ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

Nguyễn Trọng Nhân

MẠNG ĐỐI NGHỊCH TẠO SINH TRONG CHUYỂN ĐỔI ẢNH CHÂN DUNG SANG PHONG CÁCH ANIME

THỰC HÀNH NGHIÊN CỨU 2

Ngành: Khoa Học Máy Tính

HÀ NỘI - 2025

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

Nguyễn Trọng Nhân

MẠNG ĐỐI NGHỊCH TẠO SINH TRONG CHUYỂN ĐỔI ẢNH CHÂN DUNG SANG PHONG CÁCH ANIME

THỰC HÀNH NGHIÊN CỨU 2

Ngành: Khoa học máy tính

Cán bộ hướng dẫn: TS. Ma Thị Châu

HÀ NỘI - 2025

VIETNAM NATIONAL UNIVERSITY, HANOI
UNIVERSITY OF ENGINEERING AND TECHNOLOGY

Nguyen Trong Nhan

GENERATIVE ADVERSARIAL NETWORKS FOR CONVERTING PORTRAITS TO ANIME STYLE

RESEARCH PRACTICE 2

Major: Computer Science

Supervisor: Dr. Ma Thi Chau

HANOI - 2025

LỜI CẢM ƠN

Trước hết, tôi xin chân thành cảm ơn TS. Ma Thị Châu, người hướng dẫn đề tài nghiên cứu lần này, đã tận tình hỗ trợ, chỉ bảo và động viên tôi trong suốt quá trình nghiên cứu và hoàn thành. Nhờ có sự hướng dẫn của cô, tôi đã có được những kiến thức quý báu và kinh nghiệm thực tiễn trong lĩnh vực nghiên cứu của mình.

Tôi cũng xin gửi lời cảm ơn đến các thầy cô giáo của Trường Đại học Công Nghệ, đã truyền đạt cho tôi những kiến thức cơ bản và nâng cao về ngành Khoa học Máy Tính. Tôi xin chúc các thầy cô luôn mạnh khỏe và thành công trong công tác giảng dạy và nghiên cứu khoa học.

Sau cùng, tôi muốn bày tỏ lòng biết ơn sâu sắc đến gia đình, người thân và bạn bè, những người luôn ở bên cạnh tôi trong những thời điểm khó khăn nhất, luôn động viên và khích lệ tôi trong cuộc sống và công việc. Thành tựu này của tôi không thể có được nếu thiếu sự giúp đỡ của họ.

Mặc dù tôi đã cố gắng hoàn thành báo cáo nhưng không thể tránh khỏi những sai sót, tôi rất mong nhận được nhận xét và sự hướng dẫn từ phía giáo viên và hội đồng.

Hà Nội, ngày ... tháng ... năm 2023

Học viên

Nguyễn Trọng Nhân

LỜI CAM ĐOAN

Tôi xin cam đoan rằng tất cả các nội dung trong tài liệu này đều là kết quả của công trình nghiên cứu cá nhân của tôi dưới sự hướng dẫn của TS.Ma Thị Châu. Tôi không sao chép bất kỳ tài liệu hay công trình nghiên cứu nào của người khác mà không chỉ rõ nguồn gốc trong phần tài liệu tham khảo. Tôi hiểu rằng việc sao chép trái phép là một hành vi vi phạm đạo đức học thuật và tôi sẽ chịu trách nhiệm về những hành vi này.

Tôi cam kết rằng, bản báo cáo của tôi không vi phạm bản quyền của bất kỳ ai và không vi phạm bất kỳ quyền sở hữu nào, cũng như bất kỳ ý tưởng, kỹ thuật, trích dẫn hoặc bất kỳ tài liệu nào từ công trình nghiên cứu của người khác. Tôi xin nhận đầy đủ trách nhiệm và sẵn sàng chấp nhận mọi biện pháp kỷ luật nếu vi phạm cam kết.

