

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH  
TRƯỜNG ĐẠI HỌC BÁCH KHOA  
KHOA KHOA HỌC VÀ KỸ THUẬT MÁY TÍNH



Đồ án tổng hợp  
HƯỚNG TRÍ TUỆ NHÂN TẠO (CO3101)

Đề tài: PHỤC DỤNG ẢNH CHÂN DUNG CŨ

Giảng viên hướng dẫn:	Nguyễn Quang Đức
Thành viên:	Nguyễn Hữu Huy Thịnh 2213291
	Nguyễn Gia Thịnh 2213286
	Vũ Đình Hoàn 2211062
Lớp:	L12
Nhóm:	03

TP. Hồ Chí Minh, tháng 12 năm 2024



## Mục lục

<b>Phân công nhiệm vụ</b>	<b>3</b>
<b>1 Giới thiệu đề tài</b>	<b>4</b>
<b>2 Mô hình</b>	<b>4</b>
2.1 Tổng quan về GFP-GAN . . . . .	4
2.2 Module xóa hiệu ứng gây tổn hại ảnh . . . . .	4
2.3 Module GAN dùng để sinh ảnh . . . . .	5
2.4 Phép biến đổi đặc trưng không gian chia kênh . . . . .	5
2.5 Hàm mất mát . . . . .	6
2.5.1 Reconstruction Loss . . . . .	6
2.5.2 Adversarial Loss . . . . .	6
2.5.3 Facial Component Loss . . . . .	7
2.5.4 Identity Preserving Loss . . . . .	8
2.5.5 Hàm mất mát tổng thể . . . . .	9
2.5.6 Real Score và Fake Score . . . . .	9
<b>3 Thực thi</b>	<b>10</b>
3.1 Dataset và Hiện thực . . . . .	10
3.2 Dánh giá mô hình . . . . .	11
3.2.1 PSNR . . . . .	11
3.2.2 SSIM . . . . .	11
3.3 So sánh với các mô hình khác . . . . .	12
<b>4 Kết luận</b>	<b>14</b>
<b>Tài liệu tham khảo</b>	<b>15</b>



## Danh sách hình vẽ

1	Tổng quan về mô hình GFPGAN . . . . .	4
2	Reconstruction Loss . . . . .	6
3	Minh họa về ma trận Gram . . . . .	7
4	Facial Component Loss . . . . .	8
5	Identity Preserving Loss . . . . .	9
6	Real Score và Fake Score . . . . .	10
7	Điểm đánh giá . . . . .	12
8	So sánh kết quả mô hình GFPGAN với mô hình khác . . . . .	14

## Danh sách bảng

1	So Sánh đặc điểm của GFPGAN với 2 mô hình khác . . . . .	13
---	--	----



## Phân công nhiệm vụ

STT	MSSV	HỌ VÀ TÊN	CÔNG VIỆC
1	2213291	Nguyễn Hữu Huy Thịnh	Lý thuyết và hiện thực mô hình
2	2213286	Nguyễn Gia Thịnh	Lý thuyết và hiện thực mô hình
3	2211062	Vũ Dinh Hoàn	Dánh giá và so sánh mô hình GFPGAN

## 1 Giới thiệu đề tài

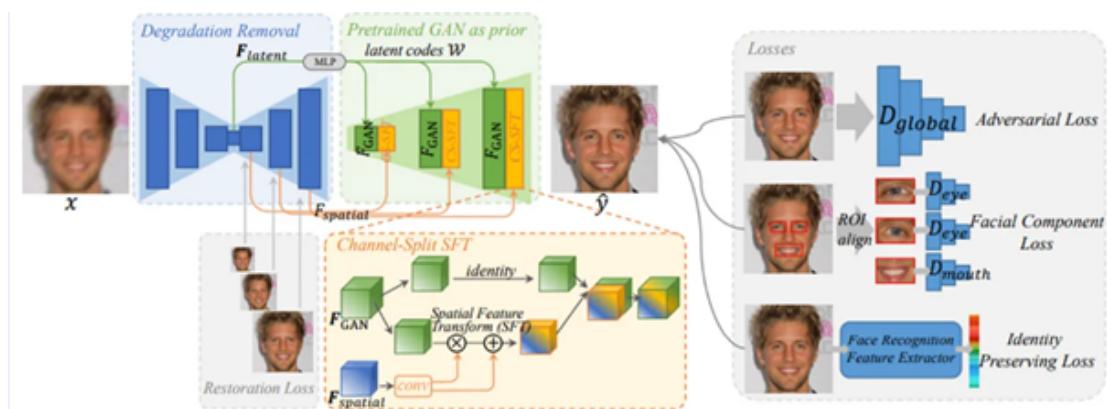
Đồ án này thực hiện đề tài phục dựng ảnh chân dung cũ dựa trên mô hình trí tuệ nhân tạo GFP-GAN.

GFP-GAN được huấn luyện để phục hồi các bức ảnh cũ, bị hư hỏng do thời gian và nhiều lý do khác, kể cả ảnh màu và trắng đen, bị nhạt màu... GFP-GAN đều có thể xử lý được về nguyên bản hoặc tốt hơn ban đầu.

