

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



**XỬ LÝ ẢNH SỐ VÀ
THỊ GIÁC MÁY TÍNH (CO3057)**
Báo cáo tiểu luận

Đề tài: Siêu phân giải trên ảnh đơn

Sinh viên thực hiện: Phan Trần Minh Đạt - 2111025
Trần Minh Khoa - 2110278

HO CHI MINH CITY, DECEMBER 2024



Mục lục

1	TỔNG QUAN VỀ ĐỀ TÀI	2
1.1	Lý do chọn đề tài	2
1.2	Phạm vi nghiên cứu	2
1.2.1	Đối tượng nghiên cứu	2
1.2.2	Phạm vi nghiên cứu	2
1.3	Phương pháp nghiên cứu	3
1.4	Mục tiêu nghiên cứu	3
1.4.1	Mục tiêu tổng quát	3
1.4.2	Mục tiêu cụ thể	3
2	CƠ SỞ LÝ THUYẾT	4
2.1	Bài toán siêu phân giải	4
2.1.1	Giới thiệu bài toán	4
2.1.2	Các phương pháp đánh giá	5
2.2	Các phương pháp siêu phân giải	6
2.2.1	Phương pháp nội suy	6
2.2.2	Phương pháp học máy	7
2.2.3	Phương pháp học sâu	8
3	HƯỚNG TIẾP CẬN	9
3.1	Super-Resolution (SR) với SRCNN	9
3.2	Super-Resolution (SR) với SRGAN	10
3.2.1	Tổng quan về GAN	10
3.2.2	Các thành phần chính trong SRGAN	11
3.2.3	Hàm mất mát trong SRGAN	12
4	THỰC NGHIỆM	14
4.1	Tập dữ liệu	14
4.2	Thiết lập thí nghiệm	15
4.3	Kết quả	15
4.3.1	Chỉ số PSNR và SSIM	15
4.3.2	Đánh giá trực quan	16
4.3.3	Tốc độ xử lý	16
5	TỔNG KẾT	17
	Tài liệu tham khảo	18

1 TỔNG QUAN VỀ ĐỀ TÀI

1.1 Lý do chọn đề tài

Bài toán siêu phân giải trên ảnh đơn (Image Super-Resolution) là một trong những thách thức quan trọng trong xử lý ảnh số và thị giác máy tính. Việc nâng cao chất lượng hình ảnh không chỉ mang lại lợi ích to lớn trong các ứng dụng thực tiễn như giám sát an ninh, y tế và truyền thông, mà còn góp phần nâng cao trải nghiệm người dùng trong các nền tảng giải trí và truyền thông kỹ thuật số.

Các phương pháp học sâu hiện nay đã tạo ra bước đột phá đáng kể trong việc giải quyết bài toán này. Các mô hình như Super-Resolution Convolutional Neural Network (SRCNN) và Generative Adversarial Networks (SRGAN) đã chứng minh khả năng vượt trội trong việc tái tạo chi tiết hình ảnh. Tuy nhiên, mỗi phương pháp lại có những điểm mạnh và hạn chế riêng, từ hiệu quả phục hồi chi tiết đến thời gian huấn luyện và triển khai.

Bằng cách hiện thực và so sánh các phương pháp Super-Resolution từ cơ bản như Interpolation đến học sâu như SRCNN và SRGAN trên tập dữ liệu CelebA-HQ, đề tài nhằm cung cấp một cái nhìn toàn diện về hiệu quả và tiềm năng của từng phương pháp.

1.2 Phạm vi nghiên cứu

1.2.1 Đối tượng nghiên cứu

Đối tượng nghiên cứu của tiểu luận bao gồm:

- Đề tài tập trung vào các mô hình và thuật toán Super-Resolution bao gồm: Nội suy, SRCNN, SRGAN
- Dataset CelebA-HQ: Một tập dữ liệu hình ảnh chất lượng cao của các khuôn mặt nổi bật, phù hợp cho bài toán Super-Resolution.

1.2.2 Phạm vi nghiên cứu

- Tập trung vào tăng độ phân giải cho ảnh khuôn mặt trên tập dữ liệu CelebA-HQ.
- So sánh các phương pháp dựa trên các tiêu chí:
 - Chỉ số Peak Signal-to-Noise Ratio (PSNR).

- Structural Similarity Index Measure (SSIM).
- Đánh giá chủ quan về chất lượng hình ảnh.

1.3 Phương pháp nghiên cứu

Phương pháp nghiên cứu được sử dụng trong đề tài được thực hiện qua các bước sau:

- Lên ý tưởng thực hiện đề tài
- Tham khảo các tài liệu liên quan
- Tái hiện thực thuật toán được sử dụng trong tài liệu tham khảo
- Đánh giá, so sánh và nhận xét kết quả đạt được

1.4 Mục tiêu nghiên cứu

1.4.1 Mục tiêu tổng quát

Hiện thực và so sánh các phương pháp Super-Resolution từ cơ bản đến nâng cao nhằm đánh giá hiệu quả và tiềm năng ứng dụng của từng phương pháp.