Hà Nội, ngày ... tháng ... năm 2023

Học viên

Nguyễn Trọng Nhân

TÓM TẮT

Tóm tắt: Sự bùng nổ của nội dung số và nghệ thuật kỹ thuật số, đặc biệt trong lĩnh vực Anime/Manga, kéo theo nhu cầu mạnh mẽ về các công cụ hỗ trợ sáng tạo dựa trên trí tuệ nhân tạo. Trong đó, bài toán chuyển đổi ảnh chân dung người thật sang phong cách anime 2D (photo-to-anime) vừa giàu tiềm năng ứng dụng, vừa đặt ra nhiều thách thức về chất lượng hình ảnh và khả năng giữ lại đặc điểm nhận dạng khuôn mặt. Các phương pháp truyền thống dựa trên chuyển phong cách (style transfer) hoặc các kiến trúc GAN thế hệ đầu thường gặp vấn đề nhiễu, tạo tác (artifacts) và khó đảm bảo tính ổn định khi huấn luyện.

Luận văn này đề xuất một mô hình chuyển đổi ảnh chân dung người sang phong cách anime dựa trên kiến trúc Mạng đối nghịch tạo sinh (GAN), với nền tảng là kiến trúc Double-Tail GAN (DTGAN) của AnimeGAN thế hệ mới. Mô hình sử dụng kiến trúc encoder–decoder dựa trên ResNet với generator hai nhánh (double-tail) gồm nhánh hỗ trợ tạo ảnh anime thô và nhánh chính tinh chỉnh để tạo ra ảnh anime chất lượng cao. Bên cạnh đó, luận văn áp dụng kỹ thuật chuẩn hóa Linearly Adaptive Denormalization (LADE) nhằm giảm nhiễu và làm mượt vùng màu theo phong cách anime, đồng thời xây dựng hệ thống hàm mất mát (loss) chuyên biệt, kết hợp giữa mất mát đối nghịch, mất mát nội dung/nhận dạng (dựa trên đặc trưng VGG19), mất mát phong cách và các mất mát làm mượt–chỉnh chi tiết để cân bằng giữa bảo toàn nội dung và thể hiện phong cách.

Mô hình được huấn luyện trên hai tập dữ liệu không ghép cặp gồm ảnh chân dung người thật và ảnh chân dung anime 2D, với các bước tiền xử lý như phát hiện và căn chỉnh khuôn mặt, chuẩn hóa kích thước và tăng cường dữ liệu. Kết quả thực nghiệm cho thấy mô hình đề xuất tạo ra ảnh anime có tính thẩm mỹ cao, giữ được đặc điểm khuôn mặt của đối tượng gốc và giảm đáng kể nhiễu, được xác nhận qua các chỉ số định lượng (FID, LPIPS) cũng như đánh giá định tính từ người dùng. Cuối cùng, luận văn thảo luận các hạn chế và đề xuất hướng phát triển trong tương lai như tăng mức độ điều khiển phong cách chi tiết (màu mắt, kiểu tóc, biểu cảm) và mở rộng sang bài toán chuyển đổi video–to–anime.

Từ khóa: *GAN, AnimeGAN, DTGAN, chuyển phong cách ảnh, ảnh chân dung, phong cách anime, LADE, VGG19.*

ABSTRACT

Abstract: The explosion of digital content and digital art, especially in the Anime/Manga field, has led to a strong demand for AI-powered creative tools. Among these, the task of converting real-person portraits to 2D anime style (photo-to-anime) offers significant application potential but also poses many challenges regarding image quality and the ability to retain facial identity features. Traditional methods based on style transfer or early-generation GAN architectures often suffer from noise, artifacts, and difficulty in ensuring training stability.

