

# DL-Seminar Season#5 AI Lab

유 재창

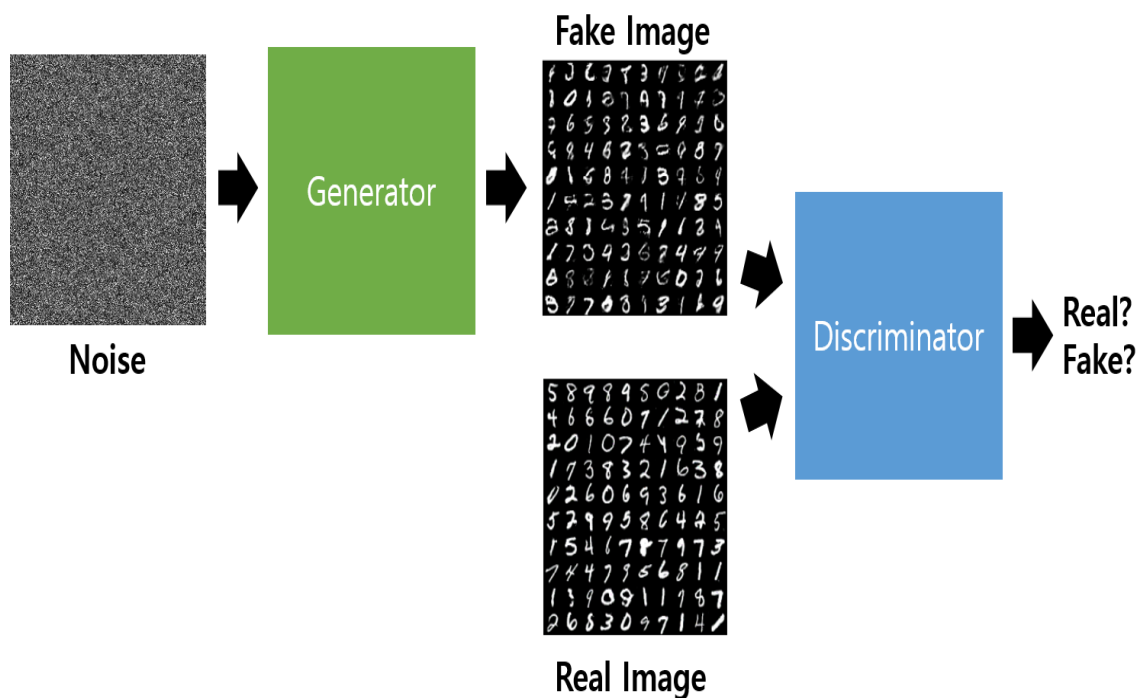
# **StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks**

Han Zhang<sup>1</sup>, Tao Xu<sup>2</sup>, Hongsheng Li<sup>3</sup>,  
Shaoting Zhang<sup>4</sup>, Xiaogang Wang<sup>3</sup>, Xiaolei Huang<sup>2</sup>, Dimitris Metaxas<sup>1</sup>

<sup>1</sup>Rutgers University   <sup>2</sup>Lehigh University   <sup>3</sup>The Chinese University of Hong Kong   <sup>4</sup>Baidu Research

{han.zhang, dnm}@cs.rutgers.edu, {tax313, xih206}@lehigh.edu  
{hsli, xgwang}@ee.cuhk.edu.hk, zhangshaoting@baidu.com

# Generalized Adversarial Networks (GAN)



## Generator(생성자)

- Random noise  $z$  를 입력으로 받아 Discriminator가 real data와 구분이 어렵도록 학습.

## Discriminator(분류기)

- Real data와 fake data를 잘 구분하도록 학습.  
- Real  $\rightarrow$  1, fake  $\rightarrow$  0

## Generalized Adversarial Networks (GAN)

### Loss Function

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

Problem

Text  
description

A small bird  
with varying  
shades of  
brown with  
white under the  
eyes



A small yellow  
bird with a  
black crown  
and a short  
black pointed  
beak



StackGAN

?

# Problem

Text  
description

A small bird  
with varying  
shades of  
brown with  
white under the  
eyes

A small yellow  
bird with a  
black crown  
and a short  
black pointed  
beak

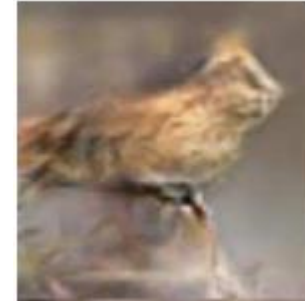


StackGAN



# Stack Generalized Adversarial Networks (GAN)

**The Stage-I GAN** sketches the primitive shape and colors of the object based on the given text description, yielding Stage-I low-resolution images.



# Stack Generalized Adversarial Networks (GAN)

**The Stage-I GAN** sketches the primitive shape and colors of the object based on the given text description, yielding Stage-I low-resolution images.