1.4.2 Mục tiêu cụ thể

- Nghiên cứu và tổng hợp lý thuyết về các phương pháp Interpolation, SRCNN và SRGAN.
- Thực nghiệm:
 - Huấn luyện các mô hình trên tập dữ liệu CelebA-HQ.
 - So sánh chất lượng hình ảnh phục hồi qua các chỉ số PSNR, SSIM.
- Đánh giá và phân tích:
 - Đánh giá ưu điểm và hạn chế của từng phương pháp.
 - Đề xuất cải tiến và khả năng ứng dụng vào các lĩnh vực thực tiễn.

2 CƠ SỞ LÝ THUYẾT

2.1 Bài toán siêu phân giải

2.1.1 Giới thiệu bài toán

Bài toán siêu phân giải (*Super-Resolution* - SR) là một vấn đề trong xử lý ảnh và thị giác máy tính, nhằm tái tạo một hình ảnh có độ phân giải cao (*High-Resolution* - HR) từ một hình ảnh có độ phân giải thấp (*Low-Resolution* - LR). Đây là một bài toán bất định (*ill-posed problem*), vì có vô số ảnh HR có thể phù hợp với ảnh LR đã cho.

Để biểu diễn dưới dạng toán học, bài toán siêu phân giải có thể được phát biểu như sau: Cho một hình ảnh đầu vào $I_{LR} \in \mathbb{R}^{h \times w}$, trong đó h và w lần lượt là chiều cao và chiều rộng của ảnh LR, mục tiêu là tìm một ảnh $I_{HR} \in \mathbb{R}^{s \cdot h \times s \cdot w}$ (với s là hệ số phóng đại) thỏa mãn:

$$I_{LR} = D(I_{HR}) + n$$

Trong đó:

- I_{LR} : Hình ảnh đầu vào có độ phân giải thấp.
- I_{HR} : Hình ảnh đầu ra có độ phân giải cao.
- $D(\cdot)$: Hàm ánh xạ thể hiện quá trình xuống mẫu (*downsampling*), bao gồm các bước như làm mờ (*blurring*) và giảm kích thước.
- n : Nhiễu (*noise*) trong hình ảnh.

Mục tiêu của bài toán là xây dựng một hàm ánh xạ $F(\cdot)$ để tái tạo I_{HR} từ I_{LR} , sao cho:

$$I_{HR} \approx F(I_{LR}; \theta)$$

Trong đó θ là tập các tham số của mô hình tái tạo $F(\cdot)$.

Các phương pháp tiếp cận bài toán SR

Để giải bài toán này, có nhiều cách tiếp cận khác nhau, bao gồm:

1. **Phương pháp nội suy (*Interpolation*)**: Sử dụng các kỹ thuật nội suy để tăng kích thước ảnh. Đây là cách tiếp cận trực tiếp nhưng không thể khôi phục chi tiết phức tạp.

2. **Phương pháp dựa trên học máy (*Machine Learning*):** Sử dụng các kỹ thuật học máy truyền thống, được nghiên cứu từ lâu trước khi các phương pháp học sâu trở nên phổ biến.
3. **Phương pháp học sâu (*Deep Learning*):** Sử dụng các mạng nơ-ron sâu, như SRCNN, VDSR, hoặc GAN, để học ánh xạ $F(\cdot)$ từ các cặp dữ liệu LR-HR.

2.1.2 Các phương pháp đánh giá

Để đánh giá hiệu quả của các phương pháp siêu phân giải, cần sử dụng các tiêu chí định lượng và định tính nhằm so sánh chất lượng hình ảnh được tái tạo. Các phương pháp đánh giá phổ biến bao gồm:

1. Đánh giá định lượng:

- **PSNR (*Peak Signal-to-Noise Ratio*):** Chỉ số này đo độ tương đồng giữa hình ảnh tái tạo và hình ảnh gốc. Giá trị PSNR cao hơn biểu thị chất lượng tái tạo tốt hơn. PSNR được tính như sau:

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

Trong đó:

- MAX: Giá trị cực đại của pixel (thường là 255 đối với ảnh 8-bit).
- MSE: Sai số bình phương trung bình giữa ảnh tái tạo và ảnh gốc, được tính bởi:

$$\text{MSE} = \frac{1}{h \cdot w} \sum_{i=1}^h \sum_{j=1}^w \left(I_{HR}(i, j) - \hat{I}_{HR}(i, j) \right)^2$$

Trong đó \hat{I}_{HR} là ảnh tái tạo.

