

# MULTI-SCALE LEARNING A GAN FROM A SINGLE REPRESENTATIVE

**Sinh viên thực hiện**

Lý Phi Long – 1611146

**Giảng viên hướng dẫn**

TS. Huỳnh Thế Đăng

# Mục lục

1. Lý do chọn đề tài
2. Các phương pháp trước đó
3. Phương pháp nghiên cứu
4. Kết quả của mô hình
5. Tổng kết

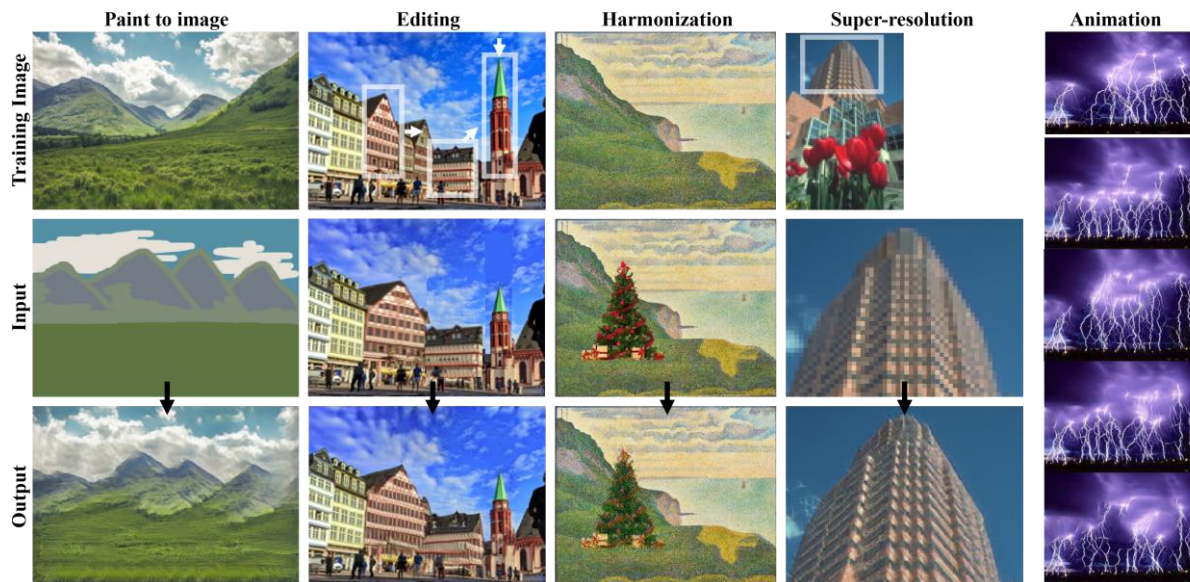
# Lý do chọn đề tài

- GAN ngày càng được ứng dụng rộng rãi.
- Huấn luyện GAN cần rất nhiều dữ liệu.
- Trong vài năm gần đây, giới nghiên cứu quan tâm nhiều hơn quá trình huấn luyện GAN trên một tấm ảnh.

# Các phương pháp trước đó

## Mô hình SinGAN

Vào năm 2019, Shaham cùng cộng sự [2] đã đề xuất mô hình SinGAN, mô hình đã được giải "Best paper award" tại ICCV'19



# Các phương pháp trước đó

Các mô hình liên quan

## ConSinGAN [3]

Số lượng tham số mô hình giảm đi  
nhưng kết quả vẫn như SinGAN [2]

