**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT THÀNH PHỐ HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**

Text

Description automatically generated

**BÁO CÁO TIỂU LUẬN CHUYÊN NGÀNH**

**NGÀNH: CÔNG NGHỆ PHẦN MỀM**

**ĐỀ TÀI: Ứng dụng Deep Learning xây dựng model tạo hình ảnh khuôn mặt người dựa trên miêu tả**

**GVPB: Ts Huỳnh Xuân Phụng**

**GVHD: PGS Ts. Hoàng Văn Dũng**

**SVTH:**

**Phạm Nguyễn Hải Dương - 19110343**

**Nguyễn Tấn Hào - 19110358**

**TP. HCM, tháng 12 năm 2022**

**BỘ GIÁO DỤC VÀ ĐÀO TẠO**

**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT THÀNH PHỐ HỒ CHÍ MINH**

**KHOA CÔNG NGHỆ THÔNG TIN**

Text

Description automatically generated

**BÁO CÁO TIỂU LUẬN CHUYÊN NGÀNH**

**NGÀNH: CÔNG NGHỆ PHẦN MỀM**

**ĐỀ TÀI: Ứng dụng Deep Learning xây dựng model tạo hình ảnh khuôn mặt người dựa trên miêu tả**

**GVPB: Ts Huỳnh Xuân Phụng**

**GVHD: PGS Ts. Hoàng Văn Dũng**

**SVTH:**

**Phạm Nguyễn Hải Dương - 19110343**

**Nguyễn Tấn Hào - 19110358**

**TP. HCM, tháng 12 năm 2022**

**NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN**

**LỜI CẢM ƠN**

**MỤC LỤC**

[Mở đầu 1](#_Toc121426609)

[1. Lý do chọn đề tài 1](#_Toc121426610)

[2.Đối tượng và phạm vi nghiên cứu 1](#_Toc121426611)

[3.Phương pháp nghiên cứu 2](#_Toc121426612)

[4.Mục đích của đề tài 2](#_Toc121426613)

[5.Bố cục của đề tài 2](#_Toc121426614)

[Chương 1.Cơ sở lý thuyết 3](#_Toc121426615)

[1.1 Genrative Adversarial networks - GANs 3](#_Toc121426616)

[1.1.1 Hàm mất mát 5](#_Toc121426617)

[1.1.2 BCE hàm chi phí 7](#_Toc121426618)

[1.2 StyleGAN 8](#_Toc121426619)

[1.3 Kỹ thuật đánh giá mô hình 9](#_Toc121426620)

[1.4 Encoder 10](#_Toc121426621)

[Chương 2. Áp dụng kiến thức xây dựng mô hình 12](#_Toc121426622)

[2.1 Mô hình 12](#_Toc121426623)

[2.1.1 Kiến trúc 12](#_Toc121426624)

[2.1.2 Mạng mapping f 13](#_Toc121426625)

[2.1.3 Khối mạng generator 13](#_Toc121426626)

[2.2 Tập dữ liệu huấn luyện 15](#_Toc121426627)

[Chương 3. Kết quả 16](#_Toc121426628)

[3.1 Hình ảnh được sinh ra từ mô hình 16](#_Toc121426629)

[3.2 Đánh giá mô hình 19](#_Toc121426630)

[Chương 4. Kết luận 21](#_Toc121426631)

[Chương 5. Tài liệu tham khảo 22](#_Toc121426632)

# Mở đầu

## Lý do chọn đề tài

Một trong những xu hướng nghiên cứu thu hút được đông đảo các nhà khoa học, có tính ứng dụng cao và phát triển mạnh mẽ trong những năm gần đây trong Deep Learning có lẽ là GAN.

Kiến trúc GAN đầu tiên được Ian GoodFellow giới thiệu vào năm 2014, để dễ hiểu hơn cho người bắt đầu. Các kiến trúc GAN nâng cao hơn sẽ được mình trình bày ở các bài sau. Những kiến trúc này vẫn dựa trên ý tưởng chủ đạo của model GAN đầu tiên nhưng có sự cải tiến đầu vào, phương pháp huấn luyện, hàm loss function một chút để kết quả học được tốt hơn.

Công nghệ trong thời đại số còn có thể ứng dụng trong nhiều ngành nghề khác, ví dụ như Gaming chẵn hạn. Thay vì thiết kế từng nhân vật sẽ tốn nhiều thời gian và nguồn lực, chỉ cần nhập mô tả nhân vật và hệ thống sẽ tự động tạo ra một nhân vật đúng với mô tả.