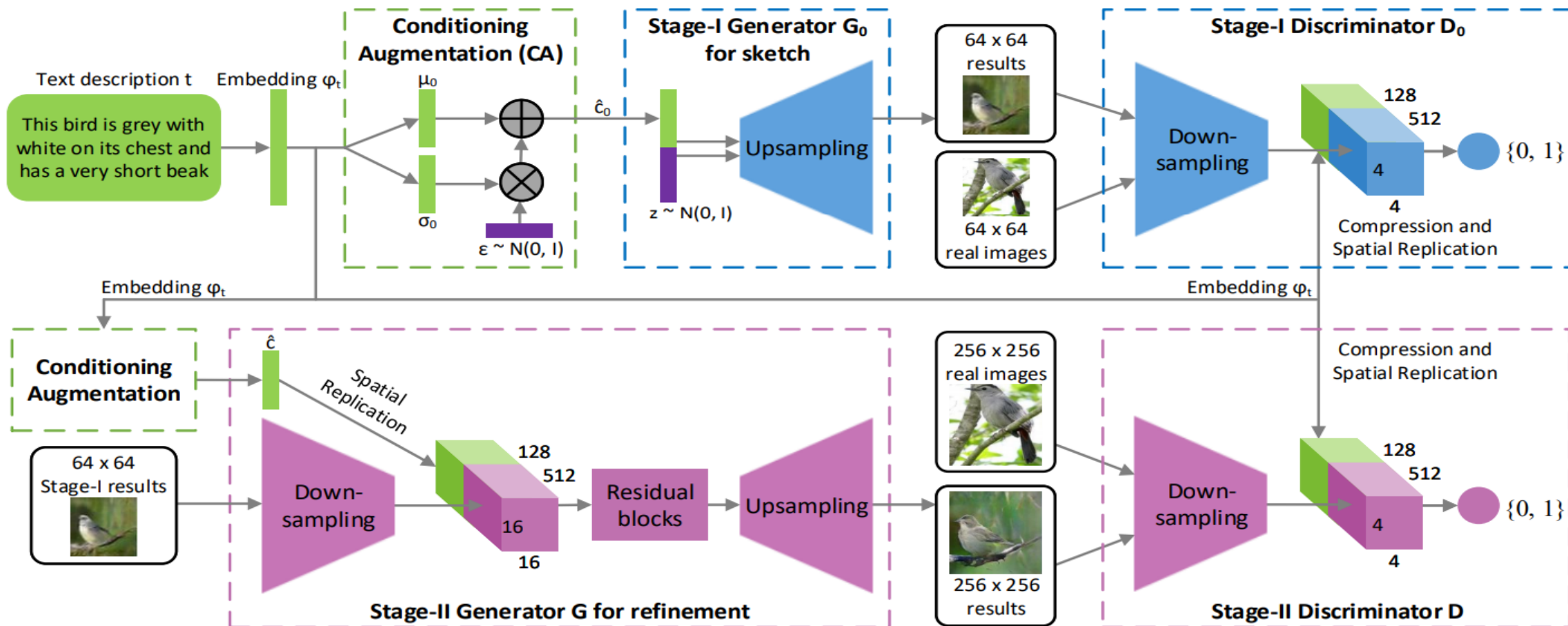


**The Stage-II GAN** takes Stage-I results and text descriptions as inputs, and generates high-resolution images with photo-realistic details.