This thesis proposes a model for converting real-person portraits to anime style based on the Generative Adversarial Network (GAN) architecture, specifically building upon the Double-Tail GAN (DTGAN) architecture of the new generation AnimeGAN. The model utilizes a ResNet-based encoder–decoder architecture with a double-tail generator, consisting of a branch that assists in generating rough anime images and a main branch that refines them to produce high-quality anime images. Additionally, the thesis applies the Linearly Adaptive Denormalization (LADE) technique to reduce noise and smooth color regions in anime style, while also developing a specialized loss function system that combines adversarial loss, content/identity loss (based on VGG19 features), style loss, and smoothing–detail adjustment losses to balance content preservation and style expression.

The model was trained on two unpaired datasets of real-person portraits and 2D anime portraits, with preprocessing steps such as face detection and alignment, size normalization, and data augmentation. Experimental results show that the proposed model generates aesthetically pleasing anime images, preserves the facial characteristics of the original subject, and significantly reduces noise, as confirmed by quantitative metrics (FID, LPIPS) as well as qualitative evaluation from users. Finally, the thesis discusses limitations and proposes future development directions such as increasing the level of control over detailed styles (eye color, hairstyle, expression) and extending to the video-to-anime conversion problem.

Keywords: *GAN, AnimeGAN, DTGAN, image style transfer, portrait, anime style, LADE, VGG19.*

MỤC LỤC

DANH MỤC HÌNH VẼ

DANH MỤC BẢNG BIỂU

DANH MỤC KÝ HIỆU VÀ CHỮ VIẾT TẮT

Danh mục ký hiệu		
STT	**Ký hiệu**	**Giải thích**
1	in thường	Vô hướng
2	in thường, đậm	Vector

Danh mục chữ viết tắt			
STT	**Chữ viết tắt**	**Giải thích tiếng Anh**	**Giải thích tiếng Việt**
1	ADC	Analog Digital Converter	Bộ chuyển đổi tương tự sang số
2	AM	Amplitude Modulation	Điều chế biên độ

MỞ ĐẦU

Tính cấp thiết và Ý nghĩa của Đề tài

Sự phát triển mạnh mẽ của nghệ thuật kỹ thuật số cùng với thị trường Anime/Manga toàn cầu đã tạo động lực lớn cho các nghiên cứu về chuyển đổi phong cách hình ảnh. Theo thống kê năm 2024, quy mô thị trường anime toàn cầu đạt khoảng 81,96 tỷ USD và dự kiến vượt 200 tỷ USD vào năm 2034. Công nghệ AI hiện nay cũng đang dần xâm nhập vào quy trình sản xuất anime – ví dụ, các công cụ Generative AI đã có thể tự động hoá việc vẽ phông nền và tô màu, giúp giảm bớt công việc lặp lại cho các hoạ sĩ. Trong bối cảnh đó, nhu cầu tự động hoá chuyển đổi ảnh thực sang phong cách anime trở nên cấp thiết, nhằm phục vụ cộng đồng người hâm mộ khổng lồ và ngành công nghiệp sáng tạo nội dung đang bùng nổ.

Tuy nhiên, việc chuyển một ảnh chân dung người thật sang tranh anime là thách thức không nhỏ. Phong cách anime có đặc trưng rất khác biệt (đường nét đơn giản, mắt to, màu sắc phẳng, v.v.), trong khi ảnh chụp chứa nhiều chi tiết thực tế phức tạp. Các phương pháp truyền thống dựa trên Neural Style Transfer thường gặp khó khăn trong việc giữ được nội dung gốc và dễ sinh ra nhiễu, tạo tác (artifacts) khi khác biệt phong cách quá lớn. Mặt khác, vẽ tay thủ công bởi các hoạ sĩ tuy chất lượng cao nhưng tốn kém thời gian và công sức. Do đó, bài toán đặt ra là làm thế nào để tự động hoá quá trình này mà vẫn đảm bảo chất lượng cao và bảo toàn được đặc điểm nhận dạng của đối tượng trong ảnh gốc.