- **SSIM (*Structural Similarity Index*):** Chỉ số này đo độ tương đồng về cấu trúc giữa ảnh tái tạo và ảnh gốc. SSIM được định nghĩa trong khoảng từ 0 đến 1, với giá trị gần 1 biểu thị sự tương đồng cao hơn. Công thức SSIM:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

Trong đó:

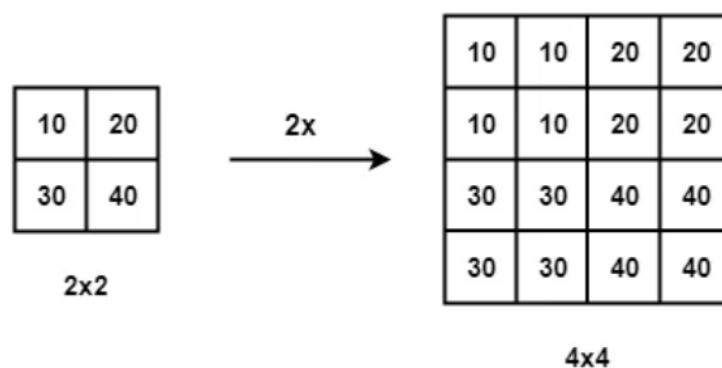
- μ_x, μ_y : Giá trị trung bình của ảnh x và y .
- σ_x^2, σ_y^2 : Phương sai của x và y .
- σ_{xy} : Hiệp phương sai giữa x và y .
- C_1, C_2 : Hằng số để tránh chia cho 0, thường được đặt là $C_1 = (k_1 \cdot L)^2$ và $C_2 = (k_2 \cdot L)^2$, với L là giá trị cực đại của pixel, k_1 và k_2 là các hằng số nhỏ (thường là 0.01 và 0.03).

2.2 Các phương pháp siêu phân giải

2.2.1 Phương pháp nội suy

Một trong những cách đơn giản nhất để thực hiện tăng cường độ phân giải của ảnh (Super Resolution) là sử dụng phương pháp nội suy (interpolation). Nội suy là cách trực tiếp để tăng kích thước của ảnh bằng cách thêm các điểm ảnh mới (pixel) hoặc các điểm dữ liệu vào ảnh có độ phân giải thấp. Có 3 kỹ thuật nội suy phổ biến, bao gồm:

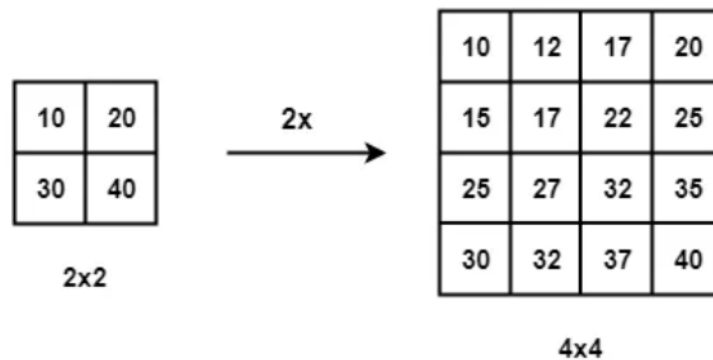
1. **Nội suy gần nhất (Nearest Neighbour Interpolation)**: Đây là phương pháp đơn giản và nhanh chóng nhất vì yêu cầu ít tính toán. Các điểm ảnh mới được thêm vào dựa trên giá trị cường độ (intensity value) của điểm ảnh gần nhất. Mặc dù kết quả tạo ra hình ảnh có độ phân giải cao hơn, nhưng hình ảnh có thể bị vỡ hạt (blocky) và trông không tự nhiên.



Hình 1: Ví dụ nội suy gần nhất với hệ số tăng là 2, áp dụng cho mảng 2x2.

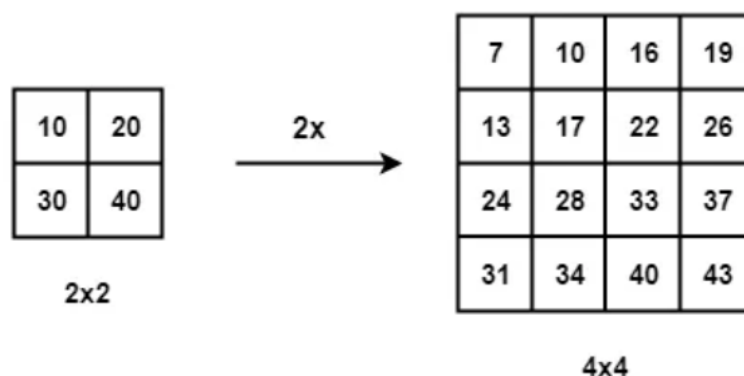
2. **Nội suy song tuyến tính (Bilinear Interpolation)**: Phương pháp này sử dụng nội suy tuyến tính. Giá trị trung bình có trọng số của cường độ từ 4 điểm ảnh lân cận gần nhất được tính để tạo ra giá trị cường độ cho điểm

ảnh mới. Kỹ thuật này cho kết quả mượt mà hơn so với nội suy gần nhất, nhưng không lý tưởng với các cạnh sắc nét (sharp edges).



