CẢI THIỆN ĐỘ SIÊU PHÂN GIẢI ẢNH, SỬ DỤNG GAN KẾT HỢP VGG19 VÀ RESNET

IMPROVING SUPER-RESOLUTION OF IMAGE USING GAN COMBINE VGG19 AND RESNET

Lương Thị Thảo Hiếu

Khoa Công nghệ thông tin, Trường Đại học Kinh tế - Kỹ thuật Công nghiệp

Đến Tòa soạn ngày 20/4/2025, chấp nhận đăng ngày 19/5/2025

Tóm tắt: Trong bài báo này trình bày phương pháp cải thiện độ phân giải hình ảnh sử dụng mô hình GAN cải tiến. Mô hình cải tiến được đề xuất bằng cách kết hợp VGG19 vào mạng phân biệt (Discriminator - D) để tăng khả năng trích xuất đặc trưng và sử dụng ResNet trong mạng sinh (Generator - G) nhằm giảm thiểu sự thay đổi đột ngột trong quá trình huấn luyện, giúp mô hình ổn định, đồng thời cải thiện khả năng tái tạo ảnh chi tiết. Những thay đổi này giúp mô hình đề xuất có khả năng khôi phục hình ảnh chính xác tự nhiên, giảm hiện tượng mờ và biến dạng thường thấy trong các phương pháp trước đây. Kết quả thực nghiệm cho thấy mô hình đề xuất học ổn định, hình ảnh được tạo ra có chất lượng vượt trội so với mô hình GAN gốc.

Từ khóa: Mạng sinh đối kháng, mô hình GAN, thị giác máy tính, VGG19, RestNet.

Abstract: This paper presents a method to enhance image resolution using an improved GAN model. The proposed improvement involves integrating VGG19 into the Discriminator (D) to increase feature extraction capabilities, and employing ResNet in the Generator (G) to minimize changes during training, thereby enabling faster and more stable convergence while improving the model's ability to reconstruct detailed images. These improvements enable improved model to restore image more accurately and naturally, while reducing the blurring and distortion commonly seen in previous methods. Experimental results show that the proposed model not only improves image resolution but also produces images with superior quality compared to previous method.

Keywords: Generative Adversarial Networks, Computer Vision, VGG19, RestNet.

1. GIỚI THIỆU

Mạng đối kháng sinh (Generative Adversarial Network-GAN) là mạng nơron nhân tạo được giới thiệu vào năm 2014 bởi Ian GoodFellow. Mô hình GAN gồm hai mạng nơron đối kháng: Mạng sinh (Generator - G): Tạo ra mẫu giả dựa trên phân phối dữ liệu và mạng phân biệt (Discriminator - D): Phân biệt giữa dữ liệu thực từ tập huấn luyện và dữ liệu giả do G tạo ra [1]. Do thành công của mô hình, số lượng nghiên cứu liên quan đến GAN tăng theo cấp số nhân. Những nghiên cứu này tập trung vào các khía cạnh khác nhau, từ tối ưu hóa quá trình huấn luyện [2] đến áp dụng GAN vào các lĩnh vực như: tái tạo ngôn ngữ, tái tạo hình ảnh, video [3].

Những mô hình mạng nơron sâu như ResNet [4] và VGG19[5] đã chứng minh hiệu quả cao trong xử lý hình ảnh. ResNet giúp cải thiện quá trình huấn luyện nhờ cơ chế residual learning, trong khi VGG19 với kiến trúc sâu có thể trích xuất đặc trưng hiệu quả. Do đó, kết hợp GAN với ResNet và VGG19 giúp cải thiện chất lượng hình ảnh tổng hợp.

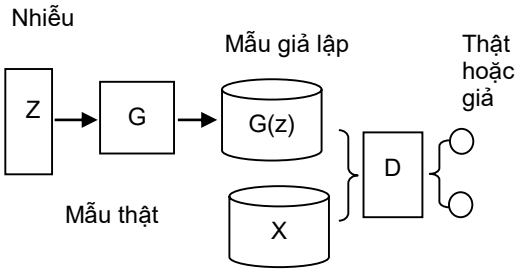
Trong bài báo này, chúng tôi đề xuất một cách tiếp cận mới nhằm cải thiện hiệu suất của GAN bằng cách kết hợp ResNet vào kiến trúc của mạng sinh (G) và VGG19 vào mạng phân biệt (D). Kết quả thực nghiệm cho thấy mô hình đề xuất học ổn định và mang lại chất lượng hình ảnh tốt.

