nghiệm
- ⑤ Demo
- ⑥ Kết luận

1

## Giới thiệu

# Giới thiệu bài toán



Tô màu ảnh độ xám là quá trình dự đoán các giá trị màu cho một ảnh độ xám, sao cho ảnh kết quả phù hợp nhất với nhu cầu sử dụng.

# Bối cảnh và động lực nghiên cứu



Phục hồi ảnh lịch sử, ảnh chân dung, ảnh trắng đen lúc trước,...

# Bối cảnh và động lực nghiên cứu



Thay đổi màu của các đối tượng như đồ nội thất, trang phục, trang sức,...

# Phát biểu bài toán



+

c: two ladybugs are lying on a leaf



Ảnh xám ( $I_g$ ) và điều kiện (c)

Ảnh được tô màu ( $I_{rgb}$ )

Vì đây là bài toán không đơn trị nên ta có thể mô hình hóa quá trình tô màu như một bài toán sinh mẫu từ một phân phối xác suất có điều kiện:

$$I_{rgb} \sim P(I_{rgb}|I_g, c),$$

với c là điều kiện điều khiển quá trình tô màu (nếu có).

# Các thách thức của bài toán



- Đây là bài toán không đơn trị.
- Dễ gặp các lỗi lem màu, tô màu sai.
- Yêu cầu tương tác, tốn nhiều thời gian công sức.

# Mục tiêu

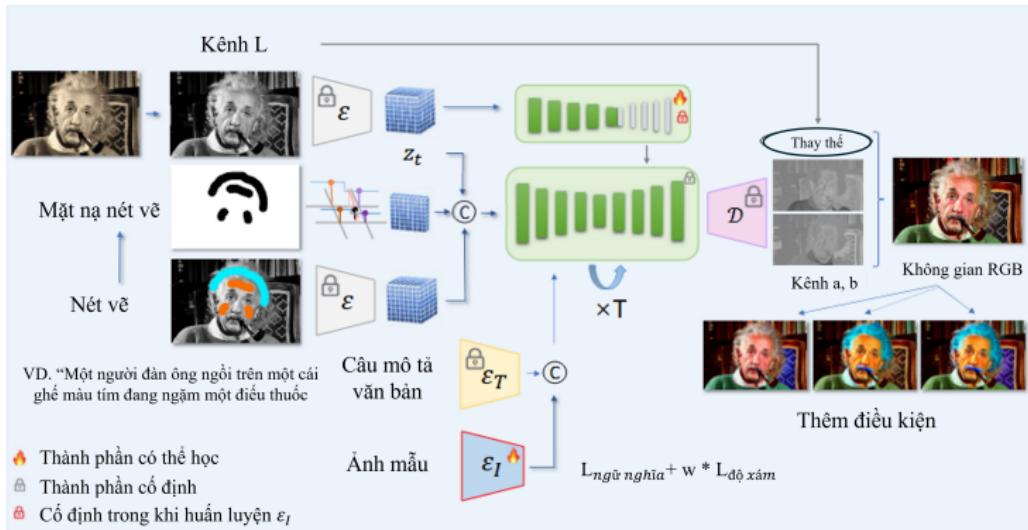
- Xây dựng một mô hình tô màu ảnh độ xám đa phương thức dựa trên mô hình khuếch tán sao cho mô hình ít tiêu tốn tài nguyên trong quá trình huấn luyện và thực thi đồng thời cho ra ảnh màu có chất lượng tốt.
- Tìm hiểu sâu hơn về khả năng của các mô hình khuếch tán. Một mô hình đầy mạnh mẽ nhưng chưa được áp dụng quá nhiều vào lĩnh vực tô màu ảnh độ xám.
- Tạo ra các mô hình tô màu phục vụ cho các ứng dụng cụ thể.
- Tạo ra một giao diện đơn giản có thể giúp người dùng có thể dễ dàng sử dụng các mô hình được xây dựng.