Hình 2: Ví dụ nội suy song tuyến tính với hệ số tăng là 2, áp dụng cho mảng 2×2 .

3. **Nội suy bicubic (Bicubic Interpolation):** Nội suy bicubic cũng sử dụng giá trị trung bình có trọng số của các điểm lân cận để tạo ra kết quả. Tuy nhiên, khác với nội suy song tuyến tính, phương pháp này sử dụng 16 điểm lân cận (trong vùng 4×4) và áp dụng cách tính phân phối trọng số khác. Phương pháp này thường tạo ra hình ảnh sắc nét hơn so với hai phương pháp trên và thường được sử dụng trong các phần mềm chỉnh sửa ảnh.



Hình 3: Ví dụ nội suy bicubic với hệ số tăng là 2, áp dụng cho mảng 2×2 .

2.2.2 Phương pháp học máy

Mặc dù các phương pháp học sâu hiện nay đã chiếm ưu thế, học máy truyền thống vẫn đóng một vai trò quan trọng trong quá trình nghiên cứu và ứng dụng SR, đặc

biệt là trong các tình huống có dữ liệu hạn chế hoặc khi các mô hình phức tạp không thể triển khai được. Dưới đây là các giải pháp học máy truyền thống thường được sử dụng trong Super Resolution:

1. **Mã hóa thưa (Sparse Coding):** Mã hóa thưa tìm cách biểu diễn hình ảnh hoặc các patch (mảnh nhỏ của hình ảnh) bằng một tổ hợp tuyến tính của các cơ sở cơ bản (basis functions), sao cho số lượng các cơ sở cơ bản được sử dụng là tối thiểu. Mã hóa thưa cho phép phục hồi chi tiết bằng cách tận dụng các đặc trưng đã học từ một tập dữ liệu lớn, nhưng phương pháp này có thể bị hạn chế về hiệu suất khi đối mặt với các tình huống phức tạp hoặc khi hình ảnh có nhiễu.
2. **Rừng ngẫu nhiên (Random Forests):** Rừng ngẫu nhiên tạo ra nhiều cây quyết định (decision trees) và kết hợp kết quả của tất cả các cây để đưa ra dự đoán cuối cùng. Một mô hình Random Forest cho SR sẽ được huấn luyện với các cặp hình ảnh LR và HR, nơi mỗi cây quyết định trong rừng học cách kết nối các đặc trưng của hình ảnh LR với hình ảnh HR tương ứng.

2.2.3 Phương pháp học sâu

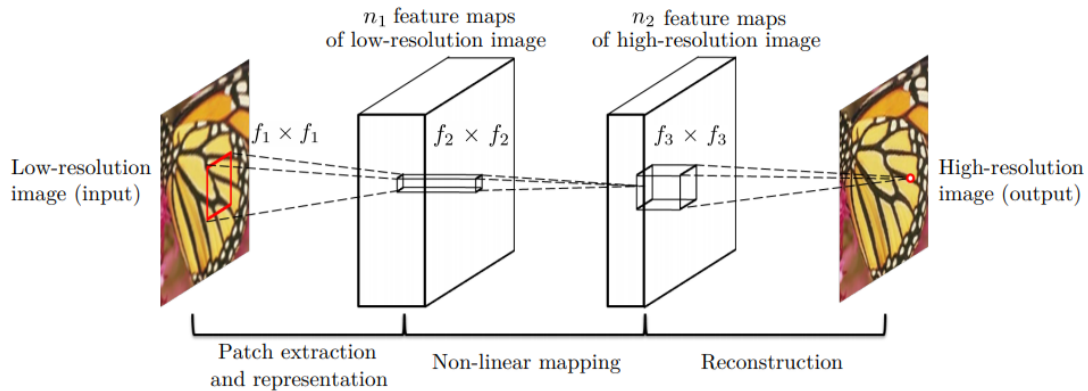
Học sâu đã mang lại những bước đột phá lớn trong Super Resolution (SR) với khả năng học tự động các đặc trưng từ dữ liệu hình ảnh. Một trong những mô hình học sâu đầu tiên cho bài toán SR được gọi là **SRCNN (Super-Resolution Convolutional Neural Network)**, được giới thiệu bởi *Dong et al.* vào năm 2014, và **SRGAN (Super-Resolution Generative Adversarial Network)**, được giới thiệu bởi *Ledig et al.* vào năm 2017. Đây cũng là hướng tiếp cận mà nhóm đã lựa chọn để hiện phân tích và hiện thực trong tiểu luận.