2. CÁC NGHIÊN CỨU LIÊN QUAN

Trong những năm gần đây nhiều nghiên cứu tập trung giải quyết các vấn đề bất ổn mà GAN gặp phải [6]. Kết quả cho thấy một số nghiên cứu mới cố gắng kiểm soát quá trình huấn luyện GAN bằng cách tinh chỉnh siêu tham số. Nghiên cứu [7] tập trung vào việc phân tích các mô hình GAN tiên tiến và đánh giá hiệu suất. Một số nghiên cứu khác tập trung vào nghiên cứu kiến trúc của các mô hình GAN, áp dụng trong thị giác máy tính [8]. Cùng với sự bùng nổ của các mô hình học sâu, các nghiên cứu kết hợp mô hình nơron sâu vào GAN giải quyết các vấn đề gặp phải trong quá trình huấn luyện GAN để cải thiện chất lượng hình ảnh thu hút nhiều nghiên cứu điển hình trong [9] xây dựng mô hình học sâu lai dựa trên kết hợp GAN và RestNet50 giúp cải thiện độ chính xác trong phát hiện khuôn mặt giả mạo, mô hình đạt hiệu quả cao trên bộ dữ liệu mẫu tuy nhiên nếu dữ liệu không đa dạng thì mô hình không hoạt động tốt. Trong nghiên cứu [10] đề xuất SRGAN, đây là mô hình sử dụng kiến trúc GAN cho bài toán tăng cường độ phân giải ảnh đơn, SRGAN sử dụng mạng G có cấu trúc dựa trên Restnet, hàm kích hoạt PreLu, mạng D học cách phân biệt ảnh thật và giả sinh ra. Hàm mất mát pixel - L_2 trong SRGAN được tính dựa trên đặc trưng trung gian của mô hình VGG đã được tiền huấn luyện trên tập Image net. Cấu trúc gồm 8 lớp tích chập, kernel size 3*3, hàm kích hoạt LeakyReLU, tổng số tham số khoảng 5.4 triệu. Tuy nhiên SRGAN tồn tại một số hạn chế như huấn luyện khó ổn định, dễ sinh nhiễu hoặc chi tiết giả. Do đó nghiên cứu này đề xuất theo hướng cải tiến kiến trúc G bằng cách tăng khối residual block, sử dụng hàm kết nối tắt-scale skip connection với hệ số học 0.1, hàm kích hoạt reLu. Mạng D cải tiến sử dụng VGG19 với 16 lớp tích chập, hàm kích hoạt Relu, được huấn luyện và tối ưu hàm mất mát, số tham số khoảng 20 triệu. Kiến trúc tổng thể có khả năng phục hồi chi tiết ảnh tốt hơn đồng thời duy trì sự ổn định trong quá trình huấn luyện.

3. MÔ HÌNH GAN VÀ MỘT SỐ VẤN ĐỀ TRONG QUÁ TRÌNH HUẤN LUYỆN

3.1. Mô hình mạng GAN



Hình 1. Mô hình GAN

GAN là mô hình tạo sinh dựa trên lý thuyết trò chơi, gồm hai mạng nơron đối kháng G và D. Xét về mặt toán học, D ước lượng xác suất $p(y|x)$ trong khi G tạo ra mẫu dựa trên không gian tiềm ẩn z, ký hiệu $G(z)$. Công thức hàm sai số [1]:

$$L(D,G) = \\ = E_{x\sim px}\log[D(x)] + E_{z\sim pz}\log[1-D(G(x))] \quad (1)$$

Trong đó $x \sim pr$ là phân phối của dữ liệu thực và $z \sim pz$ phân phối xác suất không gian tiềm ẩn của G, thường là nhiễu Gauss hoặc nhiễu đồng nhất mà G sử dụng để mô hình hóa các mẫu dữ liệu mới, ký hiệu là $G(z)$.

Hàm D: tính độ sai khác giữa phân phối thực $D(x)$ và phân phối tổng hợp $D(G(Z))$.

Trong khi G cố gắng tạo ra các kết quả chân thực hơn, D cải thiện độ chính xác bằng cách phát hiện xem một mẫu là thật hay giả. Trong quá trình này, cả hai đối thủ đều được đồng bộ hóa: nếu G tạo ra đầu ra tốt hơn, D khó phân biệt chúng hơn. Ngược lại, nếu D chính xác hơn, G sẽ khó đánh lừa D hơn [10].

3.2. Một số vấn đề trong huấn luyện GAN

3.2.1. Triệt tiêu đạo hàm

Quá trình huấn luyện GAN phải được cân bằng, cả G và D cần được đồng bộ hóa để học cùng nhau một cách tiến bộ. Một D quá chính xác, nghĩa là $D(x)=1$ và $D(G(z))=0$. Khi đó, hàm mất mát tiến đến 0, làm cho đạo hàm gần như bằng 0 và cung cấp ít phản hồi cho G. Ngược lại, một D kém chính xác, không thể phân biệt giữa dữ liệu thật và dữ liệu tổng hợp, điều này khiến G không nhận được thông tin hữu ích.

