

VIETNAM NATIONAL UNIVERSITY HO CHI MINH CITY
UNIVERSITY OF INFORMATION TECHNOLOGY



PROJECT FINAL REPORT

IMAGE RESTORATION

LỚP: CS114.O11.KHCL

Giảng viên: Phạm Nguyễn Trường An

Thành viên:

Trần Đức Long – 20521574

HO CHI MINH CITY, JANUARY 2024

TÓM TẮT NỘI DUNG ĐỒ ÁN

- Tìm hiểu bài toán Image Restoration
- Training lại mô hình huấn luyện bài toán trên bộ dataset có sẵn
- Đánh giá mô hình trên bộ dataset tự thu thập
- So sánh số liệu với mô hình huấn luyện sẵn.
- Link github: https://github.com/Longtd1605/Project_CS114

MỤC LỤC

TÓM TẮT NỘI DUNG ĐỒ ÁN	3
LỜI MỞ ĐẦU.....	5
CHƯƠNG 0. UPDATE SAU KHI VẤN ĐÁP.....	6
CHƯƠNG 1. GIỚI THIỆU BÀI TOÁN.....	7
1.1. MỤC ĐÍCH.....	7
1.2. MÔ TẢ BÀI TOÁN.....	7
1.3. THÁCH THỨC	8
CHƯƠNG 2. TỔNG QUAN VỀ MÔ HÌNH	9
2.1 BRINGING OLD PHOTOS BACK TO LIFE	9
2.1.1 Semi-supervised	9
2.1.2 Architecture	9
2.1.3 Huấn Luyện VAEs.....	10
2.1.4 Giảm Thiểu Khoảng Cách Domain	10
2.1.5 Học và khôi phục ảnh thông qua không gian latent	11
CHƯƠNG 3. THỰC NGHIỆM VÀ ĐÁNH GIÁ.....	12
3.1 Tổng quan về Dataset.....	12
3.2 Độ đo	13
3.3 Hiệu suất mô hình dự đoán.....	14
CHƯƠNG 4: KẾT LUẬN.....	15
4.1 Ưu điểm.....	15
4.2 Hạn chế.....	15
4.3 Hướng phát triển	15
TÀI LIỆU THAM KHẢO.....	16

LỜI MỞ ĐẦU

Trong thời đại ngày nay, công nghệ ngày càng chứng kiến sự tiến bộ đáng kể, mở ra những khả năng mới và tạo ra những giải pháp sáng tạo cho nhiều vấn đề. Một trong những ứng dụng thú vị của công nghệ hình ảnh là khả năng phục hồi và làm mới những bức ảnh cũ, những ký ức quý giá mà thời gian có thể làm mờ hoặc làm mất đi.

Bài toán Image Restoration không chỉ là việc tái tạo hình ảnh cũ thành hình ảnh chất lượng cao, mà còn là sự kết hợp tinh tế giữa trí tuệ nhân tạo và xử lý ảnh. Những công nghệ hiện đại đã mở ra những khả năng đáng kinh ngạc, từ việc loại bỏ vết nứt, nám, đến việc cải thiện màu sắc và độ phân giải. Qua quá trình này, chúng ta không chỉ làm mới những bức ảnh, mà còn tái tạo lại câu chuyện và hồi sinh những khoảnh khắc quý báu của quá khứ.

Trong bài toán này, chúng ta sẽ khám phá những ưu điểm và thách thức của việc phục hồi hình ảnh cũ, cùng với những hướng phát triển tiềm năng của công nghệ. Từ việc bảo tồn di sản gia đình đến ứng dụng trong lưu trữ văn hóa, Image Restore không chỉ là một công cụ kỹ thuật, mà còn là cầu nối kỳ diệu giữa quá khứ và tương lai, nơi mà mỗi bức ảnh trở nên sống động và đầy ý nghĩa hơn.

CHƯƠNG 0. UPDATE SAU KHI VẤN ĐÁP

CHƯƠNG 1. GIỚI THIỆU BÀI TOÁN

1.1. MỤC ĐÍCH

Phục hồi hình ảnh (**Image Restoration**) không chỉ khôi phục và cải thiện chất lượng ảnh khuôn mặt mà còn sử dụng các kỹ thuật xử lý ảnh để tái tạo hình ảnh gốc ban đầu.

Ngoài ra, phục hồi hình ảnh được ứng dụng cho việc nâng cao trải nghiệm người dùng trong lĩnh vực xử lý ảnh và làm đẹp ảnh cá nhân.