Ví dụ trên là một ví dụ điển hình của đề tài nghiên cứu. Nhóm nghiên cứu về công nghệ để giải quyết một bài toán nhỏ hơn là tạo ra hình ảnh khuôn mặt con người dựa trên các đặc tính mô tả bằng ngôn ngữ tự nhiên.

## Đối tượng và phạm vi nghiên cứu

Đối tượng nghiên cứu: Nghiên cứu thuật toán để tổng hợp hình ảnh.

Phạm vi nghiên cứu:

Thời gian: thời gian nghiên cứu trong khuôn khổ thời gian của học phần Tiểu Luận Chuyên Ngành.

Nội dung: trong tiểu luận này nhóm thực hiện chủ yếu quan tâm tới vấn đề có thể tạo được hình ảnh từ miêu tả. Chứ không tập trung vào phần mã hóa các câu/từ ngữ của hình ảnh.

## Phương pháp nghiên cứu

Trong tiểu luận này, nhóm em sử dụng phương pháp phân tích, phương pháp nghiên cứu thực tiễn làm phương pháp nghiên cứu chủ đạo nhằm làm rõ quy trình huấn luyện và tạo ra hình ảnh.

## Mục đích của đề tài

Mục đích của đề tài là huấn luyện được một mô hình có thể đạt được những yêu cầu sau:

Sinh ra được hình ảnh khuôn mặt người với đầu vào là miêu tả bằng ngôn ngữ tự nhiên.

Những hình ảnh phải nằm trong khoản chấp nhận được của người nhìn.

Đạt được kết quả tốt khi đánh giá bằng thuật toán khác.

## Bố cục của đề tài

Bài báo cáo được chia làm 5 phần chính:

Phần 1: Cơ sở lý thuyết: những kiến thức lý thuyết mà nhóm đã nghiên cứu trong lúc làm tiểu luận.

Phần 2: Áp dụng kiến thức vào quá trình xây dựng mô hình trình bày cách nhóm thực hiện và kết quả của quá trình nghiên cứu.

Phần 3: kết quả đạt được

Phần 4: Kết luận: kết luận của quá trình nghiên cứu, những cái đã đạt được và phần chưa được, cũng như là rút kinh nghiệm cho bài Khóa Luận Tốt Nghiệp.

Phần 5:  Danh mục tài liệu tham khảo.

# Cơ sở lý thuyết

## Genrative Adversarial Networks - GANs

Genrative Adversarial Networks (mạng sáng tạo đối nghịch) với tên ngắn phổ biến là GANs là một trong các mô hình sinh mẫu sử dụng kiến trúc của mạng nơ-ron nhân tạo của học sâu.

GANs bao gồm hai mạng no-ron cạnh tranh với nhau gọi là generator và discriminator.

Mỗi mô hình đảm nhận một chức năng khác nhau. Với mô hình sinh sẽ sinh ra thực thể mới từ dữ liệu đầu vào và mô hình phân biệt sẽ phân biệt thực thể được tạo bởi mô hình sinh là thật hay giả.

Khởi đầu là bài toán về học không giám sát nhưng với kiến trúc được thiết kế thông minh nên GANs được huấn luyện như một bài toán học có giám sát. Hai mô hình được huấn luyện đồng thời và mô hình này sẽ học được từ mô hình kia và ngược lại.

Graphical user interface, qr code

Description automatically generated

* Mô hình phân biệt

Mô hình phân biệt: mục tiêu của mô hình phân biệt là giải bài toán phân lớp cho đầu vào là các thực thể thật và giả. Thực thể thật được lấy từ tập dữ liệu huấn luyện còn thực thể giả được lấy từ mô hình sinh.

Diagram

Description automatically generated

(backpropagation trong mô hình phân biệt)

Quy trình huấn luyện mô hình phân biệt:

Bước 1: Mô hình  sẽ phân loại hình ảnh thật và hình ảnh giả từ mô hình sinh.

Bước 2: Mô hình sẽ tính toán độ sai cho những cái phân loại ví dụ thật của giả hoặc ví dụ giả của thật.

Bước 3: Mô hình phân loại sẽ cập nhật những trọng số thông qua cơ chế backpropagation từ sự mất mát của mô hình phân biệt.