3.2.2. Sự bất ổn định

Do đặc thù của GAN, việc kết hợp hai mô hình học từ nhau là một nhiệm vụ phức tạp. Quá trình huấn luyện GAN dựa trên một trò chơi tổng bằng không, nơi cả hai mạng cạnh tranh để tìm ra nghiệm tối ưu. Kiến trúc này dựa trên sự hợp tác để tối ưu hóa hàm mất mát toàn cục, nhưng các bài toán tối ưu của D và G là trái ngược nhau. Do tính chất của hàm mục tiêu của các mạng trong quá trình huấn luyện, một thay đổi nhỏ của một trong hai mạng có thể dẫn đến một thay đổi lớn trong mạng kia. Những khoảng thời gian mà cả hai mạng bắt đầu mất đồng bộ rất nhạy cảm vì sự thay đổi lớn trong đạo hàm có thể khiến một mạng mất khả năng học.

Do đó cần lựa chọn mạng học sâu làm mạng xương sống G và D trong GAN, cải thiện sự bất ổn định và mất cân bằng trong quá trình huấn luyện.

3.3. Hàm mất mát

Hàm mất mát pixel - L_2, được tính là sai số bình phương giữa ma trận pixel của hình ảnh dự đoán và hình ảnh đích nhưng L_2 không thể biểu diễn hoàn toàn chất lượng của quá trình tái tạo [11]. Công thức như sau:

$$L_2\left(\hat{Y},Y\right)=\sum_{i=1}^{n}(\hat{Y}_i-Y_i)^2 \quad (2)$$

Trong đó :

$\hat{Y}$: giá trị dự đoán;

Y_i: giá trị thực tế (Ground Truth- GT);

n: tổng số pixel.

Hàm mất mát cảm nhận - L_{percep} [11] hoạt động bằng cách truyền ảnh gốc và ảnh đầu ra qua một CNN được đào tạo trước và so sánh các bản đồ đặc trưng ở một hoặc nhiều lớp. Sau đó, mất mát được tính là khoảng cách Euclid giữa các bản đồ đặc trưng. Tiếp cận này tập trung vào sự khác biệt về ngữ nghĩa giữa ảnh gốc và ảnh sinh ra. Công thức [11]:

$$\frac{1}{N}\sum_{i=1}^{N}\| \phi(SR_i-\phi(GT_i)\|^2 \quad (3)$$

Trong đó :

N: số lượng mẫu trong tập dữ liệu;

SR_i: ảnh siêu phân giải thứ i;

GT_i: ảnh gốc thứ i;

Φ: mạng nơron tích chập huấn luyện trước được sử dụng để trích xuất các đặc trưng ở một hoặc nhiều lớp ẩn.

Hàm mất mát biến thiên toàn phần - L_{TV} - nhằm giảm thiểu sự khác biệt giữa các pixel lân cận. Mục tiêu của L_{TV} là giảm nhiễu trong khi vẫn bảo toàn các cạnh và chi tiết quan trọng trong ảnh, đảm bảo khi ảnh được phóng to không xuất hiện các hiện tượng giả tạo hoặc nhiễu [12].

Cho 1 ảnh SR $\in \square^{H \times W}$, L_{TV} có công thức:

$$L_{TV} = \sum_{ij}((SR_{i+1,j} - SR_{i,j})^2 + (SR_{i,j+1} - SR_{i,j})^2) \quad (4)$$