+

## Mô hình của

## Benamin cùng cộng sự [1]

Đưa ra những ứng dụng mới cho  
SinGAN [2]: image-to-image

Mô hình giữ được các đặc tính tốt của cả hai

# Phương pháp nghiên cứu

Phát biểu lại bài toán

Bài toán ban đầu



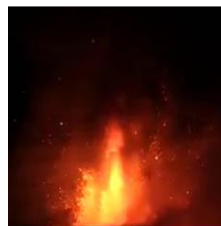
Hình ảnh  
mục tiêu



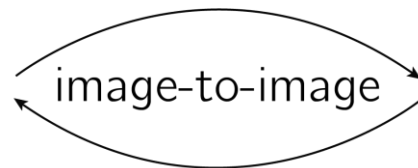
Video gốc



Bài toán được đưa về



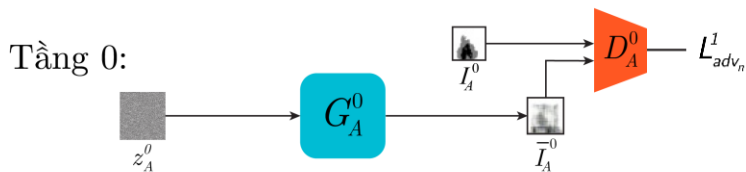
$I_A = f_i$



$I_B$

# Phương pháp nghiên cứu

Kiến trúc mô hình & Hàm mất mát



## Chiều từ A sang B:

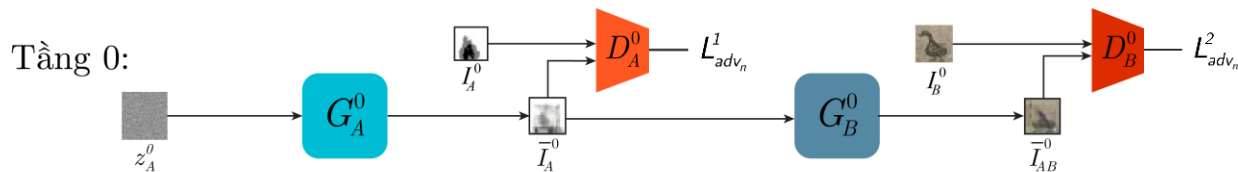
Tại tầng đầu tiên  $n = 0$ , ta sẽ đưa nhiễu  $z_A^0$  qua  $G_A^0$  và tạo được hình ảnh  $\bar{I}_A^0 = G_A^0(z_A^0)$ . Hình ảnh  $\bar{I}_A^0$  sẽ được đưa qua  $D_A^0$  cùng với  $I_A^0$  để đánh giá:

$$\mathcal{L}_{adv}^1(D_A^n, G_A^n) = D_A^n(\bar{I}_A^n) - D_A^n(I_A^n) + \lambda_{pen} \left( \left\| \nabla_{\hat{I}_A^n} D_A^n(\hat{I}_A^n) \right\|_2 - 1 \right)^2$$

Trong đó,  $\hat{I}_A^n = \alpha I_A^n + (1 - \alpha) \bar{I}_A^n$  với  $\alpha \sim U(0,1)$  và  $\lambda_{pen}$  là hệ số của gradient penalty

# Phương pháp nghiên cứu

Kiến trúc mô hình & Hàm mất mát



## Chiều từ A sang B:

Sau đó, ta sẽ đưa hình ảnh  $\bar{I}_A^0$  qua  $G_B^0$  và tạo được hình ảnh  $\bar{I}_{AB}^0 = G_B^0(\bar{I}_A^0)$ .  
Hình ảnh  $\bar{I}_{AB}^0$  sẽ được đưa qua  $D_B^0$  cùng với  $\bar{I}_B^0$  để đánh giá.

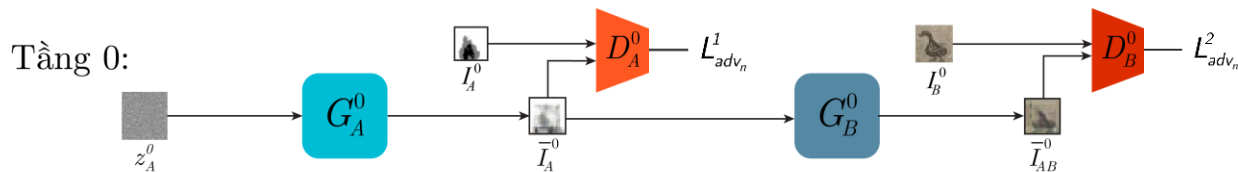
$$\mathcal{L}_{adv}^2(D_B^n, D_B^n) = D_A^n(\bar{I}_{AB}^n) - D_A^n(I_B^n) + \lambda_{pen} \left( \left\| \nabla_{\hat{I}_{AB}^n} D_A^n(\hat{I}_{AB}^n) \right\|_2 - 1 \right)^2$$

Trong đó,  $\hat{I}_{AB}^n = \alpha I_B^n + (1 - \alpha) \bar{I}_{AB}^n$  với  $\alpha \sim U(0,1)$  và  $\lambda_{pen}$  là hệ số của gradient penalty