* Mô hình sinh

Mô hình sinh: Mục tiêu của mô hình này là sinh ra thực thể mới dựa trên đầu vào là tập dữ liệu huấn luyện. Và mô hình này sẽ cố gắng đánh lừa mô hình phân biệt thực thể sinh ra là thật.

Diagram

Description automatically generated

 Quá trình huấn luyện:

Bước 1: chương trình sẽ tạo ngẫu nhiên *noise*

Bước 2: sinh hình ảnh từ noise ở bước 1.

Bước 3: Nhận về giá trị thật hay giả của các hình ảnh ở bước 2 của quá trình phân loại trong mô hình phân biệt.

Bước 4: Tính toán mất mát từ sự phân loại trên.

Bước 5: Sử dụng cơ chế backpropagate cho cả mô hình sinh và mô hình phân biệt để lấy được gradient.

Bước 6: Sử dụng gradient ở bước 5 và thay đổi các trọng số trong mô hình sinh.

### Hàm mất mát

Z - là nhiễu đầu vào của mô hình sinh

X - là giữ liệu thật từ data set

G - là mô hình sinh

D - là mô hình phân biệt

G(z) là ảnh được sinh ra từ mô hình sinh

D(x) là giá trị dự đoán của hàm phân biệt xem x là thật hay không

D(G(z)) là giá trị dự đoán thực thể được sinh ra từ mô hình sinh là thật hay không.

Diagram

Description automatically generated

Vì có tới hai mô hình khác nhau nên có 2 hàm mất mát dành riêng cho mỗi mô hình.

Với hàm phân biệt thì cố gắng phân biệt ảnh thật ảnh giả. Nên có thể quy về bài toán phân loại nhị phân nên hàm mất mát là BCE. Có thể dễ dàng đoán được giá trị đầu ra của mô hình phân loại gần tới 1 khi đầu vào là tập huấn luyện còn đầu ra sẽ gần 0 khi tập huấn luyện từ mô hình phân loại. Hay D(x) 🡪 1 còn D(G(z)) 🡪 0

Hay nói một cách khác là hàm mất mát sẽ muốn maximize D(z) và minimize D(G(z)), mà minimize D(G(z)) tương đương với maximize(1 – D(G(z))).

Text

Description automatically generated with medium confidence

 Mô hình sinh sẽ cố đánh lừa mô hình phân biệt rằng thực thể nó sinh ra là thật hay D(G(z)) 🡪 1.

Hay hàm chi phí muốn D(G(z)) đạt được kết quả lớn nhất tương đương với 1 - D(G(z)) sẽ là nhỏ nhất.

A picture containing text

Description automatically generated

Từ hai công thức trên có thể thấy được hai mô hình đối nghịch và học hỏi lẫn nhau giống như trò chơi minimax. và công thức tổng quát cho quá trình này là:

Text

Description automatically generated

Quá trình huấn luyện GANs kết thúc khi mô hình GANs đạt đến trạng thái cân bằng của hai mô hình sinh và phân loại gọi là Nash Equilibrium

### BCE hàm chi phí

BCE - Binary cross entropy là một hàm chi phí thường được sử dụng với GANs với hai lớp đặc biệt là 1 - cho thực thể là thật và  0 - cho thực thể là giả.

Công thức của hàm chi phí BCE:

Text, letter

Description automatically generated

Với :

|  |  |
| --- | --- |
| A picture containing text, clock  Description automatically generated | cho ra giá trị trung bình của tất cả giá trị trong mỗi batch . |
|  | khi so sánh nhãn thật với cái được dự đoán là thật |
|  | khi so sánh nhãn thật với cái được dự đoán là giả. |

Khi nhãn và phần dự đoán giống nhau giá trị của hàm chi phí sẽ tiến gần đến 0 và sẽ đến vô cực nếu khác nhau

## StyleGAN

StyleGan là mô hình được cải tiến so với GANS. StyleGAN được giới thiệu lần đầu bởi NVIDIA vào năm 2018.