1.2. MÔ TẢ BÀI TOÁN

Input: 1 hình ảnh cần được khôi phục và cải thiện chất lượng. Ảnh này có thể bao gồm nhiễu, mất mát chi tiết và ánh sáng không đồng đều.

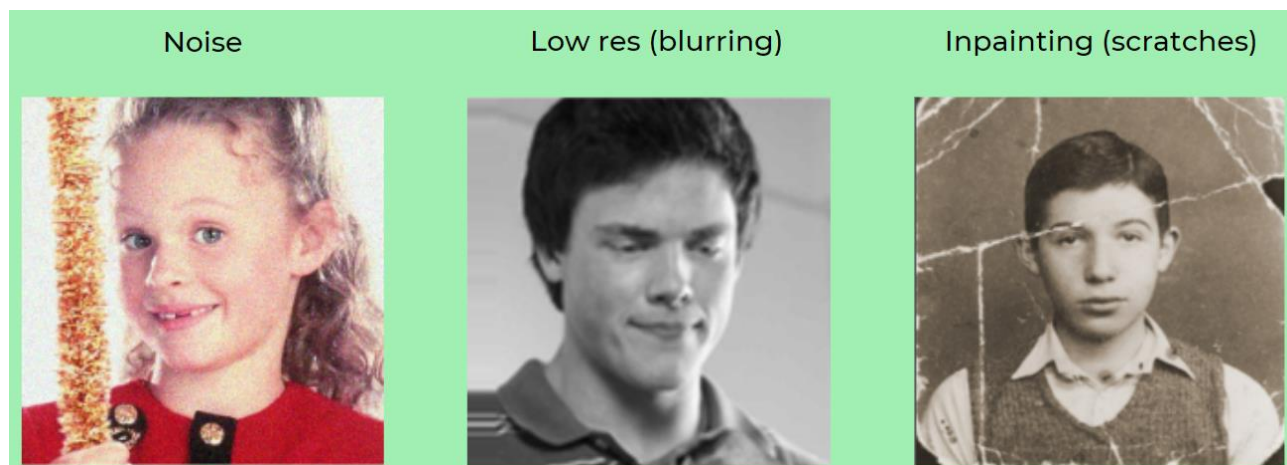
Output: 1 hình ảnh đã được làm mới và cải thiện chất lượng. Nó sẽ loại bỏ nhiễu, tái tạo chi tiết và có thể cải thiện độ tương phản, ánh sáng, hay màu sắc để trở nên rõ ràng và chất lượng hơn.



1.3. THÁCH THỨC

Nhiệm vụ này đặt ra hai thách thức chính như sau:

- Hình ảnh có thể bị ảnh hưởng bởi nhiều loại méo mó như nhiễu, làm mờ, vết trầy,... Mỗi loại méo mó đều đòi hỏi các phương pháp và mô hình phục hồi riêng biệt.
- Các mô hình phục hồi hình ảnh thường yêu cầu lượng lớn dữ liệu đào tạo, và việc tạo ra tập dữ liệu đào tạo chất lượng có thể là một thách thức đặc biệt, đặc biệt là khi cần mô phỏng nhiều loại méo mó.



Hình minh họa các loại méo mó lần lượt là nhiễu (noise), làm mờ (blurring), trầy xước (scratches)