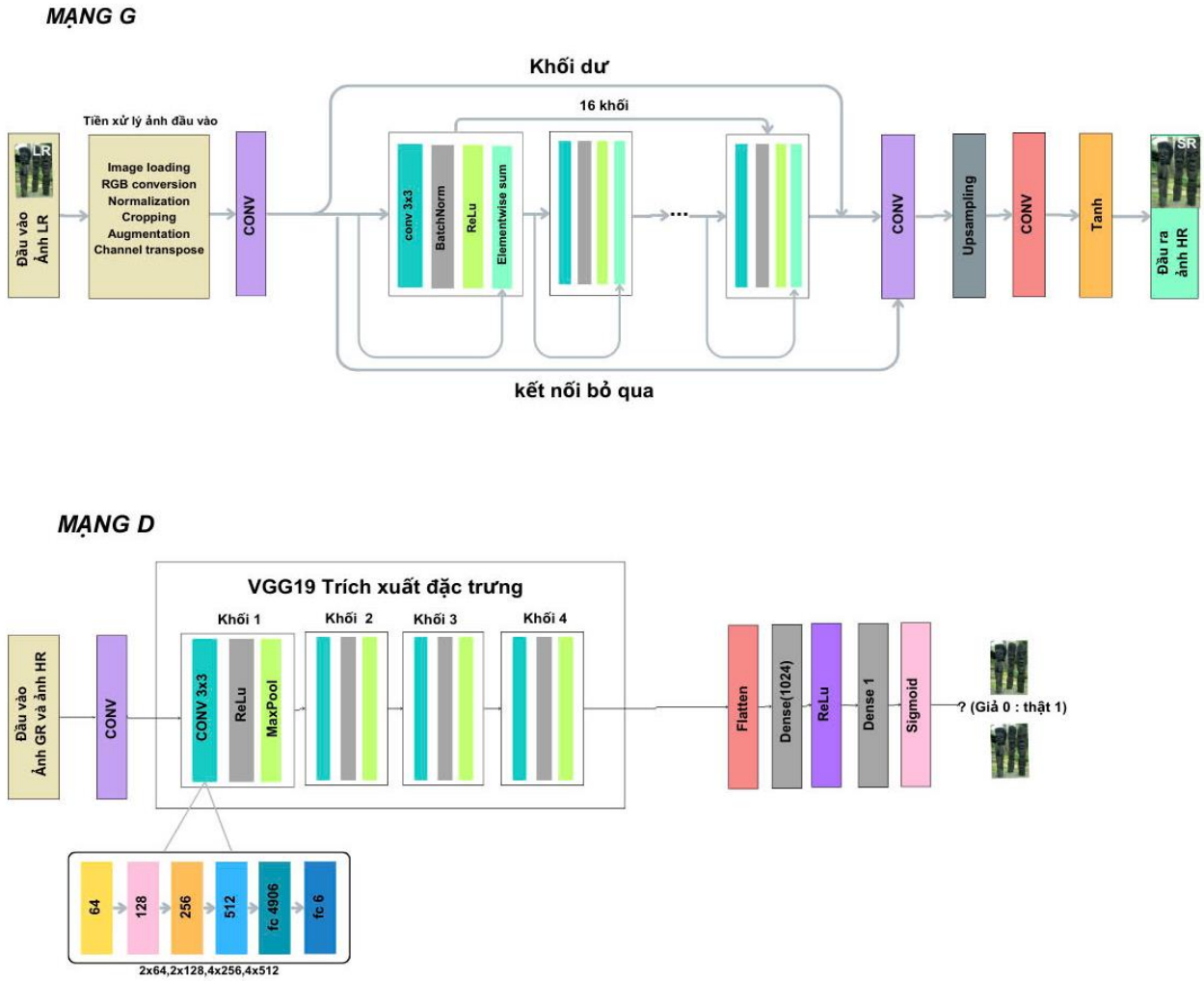
4. ĐỀ XUẤT MÔ HÌNH GAN CẢI TIẾN

4.1. Kết hợp GAN với Restnet và VGG19

Mô hình cải tiến được đề xuất bằng cách kết hợp VGG19 vào D để tăng khả năng trích xuất đặc trưng và sử dụng ResNet trong G để cải thiện khả năng tái tạo ảnh chi tiết. Việc sử dụng VGG19 giúp D học được đặc trưng cấp cao, thay vì chỉ phân biệt thật/giả đơn thuần, trong khi đó ResNet với độ sâu tăng cường, và tính năng scale skip connection giúp G tạo ảnh có độ phân giải cao và giảm mất thông tin trong quá trình truyền tín hiệu đồng thời giảm thiểu sự thay đổi đột ngột trong quá trình huấn luyện, giúp mô hình hội tụ và ổn định.

Ảnh được tạo ra từ G sử dụng làm đầu vào cho D, sau đó G được huấn luyện dùng hàm mất mát: $L_G = L_2 + L_{percep} + L_{GAN} + L_{TV}$

4.2. Kiến trúc mạng GAN cải tiến



Hình 2. Mô hình GAN cải tiến

Kiến trúc GAN cải tiến được minh họa trong hình 2. Ảnh đầu vào kích thước 512×512×3 được xử lý qua:

- Lớp tích chập đầu vào (CONV), kernel size 3×3, trích xuất đặc trưng ban đầu.
- Sau đó đi qua 16 khối Residual Blocks (khối dư), một khối bao gồm:

- Một chuỗi các lớp học sâu (lớp Conv 3×3+ BatchNorm + ReLU).
- Một kết nối bỏ qua có hệ số học 0.1 (scale

skip connection) được nhân với *x*, trước khi đưa thẳng tới phần đầu ra sau lớp học sâu giúp giữ lại các đặc trưng quan trọng trong quá trình tái tạo ảnh đồng thời ổn định huấn luyện khi độ sâu lớn.

- Lớp CONV trung gian, duy trì tính ổn định trong quá trình huấn luyện.
- Lớp upsampling (nội suy) với kernel size 3×3, tăng kích thước ảnh lên 2 lần, nâng tổng hệ số phóng đại ảnh lên 4 lần.
- Lớp CONV cuối cùng, sử dụng hàm kích hoạt Tanh, đưa ảnh đầu ra về [-1,1].