# Phương pháp nghiên cứu

Kiến trúc mô hình & Hàm mất mát



**Chiều từ B sang A:**

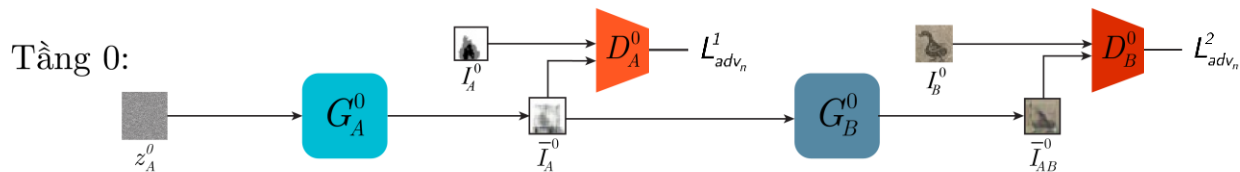
$$\mathcal{L}_{adv}^1(D_B^n, D_B^n) = D_B^n(\bar{I}_B^n) - D_B^n(I_B^n) + \lambda_{pen} \left( \left\| \nabla_{\hat{I}_B^n} D_B^n(\hat{I}_B^n) \right\|_2 - 1 \right)^2$$

$$\mathcal{L}_{adv}^2(D_A^n, D_A^n) = D_A^n(\bar{I}_{BA}^n) - D_A^n(I_A^n) + \lambda_{pen} \left( \left\| \nabla_{\hat{I}_{BA}^n} D_A^n(\hat{I}_{BA}^n) \right\|_2 - 1 \right)^2$$

Trong đó,  $\hat{I}_B^n = \alpha I_B^n + (1 - \alpha) \bar{I}_B^n$ ,  $\hat{I}_{BA}^n = \alpha I_A^n + (1 - \alpha) \bar{I}_{BA}^n$  với  $\alpha \sim U(0,1)$  và  $\lambda_{pen}$  là hệ số của gradient penalty

# Phương pháp nghiên cứu

Kiến trúc mô hình & Hàm mất mát

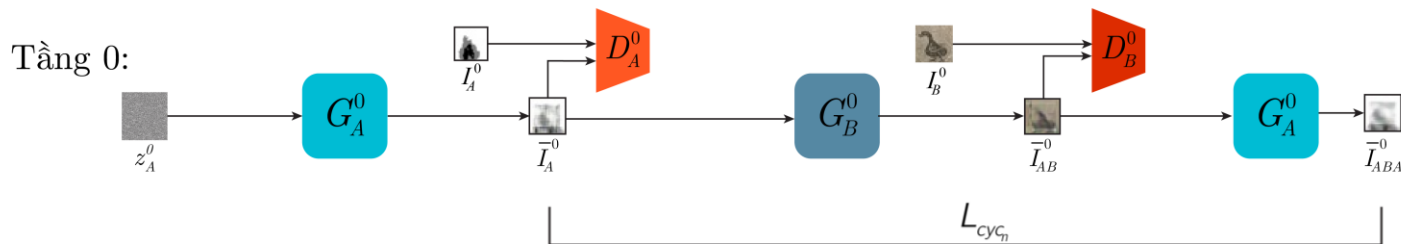


Mất mát đối kháng của mô hình tại tần thứ  $n$ :

$$\mathcal{L}_{adv_n} = \mathcal{L}_{adv}^1(D_A^n, G_A^n) + \mathcal{L}_{adv}^2(D_B^n, D_B^n) + \mathcal{L}_{adv}^1(D_B^n, D_B^n) + \mathcal{L}_{adv}^2(D_A^n, D_A^n)$$