## ② Các công trình nghiên cứu liên quan

# Một số mô hình tô màu trước đó

Mô hình	Kiến trúc	Phương pháp	Ưu điểm	Hạn chế
UniColor	Bộ biến đổi	Chuyển các điều kiện đa dạng thành điểm màu gợi ý và tô màu thông qua bộ chuyển đổi	Hỗ trợ đa điều kiện; Cho phép chỉnh sửa lại nhiều lần; Kết quả điều khiển khá tốt	Phức tạp trong triển khai; Cần xử lý trước mỗi điều kiện riêng biệt
BigColor	Mạng đôi kháng tạo sinh	Học cách tạo sinh màu bằng GAN dựa trên cấu trúc không gian sẵn có của ảnh xám đầu vào	Có thể sinh ra nhiều kết quả tô màu khác nhau; Hỗ trợ độ phân giải ảnh kết quả linh hoạt	Kém hiệu quả với các ảnh phức tạp hoặc chi tiết nhỏ; Không hỗ trợ tương tác trực tiếp
Palette	Khuếch tán	Học một mô hình khuếch tán được huấn luyện bằng điều kiện ảnh độ xám	Tổng quát, hiệu năng cao; Kết quả vượt GAN trong nhiều thí nghiệm	Chậm do sinh mẫu qua nhiều bước; Không hỗ trợ điều kiện; Tốn nhiều tài nguyên

# Ý tưởng từ mô hình ControlColor

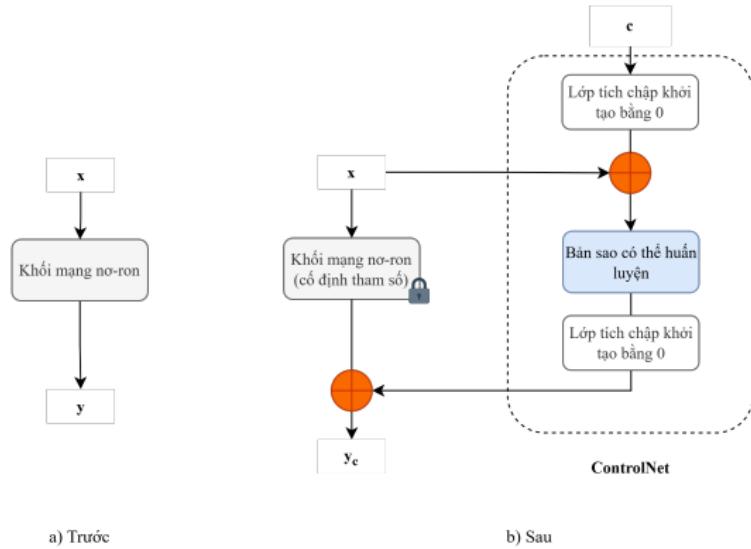


**Khuyết điểm:** Mô hình quá phức tạp cũng như yêu cầu quá nhiều tài nguyên cho quá trình huấn luyện và thực thi để đạt được kết quả tốt.