## 2 Mô hình

### 2.1 Tổng quan về GFP-GAN

Với một tấm ảnh đầu vào  $x$  bị những tổn hại không biết, quá trình khôi phục ảnh cũ sẽ sinh ra một tấm ảnh chất lượng  $\hat{y}$  sao cho gần với tấm ảnh gốc  $y$  không chịu tổn hại nhất.



Hình 1: Tổng quan về mô hình GFPGAN

Tổng quan về mô hình GFP-GAN được miêu tả như hình trên. GFP-GAN được tạo thành từ module dùng để xóa những hiệu ứng gây tổn hại ảnh (U-Net) và một mô hình GAN tập trung khôi phục khuôn mặt được huấn luyện trước (ví dụ như StyleGAN2). Hai module được kết nối với nhau thông qua một phép ánh xạ các vector dữ liệu nén và một vài lớp tích chập được sử dụng cho phép biến đổi đặc trưng không gian kết hợp chia kenh (CS-SFT).

### 2.2 Module xóa hiệu ứng gây tổn hại ảnh

Mô hình sử dụng kiến trúc U-net[1] cho module xóa hiệu ứng gây tổn hại ảnh, vì mô hình có thể tăng cường loại bỏ các chi tiết gây tổn hại với ảnh và sinh ra các đặc trưng ở nhiều độ phân giải khác nhau. Module sẽ sinh ra 2 vector đặc trưng theo công thức như sau:

$$F_{latent}, F_{spatial} = U - Net(x) \quad (1)$$

- Vector đặc trưng dữ liệu nén  $F_{latent}$  chứa những thông tin quan trọng của tấm ảnh đầu vào  $x$ , được sử dụng để ánh xạ thành vector đặc trưng làm đầu vào cho module sinh ảnh StyleGAN2 (Phần 2.3).



- Vector đặc trưng không gian ở nhiều độ phân giải khác nhau  $F_{spatial}$ , chứa các dữ liệu về phân vùng ảnh, được sử dụng để cải thiện chất lượng và độ trung thực của tấm ảnh bằng cách thực hiện phép biến đổi SFT lên các vector đặc trưng đầu ra của StyleGAN2 (Phần 2.4).

### 2.3 Module GAN dùng để sinh ảnh

Mô hình GAN (Generative Adversarial Network)[2] là một loại mô hình học sâu được sử dụng rộng rãi trong bài toán sinh ảnh. StyleGAN2[3] là phiên bản cải tiến của mô hình GAN, có thể sinh ra các đặc trưng ở nhiều độ phân giải khác nhau, đồng thời khắc phục được các hiện tượng bất thường trong hình ảnh sinh ra. Vì khả năng tạo ra các ảnh chân dung giống người thật đến mức khó phân biệt bằng mắt thường, nhóm sử dụng StyleGAN2 cho module sinh ảnh trong bài toán khôi phục ảnh cũ.

Với vector đặc trưng dữ liệu nén  $F_{latent}$  (được sinh từ U-net, phương trình 1), đầu tiên ta sẽ ánh xạ nó thành vector dữ liệu nén cấp cao W mang thông tin có ý nghĩa hơn, phù hợp làm dữ liệu đầu vào cho mô hình GAN. Phép ánh xạ thành không gian dữ liệu nén cấp cao sẽ được tiến hành bằng cách đưa qua các lớp perceptron đa lớp (MLP). Vector nén W sau đó sẽ được đưa mỗi lớp tích chập trong mô hình GAN được huấn luyện trước, và sinh ra các đặc trưng GAN ở nhiều độ phân giải khác nhau.

$$\begin{aligned} W &= MLP(F_{latent}), \\ F_{GAN} &= StyleGAN(W) \end{aligned} \quad (2)$$

Thay vì sinh ra tấm ảnh cuối cùng, mô hình sẽ sinh ra các đặc trưng tích chập GAN  $F_{GAN}$  ở các độ phân giải khác nhau để có thể sử dụng phép biến đổi SFT nhằm tăng cường chất lượng và độ trung thực của tấm ảnh đầu ra.

### 2.4 Phép biến đổi đặc trưng không gian chia kẽ

Để có thể lưu lại tính trung thực của tấm ảnh, nhóm sử dụng vector đặc trưng không gian  $F_{spatial}$  để cải thiện các đặc trưng GAN  $F_{GAN}$  từ phương trình 2. Để có thể thực hiện được điều đó, nhóm áp dụng phép biến đổi SFT[4], SFT sẽ sử dụng các lớp tích chập để sinh ra các tham số để sử dụng cho phép biến đổi affine, từ đó giúp cải thiện các vector đặc trưng nhằm cải thiện chất lượng tấm ảnh và lưu lại các thông tin về không gian. Với mỗi độ phân giải khác nhau, nhóm tiến hành sinh ra cặp tham số cho phép biến đổi affine ( $\alpha, \beta$ ) từ đặc trưng đầu vào  $F_{spatial}$  bằng cách đưa qua các lớp tích chập. Sau đó phép biến đổi sẽ thực hiện scale và shift các pixel với đặc trưng GAN  $F_{GAN}$ .