3 HƯỚNG TIẾP CẬN

3.1 Super-Resolution (SR) với SRCNN

SRCNN (Super-Resolution Convolutional Neural Network) được giới thiệu bởi *Chao Dong et al.* vào năm 2014, là một trong những phương pháp đầu tiên áp dụng mạng nơ-ron tích chập (CNN) vào bài toán siêu phân giải. SRCNN bao gồm ba bước chính:

1. **Interpolated Input:** Hình ảnh LR đầu vào được nội suy (bằng Bicubic Interpolation) để tăng kích thước về mức HR mong muốn.
2. **Feature Extraction:** Một tầng tích chập đầu tiên trích xuất các đặc trưng từ hình ảnh đã nội suy.
3. **Non-Linear Mapping:** Một tầng tích chập thứ hai ánh xạ các đặc trưng vừa trích xuất thành các đặc trưng cấp cao hơn.
4. **Reconstruction:** Một tầng tích chập cuối cùng tái tạo hình ảnh HR từ các đặc trưng cấp cao.



Hình 4: Kiến trúc của SRCNN

Hàm mất mát của SRCNN là hàm MSE (*Mean Squared Error*), được định nghĩa như sau:

$$\mathcal{L}_{MSE} = \frac{1}{N} \sum_{i=1}^N \|F(Y_i) - X_i\|^2$$

Trong đó:

- Y_i : Hình ảnh LR đầu vào.

- X_i : Hình ảnh HR gốc.
- $F(Y_i)$: Hình ảnh HR được tái tạo.

SRCNN đơn giản nhưng hiệu quả, đánh dấu việc mở ra một hướng đi mới (hướng tiếp cận học sâu) cho bài toán siêu phân giải. Tuy nhiên, SRCNN có hạn chế trong việc tái tạo các chi tiết phức tạp và tính chân thực của hình ảnh.

3.2 Super-Resolution (SR) với SRGAN

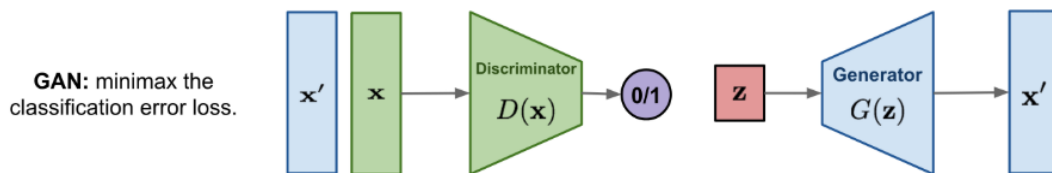
SRGAN, được giới thiệu bởi *Ledig et al.* vào năm 2017, là một bước tiến quan trọng trong Super Resolution, sử dụng Generative Adversarial Networks (GANs) để tạo ra hình ảnh HR có tính thực tế cao, vượt xa các phương pháp truyền thống và cả SRCNN.

3.2.1 Tổng quan về GAN

GAN (Generative Adversarial Network) là một phương pháp học sâu sáng tạo được đề xuất bởi *Ian Goodfellow et al.* vào năm 2014. GAN sử dụng hai mạng nơ-ron nhân tạo đối kháng nhau — **Generator** và **Discriminator** — để sinh ra dữ liệu giả (như hình ảnh, âm thanh) trông giống với dữ liệu thật. Các thành phần và cơ chế của GAN như sau:

- **Generator (G):**
 - **Nhiệm vụ:** Sinh ra dữ liệu giả $G(z)$ từ một phân phối ngẫu nhiên z (thường là Gaussian hoặc Uniform).
 - **Mục tiêu:** Làm cho dữ liệu giả trông giống dữ liệu thật đến mức khó phân biệt.
 - **Hàm mất mát:** Khuyến khích $G(z)$ đánh lừa được Discriminator.
- **Discriminator (D):**
 - **Nhiệm vụ:** Phân biệt dữ liệu thật và dữ liệu giả do G sinh ra.
 - **Đầu ra:** Giá trị xác suất $D(x)$ cho biết dữ liệu là thật (1) hay giả (0).
 - **Mục tiêu:** Phát hiện chính xác dữ liệu thật và giả.
- **Cơ chế học đối kháng (Adversarial Learning):**
 - Discriminator cố gắng cải thiện khả năng phân biệt thật/giả.
 - Generator cố gắng tạo dữ liệu ngày càng giống thật hơn để đánh lừa Discriminator.