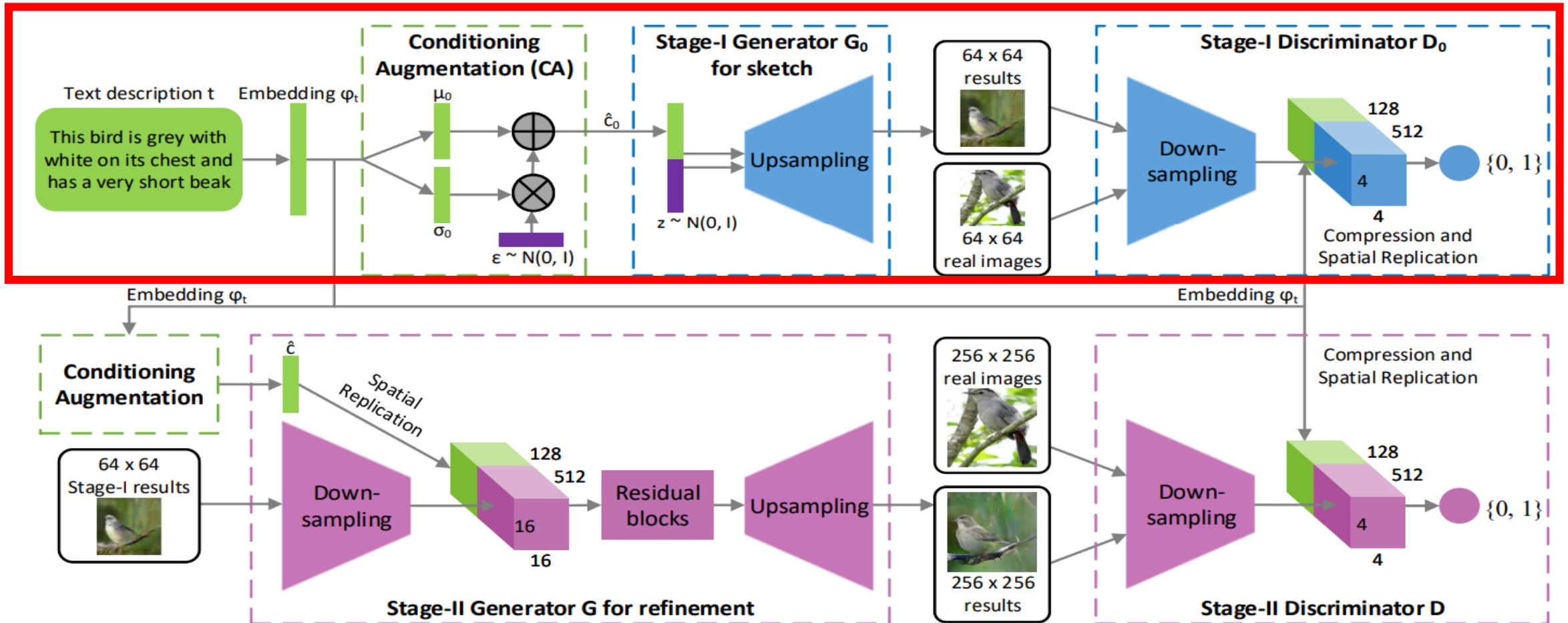


# Stack GAN Model Architecture

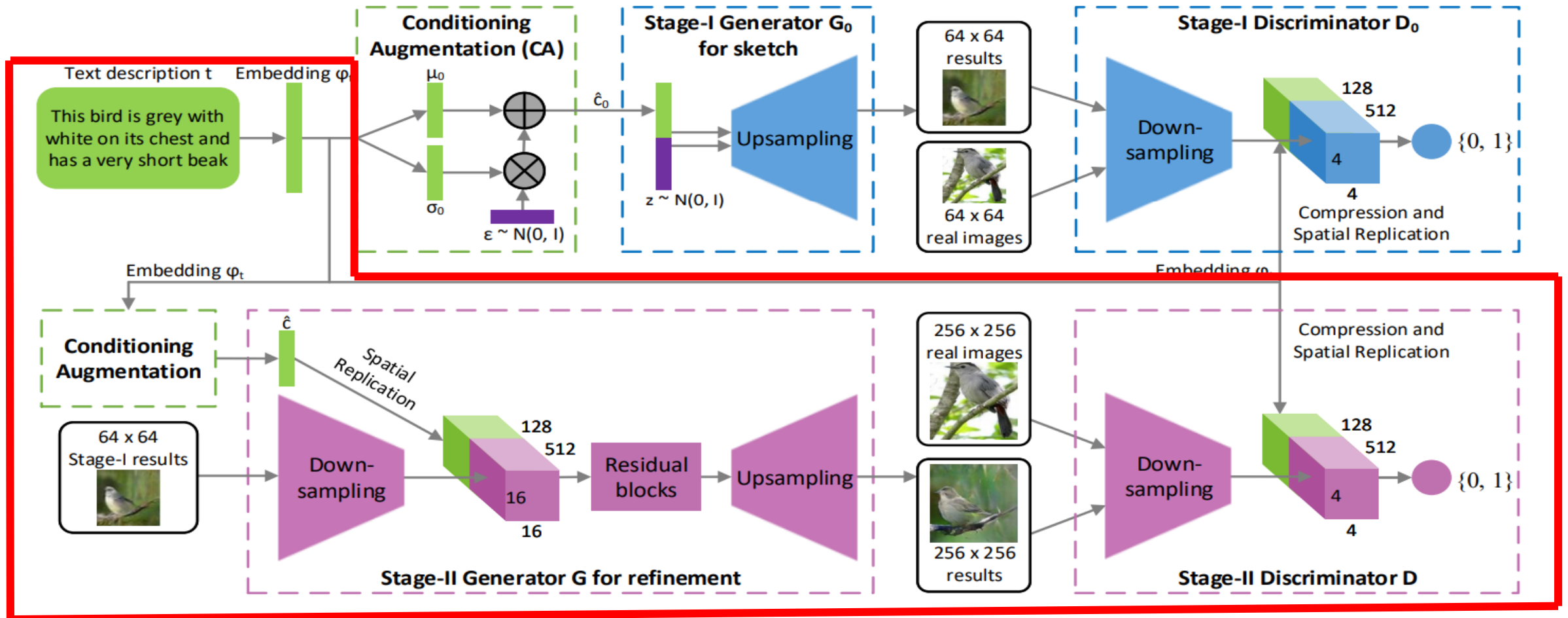


# Stack GAN Model Architecture

## Stage-1 GAN



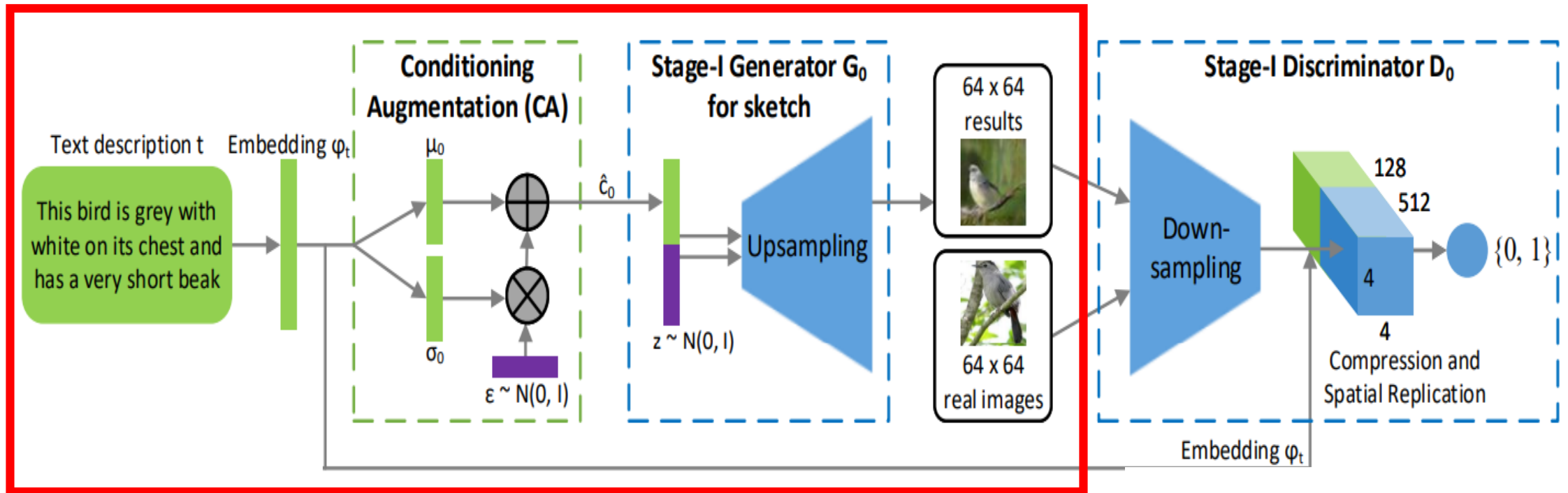
# Stack GAN Model Architecture



Stage-2 GAN

## Stage-1 GAN / Generator

Pre-trained Encoder -> Embedding vector 생성.  
CA를 통해 conditional vector  $\hat{c}_0$ 를 구함.



the element-wise multiplication

