

**BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC SƯ PHẠM TP HỒ CHÍ MINH**

**BÁO CÁO TỔNG KẾT  
ĐỀ TÀI NGHIÊN CỨU KHOA HỌC CỦA SINH VIÊN**

**XÂY DỰNG MÔ HÌNH THIẾT KẾ TÚI XÁCH BẰNG  
GENERATIVE ADVERSARIAL NETWORKS**

**<Mã số đề tài>**

Thuộc nhóm ngành khoa học: Công nghệ thông tin

**TP Hồ Chí Minh, 4/2021**

**BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC SƯ PHẠM TP HỒ CHÍ MINH**

**BÁO CÁO TỔNG KẾT  
ĐỀ TÀI NGHIÊN CỨU KHOA HỌC CỦA SINH VIÊN**

**XÂY DỰNG MÔ HÌNH THIẾT KẾ TÚI XÁCH BẰNG  
GENERATIVE ADVERSARIAL NETWORKS**  
<Mã số đề tài>

Thuộc nhóm ngành khoa học: Công nghệ thông tin

SV thực hiện: Nguyễn Xuân Tính    Nam, Nữ: Nam  
Dân tộc: Kinh  
Lớp, khoa:    44.CNTT.C    Năm thứ: 3 /Số năm đào tạo: 4  
Ngành học:    Công nghệ thông tin

SV thực hiện: Lâm Phát Tài    Nam, Nữ: Nam  
Dân tộc: Kinh  
Lớp, khoa:    44.CNTT.C    Năm thứ: 3 /Số năm đào tạo: 4  
Ngành học:    Công nghệ thông tin

Người hướng dẫn: ThS. Trần Thanh Nhã

**TP Hồ Chí Minh, 4/2021**

# MỤC LỤC

MỤC LỤC .....	2
DANH MỤC CÁC KÝ HIỆU VÀ CHỮ VIẾT TẮT .....	3
DANH MỤC CÁC HÌNH VẼ, BIỂU ĐỒ .....	4
<b>Chương I. GIỚI THIỆU .....</b>	<b>5</b>
1. Tổng quan, động lực nghiên cứu và tính cấp thiết của đề tài.....	5
2. Mục tiêu nghiên cứu .....	6
3. Đối tượng nghiên cứu .....	7
4. Cách tiếp cận, phương pháp nghiên cứu .....	7
a. Cách tiếp cận .....	7
b. Phương pháp nghiên cứu.....	8
5. Nội dung nghiên cứu .....	8
a. Tổng quan mô hình Generative Adversarial Networks: .....	8
b. Hàm mục tiêu .....	11
<b>Chương II. TÌNH TRẠNG NGHIÊN CỨU .....</b>	<b>13</b>
1. Ứng dụng tạo hình nhân vật Anime .....	13
2. CycleGAN .....	13
3. Super resolution.....	14
4. StyleGAN2 .....	15
5. Text to Image (StackGAN) .....	15
<b>Chương III. XÂY DỰNG BỘ DỮ LIỆU .....</b>	<b>17</b>
1. Tình trạng cơ sở dữ liệu.....	17
2. Quá trình thu thập dữ liệu.....	18
<b>Chương IV. XÂY DỰNG MÔ HÌNH THIẾT KẾ TÚI XÁCH BẰNG GENERATIVE ADVERSARIAL NETWORKS.....</b>	<b>22</b>
1. Quá trình sinh .....	22
2. Quá trình phân biệt .....	24
<b>Chương V. THỰC NGHIỆM VÀ ĐÁNH GIÁ .....</b>	<b>26</b>
1. Kết quả thực nghiệm.....	26
2. Đánh giá.....	28
<b>Chương VI. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN .....</b>	<b>29</b>
<b>Tài liệu tham khảo .....</b>	<b>30</b>

## DANH MỤC CÁC KÝ HIỆU VÀ CHỮ VIẾT TẮT

Chữ viết tắt	Nguyên mẫu	Diễn giải
GAN	Generative Adversarial Networks	Mạng sinh đối kháng
CNTT		Công nghệ thông tin
G	Generator	Mô hình sinh
D	Discriminator	Mô hình phân biệt

## DANH MỤC CÁC HÌNH VẼ, BIỂU ĐỒ

Hình 1. Sơ đồ giải thích kiến trúc Generative Adversarial Network [Thalles Silva]	8
Hình 2. Quá trình cải thiện của mô hình sinh .....	9
Hình 3. Quá trình cải thiện của mô hình phân biệt .....	10
Hình 4. Quá trình cải thiện của mô hình sinh và mô hình phân biệt [Goodfellow, 2014] .....	12
Hình 5. Ảnh nhân vật Anime được sinh ra [Jin, 2017].....	13
Hình 6. Sơ đồ chuyển đổi một bức ảnh thành bức tranh mang phong cách của Van Gogh [Zhu, 2017].....	14
Hình 7. Ứng dụng của GAN trong việc cải thiện chất lượng ảnh [Ledig, 2017] ....	14
Hình 8. Ảnh mặt người được sinh ra bằng StyleGAN [Karras, 2020] .....	15
Hình 9. Những bông hoa được tạo ra từ câu mô tả [Zhang, 2017].....	16
Hình 10. Các loại túi có trong dataset.....	17
Hình 11. Các nguồn dữ liệu chính .....	18
Hình 12. Một số ảnh nhiễu và ảnh đạt yêu cầu.....	19
Hình 13. Ảnh túi xách đã xử lý Padding.....	20
Hình 14. Cấu trúc lưu trữ của bộ dữ liệu .....	21
Hình 15. Cách đặt tên ảnh trong bộ dữ liệu .....	21
Hình 16. Hình minh họa cho ảnh nhiễu .....	22
Hình 17. Cấu trúc của mô hình sinh .....	23
Hình 18. Một số ảnh minh họa cho kết quả đầu ra .....	26
Hình 19. Giao diện của hệ thống thiết kế túi xách.....	27
Hình 20. Ảnh đầu ra bị khuyết điểm.....	27
Hình 21. Hình minh họa giá trị của hàm mục tiêu trong quá trình huấn luyện .....	28

# CHƯƠNG I. GIỚI THIỆU

## 1. Tổng quan, động lực nghiên cứu và tính cấp thiết của đề tài

Hiện nay, ngành công nghệ thông tin đang rất phát triển, đặc biệt là trong cuộc cách mạng công nghiệp 4.0, công nghệ số góp mặt trong nhiều lĩnh vực, xuất hiện trong cuộc sống thường nhật của mỗi chúng ta. Bên cạnh đó, ngành công nghiệp thời trang cũng phát triển không kém do nhu cầu nâng cao vẻ đẹp ngoại hình của phần lớn người trong xã hội ngày nay. Công nghệ thông tin và thời trang là hai ngành công nghiệp tưởng chừng như tách biệt với nhau, nhưng với sự phát triển lớn mạnh của ngành CNTT thì việc công nghệ số tham gia vào ngành công nghiệp thời trang là hoàn toàn có thể.

Trong ngành công nghiệp thời trang nói chung và lĩnh vực thiết kế các sản phẩm thời trang nói riêng, chúng tôi nhận thấy quá trình thiết kế luôn luôn phải được thực hiện bởi bàn tay con người và quá trình này tiêu tốn rất nhiều thời gian và công sức. Các nhà thiết kế phải làm mọi việc một cách thủ công từ những nét vẽ cơ bản nhất. Với suy nghĩ này, chúng tôi mong muốn xây dựng một hệ thống thiết kế túi xách tự động có thể thay thế một phần công việc của các nhà thiết kế, giúp các nhà thiết kế nâng cao năng suất làm việc của mình, nhờ vào “vẽ” thêm các ý tưởng của riêng mình lên trên một thiết kế đã được sinh ra sẵn.