$$\begin{aligned} \alpha, \beta &= Conv(F_{spatial}), \\ F_{Output} &= SFT(F_{GAN}|\alpha, \beta) = \alpha^{\circ}F_{GAN} + \beta \end{aligned} \quad (3)$$

Tuy nhiên để tránh trường hợp tấm ảnh sinh ra quá khác so với tấm ảnh gốc, nhóm đề xuất phép biến đổi đặc trưng không gian chia kẽ (CS-SFT). Nhóm sẽ chia đặc trưng GAN  $F_{GAN}$  ra làm hai, một nửa sử dụng kỹ thuật SFT, một nửa sẽ cho phép đi thẳng qua để lưu lại các nhận dạng của tấm ảnh.

$$\begin{aligned} F_{Output} &= CS - SFT(F_{GAN}|\alpha, \beta) \\ &= Concat [Identity(F_{GAN}^{split0}), \alpha^{\circ}F_{GANs}^{split1} + \beta] \end{aligned} \quad (4)$$

Với  $F_{GAN}^{split0}$  và  $\alpha^0 F_{GAN}^{split1}$  là các đặc trưng được tách ra từ  $F_{GAN}$ . Concat là phép dùng để kết hợp 2 vector đặc trưng lại.

CS-SFT không chỉ giúp bảo toàn tính trung thực của tấm ảnh, đồng thời tránh được các trường hợp sinh tấm ảnh quá khác so với ảnh gốc. Ngoài ra, CS-SFT cũng làm giảm độ phức tạp vì nó cần ít kênh hơn để sử dụng kĩ thuật SFT.

Nhóm thực hiện phép biến đổi CS-SFT tại mỗi độ phân giải, và cuối cùng sinh ra được tấm ảnh khôi phục  $\hat{y}$ .

## 2.5 Hàm mất mát

Hàm mất mát cho phép ta xác định mức độ sai khác của ảnh được tạo ra so với ảnh thật. Mô hình này kết hợp nhiều hàm mất mát khác nhau nhằm tối ưu hóa mô hình theo nhiều mục tiêu được trình bày dưới đây.

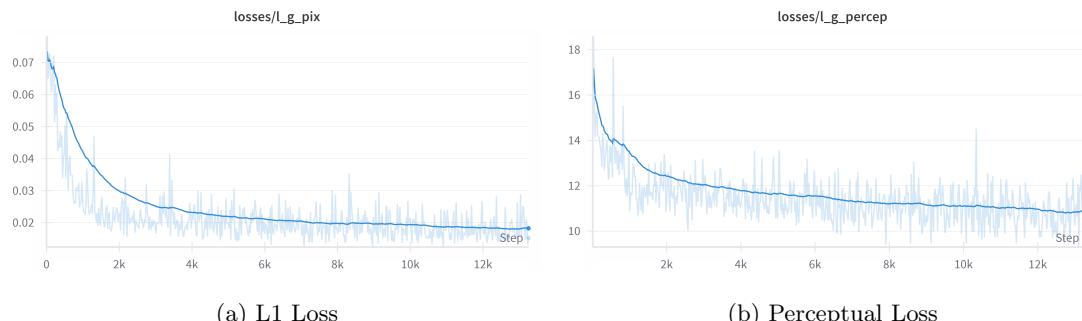
### 2.5.1 Reconstruction Loss

Reconstruction Loss kết hợp giữa L1 Loss và Perceptual Loss với mục tiêu làm giảm sự khác biệt giữa ảnh được tạo ra so với ảnh thật. L1 Loss đảm bảo ảnh tạo ra có màu sắc và chi tiết giống với ảnh gốc ở mức độ pixel. Ngược lại, Perceptual Loss đảm bảo ảnh sinh ra giữ được cấu trúc tổng thể và các chi tiết quan trọng, thậm chí trong trường hợp pixel không khớp hoàn toàn. Reconstruction Loss được định nghĩa như sau:

$$\mathcal{L}_{rec} = \lambda_{\ell1} \|\hat{\mathbf{y}} - \mathbf{y}\|_1 + \lambda_{per} \|\phi(\hat{\mathbf{y}}) - \phi(\mathbf{y})\|_1 \quad (5)$$

Trong đó:

- $\lambda_{\ell1}$  và  $\lambda_{per}$  lần lượt là trọng số (loss weight) của hàm L1 và Perceptual.
- $\phi$  là mạng VGG19 được huấn luyện sẵn. VGG19 thường được sử dụng cho các nhiệm vụ cụ thể như phân loại y tế, ảnh X-ray...



Hình 2: Reconstruction Loss

### 2.5.2 Adversarial Loss

Adversarial Loss được sử dụng để huấn luyện mô hình GAN sinh ảnh có tính chất giống với ảnh tự nhiên. Hàm này giúp generator(mô hình sinh) học cách sinh ảnh tự nhiên bằng cách "đánh lừa" discriminator (mô hình phân loại), có nghĩa là tạo ra ảnh mà discriminator khó phân

bíêt với ảnh thật.