Mạng Đối nghịch Tạo sinh (GAN) nổi lên như một công nghệ đột phá có thể giải quyết các nhiệm vụ tổng hợp hình ảnh phức tạp. Được Ian Goodfellow giới thiệu năm 2014 và được Yann LeCun ca ngợi là "ý tưởng thú vị nhất của ML trong 10 năm", GAN đã chứng tỏ hiệu quả vượt trội trong việc sinh ảnh chân thực từ dữ liệu huấn luyện. Đặc biệt trong bài toán dịch chuyển phong cách (style transfer), GANs cung cấp một khuôn khổ linh hoạt để huấn luyện mô hình tạo ảnh anime từ ảnh thật mà không đòi hỏi dữ liệu ảnh cặp một-một. Nhờ GAN, những tiến bộ gần đây cho thấy khả năng tạo các hình ảnh anime hóa ngày càng sắc nét và chính xác hơn, mở ra hướng tiếp cận mới cho bài toán này.

Mục tiêu Nghiên cứu

Mục tiêu tổng quát là xây dựng và đánh giá một mô hình chuyển đổi ảnh chân dung người thật sang phong cách anime chất lượng cao dựa trên GAN, đảm bảo giữ được các nét nhận dạng chính của khuôn mặt gốc. Mục tiêu cụ thể bao gồm:

- Lựa chọn kiến trúc GAN phù hợp: Nghiên cứu các kiến trúc GAN tiên tiến và chọn mô hình nền tảng hiệu quả nhất cho bài toán (dự kiến sử dụng kiến trúc Double-Tail GAN (DTGAN) – phiên bản AnimeGAN thế hệ mới).
- Xây dựng tập dữ liệu và tiền xử lý: Thu thập hoặc tinh chỉnh tập dữ liệu gồm ảnh chân dung thực và ảnh anime tương ứng (dạng không ghép cặp), áp dụng các bước tiền xử lý như căn chỉnh khuôn mặt, chuẩn hóa kích thước, tăng cường dữ liệu.
- Đề xuất hàm mất mát và chiến lược huấn luyện tối ưu: Thiết kế bộ hàm loss chuyên biệt (kết hợp giữa loss truyền thống của GAN và các loss phong cách/anime đặc thù) cùng lịch trình huấn luyện thích hợp nhằm tăng độ ổn định và làm nổi bật chi tiết ảnh đầu ra.
- Đánh giá mô hình: Đánh giá chất lượng ảnh anime sinh ra bằng cả phương pháp định lượng (ví dụ: FID, PSNR, LPIPS) và định tính (đánh giá cảm quan, khảo sát người dùng), so sánh với các phương pháp chuyển đổi phong cách hiện có để xác định hiệu quả của mô hình đề xuất.

Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu: Các kiến trúc mạng Generative Adversarial Networks (GANs) phục vụ cho nhiệm vụ dịch chuyển phong cách hình ảnh, đặc biệt là các mô hình chuyên dành cho chuyển ảnh người sang hoạt hình/anime. Ngoài ra, đề tài cũng liên quan đến các kỹ thuật thị giác máy tính và học sâu hỗ trợ, như mô hình encoder-decoder, các phương pháp xử lý ảnh (phát hiện, căn chỉnh khuôn mặt) và các hàm tổn thất perceptual.

Phạm vi nghiên cứu: Đề tài tập trung vào ảnh chân dung người (face portraits) và phong cách Anime 2D. Cụ thể, ảnh đầu vào giới hạn ở ảnh khuôn mặt người thật (có thể chụp từ camera hoặc ảnh selfie), và ảnh đầu ra hướng đến phong cách tranh vẽ nhân vật anime 2D (phong cách hoạt hình Nhật Bản). Mô hình được huấn luyện và đánh giá trên tập dữ liệu ảnh chân dung và ảnh anime không ghép cặp. Những khía cạnh ngoài phạm vi bao gồm chuyển đổi phong cách cho video, phong cách 3D hoặc các phong cách hoạt

hình khác (cartoon kiểu phương Tây, tranh phác hoạ, v.v.), cũng như không đi sâu vào các kỹ thuật diffusion models hay transformer mới hơn (chỉ tập trung vào GAN truyền thống trong giai đoạn 2020-2025).