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \odot \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} a_1 \cdot b_1 \\ a_2 \cdot b_2 \end{bmatrix}$$

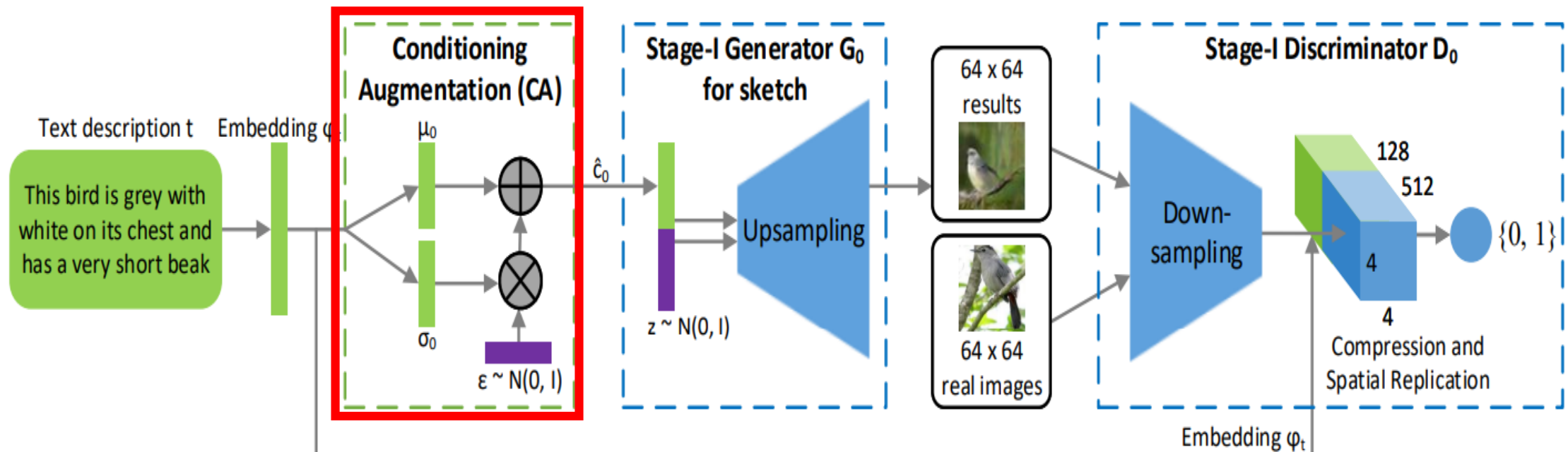
## Conditioning Augmentation technique

Conditional GAN -> fixed Condition variable  $c$ 를 input으로 사용.

Embedding vector -> Fully Connected Layer ->  $\mu_0, \sigma_0$  (independent Gaussian distribution)

$\hat{c}_0 = \mu_0 + \sigma_0 \odot \epsilon$  ( $\epsilon \sim \mathcal{N}(0, I)$ ) -> conditioning vector (not fixed)