Với tốc độ phát triển của ngành công nghiệp thời trang như hiện nay, các hãng thời trang liên tục phải cho ra các sản phẩm mới và phải bắt kịp xu hướng thời trang. Việc thiết kế đóng một vai trò vô cùng quan trọng trong việc tạo nên thành công của một sản phẩm bên cạnh các yếu tố khác như marketing, giá thành hay chất lượng của sản phẩm,...Nắm bắt được nhu cầu đó, chúng tôi đã quyết định lựa chọn một mô hình học sâu đang được các nhà phát triển rất quan tâm hiện nay là **Generative Adversarial Networks** (GAN). Mô hình GAN được biết đến như là một công cụ giúp chúng ta tạo ra những hình ảnh mới từ những hình ảnh có sẵn trước đó, trong trường hợp của chúng tôi chính là tạo nên những thiết kế của túi xách. Thực tế đã chứng minh GAN là một mô hình tốt để thực hiện tác vụ này dựa

trên những công trình đã được công bố trước đó như: nhóm các tác giả đề xuất hệ thống mô hình lấy các xu hướng thời trang mới nhất và những bộ trang phục được người dùng mua làm đầu vào và tạo ra những bộ trang phục mới. Những bộ trang phục mới này sẽ dựa trên xu hướng thời trang hiện hành nhưng đồng thời sẽ có các đặc trưng của quần áo mà người dùng đã mua trước đó [Singh, 2019].

FashionGAN đơn giản hóa quy trình thiết kế quần áo theo kiểu truyền thống, nó trực tiếp tạo ra một số hình ảnh sắc nét cụ thể từ các bản phác thảo và mẫu vải. Ngoài ra FashionGAN có tính năng tái sử dụng có nghĩa là dựa trên các mẫu thiết kế hiện có, các nhà thiết kế có thể được sử dụng lại để tạo ra các hình ảnh quần áo mới [Cui, 2018].

Những hệ thống đã được nêu trên đều giải quyết rất tốt những nút thắt về vấn đề thiết kế sản phẩm trong ngành may mặc quần áo.

Với những thành tựu mà Generative Adversarial Networks đã đạt được chúng tôi tin tưởng mô hình này sẽ là một phương pháp hiệu quả giúp chúng tôi xây dựng một hệ thống thiết kế túi xách tự động của riêng mình. Chúng tôi lựa chọn đề tài này với mong muốn giúp đỡ các nhà thiết kế tạo ra những thiết kế mới từ những sản phẩm túi xách mà họ đã tạo ra trước đây hoặc những sản phẩm đang có trên thị trường. Việc này giúp các nhà thiết kế tránh được việc làm đi làm lại một công việc gần như tương tự nhau vì thế giúp hạn chế sự nhàm chán trong công việc và giảm được thời gian tạo ra một sản phẩm mới, từ đó giúp họ có nhiều thời gian để suy nghĩ ra những ý tưởng mới góp phần tô điểm thêm cho những thiết kế của mình qua đó góp phần nâng chất lượng đầu ra của các sản phẩm túi xách.

## 2. Mục tiêu nghiên cứu

Sau khi đã lên ý tưởng về hệ thống thiết kế túi xách tự động và tìm được một mô hình thích hợp là GAN, nhóm nghiên cứu hướng đến các mục tiêu quan trọng trong đề tài nghiên cứu này là:

- Tìm hiểu và áp dụng mô hình học máy **Generative Adversarial Networks** [Goodfellow, 2014] do Ian Goodfellow và cộng sự thiết kế vào năm 2014.

- Từ mô hình GAN nguyên bản, nhóm chúng tôi tiến hành thử nghiệm và thực hiện một số biến thể khác của GAN.
- Xây dựng một bộ dữ liệu về các loại túi xách phổ biến.
- Xây dựng hệ thống thiết kế túi xách tự động dựa trên mô hình GAN và bộ dữ liệu đã được sưu tầm.

Đó là những mục tiêu chính mà nhóm đề ra trong đề tài nghiên cứu lần này. Tuy những mục tiêu được đề ra này không quá to lớn nhưng chúng tôi hy vọng sẽ đóng góp được một ít khả năng của bản thân mình để có thể giúp đỡ các nhà thiết kế trong việc tạo ra và nâng cao chất lượng sản phẩm.

### **3. Đối tượng nghiên cứu**

- Một số loại túi xách phổ biến trên thị trường
- Mô hình Generative Adversarial Networks

### **4. Cách tiếp cận, phương pháp nghiên cứu**

#### **a. Cách tiếp cận**

Sau khi được giới thiệu vào năm 2014, Generative Adversarial Networks [Goodfellow, 2014] đã được những nhà nghiên cứu đón nhận và ứng dụng thực tế. Chính vì những khả năng tuyệt vời mà GAN mang lại, mô hình đã không ngừng được các nhà khoa học phát triển thêm, từ đó tạo nên những biến thể khác của GAN, mỗi biến thể giúp cải thiện hiệu suất của mô hình hoặc tăng thêm một vài tính năng mới:

- Deep Convolutional GAN [Mehralian, 2018]
- Conditional GAN [Mirza, 2014]
- CycleGAN [Zhu, 2017]
- StyleGAN [Karras, 2019]

Đây chỉ là một vài biến thể tiêu biểu của Generative Adversarial Networks, trên thực tế GAN vẫn còn rất nhiều các biến thể khác đã được phát triển và giới thiệu trong những năm gần đây. Trong phạm vi của đề tài nghiên cứu này, chúng tôi sẽ tập trung vào mô hình GAN truyền thống.



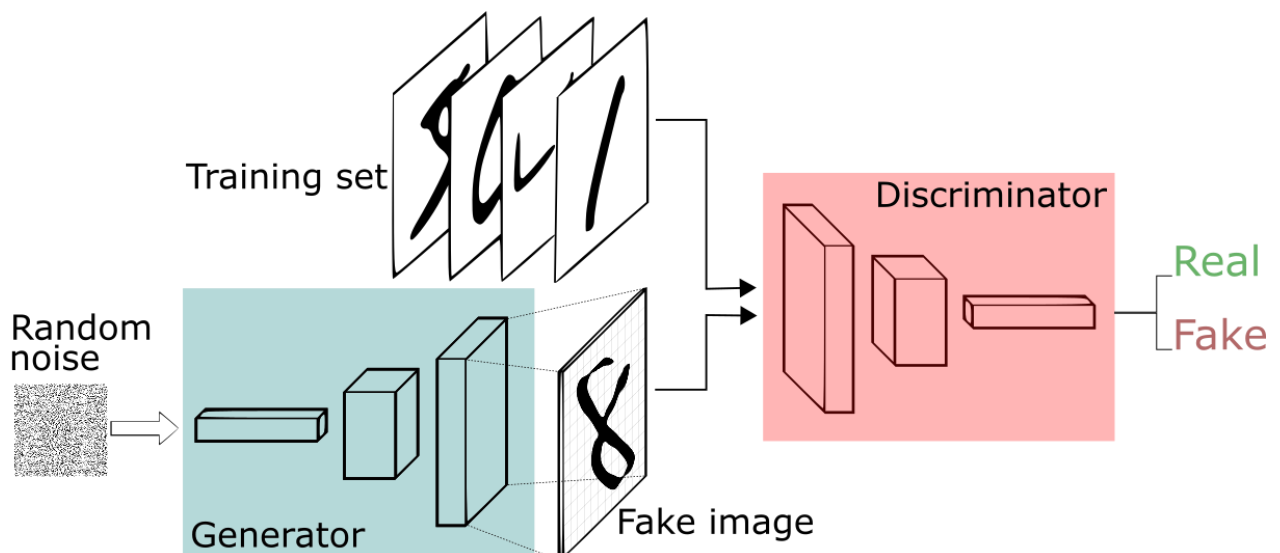
## b. Phương pháp nghiên cứu

- Về mặt lý thuyết:
  - + Tìm hiểu và tham khảo các công trình liên quan.
  - + Tìm hiểu mô hình sinh dữ liệu.
  - + Tìm hiểu mô hình học sâu cho bài toán phân lớp.
  - + Tìm nguồn dữ liệu túi xách mẫu để phục vụ cho việc nghiên cứu.
- Ứng dụng thực tiễn:
  - + Cài đặt và xây dựng mô hình.
  - + So sánh với ảnh mẫu để đánh giá kết quả.