Adversarial Loss được định nghĩa như sau:

$$\mathcal{L}_{\text{adv}} = -\lambda_{\text{adv}} \mathbb{E}_{\hat{\mathbf{y}}} \text{softplus}(D(\hat{\mathbf{y}})), \quad (6)$$

Trong đó:

- $\lambda_{\text{adv}}$  là trọng số của hàm Adversarial.
- $D$  dùng để chỉ discriminator, một thành phần của mạng GAN.  $D(\hat{\mathbf{y}})$  là đầu ra của discriminator khi nhận ảnh được tạo ra từ generator. Nếu  $D(\hat{\mathbf{y}})$  nhỏ: generator được "thưởng" (giảm hàm mất mát, ngược lại nếu  $D(\hat{\mathbf{y}})$  lớn: discriminator đã dễ dàng phát hiện được ảnh tạo ra là giả. Generator sẽ bị "phạt" và hàm mất mát sẽ buộc generator cải thiện).
- $\text{softplus}(x) = \ln(1 + e^x)$  là một biến thể của hàm ReLU. Hàm ReLU khá phổ biến trong huấn luyện các mạng lưới neuron, đơn giản lọc các giá trị  $< 0$ . Nó đảm bảo tính liên tục và không âm, giúp ổn định quá trình huấn luyện mô hình.

### 2.5.3 Facial Component Loss

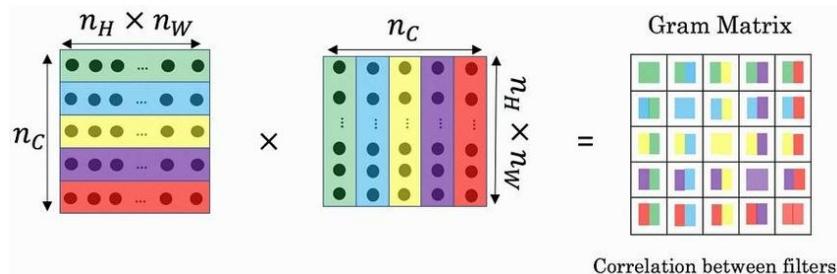
Facial Component Loss tập trung vào các thành phần quan trọng trên khuôn mặt (mắt trái, mắt phải, miệng). Hàm kết hợp giữa Adversarial Loss cho từng vùng và Feature Style Loss nhằm cải thiện chất lượng của khuôn mặt, đặc biệt ở các vùng nêu trên. Điều này giúp giảm các hiện tượng "nhân tạo" trên khuôn mặt của ảnh được sinh ra.

Facial Component Loss được định nghĩa như sau:

$$\mathcal{L}_{\text{comp}} = \sum_{\text{ROI}} \lambda_{\text{local}} \mathbb{E}_{\hat{\mathbf{y}}_{\text{ROI}}} [\log(1 - D_{\text{ROI}}(\hat{\mathbf{y}}_{\text{ROI}}))] + \lambda_{f_s} |\text{Gram}(\psi(\hat{\mathbf{y}}_{\text{ROI}})) - \text{Gram}(\psi(y_{\text{ROI}}))|_1 \quad (7)$$

Trong đó:

- $\lambda_{\text{local}}$  và  $\lambda_{f_s}$  lần lượt là trọng số của Adversarial từng phần và Feature Style Loss.
- $D_{\text{ROI}}$  là discriminator dành riêng cho từng vùng ROI (Region of Interest, cụ thể là mắt trái, mắt phải và miệng)
- Ma trận Gram được sử dụng trong Feature Style Loss với  $\psi$  là các đặc trưng được trích xuất từ discriminator cho từng vùng.