적은 text, image pair에도 많은 training pair 효과. -> 학습 good~



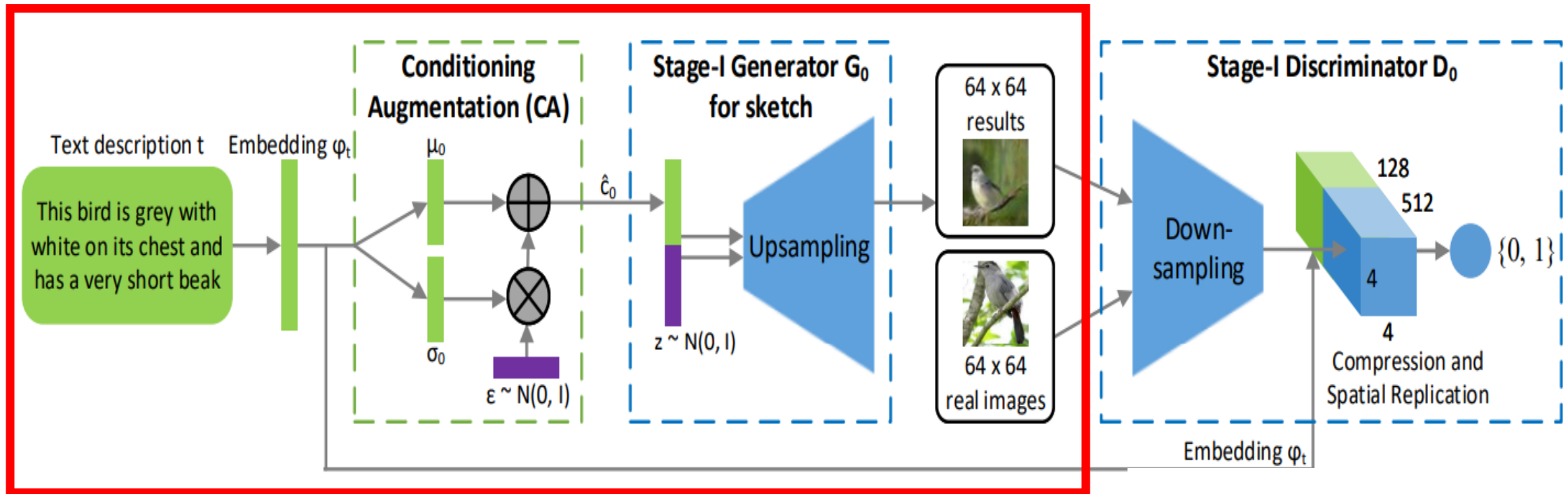
## Stage-1 GAN / Generator

Pre-trained Encoder -> Embedding vector 생성.

CA를 통해 conditional vector  $\hat{c}_0$ 를 구함.

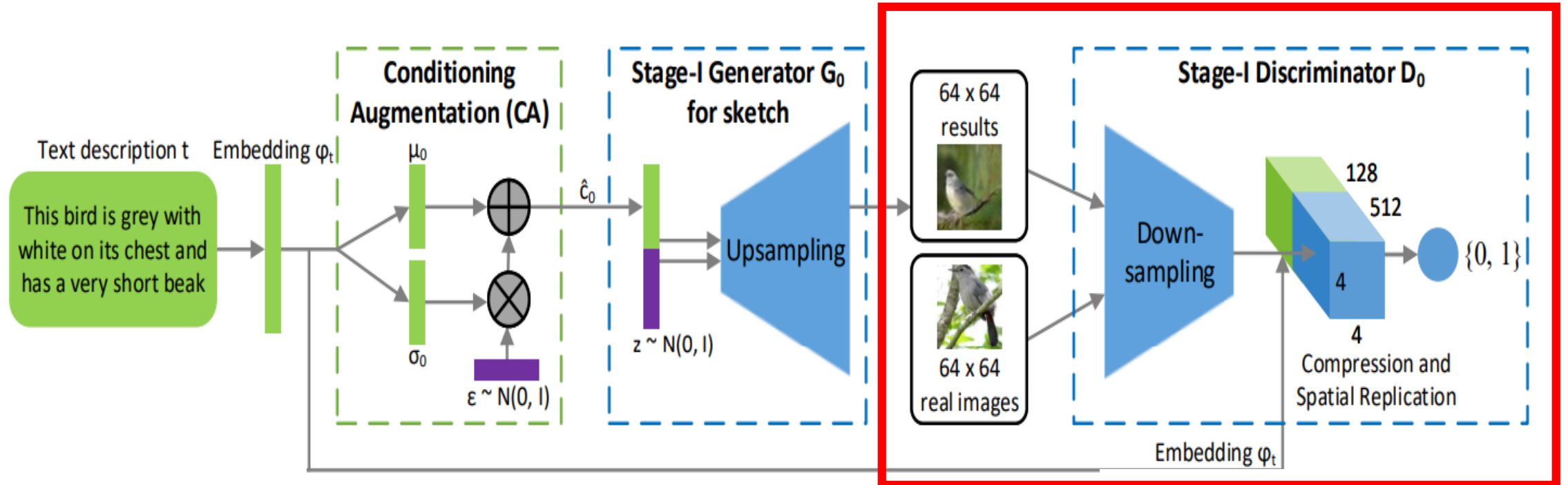
$z$ (random noise vector)와 concatenate.

Deconvolution(upsampling) -> 64x64 low image-resolution.



## Stage-1 GAN / Discriminator

$G_0$ 로부터 생성된 64x64 low image-resolution  
Input 으로 들어온 image를 Down-sampling -> 4x4x512 image map 생성  
Embedding vector(128 dimension)를 Replication -> 4x4x128의 tensor 생성  
1x1 convolutional layer -> image, text feature 관계 학습.  
FCL with 1 node