3

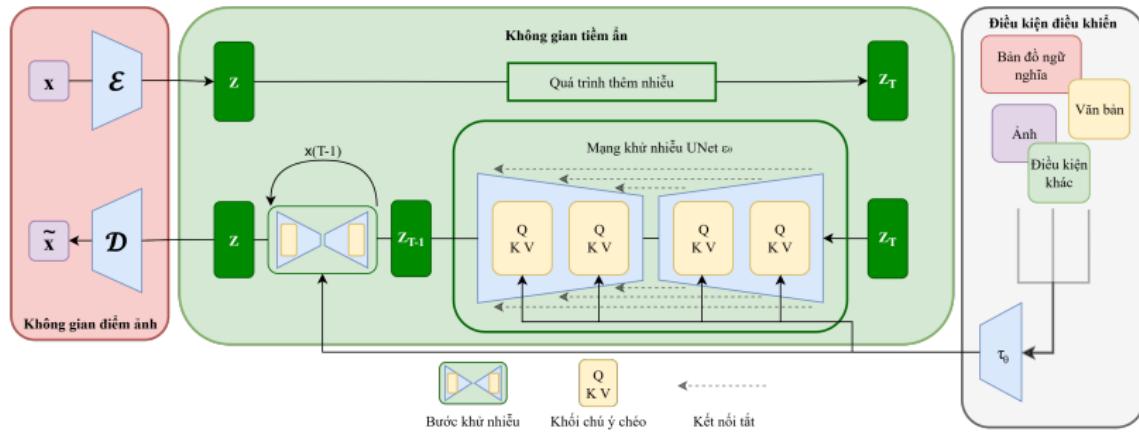
### Mô hình đề xuất

# Kiến trúc ControlNet



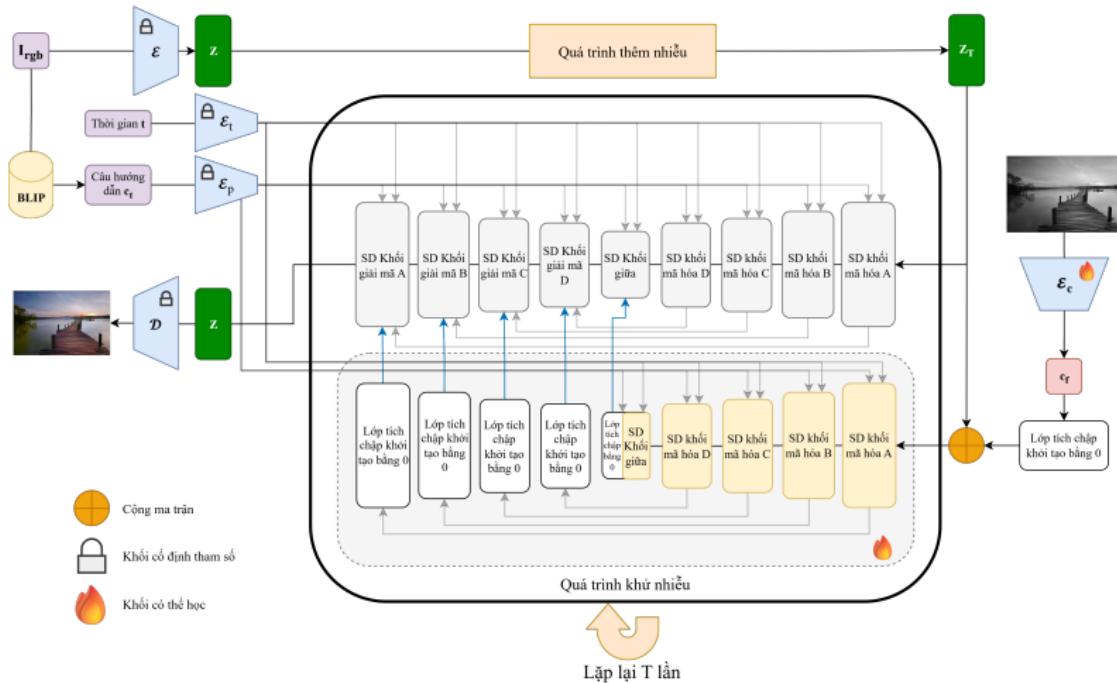
Kiến trúc ControlNet giúp thêm điều kiện điều khiển không gian vào các mô hình tạo sinh ảnh từ văn bản được huấn luyện sẵn.

# Mô hình khuếch tán trong không gian tiềm ẩn



Sử dụng mô hình tiền huấn luyện Stable Diffusion được huấn luyện trên 2 tỉ cặp ảnh và văn bản của tập dữ liệu LAION.