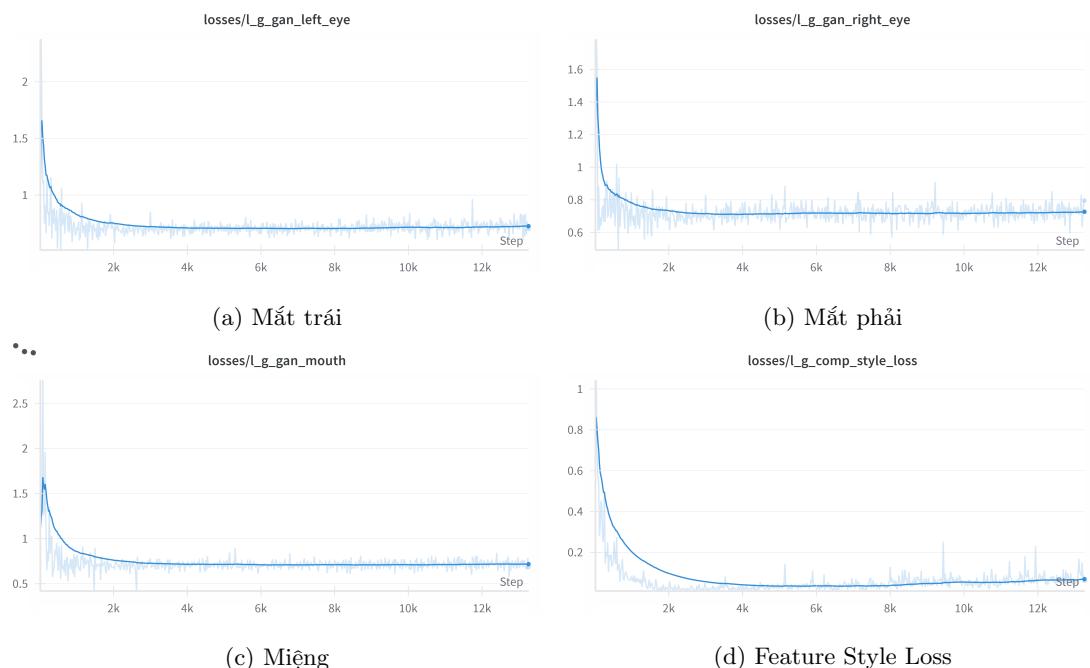


Hình 3: Minh họa về ma trận Gram

Ma trận Gram được xây dựng từ tập hợp các vectơ bằng cách tính tích vô hướng giữa từng cặp vectơ. Ma trận đặc trưng có kích thước  $n_C \times n_H \times n_W$  với  $n_C$  là số kênh,  $n_H$  và  $n_W$  là chiều

cao và chiều rộng của đặc trưng. Trong một lớp tích chập (Convolution Layer), mỗi màu sắc biểu thị cho một lớp kích hoạt của bộ lọc khi nó áp dụng lên ảnh. Nơi bộ lọc phát hiện được các đặc trưng cụ thể như kết cấu hoặc các chi tiết phức tạp trong ảnh.

Trong xử lý ảnh, ma trận Gram biểu diễn sự tương quan giữa các đặc trưng (feature maps) ở các kênh, giúp học được thông tin về phong cách và cấu trúc tổng quát. Ma trận Gram thường được sử dụng trong các bài toán xử lý ảnh như style transfer hay tô màu ảnh.



Hình 4: Facial Component Loss

#### 2.5.4 Identity Preserving Loss

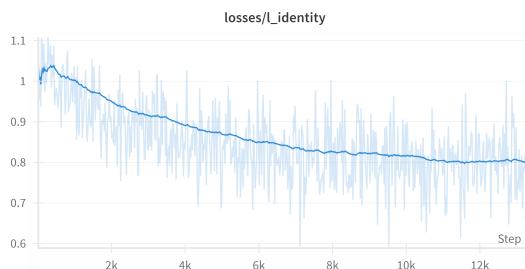
Identity Preserving Loss được sử dụng nhằm đảm bảo rằng các đặc trưng của ảnh sinh ra gần giống với ảnh gốc trong không gian đặc trưng. Vì thế khuôn mặt từ ảnh sinh ra không chỉ chân thực mà còn nhận diện được là cùng một người.

Identity Preserving Loss được định nghĩa như sau:

$$\mathcal{L}_{id} = \lambda_{id} \|\eta(\hat{y}) - \eta(y)\|_1 \quad (8)$$

Trong đó:

- $\lambda_{id}$  là trọng số của Identity Preserving Loss.
- $\eta(\hat{y})$  và  $\eta(y)$  lần lượt là đặc trưng được trích xuất từ ảnh sinh ra và ảnh gốc.
- ArcFace[5] là mô hình được sử dụng trong hàm mắt này - một mạng nhận diện khuôn mặt mạnh mẽ, được huấn luyện sẵn để trích xuất các đặc trưng về danh tính của con người (identity-discriminative features).



Hình 5: Identity Preserving Loss

### 2.5.5 Hàm mất mát tổng thể

Hàm mất mát tổng thể kết hợp những hàm trên:

$$\mathcal{L}_{total} = \mathcal{L}_{rec} + \mathcal{L}_{adv} + \mathcal{L}_{comp} + \mathcal{L}_{id} \quad (9)$$

Sau khi training nhiều lần và điều chỉnh trọng số, nhóm đã chọn ra trọng số như sau:  $\lambda_{l1} = 0.1$ ,  $\lambda_{per} = 1$ ,  $\lambda_{adv} = 0.1$ ,  $\lambda_{local} = 1$ ,  $\lambda_{fs} = 200$ ,  $\lambda_{id} = 10$

### 2.5.6 Real Score và Fake Score

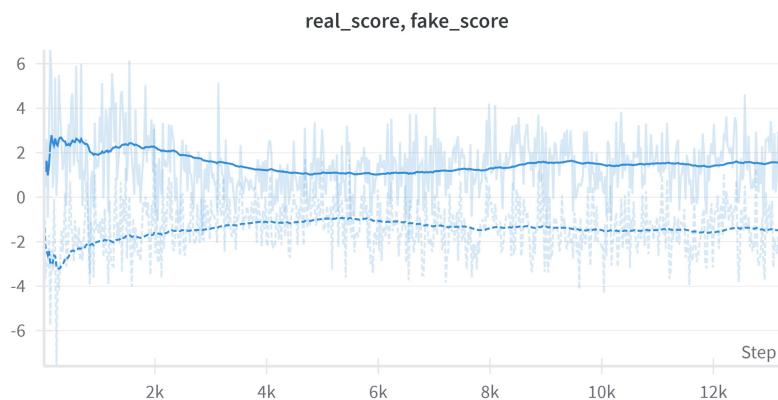
Real Score là điểm mà discriminator gán cho ảnh thật. Điểm này càng cao, discriminator càng chắc chắn rằng ảnh này là "thật".