## 5. Nội dung nghiên cứu

### a. Tổng quan mô hình Generative Adversarial Networks:

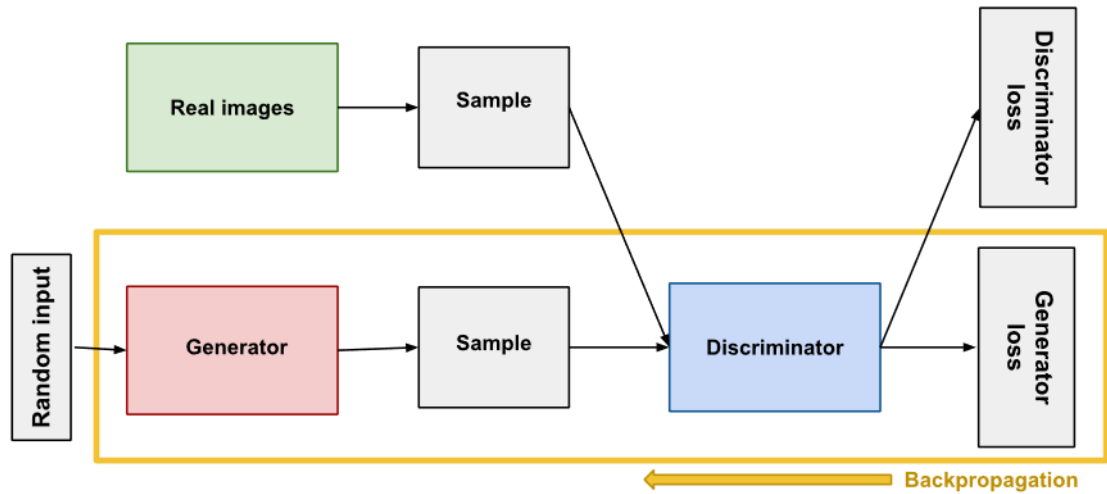
GAN là một mô hình có mục tiêu tạo ra những dữ liệu giả giống với dữ liệu thật. Generative Adversarial Networks tức là một mạng sinh đối nghịch (Generative tương ứng với sinh và Adversarial là đối nghịch). Sở dĩ GAN có tên gọi như vậy là vì kiến trúc của nó bao gồm hai mạng có mục tiêu đối nghịch nhau đó là Generator và Discriminator.



**Hình 1. Sơ đồ giải thích kiến trúc Generative Adversarial Network [Thalles Silva]**

- **Generator:** Học cách sinh ra dữ liệu giả để đánh lừa mô hình *Discriminator*. Để có thể đánh lừa được *Discriminator* thì đòi hỏi mô hình Generator phải

sinh ra dữ liệu thực sự tốt (có nghĩa là dữ liệu đầu ra phải càng tiệm cận dữ liệu thật trong tập huấn luyện). Do đó, chất lượng dữ liệu đầu vào sẽ đóng một vai trò quan trọng đến chất lượng dữ liệu được sinh ra.

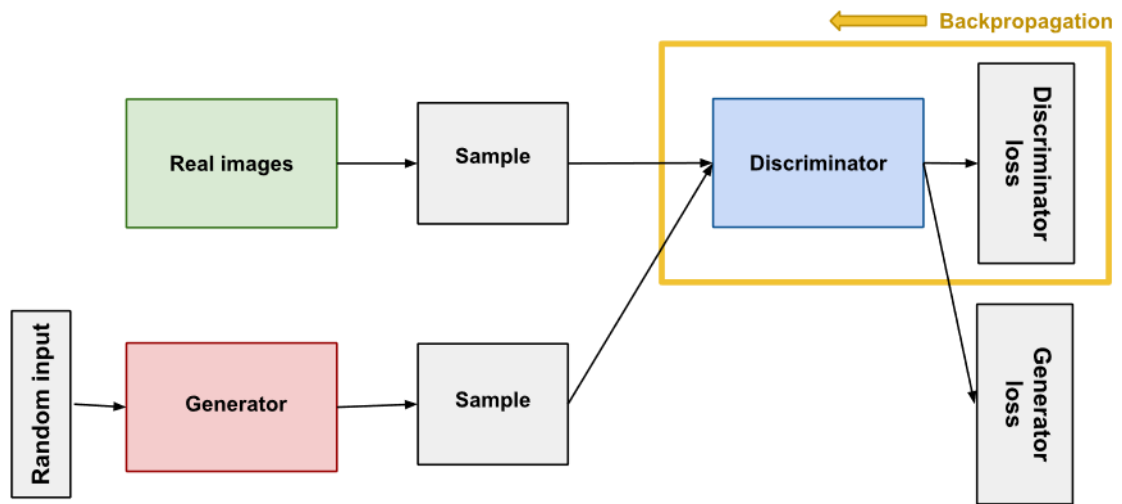


**Hình 2. Quá trình cải thiện của mô hình sinh**

(<https://developers.google.com/machine-learning/gan>)

Về bản chất là một mô hình sinh, nhận đầu vào là một tập hợp các vector nhiễu được khởi tạo ngẫu nhiên theo phân phối Gaussian. Ở một số mô hình GAN tiên tiến hơn, đầu vào có thể làm một dữ liệu chẳng hạn như bức ảnh, đoạn văn bản hoặc đoạn âm thanh. Từ tập vector đầu vào  $z$  ngẫu nhiên, mô hình Generator là một mạng học sâu có vai trò sinh ra những bức ảnh giả ở đầu ra. Những bức ảnh giả được sinh ra này sẽ được sử dụng làm đầu vào cho kiến trúc Discriminator.

- **Discriminator:** Học cách phân biệt giữa dữ liệu giả được sinh từ mô hình *Generator* với dữ liệu thật trong tập huấn luyện. Đồng thời Discriminator cũng phải cải thiện khả năng phân biệt của mình vì chất lượng ảnh được tạo ra từ *Generator* càng ngày càng giống thật hơn. Thông qua quá trình huấn luyện thì cả Generator và Discriminator đều cùng cải thiện được khả năng của mình.



**Hình 3. Quá trình cải thiện của mô hình phân biệt**

(<https://developers.google.com/machine-learning/gan>)

Mô hình Discriminator sẽ có tác dụng phân biệt dữ liệu đầu vào là thật hay giả. Nhãn của mô hình sẽ là thật nếu dữ liệu đầu vào của Discriminator được lấy tập huấn luyện và là giả nếu dữ liệu đầu vào được lấy từ đầu ra của mô hình Generator. Về bản chất đây là một bài toán phân loại nhị phân (Binary classification) thông thường. Để tính phân phối xác suất cho đầu ra của Discriminator chúng ta sử dụng hàm sigmoid.

Generator và Discriminator tương tự như hai người chơi trong bài toán zero-sum game trong lý thuyết trò chơi. Ở trò chơi này thì hai người chơi xung đột lợi ích với nhau. Hay nói cách khác, thiệt hại của người này chính là lợi ích của người kia. Mô hình Generator sẽ tạo ra những dữ liệu giả một cách chân thật hơn, vì thế làm cho Discriminator khó phân biệt hơn. Và khi Discriminator phân biệt tốt hơn thì Generator cần phải tạo ra ảnh giống thật hơn để qua mắt Discriminator. Trong zero-sum game, mỗi người chơi sẽ có chiến lược riêng của mình, đối với Generator thì đó là sinh ra ảnh giống thật và Discriminator là phân loại ảnh thật hay giả. Sau các bước ra quyết định của mỗi người chơi thì zero-sum game sẽ đạt được cân bằng Nash tại điểm cân bằng (Equilibrium Point).

## b. Hàm mục tiêu

Hàm loss của mô hình Generative Adversarial Networks là một hàm kết hợp đồng thời giữa mục tiêu của Discriminator và mục tiêu của Generator.

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

Trong đó :

$z$  là noise đầu vào của Generator.

$x$  là dữ liệu thật từ tập dữ liệu.

$G$  là mạng Generator.

$D$  là mạng Discriminator.

$G(z)$  là dữ liệu được sinh ra từ Generator.

$D(x)$  là giá trị dự đoán của Discriminator xem dữ liệu  $x$  là thật hay giả.

$D(G(z))$  là giá trị dự đoán xem dữ liệu sinh ra từ Generator là dữ liệu thật hay giả.