# Mô hình đề xuất



4

## Thực nghiệm

# Tập dữ liệu

- **ImageNet100k**: bao gồm 100K ảnh được chọn ngẫu nhiên từ tập dữ liệu huấn luyện ImageNet.
- **Fashion Product Images**: bao gồm khoảng 44K ảnh, mỗi ảnh chứa một loại phụ kiện thời trang chính, có thể được chụp với người mẫu hoặc chụp trên nền trắng.
- **Furniture Image**: bao gồm 15K ảnh của 5 loại nội thất khác nhau (tủ quần áo, ghế, tủ lạnh, bàn, tivi).
- **VN-celeb**: bao gồm khoảng 23K khuôn mặt của 1.020 người nổi tiếng ở Việt Nam có thể tìm thấy trên mạng, cụ thể là Wikipedia Việt Nam.

# Tiền xử lý dữ liệu

- ① Lọc ảnh độ xám: loại bỏ các ảnh có màu sắc tương đồng xám, chỉ giữ lại các ảnh có màu sắc ổn định cho việc huấn luyện.

$$E(Var(C_i, C_j)) = \frac{1}{3} \sum_{(i,j) \in \{(R,G), (G,B), (B,R)\}} Var(C_i - C_j) < t$$

- ② Thay đổi độ phân giải: về cùng kích thước  $512 \times 512$ .
- ③ Tạo sinh câu mô tả: sử dụng mô hình BLIP sinh câu mô tả để làm điều kiện trong quá trình huấn luyện.

Dữ liệu cho việc huấn luyện được tổ chức theo bộ ba tương ứng là: **ảnh xám, ảnh màu, câu mô tả**.

Tập dữ liệu	ImageNet100k	Fashion Product Images	Furniture Image	VN-celeb
Số lượng ảnh	97,590	38,284	13,525	22,284

# Độ đo

Có ba độ đo phổ biến trong các nghiên cứu tô màu ảnh độ xám, gồm:

- **Độ đo Fréchet Inception Distance (FID)** - kiểm tra độ chân thật của ảnh được tạo ra so với ảnh gốc. Ảnh sinh ra càng giống ảnh gốc thì FID càng thấp.
- **Độ đo Colorfulness** - đánh giá độ sắc sỡ của ảnh được tô màu. Ảnh càng sắc sỡ thì Colorfulness càng cao.
- **Điểm số CLIP** - đánh giá độ tương đồng giữa ảnh kết quả với câu mô tả trong không gian tiềm ẩn của CLIP, được dùng cho phương thức tô màu bán tự động.