Kết quả đầu ra là ảnh siêu phân giải có kích thước $2048\times2048\times3$.

Mạng D được thiết kế với kiến trúc VGG19, trích xuất đặc trưng ảnh đầu vào với kích thước $2048\times2048\times3$ được tạo ra từ mạng G.

Cấu trúc bao gồm:

- Mạng VGG19 gồm 4 khối sử dụng để trích xuất đặc trưng.
- Lớp CONV đầu vào với kernel size 3×3.
- 2 lớp fully connected (kết nối đầy đủ), trong đó lớp cuối cùng sử dụng hàm sigmoid, cho đầu ra xác suất ảnh thực hoặc giả.

Đầu ra của mạng D: Xác suất ảnh có phải là ảnh thực hay không so với ảnh gốc GT.

Mã giả: Quá trình huấn luyện

Đầu vào: Ảnh gốc - GT và ảnh độ phân giải thấp - LR.

Đầu ra: Mô hình G và D được huấn luyện, và ảnh chất lượng cao.

Khởi tạo G với tham số ngẫu nhiên θ

Khởi tạo D với tham số ngẫu nhiên φ

Khởi tạo các siêu tham số: batch size, số epoch, learning rate.

Khởi tạo bộ tối ưu Adam cho G và D

Giai đoạn 1: Tiền huấn luyện G sử dụng hàm mất mát L_2.

For *epoch = 1* to *epoch pre-train lớn nhất* do

Với mỗi batch dữ liệu (GR, LR):

Dự đoán ảnh HR bằng G

Tính toán mất mát L_2

$$L_{L_2} = \frac{1}{n}\sum GT - G(LR))^2$$

Cập nhật G bằng gradient descent.

Đánh giá PSNR trên tập kiểm tra.

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

Lưu trọng số mô hình G.

End

Giai đoạn 2: Huấn luyện G và D

For *epoch* = 1 to *epoch fine-tune lớn nhất* do

Với mỗi batch dữ liệu (GR, LR)

Huấn luyện D

Tạo ảnh HR bằng G. Ảnh sẽ được sử dụng làm đầu vào cho D.

D phân biệt ảnh của G là ảnh giả hay thật.

Tính xác suất thật/giả của ảnh

$D_{real} = D(GR)$ và $D_{fake} = D(G(LR))$

Tính hàm mất mát của D

$$L_D = -E\left[\log(D_{real})\right] - E\left[\log(1-D_{fake})\right]$$

Cập nhật trọng số D.

Huấn luyện G

Sinh ảnh giả: $SR = G(LR)$

Tính xác suất ảnh giả: $D_{fake} = D(S(R))$

Tính các hàm mất mát:

$$L_{percep}: \frac{1}{n}\sum_{i=1}^{n}\left\|\phi(SR_i - \phi(GT_i))\right\|^2$$

Với ϕ là các đặc trưng trích xuất từ VGG19.

$$L_{GAN} = -E\left[\log D_{fake}\right]$$

$$L_{TV} = \sum_{ij}\left(\left(SR_{i+1,j} - SR_{i,j}\right)^2 + \left(SR_{i,j+1} - SR_{i,j}\right)^2\right)$$

Tính tổng hàm mất mát của G:

$$L_G = L_2 + L_{percep} + L_{GAN} + L_{TV}$$

Cập nhật trọng số G.

End

4.3. Dữ liệu và quá trình tiền xử lý

4.3.1. Bộ dữ liệu DIV2K

Trong mô phỏng, chúng tôi sử dụng bộ dữ liệu DIV2K gồm 1.000 ảnh chất lượng cao (độ phân giải xấp xỉ 2K). Bộ dữ liệu được chia thành 800 ảnh để huấn luyện, 100 ảnh kiểm tra và 100 ảnh kiểm thử. Mỗi bức ảnh có nội dung đa dạng, bao gồm phong cảnh, thành phố, con người, vật thể, giúp các mô hình học sâu tổng quát hóa tốt hơn. DIV2K cung cấp ảnh HR có dung lượng khoảng 7GB, cùng với các phiên bản LR được tạo bằng phép nội suy theo các tỷ lệ X2, X3, X4 (mỗi bộ LR khoảng 1.5GB - 2GB). Đây là bộ dữ liệu tiêu chuẩn để huấn luyện và đánh giá các mô hình siêu phân giải, đồng thời được sử dụng trong xử lý ảnh, tăng cường chất lượng video và phục chế ảnh cũ.