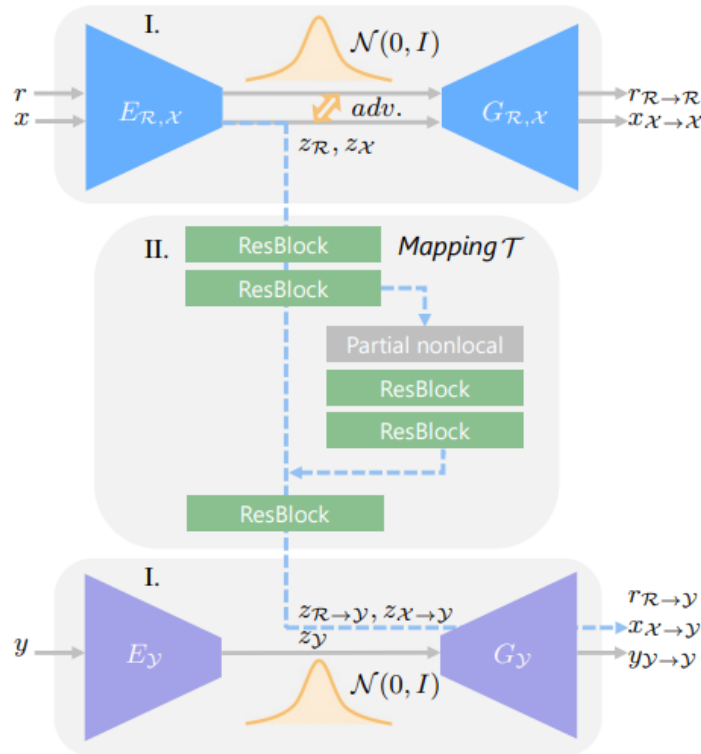
CHƯƠNG 2. TỔNG QUAN VỀ MÔ HÌNH

2.1 BRINGING OLD PHOTOS BACK TO LIFE

2.1.1 Semi-supervised

Mô hình được thiết kế theo phương pháp Semi-supervised learning. Trong đó, mô hình được huấn luyện trên cả dữ liệu có nhãn và dữ liệu không có nhãn. Dữ liệu có nhãn bao gồm các cặp hình ảnh đầu vào và đầu ra, trong khi dữ liệu không có nhãn chỉ chứa các hình ảnh đầu vào mà không có hình ảnh đầu ra tương ứng. Mô hình sẽ cố gắng tự học cách khôi phục hoặc cải thiện hình ảnh bằng cách sử dụng thông tin từ cả hai loại dữ liệu. Semi-supervised learning thường được áp dụng trong các tình huống khi việc thu thập dữ liệu có nhãn là đắt đỏ hoặc khó khăn, nhưng vẫn muốn tận dụng tối đa thông tin từ dữ liệu có sẵn để cải thiện hiệu suất của mô hình.

2.1.2 Architecture



Kiến trúc của mạng phục hồi. (I.) Trước tiên là hai mô hình Encoder-Decoder biến thể (VAE - Variational Autoencoder): VAE1 cho hình ảnh trong ảnh cũ $r \in \mathcal{R}$ và

hình ảnh tổng hợp $x \in X$, với khoảng cách giữa các miền được giảm bớt bằng cách đào tạo đồng thời một adversarial discriminator (GANs); VAE2 được đào tạo cho các hình ảnh thực $y \in Y$. Với các mô hình VAE, hình ảnh được chuyển vào không gian tiềm ẩn. (II.) Sau đó, mô hình học các ánh xạ để khôi phục hình ảnh bị hỏng về hình ảnh sạch trong không gian ẩn.

2.1.3 Huấn Luyện VAEs

- Sử dụng VAE1 cho R và X , chia sẻ cùng một encoder $E_{R,X}$ và generator $G_{R,X}$.
- Sử dụng VAE2 cho Y (ground true images), với encoder-generator E_Y, G_Y .
- VAE1 được chia sẻ cho cả R và X với mục tiêu là ánh xạ từ cả hai miền dữ liệu đều có thể được ánh xạ vào một không gian latent chung. VAE giả định phân phối Gaussian cho không gian latent, từ đó ảnh có thể được tái tạo bằng cách lấy mẫu từ không gian latent. Sử dụng thuật toán tái tham số hóa để cho phép lấy mẫu ngẫu nhiên và tối ưu hóa VAE1 với dữ liệu $\{r\}$ và $\{x\}$ tương ứng. Với $\{r\}$ được định nghĩa như sau:

$$\begin{aligned}\mathcal{L}_{VAE_1}(r) = & \text{KL}(E_{\mathcal{R},\mathcal{X}}(z_r|r) || \mathcal{N}(0, I)) \\ & + \alpha \mathbb{E}_{z_r \sim E_{\mathcal{R},\mathcal{X}}(z_r|r)} [\|G_{\mathcal{R},\mathcal{X}}(r_{\mathcal{R} \rightarrow \mathcal{R}}|z_r) - r\|_1] \\ & + \mathcal{L}_{VAE_1, \text{GAN}}(r)\end{aligned}$$

- Sử dụng KL-divergence để giữ cho phân phối latent gần với phân phối Gaussian. KL (Divergence Kullback-Leibler), là một độ đo sự chênh lệch giữa phân phối latent và phân phối Gaussian.