# Phương pháp nghiên cứu

## Kiến trúc mô hình & Hàm mất mát

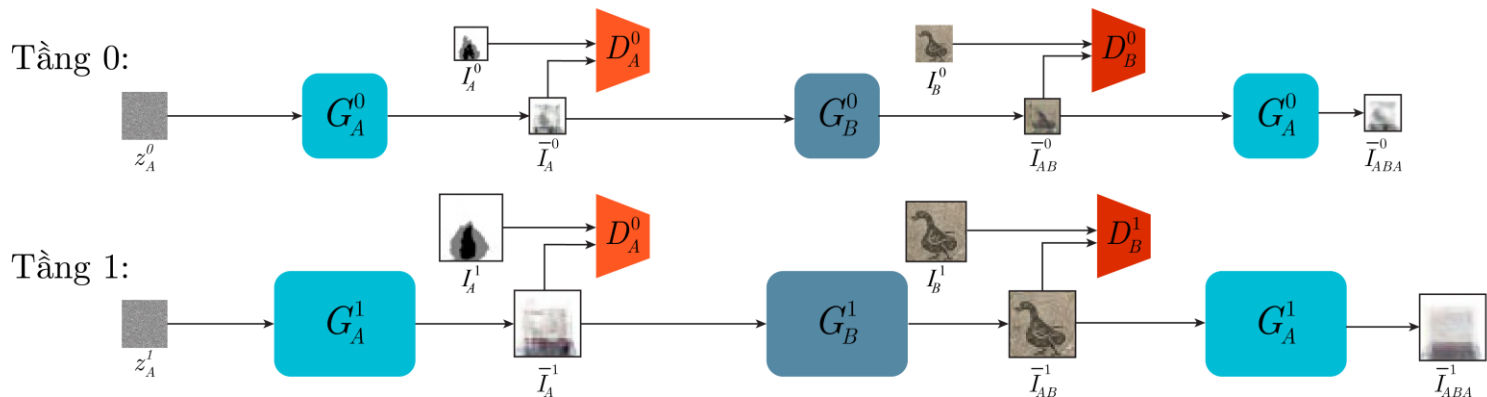


Sau đó, ta sẽ đưa hình ảnh  $\bar{I}_{AB}^0$  qua  $G_A^0$  và tạo được hình ảnh  $\bar{I}_{ABA}^0 = G_A^0(\bar{I}_{AB}^0)$  nhằm đảm bảo khoảng cách  $\bar{I}_{ABA}^0$  và  $\bar{I}_{AB}^0$  ngắn nhất:

$$\mathcal{L}_{cyc_n}(G_A^n, G_B^n) = \|\bar{I}_A^n - \bar{I}_{ABA}^n\|_2^2 + \|\bar{I}_B^n - \bar{I}_{BAB}^n\|_2^2$$