### - Mục tiêu của Discriminator:

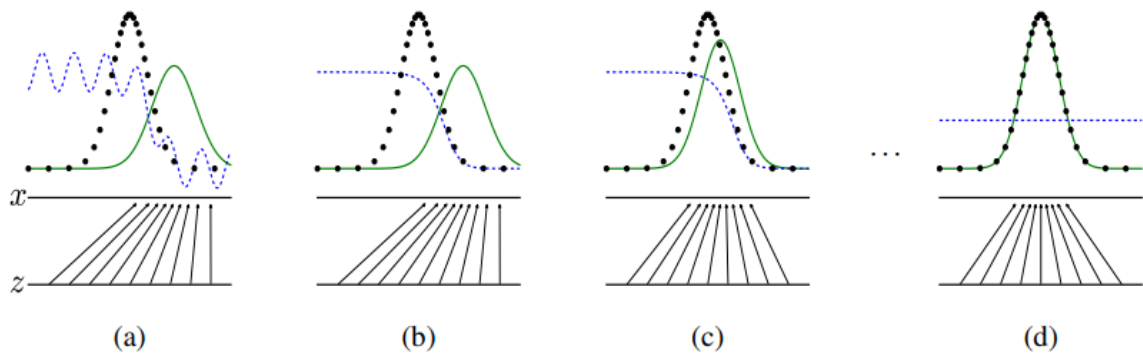
Nhiệm vụ của Discriminator cố gắng phân biệt dữ liệu là thật hay giả. Do đó đây là bài toán binary classification nên hàm loss sẽ giống với binary cross-entropy. Giá trị đầu ra của mô hình qua hàm sigmoid nên sẽ trong khoảng (0, 1), vì thế Discriminator sẽ được huấn luyện để khi nhận dữ liệu đầu vào từ tập huấn luyện thì giá trị đầu ra sẽ gần 1, còn dữ liệu đầu vào là dữ liệu được sinh ra từ Generator thì giá trị đầu ra sẽ gần 0, hay  $D(x) \rightarrow 1$  còn  $D(G(z)) \rightarrow 0$ . Hiểu một cách khác là hàm loss muốn maximize  $D(x)$  mà minimize  $D(G(z))$ . Và minimize  $D(G(z))$  tương đương với maximize  $(1 - D(G(z)))$ . Do đó, hàm loss của Discriminator được viết lại thành:

$$\max_D V(D) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

- **Mục tiêu của Generator:**

Nhiệm vụ của Generator là học để đánh lừa Discriminator, để Discriminator tin rằng dữ liệu được sinh ra từ Generator là thật, hay  $D(G(z)) \rightarrow 1$ . Hàm loss muốn maximize  $D(G(z))$ , tương đương với  $\text{minimize}(1 - D(G(z)))$ .

$$\min_G V(G) = E_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$



**Hình 4. Quá trình cải thiện của mô hình sinh và mô hình phân biệt**

[Goodfellow, 2014]

## CHƯƠNG II. TÌNH TRẠNG NGHIÊN CỨU

### 1. Ứng dụng tạo hình nhân vật Anime

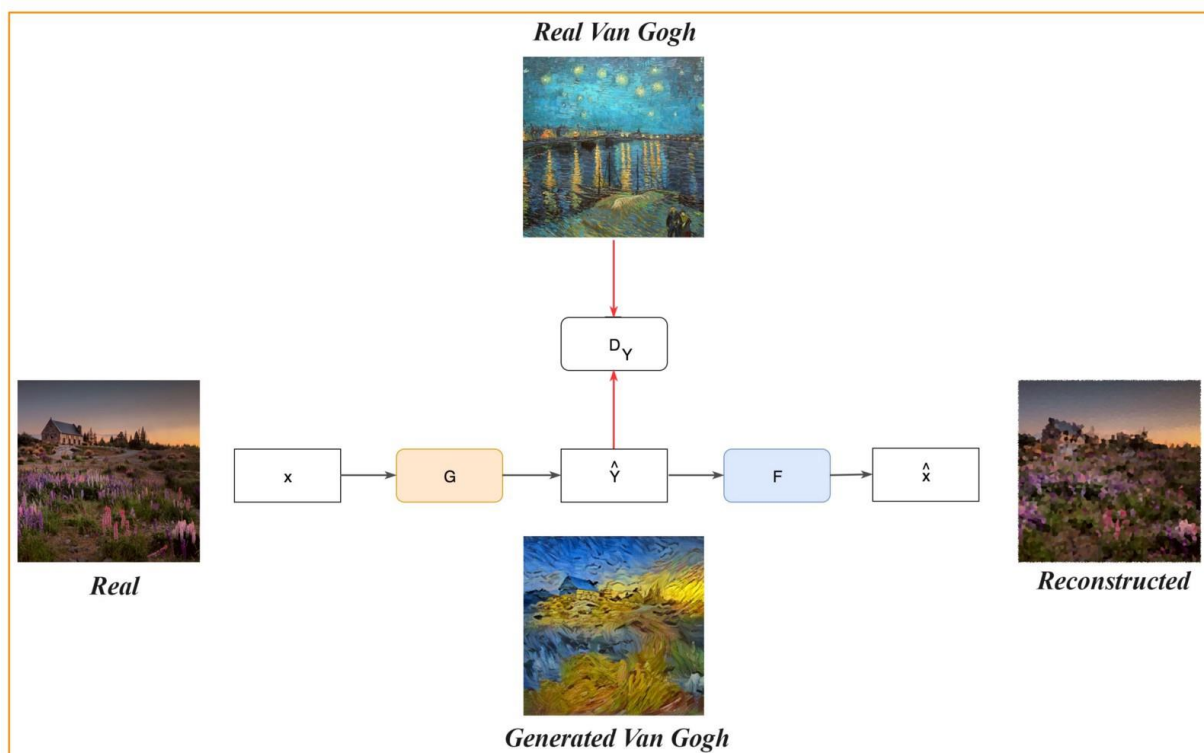
Ngành công nghiệp sản xuất phim hoạt hình (Anime) của Nhật Bản đã trở nên nổi tiếng trên toàn thế giới. Những bộ phim này tạo được ấn tượng với nhiều người xem vì kịch bản hay và đa dạng về thể loại như: trinh thám, phiêu lưu, học đường,...Nhưng bên cạnh đó, vẫn còn một yếu tố tạo nên thành công của các bộ phim Anime đó là nhân vật trong phim. Các nhân vật được tạo hình một cách rất tỉ mỉ để phù hợp với vai trò trong bộ phim. Do đó việc thuê các nghệ sĩ để thiết kế các nhân vật trong phim thường rất tốn kém. GAN có thể tự động tạo ra và tô màu cho các nhân vật Anime [Jin, 2017].



Hình 5. Ảnh nhân vật Anime được sinh ra [Jin, 2017]

### 2. CycleGAN

CycleGAN chuyển đổi những hình ảnh từ một miền sang một miền khác bằng xây dựng hai mạng G và F. CycleGAN sử dụng Discriminator D để đánh giá những hình ảnh được sinh ra tốt như thế nào. Ví dụ, G chuyển đổi hình ảnh thật thành một bức tranh theo phong cách của Van Gogh và D(y) được sử dụng để phân biệt hình ảnh là thật hay được tạo ra [Zhu, 2017].



**Hình 6. Sơ đồ chuyển đổi một bức ảnh thành bức tranh mang phong cách của Van Gogh [Zhu, 2017]**

### 3. Super resolution

Một ứng dụng khá thú vị khác của Generative Adversarial Networks là giúp nâng cao chất lượng hình ảnh (super-resolution). Super-resolution là một tác vụ liên quan đến việc nâng cấp hình ảnh từ độ phân giải thấp thành hình ảnh có độ phân giải cao.



**Hình 7. Ứng dụng của GAN trong việc cải thiện chất lượng ảnh [Ledig, 2017]**



## 4. StyleGAN2

StyleGAN2 là một biến thể khác của GAN có thể sinh ra được những hình ảnh có độ phân giải cao.



**Hình 8. Ảnh mặt người được sinh ra bằng StyleGAN [Karras, 2020]**

## 5. Text to Image (StackGAN)

Chuyển đổi văn bản thành hình ảnh là một trong những ứng dụng thú vị của GAN. Đầu vào là một câu mô hình hình ảnh và đầu ra chính là các hình ảnh phù hợp với mô tả của câu đó.