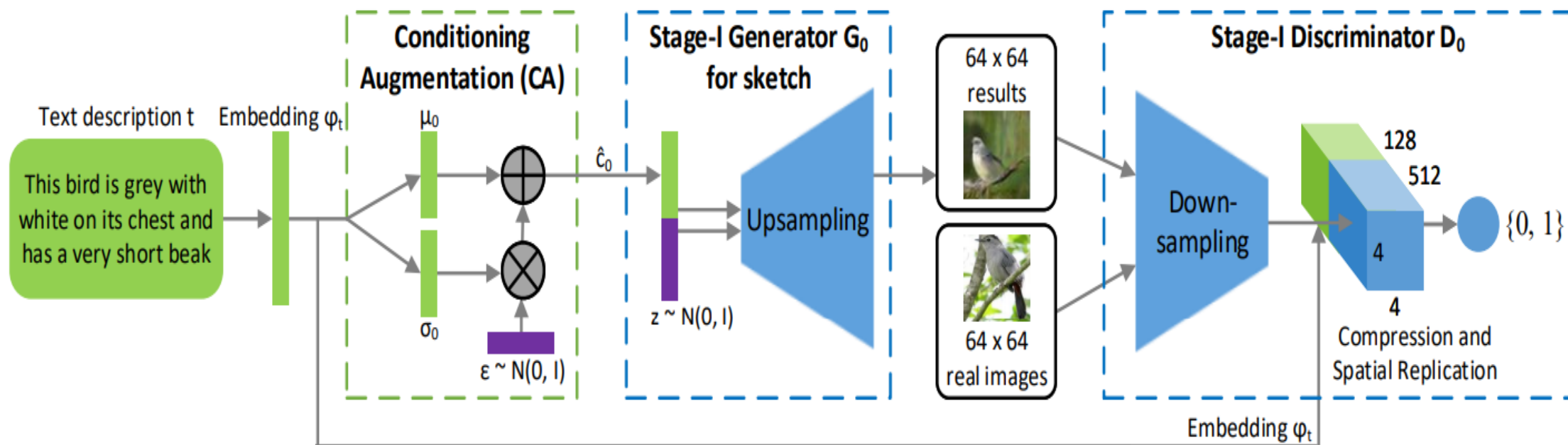




## Stage-1 GAN Loss Function

$$\mathcal{L}_{D_0} = \mathbb{E}_{(I_0, t) \sim p_{data}} [\log D_0(I_0, \varphi_t)] + \mathbb{E}_{z \sim p_z, t \sim p_{data}} [\log(1 - D_0(G_0(z, \hat{c}_0), \varphi_t))],$$

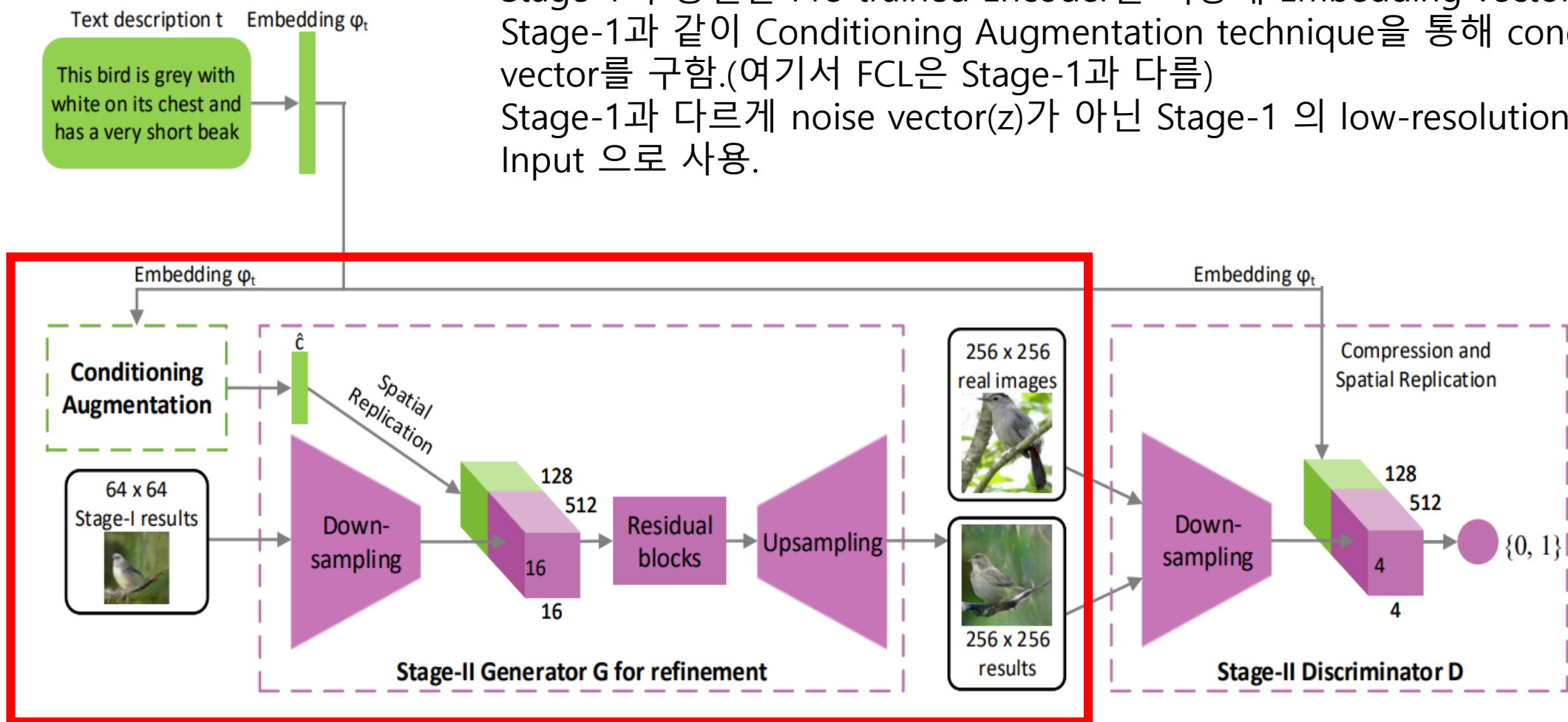
$$\mathcal{L}_{G_0} = \mathbb{E}_{z \sim p_z, t \sim p_{data}} [\log(1 - D_0(G_0(z, \hat{c}_0), \varphi_t))] + \lambda D_{KL}(\mathcal{N}(\mu_0(\varphi_t), \Sigma_0(\varphi_t)) || \mathcal{N}(0, I)),$$





## Stage-2 GAN / Generator

Stage-1과 동일한 Pre-trained Encoder를 이용해 Embedding vector 생성.  
Stage-1과 같이 Conditioning Augmentation technique을 통해 conditioning vector를 구함.(여기서 FCL은 Stage-1과 다름)  
Stage-1과 다르게 noise vector( $z$ )가 아닌 Stage-1 의 low-resolution image를 Input 으로 사용.