4.3.2. Tiền xử lý dữ liệu

Đầu tiên, ảnh được chuyển đổi từ không gian màu gốc sang RGB để đảm bảo tính đồng nhất. Sau đó, được chuẩn hóa về khoảng [-1, 1]. Điều này giúp mô hình dễ học hơn do dữ liệu được phân bố cân bằng quanh 0, giảm sai số học. Ngoài ra, để giúp mô hình học được nhiều đặc trưng khác nhau và giảm nguy cơ quá khớp, ảnh trải qua các phép tăng cường dữ liệu như lật ngang, lật dọc và xoay. Những kỹ thuật này giúp mô hình tổng quát hóa tốt hơn khi xử lý dữ liệu. Trong một số trường hợp, ảnh được cắt thành các lát cắt ảnh nhỏ để giảm yêu cầu bộ nhớ và giúp mô hình học chi tiết hơn trên từng vùng ảnh. Tất cả những bước này nhằm mục tiêu tối ưu dữ liệu đầu vào, cải thiện độ chính xác của mô hình siêu phân giải ảnh



a) Ảnh gốc b) Ảnh sau huấn luyện

Hình 3. Kết quả huấn luyện ảnh

5. KẾT QUẢ VÀ THẢO LUẬN

Thực nghiệm huấn luyện GAN đề xuất trên tập dữ liệu DIV2K với 3000 epoch pre-train và 1500 epoch fine-tune, batch size16, hàm kích hoạt ReLU cho lớp tích chập và lớp kết nối đầy đủ, các tham số:

$$learningrate = 1e-4, \beta_1 = 0.9, \beta_2 = 0.999, \varepsilon = 1e-8$$

Đánh giá, so sánh hiệu quả của GAN đề xuất so với SRGAN sử dụng độ đo:

- PSNR (Peak Signal-to-Noise Ratio): chỉ số đo lường độ tương phản giữa tín hiệu gốc và tín hiệu bị suy giảm, thường sử dụng để đánh giá chất lượng ảnh. Giá trị PSNR cao đồng nghĩa với việc ảnh tái tạo có độ méo thấp so với ảnh gốc. Theo dõi PSNR theo thời gian giúp đánh giá quá trình học của mô hình, nếu PSNR tăng, điều đó có nghĩa mô hình đang dần cải thiện chất lượng ảnh đầu ra. Ngược lại, có thể mô hình đang gặp vấn đề trong việc tái tạo chi tiết ảnh.

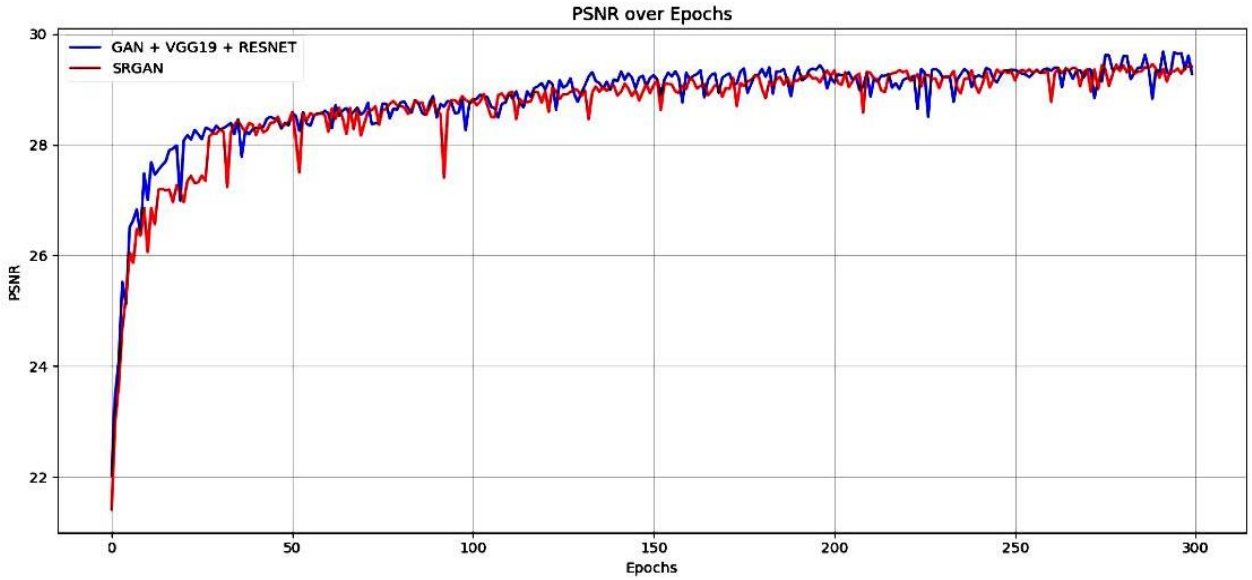
- SSIM (Structural Similarity Index Measure): chỉ số đo lường độ tương đồng về mặt cấu trúc như: sự tương đồng về độ sáng, độ tương phản giữa ảnh gốc và ảnh được tái tạo. Theo

dõi SSIM theo thời gian giúp đánh giá chất lượng ảnh đầu ra của mô hình. Nếu SSIM tăng theo thời gian, điều đó cho thấy mô hình đang dần cải thiện khả năng tái tạo, hình ảnh gần giống với ảnh gốc hơn. Ngược lại, mô hình đang gặp khó khăn trong việc giữ lại các đặc trưng quan trọng của hình ảnh.