Cấu trúc luận văn

Luận văn được tổ chức thành 5 chương như sau:

Chương 1: Mở đầu – Trình bày sự cần thiết, mục tiêu, phạm vi của đề tài và khái quát nội dung nghiên cứu.

Chương 2: Cơ sở Lý thuyết và Tổng quan Nghiên cứu – Tổng hợp nền tảng lý thuyết về học sâu, thị giác máy, kiến trúc GAN, và các nghiên cứu liên quan trong lĩnh vực chuyển ảnh chân dung sang phong cách anime.

Chương 3: Phương pháp Nghiên cứu Đề xuất – Mô tả chi tiết mô hình GAN đề xuất (kiến trúc Double-Tail GAN), các cải tiến kỹ thuật (như LADE – adaptive denormalization, hàm mất mát mới), cùng quy trình huấn luyện.

Chương 4: Thực nghiệm và Kết quả – Trình bày thiết lập thực nghiệm, quá trình huấn luyện mô hình trên tập dữ liệu thu thập, và kết quả đánh giá so sánh với các phương pháp khác. Phân tích các kết quả định lượng và định tính, minh hoạ bằng các hình ảnh đầu vào/đầu ra.

Chương 5: Kết luận và Hướng phát triển – Tóm tắt những đóng góp chính của luận văn, thảo luận những hạn chế còn tồn tại và đề xuất hướng nghiên cứu trong tương lai (mở rộng sang video, phong cách khác, ứng dụng diffusion model, v.v.).

CHƯƠNG 1
CƠ SỞ LÝ THUYẾT VÀ TỔNG QUAN NGHIÊN CỨU

1.1. Mạng Nơ-ron Tích chập

Mạng Nơ-ron Tích chập (Convolutional Neural Networks – CNNs) đã củng cố vị thế là kiến trúc nền tảng trong lĩnh vực thị giác máy tính, tạo ra những bước tiến lớn trong các tác vụ từ phân loại ảnh đơn giản đến phân đoạn thực thể phức tạp. Sự thành công của CNN bắt nguồn từ khả năng học hỏi các biểu diễn đặc trưng phân cấp, tự động trích xuất các thông tin hình học ngày càng trừu tượng từ dữ liệu đầu vào thô. [4]

1.1.1. Hoạt động của Lớp Tích chập

Lớp tích chập là khối xây dựng cơ bản của mọi kiến trúc CNN hiện đại. Cơ chế hoạt động của lớp này được thiết kế đặc biệt để xử lý dữ liệu ảnh bằng cách khai thác tính cục bộ và tính bất biến dịch chuyển thống kê của đặc trưng hình ảnh.

Lớp tích chập vận hành thông qua các kernel nhỏ (hay còn gọi là bộ lọc).3 Các kernel này trượt (slide) trên ảnh đầu vào, thực hiện phép nhân tích chập với từng phần của ảnh để tính toán tích vô hướng cục bộ.3 Điều này cho phép lớp tích chập hoạt động như các bộ lọc cục bộ, chuyên trách trích xuất các đặc trưng hình học cơ bản, chẳng hạn như các cạnh (edges), đường cong (curves), hoặc các kết cấu (textures) [User Query].