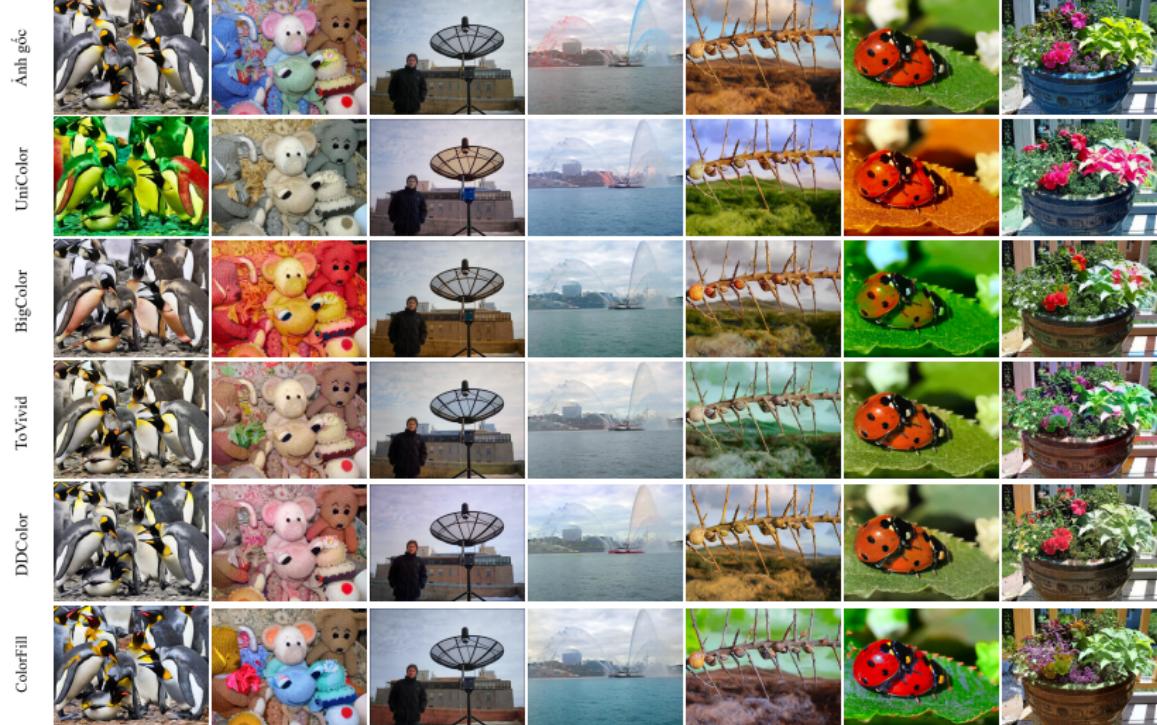
# Chiến lược đánh giá

- ① So sánh chất lượng với các mô hình SOTA
- ② So sánh Mô hình tổng quát với các mô hình chuyên biệt

# Tô màu tự động - Kết quả

Tập dữ liệu	ImageNet (val5k)		ImageNet (ctest)		COCO-Stuff	
	FID↓	Colorfulness↑	FID↓	Colorfulness↑	FID↓	Colorfulness↑
CIC	22.0860	37.0313	12.7651	37.5761	33.3418	37.6487
UGColor	15.1777	27.0966	6.5466	27.8122	21.4010	28.4487
DeOldify	10.5191	26.4827	4.2143	23.1538	13.4318	28.3779
ChromaGAN	16.4390	25.5862	9.3487	29.0895	26.4624	29.1411
InstColor	12.9455	27.5710	6.7803	28.1923	12.6844	29.2302
ToVivid	<u>5.8019</u>	37.3376	<u>2.6775</u>	37.8425	8.5452	38.8155
BigColor	7.7677	42.5364	2.8583	44.4135	10.0362	43.4104
DISCO	10.2895	40.9533	5.7196	37.4613	13.3850	39.1969
DDColor	<b>5.5726</b>	42.8370	<b>2.6294</b>	42.9575	<b>7.2718</b>	42.2919
ColorFormer	6.3831	40.5631	2.9816	41.2833	8.5623	41.0248
UniColor	9.6292	36.3415	4.9140	37.1726	<u>8.0509</u>	36.7442
Control Color	8.8749	<b>47.1680</b>	4.2915	<b>44.9256</b>	10.2651	<b>47.0501</b>
<b>ColorFill-LDM-CNet</b>	10.4125	<u>44.2871</u>	6.8341	<u>44.7854</u>	10.4632	<u>45.2589</u>