Fake Score: cho điểm mà discriminator gán cho ảnh tạo ra bởi generator. Điểm này càng thấp, discriminator càng tin rằng ảnh này là "giả".

Cả hai chỉ số này thường được sử dụng để đánh giá quá trình huấn luyện của GAN:

- Nếu real score tăng và fake score giảm dần nhưng ổn định, điều đó có nghĩa là discriminator đang học được cách phân biệt tốt hơn giữa ảnh thật và giả.
- Nếu fake score dần tăng tiệm cận với real score, điều này có thể chỉ ra rằng generator đang cải thiện và tạo ra các mẫu ảnh gần giống với ảnh thật hơn.

Trong mô hình của nhóm, sơ đồ rơi vào trường hợp thứ hai, cho thấy ảnh tạo ra giống với ảnh thật mà generator không phân biệt được.



Hình 6: Real Score và Fake Score

### 3 Thực thi

#### 3.1 Dataset và Hiện thực

**Training Dataset:** Nhóm huấn luyện mô hình sử dụng hình ảnh từ dataset FFHQ. Dữ liệu huấn luyện bao gồm tổng cộng 7000 hình ảnh chân dung chất lượng cao. Các hình ảnh này được sửa về kích thước 512 trong quá trình huấn luyện.

Hình ảnh từ dataset được xử lý, chỉnh sửa để thu được những hình ảnh thực tế, giống với ảnh chân dung chất lượng thấp ngoài đời thực nhất. Để làm được điều đó, nhóm sử dụng mô hình hạ chất lượng ảnh. Mô hình này thực thi các hành động sau:

Dầu tiên, ảnh chất lượng cao được làm mờ bằng Gaussian blur, ngẫu nhiên giữa đồng nhất hoặc dị hướng (isotropic/anisotropic). Tiếp sau đó, mô hình thực hiện phép giảm mẫu (Down-sampling) lên ảnh và nhiều Gaussian noise được thêm ngẫu nhiên vào ảnh. Tiếp theo, mô phỏng các tạp âm gây ra bởi nén JPEG được thêm vào ảnh. Cuối cùng, ảnh được áp dụng một bộ lọc áp dụng hiệu ứng xước. Với xác suất ngẫu nhiên, một số ảnh bị chuyển sang trắng đen, thay đổi độ sáng cho các kênh màu và áp dụng bộ lọc màu sepia.

**Testing Dataset:** Bao gồm tổng cộng 1000 hình ảnh, lấy ngẫu nhiên từ dataset FFHQ và CelebA. Các hình ảnh này hoàn toàn không trùng lặp với hình ảnh trong Training Dataset và cũng được đưa qua mô hình hạ chất lượng ảnh trước khi tiến hành đánh giá.

**Hiện thực:** Nhóm sử dụng mô hình StyleGAN2 đã được huấn luyện sẵn với đầu ra 512x512 làm mô hình tập trung khôi phục khuôn mặt. Hệ số nhân kênh của StyleGAN2 được thiết lập là một để thu gọn kích thước mô hình. UNet dùng cho việc khử chất lượng thấp bao gồm bảy lần giảm mẫu và bảy lần tăng mẫu, mỗi lần đều có một khối dư (residual block). Đối với mỗi lớp CS-SFT, nhóm sử dụng hai lớp tích chập để tạo ra các tham số tương ứng.

Hàm tối ưu của mô hình là hàm Adam (Adam Optimizer). Nhóm tăng cường việc huấn luyện với xoay ngang và điều chỉnh màu sắc. Mô hình tập trung khôi phục vào ba đặc điểm: mắt trái, mắt phải, và miệng vì đây là các thành phần chính của khuôn mặt. Mỗi thành phần được cắt bởi ROI align. Tốc độ học được thiết lập là  $2e^{-3}$ . Mô hình được triển khai bằng PyTorch framework và sử dụng GPU ... để huấn luyện.

Link demo mô hình huấn luyện của nhóm thực hiện trên google colab: [https://colab.research.google.com/drive/1S05xuQGCpWSL6bhKJwpZd0attoqGt\\_QZ?usp=sharing](https://colab.research.google.com/drive/1S05xuQGCpWSL6bhKJwpZd0attoqGt_QZ?usp=sharing)



### 3.2 Dánh giá mô hình

Nhóm quyết định chọn PSNR và SSIM làm hai thông số đánh giá ảnh sinh ra bởi mô hình.

#### 3.2.1 PSNR

PSNR, hay "Peak Signal-to-Noise Ratio" (Tỷ lệ tín hiệu-độ nhiễu tối đa), là một chỉ số đo lường chất lượng của hình ảnh hoặc video được tái tạo so với nguyên bản. Thường được sử dụng trong lĩnh vực xử lý ảnh và video để đánh giá hiệu quả của các thuật toán nén hoặc các kỹ thuật khác ảnh hưởng đến chất lượng hình ảnh.

PSNR được đo bằng đơn vị decibel (dB), càng cao thể hiện chất lượng của ảnh tái tạo càng gần với bản gốc. Cách tính PSNR dựa trên sự khác biệt giữa giá trị cường độ của từng điểm ảnh trong bản gốc và bản tái tạo.