# Phương pháp nghiên cứu

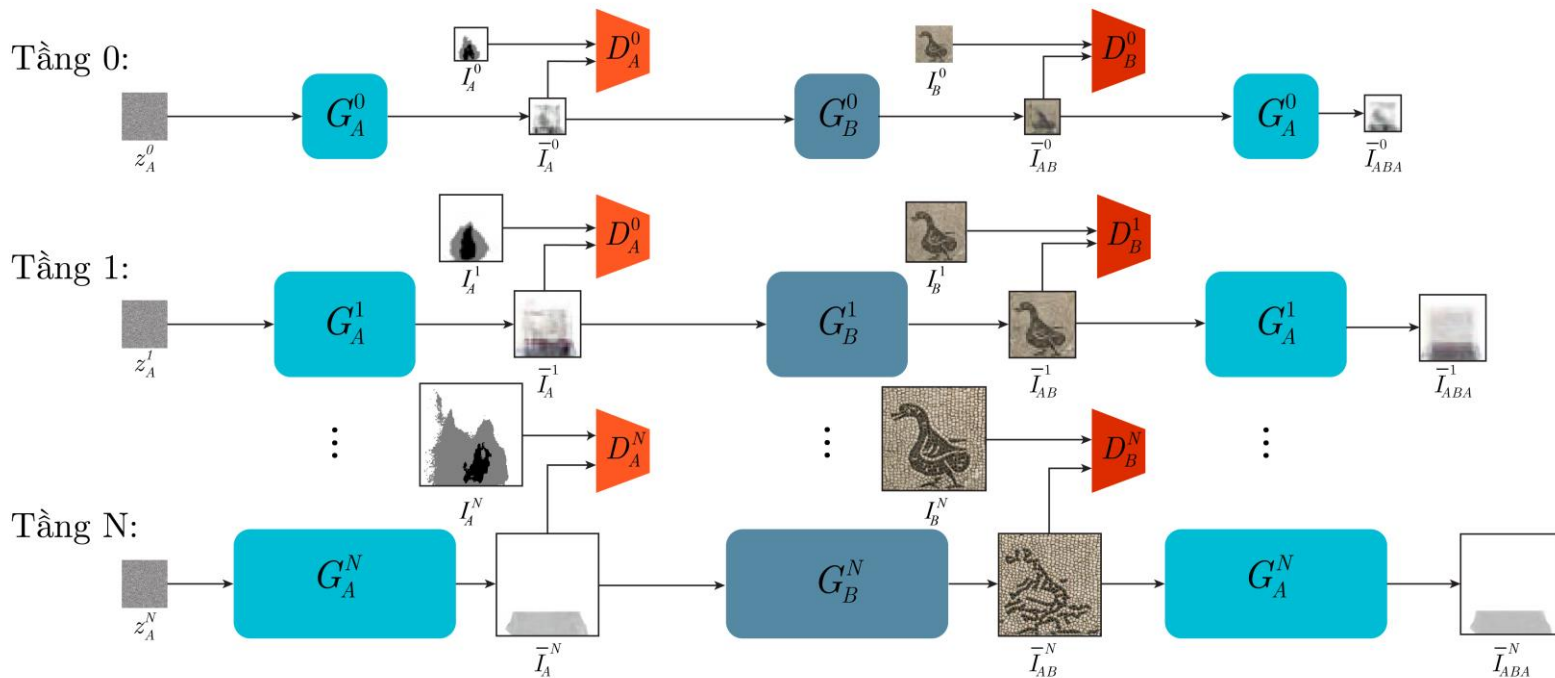
Kiến trúc mô hình & Hàm mất mát



Sau đó, feature được trích xuất từ tầng trên sẽ được đưa xuống tầng tiếp theo, và quá trình đó sẽ lặp đi lặp lại đến khi nào đạt được độ phân giải của hình như ban đầu.