**Hình 9. Những bông hoa được tạo ra từ câu mô tả [Zhang, 2017]**

# CHƯƠNG III.

## XÂY DỰNG BỘ DỮ LIỆU

### 1. Tình trạng cơ sở dữ liệu

Dữ liệu là một yếu tố vô cùng quan trọng trong việc ứng dụng một mô hình máy học và với Generative Adversarial Networks điều đó cũng không ngoại lệ. Mô hình GAN trong đề tài của chúng tôi sử dụng đầu vào là những hình ảnh về các loại túi xách. Do đó, chúng tôi quyết định lựa chọn những loại túi xách đang phổ biến trên thị trường để thuận tiện cho việc sưu tầm dữ liệu và cũng để bắt kịp xu hướng thời trang hiện giờ. Sau một thời gian tìm hiểu trên rất nhiều trang web bán hàng, nhóm đã quyết định lựa chọn được những loại túi xách tiêu biểu để ứng dụng cho mô hình. Các loại túi xách mà nhóm chúng tôi lựa chọn là:

- Balo
- Tote
- Clutch
- Satchel



Balo



Tote



Clutch



Satchel

**Hình 10. Các loại túi có trong dataset**

## 2. Quá trình thu thập dữ liệu

Như đã được đề cập ở phần giới hạn dữ liệu, nhóm chúng tôi quyết định chỉ huấn luyện mô hình trên với bốn loại túi xách. Sau quá trình tiền xử lý dữ liệu, số lượng hình ảnh ở mỗi lớp như sau:

<i>Loại túi xách</i>	<i>Số lượng ảnh</i>
<i>Balo</i>	10084
<i>Tote</i>	10625
<i>Clutch</i>	4134
<i>Satchel</i>	4827

Do đầu vào của mô hình là hình ảnh túi xách, nên tất cả dữ liệu được sử dụng trong đề tài đều được sưu tầm trên Internet mà cụ thể là các sàn thương mại điện tử và các website bán túi xách. Mục tiêu của đề tài là tạo ra những hình ảnh túi xách mới dựa trên những hình ảnh túi xách đã được huấn luyện nên chúng tôi chỉ lựa chọn những ảnh chỉ chứa túi xách (có nghĩa là chỉ chọn những ảnh chứa túi xách và ảnh nền phải màu trắng) để không ảnh hưởng đến chất lượng của hình ảnh đầu ra. Dữ liệu được sưu tầm phần lớn là trên các trang như: [zappos.com](http://zappos.com), [amazon.com](http://amazon.com), [fado.vn](http://fado.vn).



Hình 11. Các nguồn dữ liệu chính

Để có thể sưu tầm được một số lượng lớn hình ảnh trên Internet thì việc sưu tầm bằng phương pháp thủ công thì sẽ tiêu tốn rất nhiều thời gian. Vì thế, chúng tôi áp dụng một số phương pháp để việc sưu tầm dữ liệu trở nên nhanh hơn. Nhóm đã sử dụng một công cụ để sưu tầm dữ liệu khá hiệu quả - **Selenium**. Công cụ này được hiểu đơn giản là nó có thể mô phỏng hầu hết những thao tác giữa người dùng với trình duyệt.

Selenium giúp chúng tôi sưu tầm được số lượng lớn hình ảnh trên Internet một cách nhanh chóng. Nhưng ảnh sưu tầm về vẫn còn rất nhiều nhiễu không phù hợp để bắt đầu huấn luyện (ảnh không phải là túi xách, bị lẫn với vật khác, có nền khác màu trắng,...) để giải quyết vấn đề này nhóm chỉ có thể lọc lại các ảnh bằng phương pháp thủ công, việc này tuy mất thời gian nhưng lại đảm bảo được chất lượng của bộ dữ liệu trong lúc thực hiện đề tài.

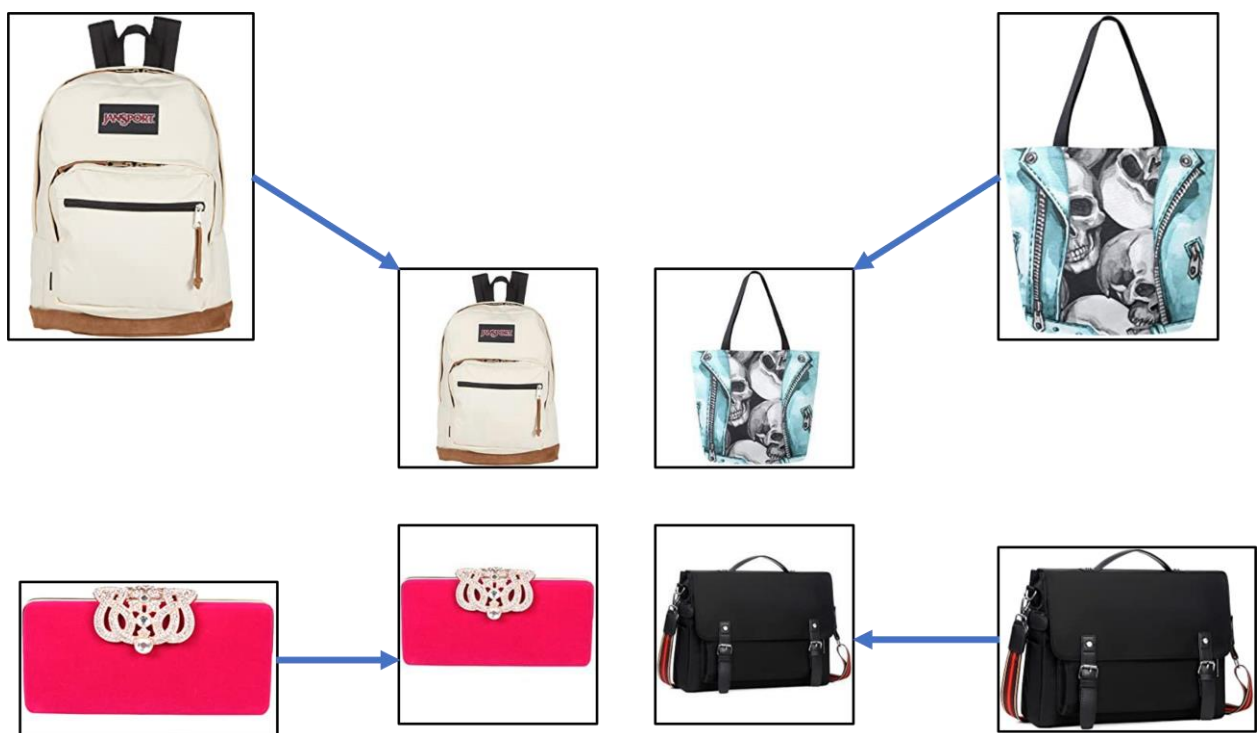


Hình 12. Một số ảnh nhiễu và ảnh đạt yêu cầu

Sau khi hoàn thành loại bỏ nhiễu khỏi bộ dữ liệu, công việc tiếp theo là phải đưa ảnh về một kích thước cố định như nhau để mô hình GAN học hiệu quả hơn.