2.1.4 Giảm Thiểu Khoảng Cách Domain

$$\begin{aligned}\mathcal{L}_{VAE_1, \text{GAN}}^{\text{latent}}(r, x) = & \mathbb{E}_{x \sim \mathcal{X}} [D_{\mathcal{R},\mathcal{X}}(E_{\mathcal{R},\mathcal{X}}(x))^2] \\ & + \mathbb{E}_{r \sim \mathcal{R}} [(1 - D_{\mathcal{R},\mathcal{X}}(E_{\mathcal{R},\mathcal{X}}(r)))^2].\end{aligned}$$

- Sử dụng mạng adversarial để kiểm tra khoảng cách latent dư thừa
- Huấn luyện một discriminator $D_{R,X}$ để phân biệt giữa Z_R và Z_X (không gian latent của R và X).
- Mô hình encoder $E_{R,X}$ của VAE1 cố gắng đánh lừa discriminator bằng một contradictory loss để đảm bảo R và X được ánh xạ vào cùng một không gian.
- Tổng hợp loss latent adversarial để định nghĩa hàm mục tiêu tổng cho VAE1

$$\min_{E_{R,X}, G_{R,X}} \max_{D_{R,X}} \mathcal{L}_{VAE1}(r) + \mathcal{L}_{VAE1}(x) + \mathcal{L}_{VAE1, GAN}^{latent}(r, x).$$

2.1.5 Học và khôi phục ảnh thông qua không gian latent

- Trong giai đoạn này, mô hình chỉ huấn luyện các tham số của mạng ánh xạ latent T , trong khi giữ nguyên hai mô hình VAE (VAE1 và VAE2).
- Hàm mất mát LT được áp dụng cả trong không gian latent và phần cuối của generator GY , bao gồm ba thành phần chính:

$$\mathcal{L}_T(x, y) = \lambda_1 \mathcal{L}_{T, \ell_1} + \mathcal{L}_{T, GAN} + \lambda_2 \mathcal{L}_{FM}$$

$\mathcal{L}_{T, \ell_1} = \mathbb{E} \|\mathcal{T}(z_x) - z_y\|_1$: penalizes khoảng cách L1 của các mã tiềm ẩn tương ứng.

$\mathcal{L}_{T, GAN}$: encourage hình ảnh tổng hợp được dịch cuối cùng $xX \rightarrow Y$ trông như thật.

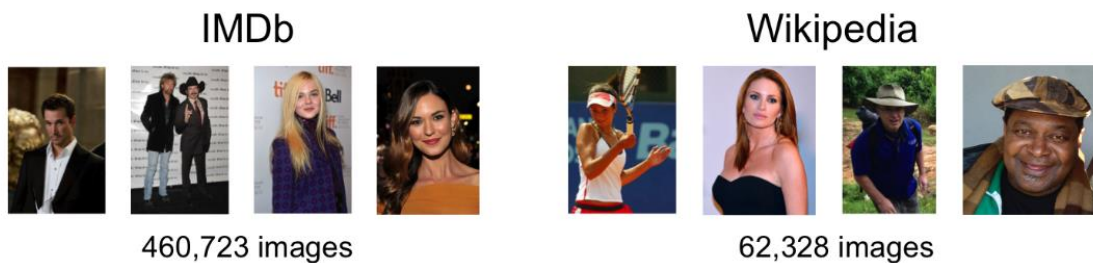
$$\mathcal{L}_{FM} = \mathbb{E} \left[\sum_i \frac{1}{n_{D_T}^i} \|\phi_{D_T}^i(xX \rightarrow Y) - \phi_{D_T}^i(yY \rightarrow Y)\|_1 + \sum_i \frac{1}{n_{VGG}^i} \|\phi_{VGG}^i(xX \rightarrow Y) - \phi_{VGG}^i(yY \rightarrow Y)\|_1 \right],$$

kết hợp giữa các activation đa cấp độ của mạng DM và mạng VGG, nhằm ổn định quá trình huấn luyện GAN.