Công thức tính PSNR là:

$$PSNR = 10 \log_{10} \left( \frac{MAX^2}{MSE} \right) \quad (10)$$

Trong đó:

- MAX là giá trị cực đại có thể có của tín hiệu, ví dụ trong ảnh màu sắc được biểu diễn bằng 8-bit mỗi kênh thì giá trị này là 255.
- MSE là "Mean Squared Error" (lỗi bình phương trung bình) giữa hai ảnh (bản gốc và bản tái tạo), tính bằng công thức:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (I(i,j) - K(i,j))^2 \quad (11)$$

Trong đó I là ảnh gốc, K là ảnh tái tạo, m và n là kích thước của ảnh.

Do PSNR chủ yếu thể hiện sự khác nhau giữa màu sắc và độ sáng, nên nó sẽ khó có thể phân biệt được các hình ảnh bị mờ, bị nhiễu. Vì vậy, ta cần thêm thông số SSIM để có thể đánh giá các hình ảnh đó.

#### 3.2.2 SSIM

SSIM, hay "Structural Similarity Index Measure," là một phương pháp đo lường sự tương đồng về cấu trúc giữa hai ảnh được sử dụng để đánh giá chất lượng hình ảnh. Khác với PSNR, SSIM không chỉ xét đến sự khác biệt về độ sáng và màu sắc giữa các điểm ảnh mà còn đánh giá sự tương quan về cấu trúc, độ sáng và độ tương phản giữa các vùng tương ứng của hai ảnh. Do đó, SSIM cung cấp một đánh giá toàn diện hơn về sự giống nhau về mặt thị giác giữa bản gốc và bản tái tạo.

Công thức tính SSIM:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (12)$$

Trong đó:

- x, y là hai cửa sổ ảnh có kích thước tương đồng (thường là 8x8 hoặc 11x11 điểm ảnh) trong ảnh bản gốc và ảnh tái tạo.
- $\mu_x$ ,  $\mu_y$  là giá trị trung bình của x và y.



- $\sigma_x^2, \sigma_y^2$  là phương sai của x và y.
- $\sigma_{xy}$  là độ lệch chuẩn chéo giữa x và y.
- $c1, c2$  là các hằng số nhỏ để tránh chia cho số rất nhỏ, thường được tính từ động cơ động của các giá trị điểm ảnh.

Các thành phần của SSIM:

- Độ sáng (luminance): Dánh giá sự khác biệt về độ sáng trung bình giữa hai cửa sổ ảnh.
- Độ tương phản (contrast): Dánh giá sự khác biệt về độ tương phản, hay biên độ dao động của giá trị điểm ảnh.
- Cấu trúc (structure): Dánh giá sự tương quan về mô hình hoặc cấu trúc của các điểm ảnh trong hai cửa sổ.

SSIM thường dao động từ -1 đến 1, với 1 chỉ ra sự tương đồng hoàn hảo về cấu trúc, độ sáng và độ tương phản giữa hai ảnh, và các giá trị thấp hơn cho thấy sự khác biệt lớn hơn. SSIM được coi là một chỉ số đáng tin cậy để đánh giá chất lượng hình ảnh, đặc biệt trong các ứng dụng mà sự giống nhau về cấu trúc và hình thức là quan trọng.

SSIM	PSNR
Trong khoảng (-1,1), càng gần 1 càng tốt	Càng cao càng tốt, PSNR $\geq 40$ dB sẽ giống thật nhất.

Hình 7: Điểm đánh giá

#### Tiến hành đánh giá:

- PSNR: 24.955
- SSIM: 0.833

### 3.3 So sánh với các mô hình khác

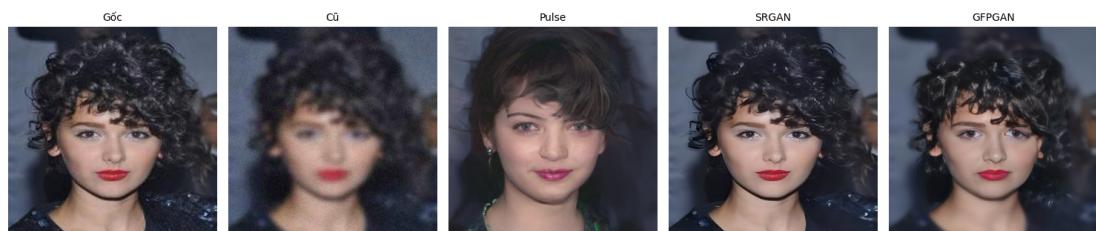
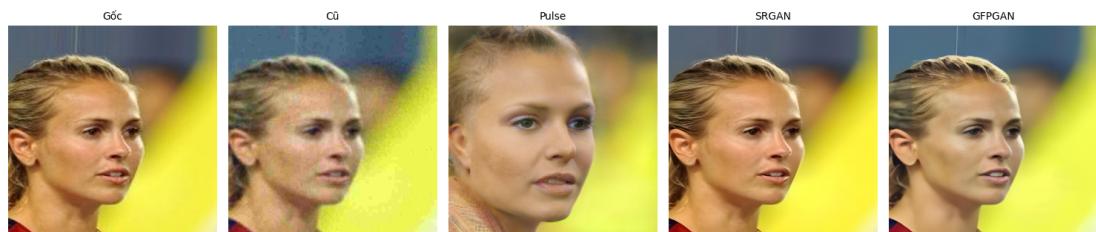
Nhóm sẽ tiến hành so sánh với hai mô hình phục dựng ảnh khác là Pulse và SRGAN. Để so sánh một cách cân bằng nhất thì Testing Dataset của mô hình của nhóm sẽ được sử dụng.