- Quá trình này tạo thành một trò chơi đối kháng, nơi cả hai mạng học từ nhau và cùng tiến hóa.



Hình 5: Kiến trúc tổng quát của GAN

Hàm mất mát tổng quát của GAN:

$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

3.2.2 Các thành phần chính trong SRGAN

1. Tầng trích xuất đặc trưng (Feature Extractor Layer):

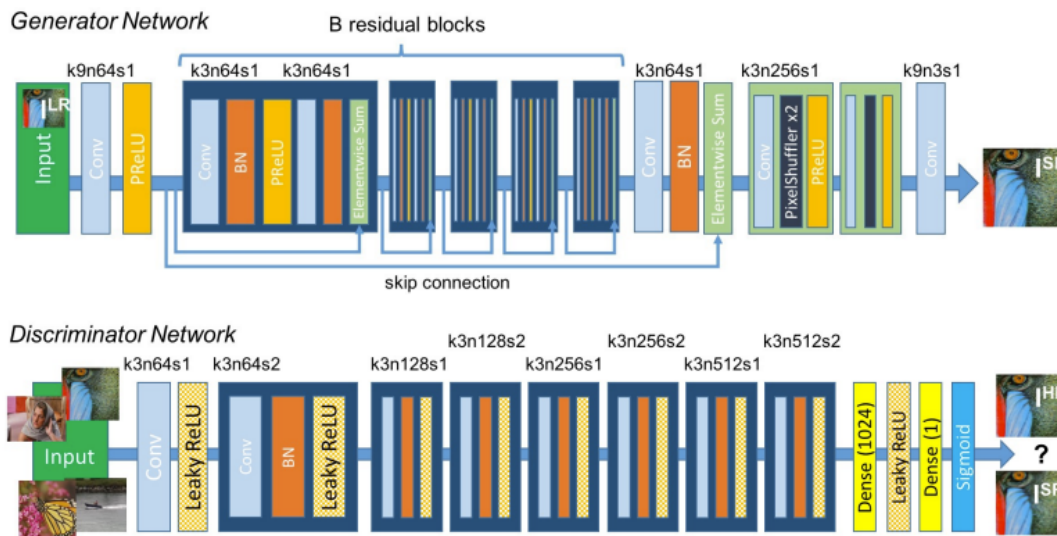
- **Vai trò:** Trích xuất các đặc trưng quan trọng từ hình ảnh LR đầu vào.
- **Cách hoạt động:** Một lớp tích chập (convolutional layer) được áp dụng với các bộ lọc để trích xuất thông tin cốt lõi từ hình ảnh LR.
- **Kết quả:** Tạo ra các đặc trưng ban đầu để đưa vào các tầng tiếp theo nhằm học thêm chi tiết.

2. Khối dư (Residual Blocks):

- **Vai trò:** Tăng cường khả năng học các đặc trưng cấp cao và tránh mất mát thông tin trong quá trình huấn luyện.
- **Cách hoạt động:** Được thiết kế dựa trên **ResNet**, bao gồm:
 - Hai lớp tích chập 3×3 với số lượng bộ lọc cố định.
 - Hàm kích hoạt ReLU được sử dụng sau mỗi tầng tích chập để tăng tính phi tuyến.
 - Batch Normalization để ổn định quá trình huấn luyện.
 - Một đường tắt (skip connection) cộng trực tiếp đầu vào với đầu ra của khối dư, giúp giảm gradient vanishing và tăng hiệu quả học.
- **Kết quả:** Bảo toàn thông tin gốc từ hình ảnh LR trong khi học thêm các đặc trưng phức tạp, như kết cấu hoặc biên.

3. Tầng tăng độ phân giải (Upsampling Layers):

- **Vai trò:** Tăng kích thước không gian (spatial resolution) của các đặc trưng để tạo ra hình ảnh HR.
- **Cách hoạt động:**
 - **Pixel Shuffle:** Một kỹ thuật thay thế cho các phương pháp truyền thống như deconvolution (transposed convolution) hoặc nội suy. Pixel Shuffle tái sắp xếp các đặc trưng từ không gian kênh (channel) sang không gian hình ảnh (spatial).
 - Một số tầng tăng độ phân giải được áp dụng liên tiếp để đạt được tỷ lệ SR mong muốn, chẳng hạn tăng kích thước từ $2\times$ đến $4\times$.
 - **Hàm kích hoạt:** Hàm ReLU thường được sử dụng trong các tầng này để duy trì tính phi tuyến.
- **Kết quả:** Tạo ra một hình ảnh HR với kích thước mong muốn.