Quan sát hình 4 cho thấy trong giai đoạn đầu của quá trình huấn luyện (epoch 0 - 30), cả hai đều có tốc độ cải thiện PSNR rất nhanh, từ mức khởi điểm khoảng 21.5 lên trên 27.5.

Từ sau epoch 30, tốc độ tăng của PSNR chậm lại và dần đạt trạng thái ổn định. Trong giai đoạn từ epoch 100 trở đi, mô hình đề xuất (GAN + VGG19 + RESNET) thể hiện sự vượt trội so với SRGAN. Cụ thể, PSNR của mô hình đề xuất dao động ổn định quanh ngưỡng 29 và đạt đỉnh tại epoch 294 với giá trị xấp xỉ 29.75. Trong khi đó, SRGAN có xu hướng hội tụ ở mức thấp hơn, với PSNR cao nhất khoảng 27.2 tại epoch 295.

Ngoài ra, SRGAN còn bộc lộ một số bất ổn trong quá trình huấn luyện, thể hiện qua các điểm sụt giảm PSNR bất thường tại các epoch như 37, 51 và 95. Trái lại, mô hình đề xuất duy trì được sự ổn định cao hơn trong suốt quá trình huấn luyện, không ghi nhận các dao động lớn, cho thấy khả năng tối ưu hóa hiệu quả và nhất quán.



Hình 4. PSNR SRGAN và GAN+VGG19+Resnet

Hình 5 thể hiện sự thay đổi của SSIM qua 300 epoch của SRGAN và GAN đề xuất. Trong giai đoạn đầu của quá trình huấn luyện (epoch 0–20), mô hình đề xuất thể hiện sự tăng trưởng nhanh và ổn định SSIM, từ mức ban đầu khoảng 0.60 lên gần 0.79. Ngược lại, SRGAN mặc dù cũng có xu hướng tăng SSIM nhưng vẫn tồn tại một số dao động cục bộ nhỏ. Từ epoch 20-60, xuất hiện vài lần suy giảm nhẹ SSIM - ví dụ tại epoch 42 và 54, SSIM giảm về dưới 0.795, phản ánh độ ổn định chưa cao so với mô hình đề xuất.

Từ epoch 50 trở đi, cả hai mô hình đều đạt trạng thái hội tụ tương đối, tuy nhiên mô hình đề xuất tiếp tục duy trì SSIM cao hơn một cách nhất quán. Cụ thể, trong khoảng epoch 100 đến 300, SSIM của mô hình đề xuất dao động quanh 0.825 đến 0.835, và đạt đỉnh tại epoch 289 với giá trị xấp xỉ 0.887. Trong khi đó, SRGAN hội tụ ở mức thấp hơn, dao động uanh 0.815 đến 0.828, và đạt cực đại tại epoch

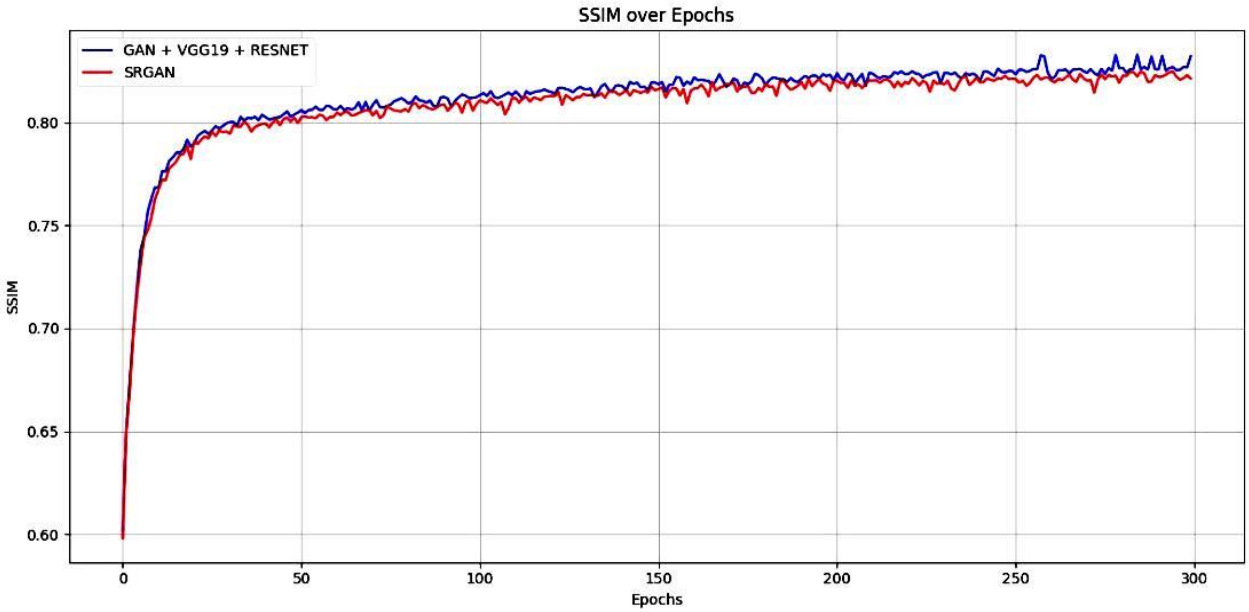
297 với SSIM khoảng 0.831. Ngoài ra, mô hình SRGAN vẫn xuất hiện một số điểm bất lợi trong giai đoạn cuối huấn luyện, khi SSIM giảm nhẹ về dưới 0.820, trong khi mô hình đề xuất giữ được độ ổn định tốt hơn và không ghi nhận các dao động lớn.