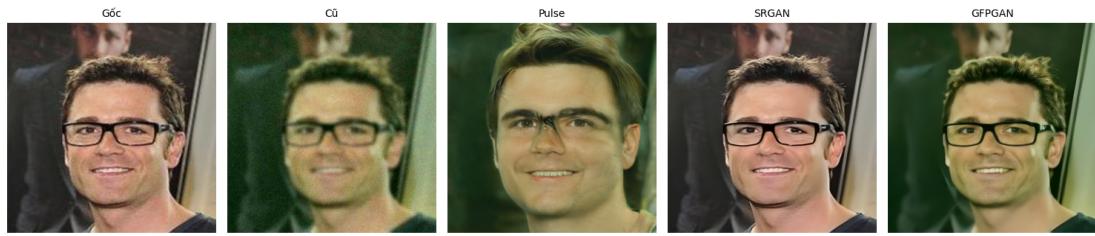
So với hai mô hình SRGAN và Pulse, có thể thấy mô hình GFP-GAN của nhóm cho ra kết quả khá tốt, đặc biệt vượt trội về chỉ số SSIM, chứng tỏ mô hình hoạt động hiệu quả trong việc bảo toàn các đặc điểm cấu trúc của ảnh gốc. Từ góc nhìn của người dùng, ảnh khôi phục của GFP-GAN sẽ có cảm giác phù hợp hơn, tự nhiên hơn.

Tuy nhiên, có thể thấy mô hình hiện tại chưa xử lý tốt các vết xước (scratches) và cũng chưa thể khôi phục màu cho ảnh cũ. Bên cạnh đó, mô hình cũng gặp khó khăn khi xử lý ảnh của những người có đặc điểm khuôn mặt không được đại diện đầy đủ trong dữ liệu huấn luyện, ví dụ như khuôn mặt của các dân tộc ít người. Các thành phần nằm ngoài khuôn mặt hiện vẫn chưa xử lý được hoặc xử lý kém.

Đặc điểm	GFPGAN	PULSE	SRGAN
Dịnh lượng	<b>PSNR: 24.955</b> <b>SSIM: 0.833</b>	PSNR: 21.337 SSIM: 0.46985	PSNR: 21.228 SSIM: 0.571
Dịnh tính	Cải thiện chi tiết khuôn mặt rõ ràng và giữ được đặc điểm nhận diện. Nhanh và hiệu quả trong các ứng dụng thực tế.	Tạo ra hình ảnh có độ phân giải cao và chi tiết tốt. Tuy nhiên, không chính xác tuyệt đối so với ảnh gốc và thời gian xử lý lâu	Cải thiện độ phân giải ảnh tự nhiên, khá chi tiết cho ảnh tổng quát nhưng không chuyên về chi tiết khuôn mặt. Vì vậy, Chi tiết ở các phần phức tạp (như mắt, tóc, miệng) thường không tự nhiên.

Bảng 1: So Sánh đặc điểm của GFPGAN với 2 mô hình khác





Hình 8: So sánh kết quả mô hình GFPGAN với mô hình khác

## 4 Kết luận

Qua đợt án lần này, nhóm đã thực hiện huấn luyện mô hình trí tuệ nhân tạo GFPGAN trong việc phục dựng lại ảnh chân dung cũ và thu về được kết quả khá tốt. Bên cạnh đó, vẫn còn có những hạn chế. Chủ đề này mang tính thực tế khá cao, mang lại nhiều ứng dụng cho cuộc sống.



## Tài liệu tham khảo

- [1] O. Ronneberger, P. Fischer, and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*, in *The International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015.
- [2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative Adversarial Nets*, in *NeurIPS*, 2014.
- [3] T. Karras, S. Laine, and T. Aila, *A Style-Based Generator Architecture for Generative Adversarial Networks*, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [4] X. Wang, K. Yu, C. Dong, and C. C. Loy, *Recovering Realistic Texture in Image Super-Resolution by Deep Spatial Feature Transform*, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [5] Deng, Jiankang and Guo, Jia and Xue, Niannan and Zafeiriou, Stefanos, *Arcface: Additive angular margin loss for deep face recognition*, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, page 4690-4699, 2019.
- [6] X. Wang, Y. Li, H. Zhang, and Y. Shan, *Towards Real-World Blind Face Restoration with Generative Facial Prior*, in *of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
- [7] S. Menon, A. Damian, S. Hu, N. Ravi, and C. Rudin, *PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models*, arXiv preprint arXiv:2003.03808, 2020. Available at: <https://arxiv.org/abs/2003.03808>.
- [8] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, *Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network*, arXiv preprint arXiv:1609.04802, 2017. Available at: <https://arxiv.org/abs/1609.04802>.