Hình 6: Kiến trúc của Generator và Discriminator của SRGAN

3.2.3 Hàm mất mát trong SRGAN

Generator G sử dụng một **hàm mất mát phức hợp** để hướng dẫn Generator trong quá trình tái tạo hình ảnh, gồm ba thành phần chính:

1. **Adversarial Loss:** Khuyến khích Generator tạo ra hình ảnh HR giống thật đến mức Discriminator không phân biệt được giữa HR giả và HR thật.

$$\mathcal{L}_{GAN} = \mathbb{E}[\log D(y)] + \mathbb{E}[\log(1 - D(G(x)))]$$

Trong đó:

- y : Hình ảnh HR thật.
- $G(x)$: Hình ảnh HR giả do Generator tạo ra.
- D : Discriminator.

2. **Perceptual Loss:** Tập trung vào các đặc trưng cao cấp của hình ảnh (như kết cấu và đường biên), thay vì chỉ tối ưu hóa các giá trị pixel, bằng cách sử dụng mạng VGG-19 đã được huấn luyện trước để trích xuất các đặc trưng cấp cao.

$$\mathcal{L}_{VGG} = \frac{1}{WHC} \sum_{i=1}^W \sum_{j=1}^H \sum_{k=1}^C \|\phi(y)_{ijk} - \phi(G(x))_{ijk}\|_2^2$$

Trong đó:

- ϕ : Biểu diễn đặc trưng từ mạng VGG.
- W, H, C : Kích thước không gian và số kênh của đặc trưng.

3. **Content Loss:** Tối ưu hóa dựa trên sai số pixel giữa hình ảnh HR thật và HR giả, tương tự như các mô hình truyền thống.

$$\mathcal{L}_{content} = \|y - G(x)\|_2^2$$

4 THỰC NGHIỆM

4.1 Tập dữ liệu

Để tiến hành thí nghiệm, nhóm sử dụng tập dữ liệu **CelebA-HQ**, một phiên bản chất lượng cao (high-quality) của tập dữ liệu gốc **CelebA (Large-scale CelebFaces Attributes Dataset)**. CelebA-HQ chứa 30.000 hình ảnh gương mặt người nổi tiếng có chất lượng cao với độ phân giải 1024×1024 pixels. Đối với thí nghiệm bên dưới, nhóm sử dụng tập dữ liệu tương tự nhưng chỉ với độ phân giải gốc là 256×256 để rút ngắn thời gian huấn luyện.

CelebA-HQ thường được sử dụng trong các nghiên cứu về tạo và tái tạo hình ảnh, chẳng hạn như trong các bài toán huấn luyện mạng GAN hoặc mô hình học sâu siêu phân giải.



Hình 7: Tập dữ liệu CelebA-HQ

4.2 Thiết lập thí nghiệm

- **Tập dữ liệu:** Tập dữ liệu CelebA-HQ gồm 30.000 cặp ảnh với kích thước gốc 256×256 và kích thước LR 64×64 được downsampling từ ảnh gốc.
- **Mục tiêu thí nghiệm:** Các giải pháp nội suy (bicubic, bilinear, nearest neighbor), SRCNN với kiến trúc 3 lớp CNN cơ bản và SRGAN được mô tả ở phần trước.
- **Chỉ số đánh giá:**
 1. **PSNR (Peak Signal-to-Noise Ratio):** Đánh giá độ chính xác theo tín hiệu hình ảnh.
 2. **SSIM (Structural Similarity Index Measure):** Đánh giá sự tương đồng về cấu trúc giữa hình ảnh tái tạo và hình ảnh gốc.
 3. **Chất lượng trực quan:** Đánh giá trực quan bởi người quan sát.
 4. **Tốc độ xử lý (Inference time):** Thời gian mô hình sau huấn luyện sinh ra ảnh HR từ ảnh LR.
- **Hiện thực:** [Kaggle Notebook](#)