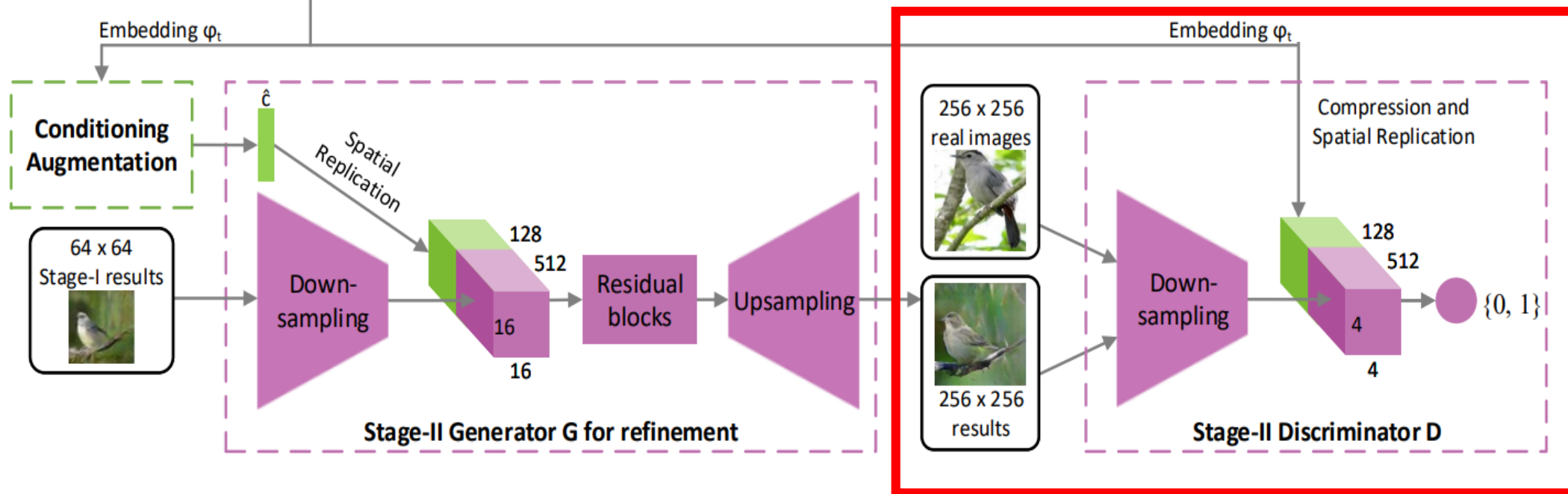
## Stage-2 GAN / Discriminator

Text description  $t$  Embedding  $\varphi_t$

This bird is grey with white on its chest and has a very short beak

**G**로부터 생성된 256x256 high image-resolution.

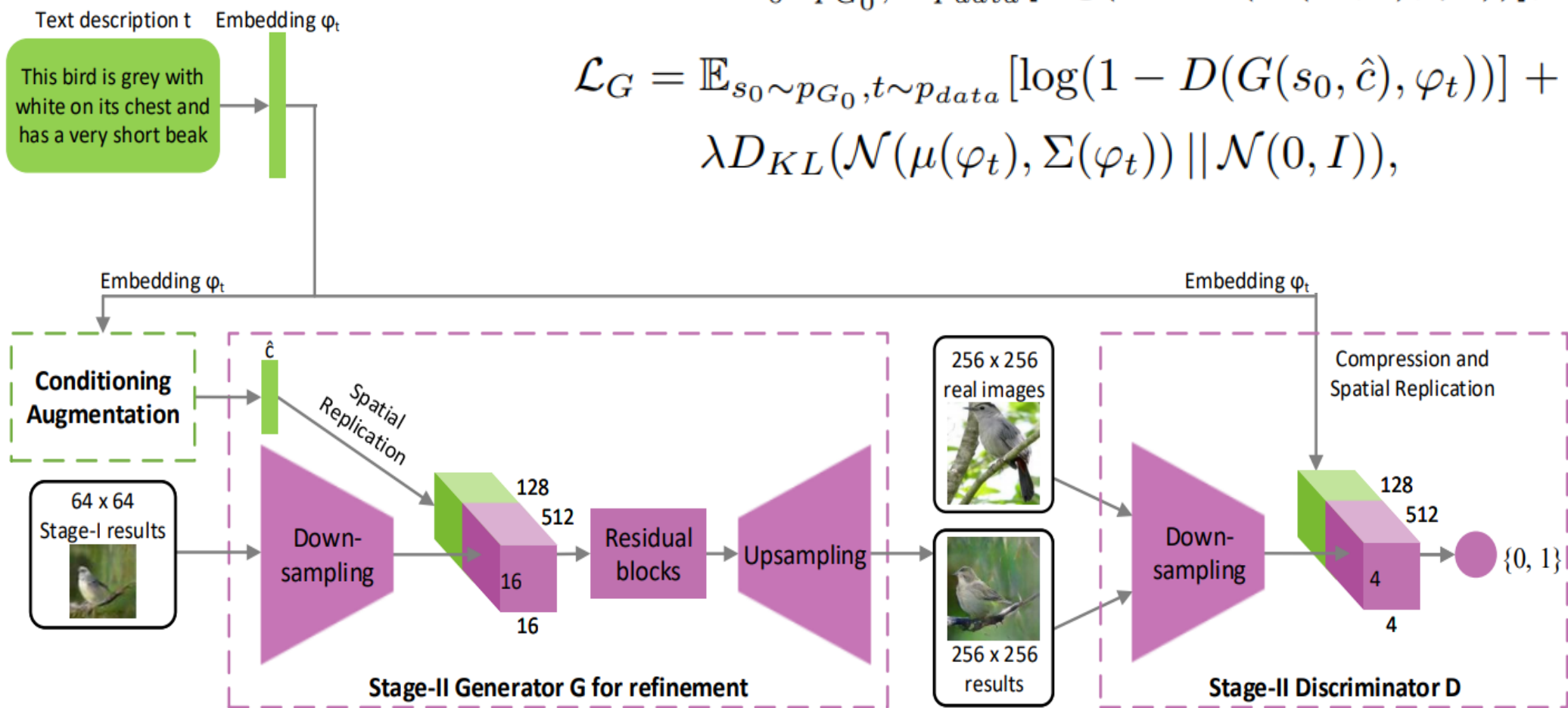
Input 으로 들어온 image를 Down-sampling -> 4x4x512 image map 생성  
Embedding vector(128 dimension)를 Replication -> 4x4x128의 tensor 생성  
1x1 convolutional layer -> image, text feature 관계  
FCL with 1 node