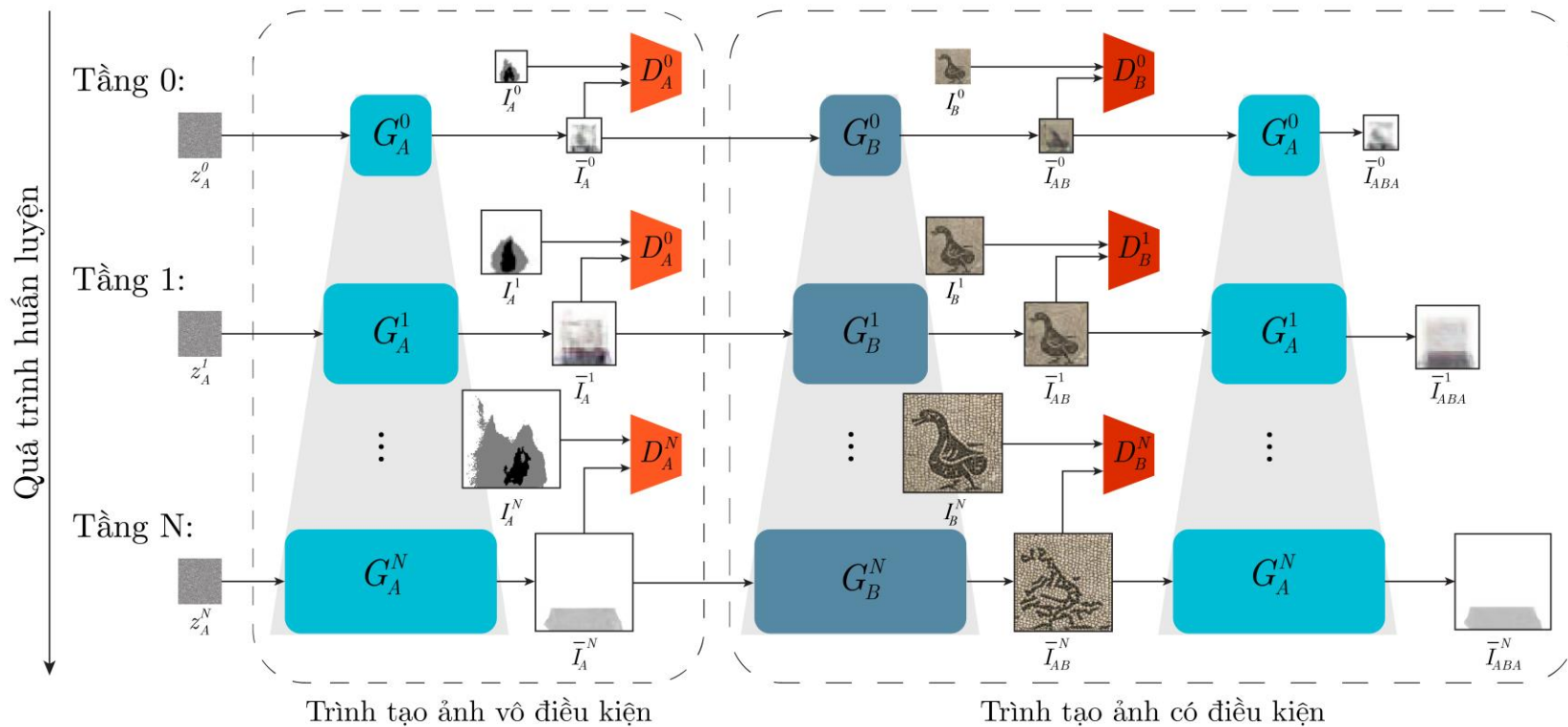
# Phương pháp nghiên cứu

Kiến trúc mô hình & Hàm mất mát



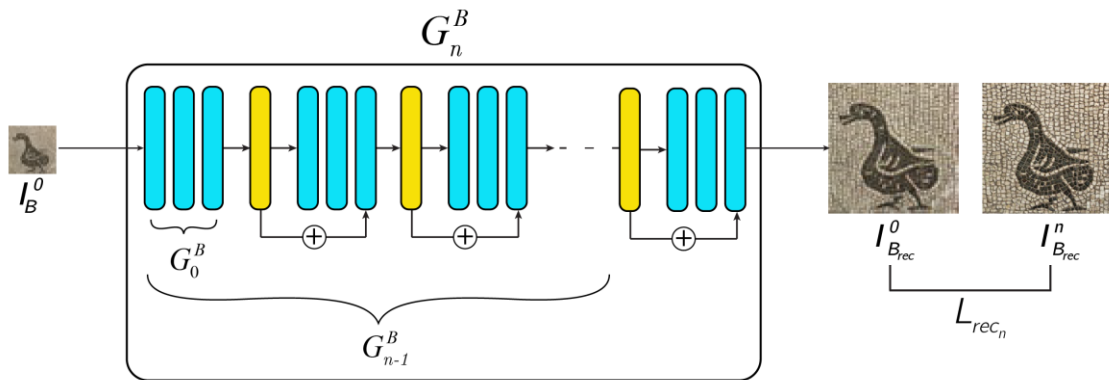
# Phương pháp nghiên cứu

Kiến trúc mô hình & Hàm mất mát



# Phương pháp nghiên cứu

Kiến trúc mô hình & Hàm mất mát



$$\mathcal{L}_{rec_n}(G_A^n, G_B^n) = \|G_A^n(I_A^0) - I_A^n\|_2^2 + \|G_B^n(I_B^0) - I_B^n\|_2^2$$

# Phương pháp nghiên cứu

Kiến trúc mô hình & Hàm mất mát

Tổng tất cả các mất mát của mô hình tại tầng thứ  $n$ :

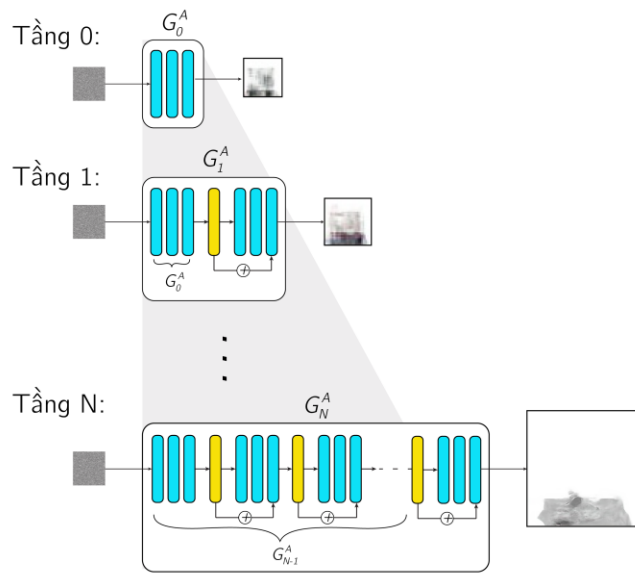
$$\mathcal{L}_n = \min_{G_A^n, G_B^n} \max_{D_B^n, D_B^n} \mathcal{L}_{adv_n} + \lambda_{rec} \mathcal{L}_{rec_n} + \lambda_{cyc} \mathcal{L}_{cyc_n}$$