# Tô màu tự động - Kết quả (tt)



# Tô màu tự động - Kết quả (tt)

Ảnh gốc



ControlColor

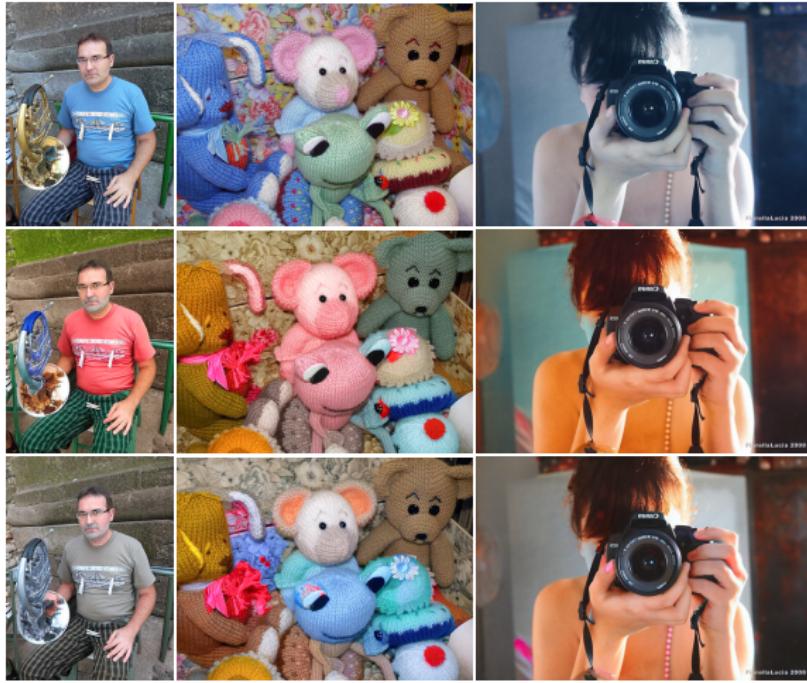


ColorFill



# Tô màu tự động - Kết quả (tt)

ColorFill      ControlColor      Ảnh gốc



# Tô màu tự động - Kết quả (tt)

Ảnh gốc



Control Color



ColorFill



# Tô màu bán tự động - Kết quả

Tập dữ liệu	COCO-Stuff		
	FID↓	Colorfulness↑	CLIP score↑
UniColor	<b>16.5833</b>	42.3788	0.2906
<b>ColorFill-LDM-CNet</b>	17.0299	<b>49.7282</b>	<b>0.3066</b>

# Tô màu bán tự động - Kết quả (tt)



"white horse with saddle  
on overlooks a river"



"a red traffic stop sign  
with a blank **white** sign  
below it"



Ảnh đầu vào

Lời nhắc văn bản

UniColor

ColorFill-LDM-CNet

# So sánh Mô hình tổng quát và các mô hình chuyên biệt

Tập dữ liệu		LFW	
Độ đo	FID↓	Colorfulness↑	
Tổng Quát	<b>40.9692</b>	15.6364	
Khuôn Mặt	41.558	<b>27.5432</b>	

Tập dữ liệu		Furniture	
Độ đo	FID↓	Colorfulness↑	
Tổng Quát	<b>3.0013</b>	21.8975	
Nội Thất	8.9494	<b>44.3698</b>	

Tập dữ liệu		Fashionpedia		Clothing	
Độ đo	FID↓	Colorfulness↑			
Tổng Quát	<b>12.3768</b>	27.528	6.7801	26.6948	
Thời Trang	18.4387	<b>27.575</b>	<b>6.6743</b>	<b>27.8723</b>	

## 5 Demo

## VIDEO DEMO

## 6 Kết luận

# Kết luận

Tổng kết nghiên cứu:

- Xây dựng một mô hình tô màu ảnh độ xám đa phương thức tốt, dựa trên mô hình khuếch tán với độ hiệu quả ổn định.
- Tạo ra các mô hình tô màu phục vụ cho các ứng dụng cụ thể.
- Xây dựng một giao diện đơn giản cho người dùng sử dụng các mô hình.

# Kết luận (tt)

Khuyết điểm của mô hình được đề xuất:

- Trong những trường hợp ảnh đầu vào có độ phân giải thấp hoặc chứa quá nhiều chi tiết, kết quả có thể không đúng như mong đợi.
- Khi mô tả văn bản xung đột với đặc trưng hình ảnh, mô hình có thể bị lúng túng bị lúng túng giữa việc tin ảnh hay tin văn bản.
- Các tập dữ liệu trên các miền chuyên biệt được chuẩn bị chưa tốt, dẫn đến kết quả đánh giá của các mô hình chuyên dụng chưa đạt được mong muốn thực tế.

Các cải tiến trong tương lai:

- Thủ nghiệm các bộ mã hóa văn bản và hình ảnh tiên tiến khác để tăng cường độ hiệu quả trên phương thức tô màu bằng câu mô tả.
- Cải thiện khả năng sinh ảnh chân thật hơn của các mô hình cải tiến sau này.

Chúng em xin cảm ơn Quý Thầy, Cô và các bạn đã lắng nghe phần trình bày.