Mỗi kernel trong lớp tích chập được học để phát hiện một loại đặc trưng cụ thể và áp dụng phát hiện đó trên toàn bộ ảnh [User Query]. Khả năng học các bộ lọc cục bộ này, thay vì dựa vào các đặc trưng được thiết kế thủ công (handcrafted features) như SIFT hay HOG trong các mô hình thị giác máy tính truyền thống, là một đổi mới cốt lõi của CNN.1 Sự thay đổi này đã giúp các mô hình CNN giảm đáng kể nhu cầu về kinh nghiệm chuyên môn trong việc tiền xử lý dữ liệu đầu vào.1 Quá trình này tạo tiền đề cho việc xây dựng các đặc trưng phân cấp phong phú (rich hierarchy of image features), nơi các lớp sâu hơn dần dần kết hợp các đặc trưng cấp thấp thành các biểu diễn trừu tượng và ngữ nghĩa hơn.2

1.1.2. Lớp Pooling và Tối ưu hóa Bản đồ Đặc trưng

Thông thường, lớp Pooling được thêm vào ngay sau một lớp tích chập.4 Chức năng chính của lớp Pooling là giảm độ phân giải không gian của bản đồ đặc trưng

(feature maps) bằng cách chia chúng thành các vùng con hình chữ nhật và giảm mẫu các đặc trưng trong mỗi vùng con thành một giá trị duy nhất, thường là giá trị trung bình (average) hoặc giá trị cực đại (maximum).5

Quá trình giảm mẫu này thực hiện hai mục đích quan trọng. Thứ nhất, nó giảm kích thước dữ liệu, làm giảm chi phí tính toán trong các lớp sau [User Query]. Thứ hai, và quan trọng hơn về mặt lý thuyết, hoạt động pooling mang lại một mức độ bất biến dịch chuyển cục bộ (local translational invariance) cho các đặc trưng.5 Tính bất biến này giúp CNN trở nên vững vàng hơn (robust) đối với các biến thể nhỏ trong vị trí hoặc biến dạng của các đặc trưng, cho phép mô hình nhận dạng các đối tượng ngay cả khi chúng hơi bị dịch chuyển trong ảnh.5

Tuy nhiên, việc giảm độ phân giải không gian thông qua pooling là một sự đánh đổi. Mặc dù pooling tạo ra tính bất biến, nó cũng dẫn đến sự mất mát thông tin vị trí chính xác (spatial location). Sự mất mát này không đáng kể trong các tác vụ phân loại hình ảnh cấp độ toàn cục, nhưng lại trở thành một rào cản kỹ thuật nghiêm trọng đối với các tác vụ định vị mật độ cao (dense prediction tasks) như phân đoạn thực thể, đòi hỏi sự căn chỉnh pixel-to-pixel.6

1.1.3. Lớp Kết nối Đầy đủ và Phân loại Quyết định

Các lớp tích chập và pooling hoạt động như các khối trích xuất đặc trưng (feature extractors). Để hoàn thành tác vụ phân loại, CNN cần các lớp kết nối đầy đủ (FC layers) ở phần cuối của kiến trúc.3

Trước khi đi vào lớp FC, bản đồ đặc trưng 2D/3D cuối cùng được làm phẳng (flattened) thành một vector 1D.4 Lớp FC là lớp cuối cùng của mạng, có chức năng tổng hợp toàn bộ các đặc trưng trừu tượng đã được trích xuất bởi các khối xử lý trước đó.3 Mỗi nơ-ron trong lớp FC này được kết nối với tất cả các đầu vào của lớp trước 3, và cuối cùng, lớp này gán một giá trị xác suất cho hình ảnh thuộc về từng lớp trong số C lớp khả dĩ.3