4.3 Kết quả

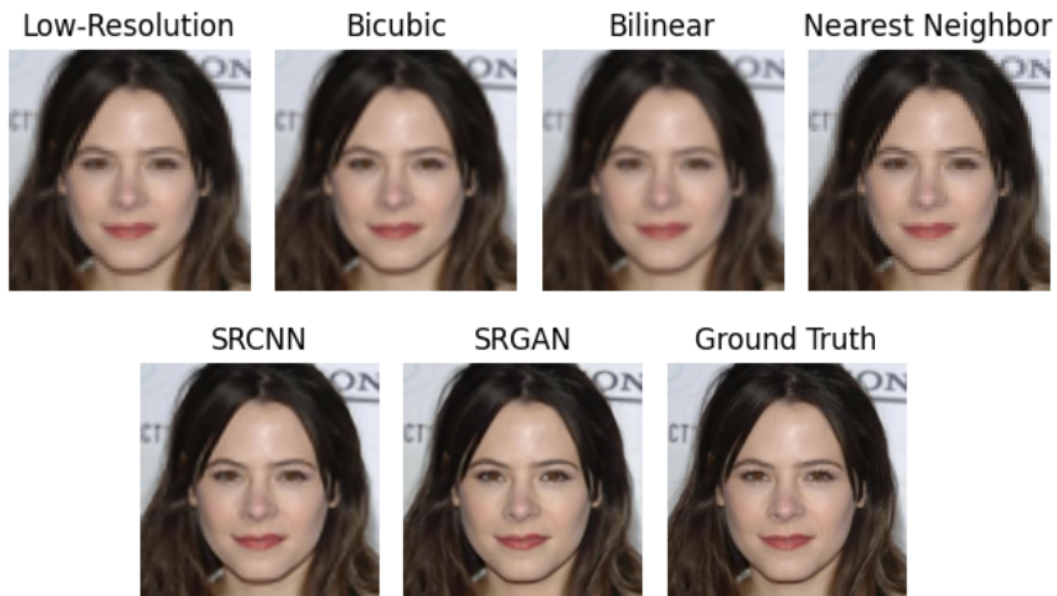
4.3.1 Chỉ số PSNR và SSIM

Giải pháp	PSNR (dB)	SSIM
Nội suy (Bicubic)	28.188	0.851
Nội suy (Bilinear)	28.319	0.833
Nội suy (Nearest)	26.722	0.777
SRCNN	30.405	0.872
SRGAN	31.613	0.893

Bảng 1: Kết quả thí nghiệm với chỉ số PSNR và SSIM

Nhận xét: SRGAN đạt điểm cao hơn ở cả hai chỉ số so với tất cả các giải pháp khác, thể hiện khả năng tái tạo hình ảnh gần hơn với hình ảnh gốc, đồng thời bảo toàn cấu trúc và các đặc trưng trong hình ảnh tốt hơn.

4.3.2 Đánh giá trực quan



Hình 8: Kết quả thí nghiệm đánh giá trực quan

Nhận xét: SRCNN tái tạo hình ảnh với các chi tiết khá rõ ràng nhưng vẫn có xu hướng làm mờ các vùng biên (edge) và giảm độ sắc nét của các đặc trưng phức tạp, như tóc và mắt. SRGAN tạo ra hình ảnh có độ sắc nét và chân thực cao hơn, đặc biệt là ở các khu vực chứa texture phức tạp. Các giải pháp nội suy tỏ ra kém chất lượng hẳn so với giải pháp học sâu.

4.3.3 Tốc độ xử lý

Giải pháp	Thời gian xử lý (s/ảnh)
Nội suy (Bicubic)	0.0001
Nội suy (Bilinear)	0.0001
Nội suy (Nearest)	0.0001
SRCNN	0.0025
SRGAN	0.0095

Bảng 2: Kết quả thí nghiệm với tốc độ xử lý trên GPU NVIDIA Tesla P100 16GB

Nhận xét: SRCNN và các giải pháp nội suy có thời gian xử lý nhanh hơn SRGAN do kiến trúc đơn giản.

5 TỔNG KẾT

Trong bài báo cáo này, nhóm đã nghiên cứu về bài toán siêu phân giải nói chung và so sánh hai phương pháp phổ biến trong bài toán siêu phân giải hình ảnh là SRCNN và SRGAN nói riêng. Các thực nghiệm được tiến hành trên tập dữ liệu CelebA-HQ với các hình ảnh có độ phân giải cao, nhằm đánh giá khả năng tái tạo hình ảnh từ đầu vào có độ phân giải thấp.

Kết quả thực nghiệm cho thấy SRCNN là một mô hình đơn giản, đạt hiệu suất tốt với chỉ số PSNR và SSIM cao, phù hợp cho các ứng dụng yêu cầu tốc độ xử lý nhanh. Tuy nhiên, SRGAN vượt trội hơn về chất lượng trực quan nhờ vào việc kết hợp mạng sinh và mạng phân biệt, đặc biệt trong việc tái tạo các chi tiết phức tạp như da và tóc.

Mặc dù SRGAN đòi hỏi nhiều tài nguyên tính toán hơn, nhưng nó mang lại kết quả sắc nét và tự nhiên hơn, khiến nó trở thành lựa chọn ưu tiên cho các ứng dụng yêu cầu hình ảnh có chất lượng cao. Tổng thể, mỗi phương pháp đều có ưu điểm riêng và phù hợp với các bài toán khác nhau trong lĩnh vực siêu phân giải hình ảnh.



Tài liệu tham khảo

- [1] Dong, C., Loy, C., He, K. & Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions On Pattern Analysis And Machine Intelligence*. **38**, 295-307 (2016)
- [2] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. & Shi, W. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *2017 IEEE Conference On Computer Vision And Pattern Recognition (CVPR)*. pp. 105-114 (2017)