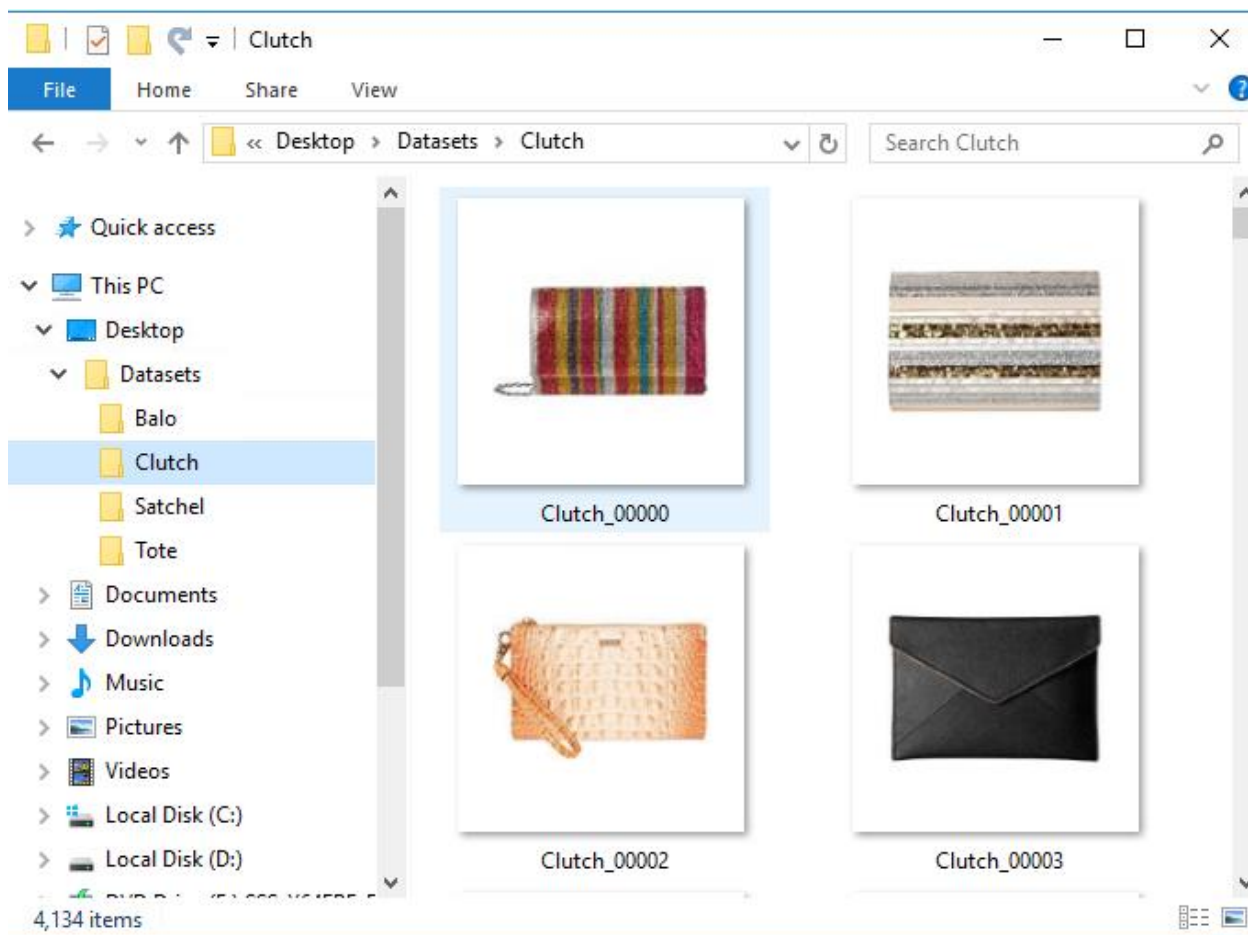
Trong hệ thống, dữ liệu đầu vào sẽ là hình ảnh và dữ liệu đầu ra cũng là hình ảnh, do đó độ phân giải có một vai trò quan trọng đến chất lượng của ảnh được sinh, để có thể tìm ra được một độ phân giải phù hợp cho mô hình của mình, nhóm đã quyết định thử nghiệm huấn luyện với nhiều độ phân giải khác nhau như: 64x64, 128x128, 256x256. Cuối cùng, chúng tôi quyết định sử dụng hình ảnh đầu vào với độ phân giải là 256x256, vì ảnh được sinh ra ở độ phân giải này có chất lượng tốt hơn hai độ phân giải trước đó và một phần quan trọng là kích thước này phù hợp cho việc huấn luyện của chúng tôi (không quá lớn).

Vì ảnh trong tập huấn luyện được sưu tầm trên Internet nên ảnh sẽ có nhiều kích thước khác nhau. Để có thể đưa ảnh về kích thước đã chọn, nhóm chúng tôi sử dụng một kỹ thuật xử lý ảnh cơ bản là padding.



**Hình 13. Ảnh túi xách đã xử lý Padding**

*Padding images*: là thêm khoảng không gian vào trong ảnh, khoảng không gian này sẽ được cộng dồn thêm vào chiều rộng hoặc chiều cao của ảnh mà không làm ảnh bị biến dạng. Sau khi padding, ảnh sẽ là một “hình vuông” với kích thước là 256x256.



**Hình 14. Cấu trúc lưu trữ của bộ dữ liệu**



TÊN LOẠI TÚI  
Clutch

SỐ THỨ TỰ  
00001

**Hình 15. Cách đặt tên ảnh trong bộ dữ liệu**



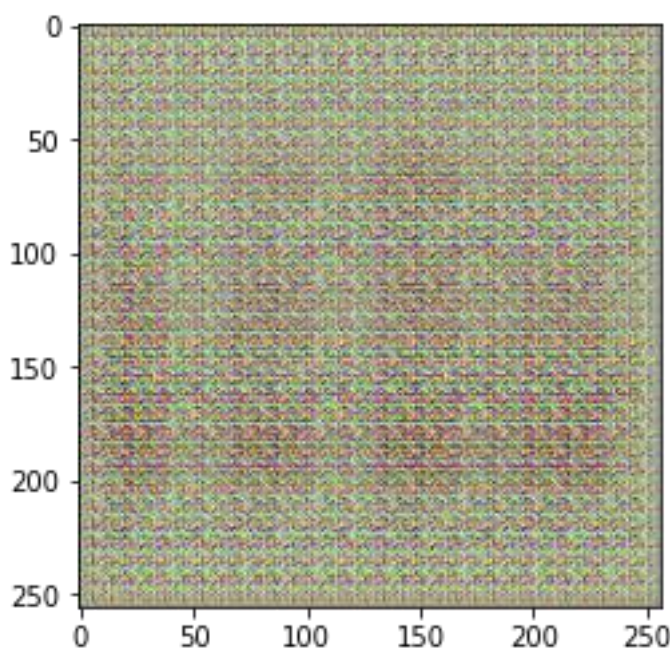
## CHƯƠNG IV.

# XÂY DỰNG MÔ HÌNH THIẾT KẾ TÚI XÁCH BẰNG GENERATIVE ADVERSARIAL NETWORKS

Sau khi thực hiện tiền xử lí tập dữ liệu, chúng tôi tiến hành xây dựng mô hình Generative Adversarial Networks.

### 1. Quá trình sinh

Generator là một mạng Convolutional Neural Networks sinh ảnh từ nhiễu. Bước đầu, chúng tôi thực hiện tạo một hình ảnh giả (một hình ảnh không có thực trong thực tế) đó chính là nhiễu. Nếu trực quan hóa hình ảnh này, thì chúng ta có thể thấy nó không mang ý nghĩa gì đối với con người.

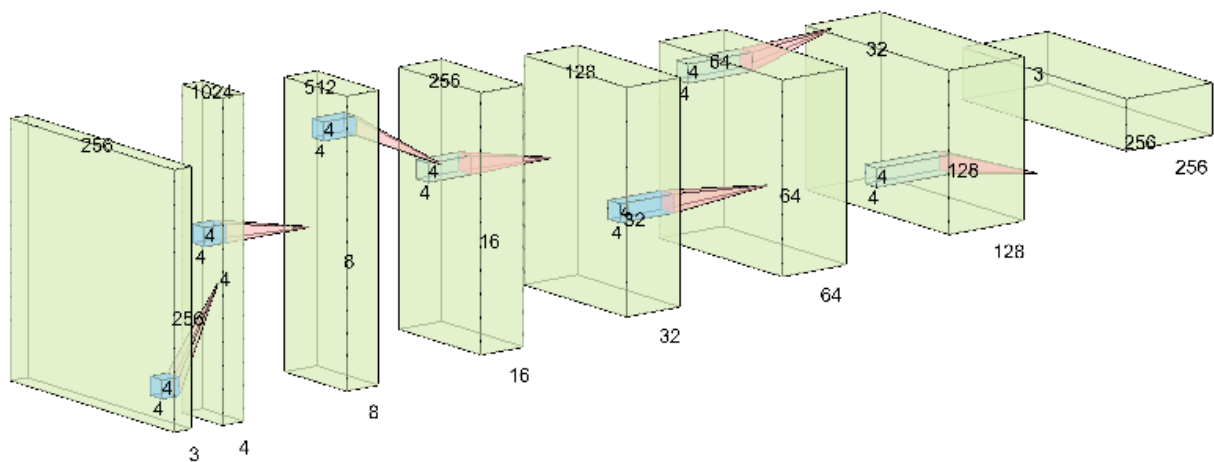


**Hình 16. Hình minh họa cho ảnh nhiễu**

Nhóm chúng tôi tạo ra một vector đặc trưng  $\mathbf{z}$  bằng cách sử dụng phân phối chuẩn. Vector đặc trưng  $\mathbf{z}$  đại diện cho các đặc trưng ẩn của hình ảnh, ví dụ như: màu sắc hay hình dạng của hình ảnh đầu ra. Đây là quá trình đề từ các đặc điểm đặc trưng, chúng tôi sẽ tái tạo lại hình ảnh gốc chứa các đặc trưng đó. Do chúng tôi xây dựng

mô hình Generator là một mạng Convolutional Neural Networks nên thay vì phải xử lí thủ công để trích xuất các đặc trưng của hình ảnh thì chúng tôi không cần kiểm soát và phân tích thủ công các đặc trưng này vì máy tính có thể tự động phân tích và học các đặc trưng đó.