  Với StyleGAN thì mô hình generator được cải tiến với một kiến trúc mới cho phép điều khiển các mức độ chi tiết của một bức ảnh với khuôn mặt thì  có thể chỉnh các chi tiết thô như màu tóc, kiểu tóc, hay những chi tiết nhỏ hơn như màu mắt, khuyên tai…

Ngoài ra cả ở StyleGAN thì quy trình huấn luyện hai mô hình có sự thay đổi đó là cả hai mạng generator và discriminator là ban đầu hai mạng này sẽ được huấn luyện trên ảnh 4x4 sau nhiều lớp sẽ dần được thêm vào kích thước ảnh cũng dần tăng lên. Từ đó có thể giảm được thời gian huấn luyện và quá trình huấn luyện sẽ ổn định hơn.

Như đã nói ở trên StyleGAN có khả năng điều khiển được các mức độ chi tiết khác nhau bằng cách sử dụng thêm một mạng mapping để mã hóa vector z (từ phân phối chuẩn nhiều chiều) thành 1 vector w. vector w sau đó sẽ được được đưa vào nhiều vị trí khác nhau trong mạng generator, tại mỗi vị trí vector w sẽ điều khiển các đặc trưng khác nhau.

Diagram

Description automatically generated

Nhìn vào kiến trúc trên thấy các lớp trên đầu có độ phân giải 4x4, 8x8, ở các vị trí này sẽ kiểm soát các đặc trưng thô như dáng đầu, kiểu tóc, khuôn mặt. Tiến dần đến cuối là các lớp có độ phân giải cao hơn 512x512, 1024x1024 sẽ kiểm soát các đặc trưng kết cấu của khuôn mặt như màu da, màu tóc, màu mắt…

## Kỹ thuật đánh giá mô hình

Mục tiêu của thuật toán StyleGAN là tạo ra được mô hình sinh ảnh giống với ảnh thật nhất. Với bình thường thì để đánh giá chất lượng của ảnh là dựa vài mắt để đánh giá. Nhưng với tập giữ liệu lớn thì việc đánh giá bằng mắt là không khách quan.Trong bài nghiên cứu này nhóm thực hiện sử dụng  Fréchet Inception Score - FID để đánh giá mô hình.

Để đánh giá chất lượng hình ảnh được sinh ra cần dựa vào 2 yếu tố:

Chất lượng hình ảnh: hình ảnh chất lượng cao (giống hình mặt người nhất)

Độ đa dạng của hình ảnh được sinh ra.

FID - Fréchet Inception Score

Diagram

Description automatically generated

Cách tính FID phụ thuộc vào mạng inception. Như vậy với mỗi ảnh qua mạng sẽ cho ra kích thước 2048x1 chứa các đặc trưng của ảnh.

 Các bước tiến hành:

Đầu tiên sẽ lấy toàn bộ ảnh chạt qua mạng Inception. Với mỗi ảnh qua thì sẽ được 1 vector 2048x1, ta cũng sẽ tìm được 1 phân phối Gauss nhiều chiều (multivariate gaussian distribution) tương ứng với tất cả những vector này với mean = 

Cho tất cả ảnh sinh ra từ mô hình generator mỗi ảnh qua mạng inception sẽ được 1 vector 2048x1, và cũng sẽ có 1 phân phối gauss nhiều chiều với



Để các ảnh sinh ra giống với các ảnh trong tập dữ liệu thì phải cho 2 cái phân phối gauss nhiều chiều tương đương với mean và phương sai gần nhau từ đó có được công thức:

Text

Description automatically generated with low confidence

Từ đó nếu FID không âm và càng nhỏ thì 2 giá trị phân phối trên càng gần nhau. Ảnh sinh ra càng giống ảnh gốc trong tập dữ liệu.

## Encoder

Những ngôn ngữ tự nhiên máy không thể hiểu được. Vì vậy cần có 1 cơ chế để chuyển đổi từ ngôn ngữ tự nhiên thành mã để máy có thể hiểu được. Vì vậy nhóm quyết định sử dụng mạng CNN để chuyển đổi.

Về cơ bản mạng dùng để encode có kiến trúc như sau:

Diagram

Description automatically generated

Với một câu miêu tả thì qua mô hình này sẽ được chuyển đổi thành một vector chữa các số thứ tự từ điển.

# Áp dụng kiến thức xây dựng mô hình

## Mô hình

### Kiến trúc

Diagram

Description automatically generated

Nhìn vào kiến trúc trên có thể thấy được StyleGAN có 2 phần riêng rẽ mapping bên trái và phần tổng hợp các mạng thành 1 khối độc lập bên phải.