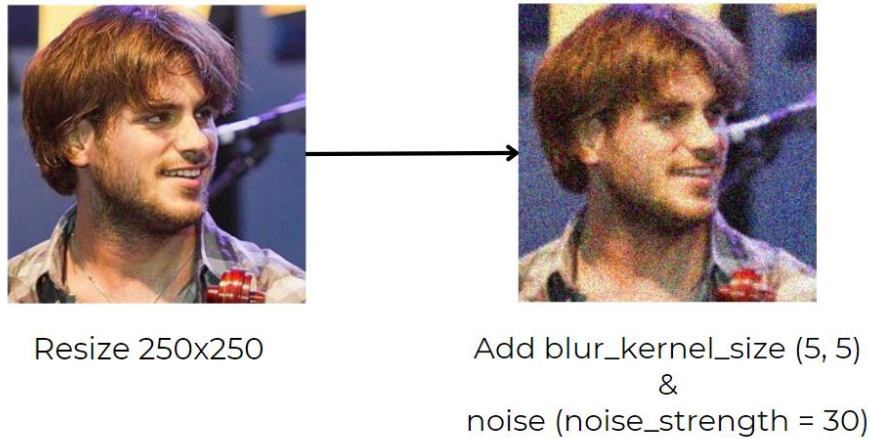
CHƯƠNG 3. THỰC NGHIỆM VÀ ĐÁNH GIÁ

3.1 Tổng quan về Dataset

- Tập train:
 - 81 bức ảnh xám cũ (sưu tầm: <https://hpcbristol.net/photographers>)
 - 8 bức ảnh màu (RGB) cũ (thư mục test_image)
<https://github.com/microsoft/Bringing-Old-Photos-Back-to-Life>
 - 1000 bức ảnh trong bộ dữ liệu Pascal-VOC
<https://www.kaggle.com/datasets/gopalbhattraipascal-voc-2012-dataset>
- Tập test: 1401 hình ảnh trong bộ dữ liệu IMDB-WIKI để test (đánh giá).
<https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/>



- Bộ dữ liệu IMDB-WIKI là một tập dữ liệu chứa thông tin về các người nổi tiếng. Bộ dữ liệu này bao gồm các thông tin như tên, giới tính, ngày sinh, quốc tịch và các hình ảnh của họ.
- Cách thức xây dựng bộ dữ liệu: Sử dụng 1401 hình ảnh trong bộ dữ liệu IMDB-WIKI để xử lý. Đầu tiên, resize ảnh về 250x250. Tiếp theo, tạo hiệu ứng làm mờ cho ảnh bằng cách áp dụng kernel blur_kernel_size = (5, 5) lên ảnh và thêm nhiễu cho ảnh bằng Gaussian noise với noise_strength = 30.



3.2 Độ đo

PSNR:

- PSNR, viết tắt của "peak signal-to-noise ratio," là một độ đo được sử dụng để đánh giá chất lượng của một tín hiệu so với nhiễu. Đối với hình ảnh, PSNR thường được sử dụng để đo lường sự chất lượng của hình ảnh nén so với hình ảnh gốc.
- Mean Squared Error (MSE) là một độ đo phổ biến để đo lường chất lượng của một ảnh hay tín hiệu so với một ảnh hay tín hiệu tham chiếu (ground truth).

SSIM:

- SSIM là viết tắt của "Structural Similarity Index." Đây là một độ đo được sử dụng trong xử lý ảnh để đo lường sự tương đồng cấu trúc giữa hai hình ảnh. SSIM cố gắng mô phỏng cách con người đánh giá sự thay đổi trong cấu trúc, màu sắc và độ tương phản của hình ảnh.

3.3 Hiệu suất mô hình dự đoán

Method	Metrics	
	PSNR	SSIM
BOPBTL_our	21.54	0.61
BOPBTL_pretrained	27.45	0.82

CHƯƠNG 4: KẾT LUẬN

4.1 Ưu điểm

- Giúp bảo tồn và duy trì giá trị của những bức ảnh cũ, làm mới chúng để chúng trở nên sống động hơn, từ đó giúp bảo vệ di sản gia đình và văn hóa.
- Khôi phục các chi tiết và thông tin mất mát trong các bức ảnh cũ, giúp tái tạo lại môi trường, người và vật dụng được chụp.
- Làm mới ảnh thường nhanh chóng hơn so với các phương pháp truyền thống như sửa ảnh bằng tay, đồng thời giảm thiểu công sức và thời gian cần thiết.

4.2 Hạn chế

- Trong quá trình tái tạo, có thể xảy ra mất mát một số thông tin gốc, đặc biệt là nếu bức ảnh ban đầu đã bị hỏng hoặc mất nhiều chi tiết.
- Hiệu suất của công nghệ phụ thuộc nhiều vào chất lượng ban đầu của bức ảnh. Nếu ảnh gốc quá mờ, nám, hoặc “tổn thương”, công nghệ có thể gặp khó khăn trong việc tái tạo.

4.3 Hướng phát triển

- Sử dụng AI để nhận biết và tái tạo chi tiết một cách tự động và chính xác hơn.
- Mở rộng khả năng tái tạo bằng cách tích hợp công nghệ 3D để tạo ra các hình ảnh có chiều sâu và hiệu ứng 3D.

TÀI LIỆU THAM KHẢO

- [1] Page: [\[2004.09484\] Bringing Old Photos Back to Life](#)
- [2] Github: [GitHub - microsoft/Bringing-Old-Photos-Back-to-Life](#)
- [3] Tools: OpenAI. (2023). ChatGPT