6. KẾT LUẬN

Trong nghiên cứu này, chúng tôi đã đề xuất một mô hình cải tiến dựa trên GAN, kết hợp với các đặc trưng trích xuất từ VGG19 và cấu trúc học sâu ResNet nhằm nâng cao chất lượng siêu phân giải hình ảnh. Mô hình không chỉ tận dụng khả năng tái tạo chi tiết ảnh, mà còn khai thác hiệu quả các đặc trưng bậc cao từ VGG19 và khả năng học thông tin dư của ResNet để tạo ra ảnh siêu phân giải sắc nét và tự nhiên hơn.

Kết quả thực nghiệm cho thấy mô hình đề xuất cải thiện đáng kể so với SRGAN.

Cụ thể, PSNR trong mô hình cải tiến cao hơn và dao động nhỏ phản ánh khả năng tái tạo chi tiết hình ảnh tốt và ít nhiễu hơn, đồng thời độ đo SSIM cũng được cải thiện rõ rệt, cho thấy mô hình bảo toàn tốt hơn cấu trúc và thông tin thị giác trong ảnh siêu phân giải.



Hình 5. SSIM SRGAN và GAN+VGG19+Restnet

TÀI LIỆU THAM KHẢO

[1] Goodfellow, I. et al. “Generative Adversarial Nets”, in Advances in Neural Information Processing Systems (NeurIPS), 27, 2672–2680, (2014).

[2] Tero Karras, Timo Aila and Samuli Laine, "progressive growing of gans for improved Quality, stability, and variation", in International Conference on Learning Representations, (2018).

[3] Yitong Li, Martin Renqiang Min, Dinghan Shen, David Carlson, Lawrence Carin, " Video Generation From Text”, CoRR, (2017).

[4] He et al., "Deep Residual Learning for Image Recognition", CVPR, (2016).

[5] Zisserman, Karen Simonyan and Andrew, "Very deep convolutional network for large-scale image

recognition", in The 3rd International Conference on Learning Representations, (2015).

[6] Wiatrak, M., and Albrecht, S. V. "*Stabilizing generative adversarial network training: A survey". arXiv,* (2019).

[7] Pan, Z., Yu, W., Wang, B., Xie, H., Sheng, V. S., Lei, J. "*Loss functions of generative adversarial networks (gans): opportunities and challenges".* IEEE Transactions on Emerging Topics in Computational Intelligence 4,4 , 500-522, (2020),.

[8] Gui, J., Sun, Z., Wen, Y., Tao, D., and Ye, J. "*A review on generative adversarial networks: Algorithms, theory, and applications",* IEEE Transactions on Knowledge and Data Engineering, 35(5), 1–21, (2020).

[9] Soha Safwat; Ayat Mahmoud; Ibrahim Eldesouky Fattoh , "Hybrid Deep Learning Model Based on GAN and RESNET for Detecting Fake Faces", in *IEEE Access*, vol. 12, pp. 86391-86402, 2024, doi: 10.1109/ACCESS.2024.3416910.

[10] C. Ledig *et al*., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 105-114, (2017). doi: 10.1109/CVPR.2017.19.

[11] Johnson et al., "Perceptual Losses for Real-Time Style Transfer and Super-Resolution", 2016.

[12] Mia Rizkinia, Nathaniel Faustine, Masahiro Okuda, "Conditional Generative Adversarial Networks with Total Variation and Color Correction for Generating Indonesian Face Photo from Sketch", Applied Sciences, (2022).

Thông tin liên hệ: **Lương Thị Thảo Hiếu**

Điện thoại: 0942160880 - Email: ltthieu@uneti.edu.vn

Khoa Công nghệ thông tin, Trường Đại học Kinh tế - Kỹ thuật Công nghiệp.