### Mạng mapping f

Mục tiêu của mạng ánh xạ là mã hóa vectơ đầu vào thành một vectơ trung gian có các yếu tố khác nhau kiểm soát các tính năng trực quan khác nhau. Đây là một quy trình khó khăn  vì khả năng kiểm soát các tính năng trực quan với vectơ đầu vào bị hạn chế, vì nó phải tuân theo mật độ xác suất của dữ liệu đào tạo. Ví dụ: nếu hình ảnh của những người có mái tóc đen phổ biến hơn trong bộ dữ liệu, thì nhiều giá trị đầu vào sẽ được ánh xạ tới tính năng đó. Do đó, mô hình không có khả năng ánh xạ các phần của đầu vào (các phần tử trong vectơ) đến các tính năng, một hiện tượng được gọi là tính năng vướng víu. Tuy nhiên, bằng cách sử dụng một mạng thần kinh khác, mô hình có thể tạo ra một vectơ không phải tuân theo phân phối dữ liệu đào tạo và có thể làm giảm mối tương quan giữa các tính năng.

Mạng ánh xạ bao gồm 8 lớp được kết nối đầy đủ và đầu ra của nó có cùng kích thước với lớp đầu vào.

Lantent space: là một không gian tiềm ẩn dùng để tập hợp và sắp xếp dữ liệu. Cho phép thuật toán hình dung và phân tích đối tượng dựa theo những thông tin  có sẵn. Nó vừa nén kích thước của dữ liệu, vừa sắp xếp các dữ liệu có độ tương đồng cao ở gần nhau dưới dạng các điểm dữ liệu hay áp dung trong bài nghiên cứu là các đặc điểm của khuôn mặt.

Trong quá trình nén dữ liệu, những thông tin không liên quan sẽ được loại bỏ, chỉ để lại những thông tin cơ bản giúp phân biệt các nhóm thông tin với nhau

### Khối mạng generator

StyleGan sử dụng nhiều khối tổng hợp những khối mà sử dụng hình ảnh tăng dần độ phân giải từ 4x4, 8x8 to 512x512...

Mỗi khối của StyleGAN gồm nhưng thành phần quan trọng sau:

* Upsampling ngoại trừ khối đầu tiên không có. Đầu ra của khối trước nó sẽ được tăng lên cho khối tổng hợp đó sử lý.
* Convolution layer.
* Adaptive Instance Normalization (AdaIN).
* Style vectors (A) and noise vectors (B).

Đầu tiên là khối Constant là nơi mà hình ảnh được xử lý đầu tiên. Đây cũng là phần làm nên sự khác biệt của StyleGan với các loại GAN khác.

Thành phần quan trọng nhất của quá trình tổng hợp là Style và noise:

* Style: hằng số này giống như là xương sống được sử dụng ở tất cả các khối tổng hợp trong mạng.Nó đáp ứng cho những đăc điểm trỏng ảnh được sinh ra.
* Noise: việc huấn luyện mô hình sinh được thực hiện lập đi lập lại bởi một tập dữ liệu duy nhất. Sau mỗi lần huấn luyện thì ta có thể nhận được một hình duy nhất. Vậy nên vector này đảm bảo các bước ảnh được tạo ra có sự ngẫu nhiên và đa dạng.

Diagram

Description automatically generated

Với GAN thì có 2 loại chi tiết:

* Thành phần có độ chi tiết thấp như hình dạng khuôn mặt, kiểu tóc...
* Thành phần có độ chi tiết cao như: mắt, miệng..

Noise vector được xác định là thành phần có độ chi tiết cao.

Adaptive Instance Normalization (AdaIN) Mô -đun chuyển thông tin được mã hóa ⱳ, được tạo bởi mạng ánh xạ, vào hình ảnh được tạo. Mô -đun được thêm vào từng mức độ phân giải của mạng tổng hợp và xác định biểu thức trực quan của các tính năng ở cấp độ đó:.

Mỗi kênh của đầu ra lớp tích chập trước tiên được chuẩn hóa để đảm bảo tỷ lệ và dịch chuyển của bước 3 có hiệu ứng dự kiến.

Vectơ trung gian được chuyển đổi bằng cách sử dụng một lớp kết nối đầy đủ khác (được đánh dấu là A) thành một thang đo và sai lệch cho mỗi kênh.

Các vectơ tỷ lệ và độ lệch thay đổi từng kênh của đầu ra tích chập, do đó xác định tầm quan trọng của từng bộ lọc trong tích chập. Điều chỉnh này dịch thông tin từ ⱳ sang biểu diễn trực quan.