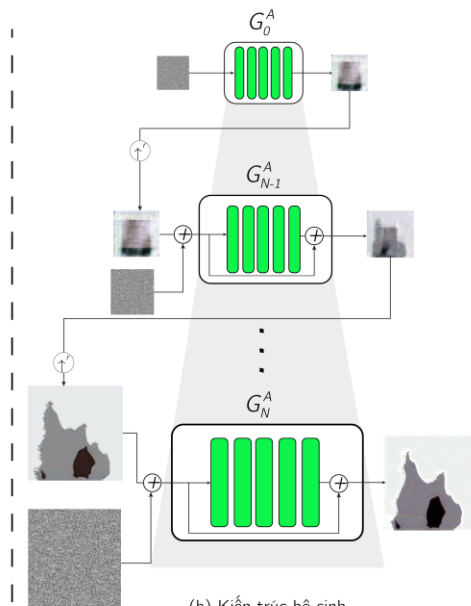
# Phương pháp nghiên cứu

## Kiến trúc Generator

Khác biệt với mô hình của Benamin cùng cộng sự, khi qua tầng tiếp theo sẽ đưa feature xuống tầng tiếp theo.



(a) Kiến trúc bộ sinh trong nghiên cứu này



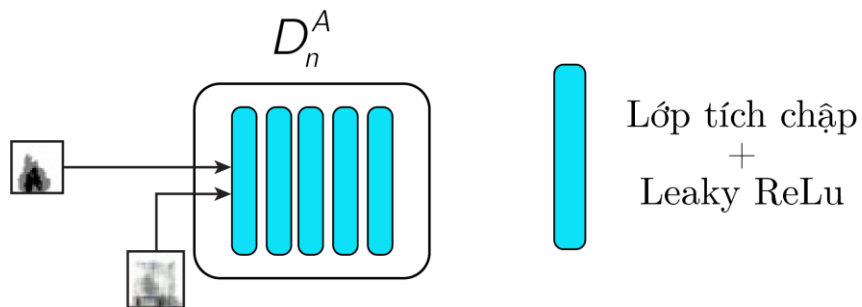
(b) Kiến trúc bộ sinh trong nghiên cứu của Benamin và cộng sự

Bảng chú thích

	Lớp tích chập + Batch Norm + Leaky ReLU
	Lớp tích chập + Leaky ReLU
	Tăng mẫu + Nhiễu
	Liên kết dư

# Phương pháp nghiên cứu

## Kiến trúc Discriminator



Gồm 5 lớp khối tích chập, mỗi khối gồm một lớp tích chập (3x3) với một hàm Leaky ReLU.