Tiếp theo, chúng tôi thực hiện nhiều lần kỹ thuật **transposed convolutionals** để có thể tăng kích thước của vector đặc trưng  $z$  để tạo ra được hình ảnh  $x$ . Chúng ta có thể hiểu nó như là một quá trình khôi phục lại hình ảnh từ các vector đặc trưng hay là một mạng giải chập.



**Hình 17. Cấu trúc của mô hình sinh**

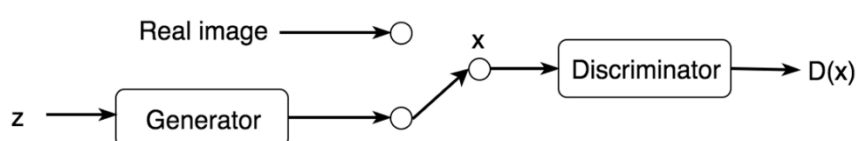
Trên thực tế Generator chỉ có khả năng tạo ra các vector đặc trưng nhiễu một cách ngẫu nhiên, và do đó các hình ảnh mà Generator tạo ra sẽ chỉ là những hình ảnh đầu ra không có ý nghĩa, hiểu một cách đơn giản là không có “ai” kiểm tra quá trình sinh ảnh của Generator. Để giải quyết vấn đề này, tiếp theo chúng tôi sẽ đi đến mô hình Discriminator.



## 2. Quá trình phân biệt

Discriminator cũng là một mạng Convolutional Neural Networks, nhận đầu vào là một bức ảnh có thể là ảnh thật hoặc ảnh giả (được sinh ra từ Generator).

Ban đầu, Discriminator sẽ “quan sát” hình ảnh trong thực tế và hình ảnh được tạo ra bởi Generator. Discriminator có nhiệm vụ là so sánh hình ảnh thực và hình ảnh được tạo xem có tương đồng với nhau hay không. Đầu ra  $D(x)$  là giá trị xác suất thể hiện hình ảnh đầu vào  $x$  là hình ảnh thực.



Chúng tôi sẽ huấn luyện Discriminator như một Classifier trong Deep Learning thông thường. Nếu đầu vào là một hình ảnh thật thì chúng tôi mong muốn đầu ra  $D(x) = 1$ . Còn trong trường hợp nó là hình ảnh được tạo ra từ Generator thì chúng tôi mong muốn đầu ra  $D(x) = 0$ .

Trong quá trình huấn luyện, chúng tôi sẽ kết hợp xen kẽ giữa hai giai đoạn.  $k$  batch đầu tiên chúng tôi sẽ huấn luyện Discriminator trước:

Huấn luyện Discriminator: lấy mẫu một mini-batch kích thước  $m$  là các nhiễu  $z$  làm đầu vào của Generator. Đồng thời lấy mẫu một mini-batch khác kích thước  $m$  là những điểm dữ liệu thật  $x$ . Những dữ liệu này sẽ được sử dụng để cập nhật gradient descent theo phương pháp mini-batch gradient descent.

$$\frac{1}{m} \nabla_{\theta_D} \sum_{i=1}^m \log D(x^{(i)}) + \log(1 - D(G(z^{(i)})))$$

Do huấn luyện trên mô hình Discriminator nên chỉ cần cập nhật các trọng số trên mô hình Discriminator. Các hệ số của Generator được “đóng băng”.

Huấn luyện Generator: sau khi kết thúc  $k$  batch huấn luyện trên Discriminator chúng tôi sẽ tiếp tục huấn luyện trên Generator. Một mini-batch kích thước  $m$  được

chọn ra từ các nhiễu là  $z$  được sử dụng như đầu vào huấn luyện. Gradient descent sẽ được tính trên  $m$  dữ liệu này theo công thức:

$$\frac{1}{m} \nabla_{\theta_G} \sum_{i=1}^m \log(1 - D(G(z^{(i)})))$$

Cập nhật gradient descent chỉ được áp dụng trên các hệ số của Generator.

Tiếp tục quá trình này cho tới khi Loss có giá trị càng nhỏ càng tốt.

Tóm lại, mục đích chính của chúng tôi là muốn Generator tạo ra hình ảnh sao cho khi truyền hình ảnh được tạo ra này vào Discriminator thì đầu ra  $D(x)$  phải hướng đến 1. Có nghĩa là chúng ta phải “đánh lừa” được Discriminator bằng cách nâng cao khả năng sinh ảnh của Generator.

# CHƯƠNG V. THỰC NGHIỆM VÀ ĐÁNH GIÁ

## 1. Kết quả thực nghiệm

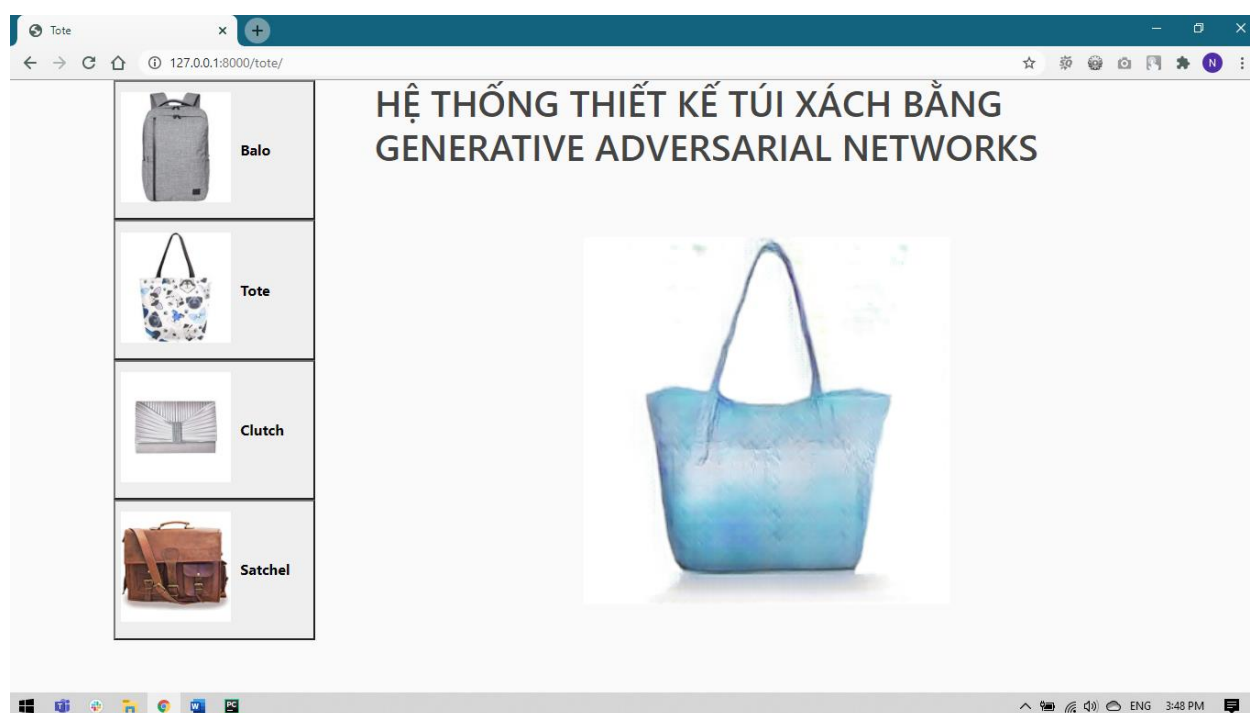
Sau khi thực hiện huấn luyện mô hình GAN trên bộ dữ liệu mà nhóm sưu tầm được, chúng tôi đã đạt một kết quả đáp ứng được mục tiêu đề ra ban đầu. Nhìn chung, hình ảnh được sinh ra từ mô hình đều mang đặc trưng và hình dáng cho mỗi loại túi xách.



Hình 18. Một số ảnh minh họa cho kết quả đầu ra

Với kết quả thu được khá khả quan, chúng tôi tiến hành xây dựng một website để giúp người dùng dễ dàng thực hiện các thao tác tạo ra thiết kế mới thay vì chạy model trên PyCharm. Nhóm đã tìm hiểu và thực hiện ứng dụng dựa vào Django framework. Thư viện này được cộng đồng đánh giá là khá tiện lợi và phù hợp để tạo ra một website nhanh chóng. Website của chúng tôi có ưu điểm là đơn giản trong việc vận hành, để sinh ra các hình ảnh túi xách mới thì người dùng chỉ cần chọn vào hình đại diện cho loại túi mà mình muốn thiết kế. Sau khi chọn, hệ thống

sẽ sinh ra một tập các hình ảnh mới và sẽ được hiển thị lần lượt ở giữa màn hình. Các nhà thiết kế có thể quan sát và lựa chọn mẫu thiết mà mình cảm thấy phù hợp.