Một phát hiện quan trọng từ các nghiên cứu gần đây là mối quan hệ giữa độ sâu kiến trúc CNN và nhu cầu thiết kế lớp FC.1 Phân tích cho thấy: Mạng Nông (Shallow CNNs): Do các đặc trưng được trích xuất ở lớp tích chập cuối cùng ít trừu tượng hơn, mạng nông cần một số lượng lớn nơ-ron và nhiều lớp FC hơn để đạt được hiệu suất phân loại tương đương.1 Mạng Sâu (Deeper CNNs): Ngược lại, mạng sâu đã trích xuất được các đặc trưng trừu tượng hóa cao hơn. Do đó, chúng cần ít nơ-ron FC hơn để tổng hợp

thông tin và đưa ra quyết định.1

Việc hình thức hóa mối quan hệ này giữa kiến trúc và tập dữ liệu (xem Phần 2.3) là một bước tiến quan trọng giúp các nhà thực hành chuyển quá trình lựa chọn kiến trúc từ kinh nghiệm (expertise) sang một quy trình thiết kế tự động và có hệ thống.1

1.2. Sự Phát triển Kiến trúc và Nguyên lý Thiết kế Sâu

1.2.1. Các Cột mốc phát triển và Ý nghĩa của Độ sâu Mạng

Sự trỗi dậy của CNN trong thị giác máy tính hiện đại được đánh dấu bằng các kiến trúc cột mốc. AlexNet, được phát triển vào năm 2012, là mô hình CNN sâu đầu tiên được công nhận rộng rãi, nổi bật qua thành tích trong Thử thách Nhận dạng Hình ảnh Quy mô Lớn ImageNet (ILSVRC).7 AlexNet đã chứng minh một cách dứt khoát rằng độ sâu của mô hình là yếu tố thiết yếu để đạt hiệu suất cao, điều này chỉ trở nên khả thi nhờ vào việc tận dụng đơn vị xử lý đồ họa (GPUs) để giảm chi phí tính toán khi huấn luyện.7

Sau thành công ban đầu, các kiến trúc tiếp theo (như VGGNet) tiếp tục đẩy giới hạn về độ sâu. Đồng thời, nghiên cứu cũng tập trung vào việc tối ưu hóa hiệu suất và tham số. Việc đơn giản hóa các kiến trúc dựa trên LeNet đã đạt được những giảm thiểu đáng kể về độ phức tạp tính toán và số lượng tham số trong khi vẫn duy trì hiệu suất cạnh tranh.8 Điều này nhấn mạnh tiềm năng của các kiến trúc hiệu quả (efficient architectures) trong việc giải quyết các ràng buộc phần cứng trong ứng dụng thực tế.8

TÀI LIỆU THAM KHẢO

Tiếng Anh

[1] Al-Waisy, A. S., et al. (2019), "Multi-Scale Inception Based Super-Resolution Using Deep Learning Approach", *Electronics*, Vol. 8(8), pp. 892.

[2] Arnaud58 (n.d.), "selfie2anime - Kaggle", *https://www.kaggle.com/datasets/arnaud58/selfie2a*

[3] Ashraf, S. M. N., Mamun, M. A., Abdullah, H. M., Alam, M. G. R. (2023), "SynthEnsemble: A Fusion of CNN, Vision Transformer, and Hybrid Models for Multi-Label Chest X-Ray Classification", *2023 International Conference on Computer and Information Technology (ICCIT)*, arXiv:2311.07750.

[4] Basha, S. H. S., Dubey, S. R., Pulabaigari, V., Mukherjee, S. (2020), "Impact of Fully Connected Layers on Performance of Convolutional Neural Networks for Image Classification", *Neurocomputing*, Vol. 378, pp. 178-189.

[5] Chen, J., Liu, G., Chen, X. (2020), "AnimeGAN: A Novel Lightweight GAN for Photo Animation", *Artificial Intelligence Algorithms and Applications*, Springer, Singapore, pp. 242-256.

[6] Chen, X., Liu, G. (2020), "AnimeGANv2", *https://tachibanayoshino.github.io/AnimeGANv2/*.

[7] Gholamalinezhad, H., Khosravi, H. (2022), "A Comparison of Pooling Methods for Convolutional Neural Networks", *Applied Sciences*, Vol. 12(17), pp. 8643.