# Kết quả và thảo luận

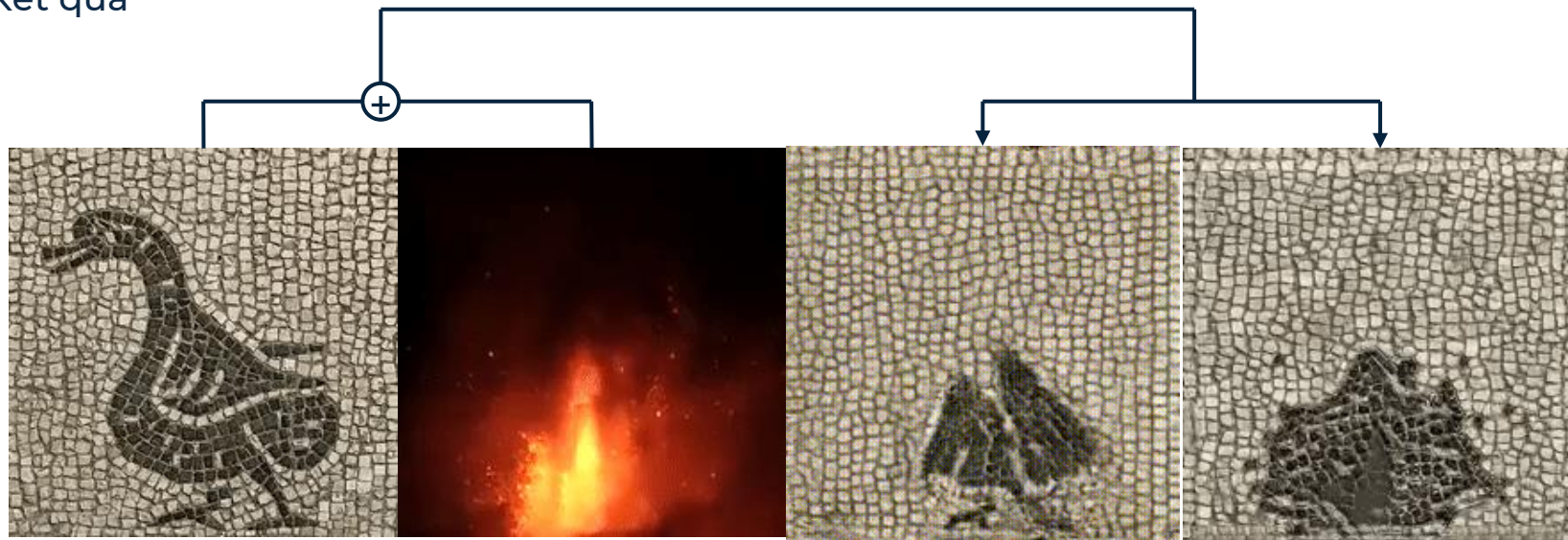
## Kết quả

Mô hình	Số lượng bộ sinh	Số lượng bộ phân biệt	Số tầng	Số lượng tham số
Nghiên cứu này	2	2	6	~ 1,320,000
Benamin và cộng sự [1]	2	2	10	~ 2,600,000

Mô hình được cải tiến có số lượng tham số giảm đi 50% so với mô hình của Benaim và cộng sự [1]

# Kết quả và thảo luận

Kết quả



Hình ảnh  
mục tiêu

Video  
gốc

Nghiên cứu này

Benaim và cộng sự [3]

Tuy nhiên, kết quả tạo ra lại không tốt bằng mô hình của Benaim và cộng sự [3]

# Kết quả và thảo luận

Dưới đây là một số nguyên nhân có thể dẫn đến việc tạo kết quả trong mô hình không tốt:

- Cấu trúc mô hình đã được rút gọn dẫn đến việc học các thông tin trở nên vừa đủ làm cho mô hình không thể có nhiều thông tin để chuyển đổi.
- Do quá trình huấn luyện  $G_A^n, D_A^n$  trên miền ảnh  $A$  và  $G_B^n, D_B^n$  trên miền ảnh  $B$  không đồng đều, dẫn đến chỉ một trong hai quá trình đã hội tụ.
- Quá trình huấn luyện chưa đủ tầng khiến cho mô hình học không đủ thông tin.

# Hướng nghiên cứu tiếp theo

## Kết quả

Tuy nhiên, với những kết quả không tốt lại đưa ra động lực để phát triển mô hình tốt hơn, bằng một số hướng cải thiện như sau:

- Mở rộng số tầng huấn luyện để có thể cho mô hình học lâu hơn và nhiều chi tiết hơn trong ảnh.
- Sử dụng thông tin của video như optical flow chứ không còn đưa riêng lẻ từng khung hình như nghiên cứu này.
- Mở rộng thêm nhiều ứng dụng cho mô hình như chỉnh sửa ảnh, siêu phân giải hình ảnh, hòa trộn ảnh.

# Tài liệu tham khảo

- [1] Sagie Benaim et al. *Structural-analogy from a Single Image Pair*. 2020.
- [2] Tamar Rott Shaham, Tali Dekel, and Tomer Michaeli. “SinGAN: Learning a Generative Model From a Single Natural Image”. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* (Oct. 2019).
- [3] Tobias Hinz et al. *Improved Techniques for Training Single-Image GANs*. 2020.

**CẢM ƠN  
THẦY VÀ CÁC BẠN ĐÃ LẮNG NGHE**