Diagram

Description automatically generated

Upsampling: những thành phần trên chủ tập trung chủ yếu vào và khối đầu tiên của quá trình tổng hợp. Từ khối thứ 2 trở đi có thêm phần Upsampling đây là thành phần để điều khiển toàn bộ khối con. Và từ những khối có thành phần này thì cách hoạt động khác với khối đầu tiên. Upsampling là một khối bao gồm nhiều kỹ thuật giúp tăng kích thước đầu ra trong mạng CNN.

Quy trình hoạt động như sau:

* Đầu tiên việc lấy mẫu song song được áp dụng để lấy mẫu hình ảnh, sap đó một lớp 2D với hạt nhân 3x3 được sử dụng để lấy mẫu đã học.
* Sau đó vector noise được đưa thêm vào sau khi AdalN được diễn ra.
* Sau đó downsampling được diễn ra với lớp 3x3 như bước 1. sau đó tiếp tục lập lại quy trình.

## Tập dữ liệu huấn luyện

Tập dữ liệu dùng để huấn luyện là CelebAMASH-HQ. Tập dữ liệu bao gồm 30000 hình ảnh người nổi tiếng. Và file miêu tả các đặc điểm trên khuôn mặt. Tập dữ liệu được chia thành 24000 hình ảnh  dùng để huấn luyện mô hình và 6000 hình ảnh để kiểm tra. Với mỗi hình ảnh sẽ có 10 câu miêu tả về các đặc trưng của khuôn mặt. 

# Kết quả

## Hình ảnh được sinh ra từ mô hình

Miêu tả: AA young man with a beard

Hình ảnh:

A person with a beard

Description automatically generated with low confidenceA person with a beard

Description automatically generated with medium confidenceA person with a beard

Description automatically generated with medium confidenceA person with a beard

Description automatically generated with medium confidence

Miêu tả: This man has bangs arched eyebrows curly hair and a small nose

A person with curly hair

Description automatically generatedA person with curly hair

Description automatically generatedA person with curly hair

Description automatically generated

Miêu tả The woman has pointy nose. She is young and is wearing lipstick.

Hình ảnh:

A picture containing person, music, blue

Description automatically generated

A picture containing person, indoor

Description automatically generated

Miêu tả: The person has arched eyebrows, wavy hair, oval face, brown hair, and mouth slightly open. She is attractive and is wearing lipstick.

Hình ảnh:

A person with red lipstick

Description automatically generated with low confidenceA person with red lipstick

Description automatically generated with medium confidenceA person with red lipstick

Description automatically generated with medium confidence

## Đánh giá mô hình

Tương quan giữa hàm chi phí và epoch

Chart, histogram

Description automatically generated

Điểm FID:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tập dữ liệu | Số lượng hình ảnh huấn luyện | GPU | Số ngày huấn luyện | FID |
| Celeb - A | 24000 | V100 / A100 Nvidia GPU. | 12 | 5.17 |

# kết luận

StyleGAN là một thuật toán đột phá không chỉ tạo ra hình ảnh chất lượng cao và thực tế mà còn cho phép kiểm soát vượt trội và hiểu biết về hình ảnh được tạo ra, làm cho nó thậm chí còn dễ dàng hơn trước để tạo ra hình ảnh giả đáng tin cậy. Các kỹ thuật được trình bày trong Stylegan, đặc biệt là mạng ánh xạ và chuẩn hóa thích ứng (ADAIN), có thể sẽ là cơ sở cho nhiều đổi mới trong tương lai trong Gans.

# tài liệu tham khảo

1. NVIDIA and Aalto University. (2020). Analyzing and Improving the Image Quality of StyleGAN.

Retrieved December 8, 2022, from <https://arxiv.org/pdf/1912.04958.pdf>

1. Source code for torchnlp.nn.cnn\_encoder. (n.d.). Pytorch.

Retrieved December 8, 2022, from <https://pytorchnlp.readthedocs.io/en/latest/_modules/torchnlp/nn/cnn_encoder.html>

1. Brownlee, J. (2019, August 30). How to Implement the Frechet Inception Distance (FID) for Evaluating GANs. Machine Learning Mastery.

Retrieved December 8, 2022, from https://machinelearningmastery.com/how-to-implement-the-frechet-inception-distance-fid-from-scratch/