[8] Girshick, R., Donahue, J., Darrell, T., Malik, J. (2014), "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[9] He, K., Gkioxari, G., Dollár, P., Girshick, R. (2017), "Mask R-CNN", *arXiv preprint arXiv:1703.06870*.

[10] Hippocampus's Garden (2021), "Awesome StyleGAN Applications", *https://hippocampus-garden.com/stylegans/*.

[11] Hussain, M., Bird, J. J., Faria, D. R. (2018), "A Study on CNN Transfer Learning for Image Classification", *Advances in Intelligent Systems and Computing*, Vol. 840, Springer, pp. 191–202.

[12] Lin, F. (2025), "Vision Language Models: A Survey of 26K Papers", *arXiv preprint arXiv:2510.09586*.

[13] Liu, G., Chen, X., Gao, Z. (2024), "A Novel Double-Tail Generative Adversarial Network for Fast Photo Animation", *IEICE Transactions on Information and Systems*, Vol. E107.D(1), pp. 72-82.

[14] Liu, G., Chen, X., Gao, Z. (2024), "AnimeGANv3: A Novel Double-Tail Generative Adversarial Network for Fast Photo Animation", *https://tachibanayoshino.github.io/AnimeGANv3/*.

[15] Liu, M.-Y., Breuel, T., Kautz, J. (2017), "Unsupervised Image-to-Image Translation Networks", *Advances in Neural Information Processing Systems (NIPS)*.

[16] Lo, S.-L., Cheng, H.-Y., Yu, C.-C. (2024), "Feature Weighted Cycle Generative Adversarial Network with Facial Landmark Recognition and Perceptual Color Distance for Enhanced Face Animation Generation", *Electronics*, Vol. 13(23), pp. 4761.

[17] Lu, Z., Zhou, Y., Chen, A. (2024), "Enhancing Photo Animation: Augmented Stylistic Modules and Prior Knowledge Integration", *Proceedings of the Asian Conference on Computer Vision (ACCV)*, pp. 1470–1485.

[18] Radford, A., Metz, L., Chintala, S. (2015), "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks", *arXiv preprint arXiv:1511.06951*.

[19] Saad, M. M., O'Reilly, R., Rehmani, M. H. (2024), "A survey on training challenges in generative adversarial networks for biomedical image analysis", *Artificial Intelligence Review*, Vol. 57(2).

[20] Stanford University (n.d.), "CS231n: Deep Learning for Computer Vision", *http://vision.stanford.edu/teaching/cs231n/*.

[21] Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A. A. (2017), "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning", *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pp. 4278-4284.

[22] Unknown (2025), "A Comparative Study of Vision Transformers and CNNs for Few-Shot Rigid Transformation and Fundamental Matrix Estimation", *arXiv preprint arXiv:2510.04794*.

[23] Wang, Y., Wen, R., Ishii, H., Ohya, J. (n.d.), "LAST: Utilizing Synthetic Image Style Transfer to Tackle Domain Shift in Aerial Image Segmentation", *SciTePress*.

[24] Wikipedia (n.d.), "AlexNet", *https://en.wikipedia.org/wiki/AlexNet*.

[25] Wikipedia (n.d.), "Convolutional neural network", *https://en.wikipedia.org/wiki/Convolutional_neural_network*.

[26] Yang, S., Jiang, L., Liu, Z., Loy, C. C. (2022), "Pastiche Master: Exemplar-Based High-Resolution Portrait Style Transfer", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11728-11737.

[27] Zhang, T., Tang, H. (2025), "Style Transfer: A Decade Survey", *arXiv preprint arXiv:2506.19278*.

[28] Zhou, Z., Siddiquee, M. R., Tajbakhsh, N., Liang, J. (2018), "UNet++: A Nested U-Net Architecture for Medical Image Segmentation", *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Vol. 11045, Springer, Cham, pp. 3-11.

[29] Zhu, J.-Y., Park, T., Isola, P., Efros, A. A. (2017), "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.