**Hình 19. Giao diện của hệ thống thiết kế túi xách**

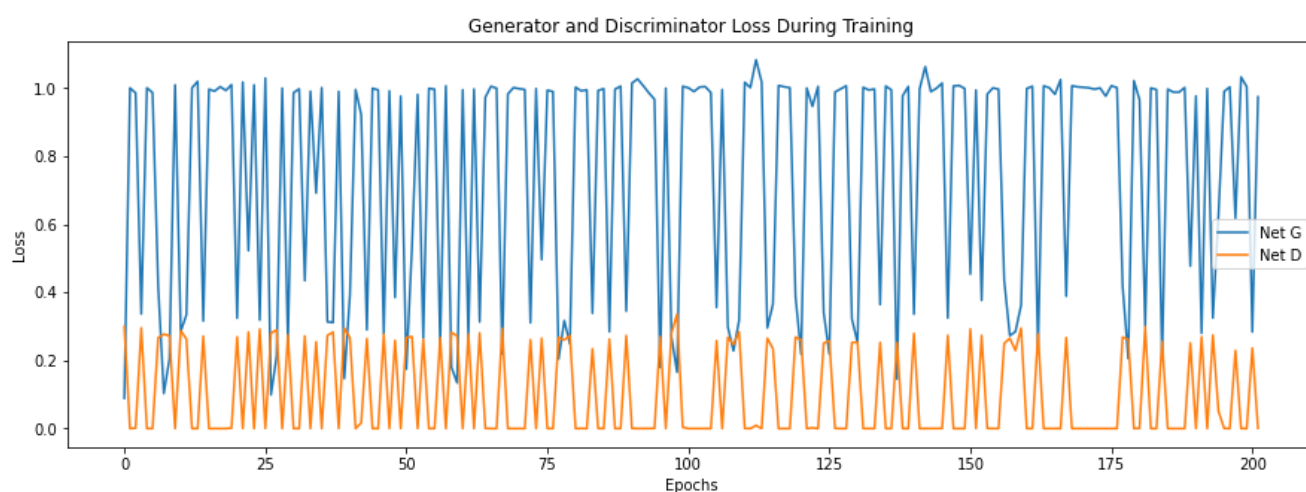
Bên cạnh đó, chúng tôi nhận thấy một số hình ảnh được sinh ra có chất lượng vẫn còn hạn chế. Hoa tiết chưa đa dạng, có vài sản phẩm còn bị trùng lặp với nhau. Thậm chí có những trường hợp sản phẩm sinh ra chưa thể hiện được hình dạng của túi xách. Đây hiện tại là một hạn chế đối với mô hình của chúng tôi.



**Hình 20. Ảnh đầu ra bị khuyết điểm**

Chúng tôi hy vọng với kết quả đạt được này sẽ là tiền đề để nhóm có thể phát triển thêm và từ đó các sản phẩm được tạo ra có chất lượng ngày càng tốt hơn.

## 2. Đánh giá



**Hình 21. Hình minh họa giá trị của hàm mục tiêu trong quá trình huấn luyện**

Biểu đồ trên thể hiện được lý thuyết của Generative Adversarial Networks. Đó là khi Loss của Generator giảm đồng nghĩa với việc ảnh sinh ra giống ảnh thật hơn và Loss của Discriminator sẽ tăng đồng nghĩa với việc Discriminator gặp khó khăn hơn khi phân biệt ảnh thật và ảnh giả.

## CHƯƠNG VI.

# KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Nhìn chung, “*Mô hình thiết kế túi xách*” đã có thể sinh ra được hình dáng của những chiếc túi xách mới. Mỗi loại túi xách được sinh ra đã có những nét đặc riêng của mình nhờ vào đó chúng ta dễ dàng phân biệt được đó là loại túi xách nào, đây là một kết quả khá tốt đối với nhóm. Qua đó, nhóm nhận thấy mô hình đã hoạt động tốt theo lý thuyết được đề ra ban đầu. Do các kiểu dáng trong tập huấn luyện vẫn chưa quá đa dạng nên chúng tôi nhận thấy các túi xách được sinh ra có kiểu dáng vẫn còn khá “cơ bản”, chưa có những thiết kế quá mới lạ so với những túi xách trong tập huấn luyện. Nhưng với mục tiêu ban đầu là tạo ra một hệ thống có thể sinh ra những bản thiết kế “thô”, sau đó các nhà thiết kế chuyên nghiệp sẽ “tô điểm” thêm cho những bản thiết kế đó thì nhóm nhận thấy kết quả này có thể chấp nhận được. Bên cạnh đó, vẫn còn một số hình ảnh được sinh ra có chất lượng còn khá tệ, đây là một hạn chế của hệ thống.

Mô hình hiện tại có thể tạo ra hình ảnh túi xách theo từng loại túi xách khác nhau như balo, tote, clutch hoặc satchel. Mỗi loại có một kiểu dáng riêng biệt của mình. Mặc dù đề tài đã thỏa mãn được những mục tiêu đã đề ra, nhưng hệ thống vẫn cần phải phát triển thêm để có thể đáp ứng được nhu cầu thực tế. Qua đó, chúng tôi có một vài hướng phát triển cho hệ thống như sau:

- **Về kiểu dáng:** Tăng số lượng các lớp trong bộ dữ liệu. Do mỗi loại túi xách có những nét đặc trưng riêng nên nhóm hướng đến việc có thể “pha trộn” những kiểu dáng của nhiều loại túi xách khác lại với nhau để có thể tạo ra những loại túi xách mới.
- **Về phương pháp:** Mô hình GAN có rất nhiều biến thể. Mỗi loại đều có ưu và khuyết điểm nhất định. Với mục tiêu sinh ra ảnh có đặc điểm mới với độ phân giải tốt, chúng tôi đề xuất nên sử dụng StyleGan để có thể đáp ứng được nhu cầu ngày càng nâng cao phù hợp thực tế.

## TÀI LIỆU THAM KHẢO

- [Cui, 2018] Cui, Y., Liu, Q., Gao, C.Y., & Su, Z. (2018). FashionGAN: Display your fashion design using Conditional Generative Adversarial Nets. *Computer Graphics Forum*, 37.
- [Goodfellow, 2014] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A.C., & Bengio, Y. (2014). Generative Adversarial Networks. *ArXiv*, *abs/1406.2661*.
- [Jin, 2017] Jin, Y., Zhang, J., Li, M., Tian, Y., Zhu, H., & Fang, Z. (2017). Towards the Automatic Anime Characters Creation with Generative Adversarial Networks. *ArXiv*, *abs/1708.05509*.
- [Karras, 2019] Karras, T., Laine, S., & Aila, T. (2019). A Style-Based Generator Architecture for Generative Adversarial Networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4396-4405.
- [Karras, 2020] Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analyzing and Improving the Image Quality of StyleGAN. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8107-8116.
- [Ledig, 2017] Ledig, C., Theis, L., Huszár, F., Caballero, J., Aitken, A.P., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 105-114.
- [Mehralian, 2018] Mehralian, M., & Karasfi, B. (2018). RDCGAN: Unsupervised Representation Learning With Regularized Deep Convolutional Generative Adversarial Networks. *2018 9th Conference on Artificial Intelligence and Robotics and 2nd Asia-Pacific International Symposium*, 31-38.
- [Mirza, 2014] Mirza, M., & Osindero, S. (2014). Conditional Generative Adversarial Nets. *ArXiv*, *abs/1411.1784*.
- [Singh, 2019] Singh, M., Bajpai, U., Vijayarajan, V., & Prasath, S. (2019). Generation of fashionable clothes using generative adversarial networks: A preliminary feasibility study. *International Journal of Clothing Science and Technology*.

- [Zhang, 2017] Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., & Metaxas, D.N. (2017). StackGAN: Text to Photo-Realistic Image Synthesis with Stacked Generative Adversarial Networks. *2017 IEEE International Conference on Computer Vision (ICCV)*, 5908-5916.
- [Zhu, 2017] Zhu, J., Park, T., Isola, P., & Efros, A.A. (2017). Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *2017 IEEE International Conference on Computer Vision (ICCV)*, 2242-2251.