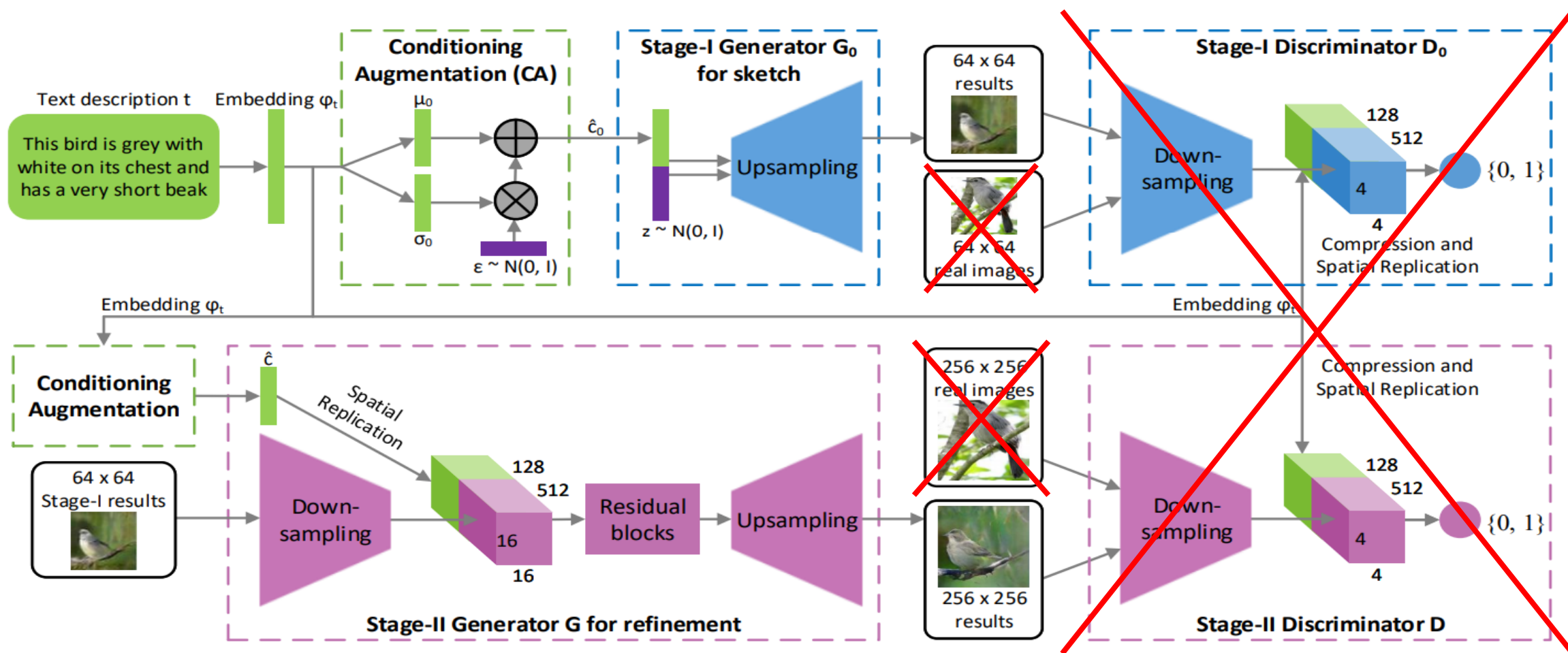
## Stage-2 GAN Loss Function

$$\mathcal{L}_D = \mathbb{E}_{(I,t) \sim p_{data}} [\log D(I, \varphi_t)] + \mathbb{E}_{s_0 \sim p_{G_0}, t \sim p_{data}} [\log(1 - D(G(s_0, \hat{c}), \varphi_t))],$$

$$\mathcal{L}_G = \mathbb{E}_{s_0 \sim p_{G_0}, t \sim p_{data}} [\log(1 - D(G(s_0, \hat{c}), \varphi_t))] + \lambda D_{KL}(\mathcal{N}(\mu(\varphi_t), \Sigma(\varphi_t)) || \mathcal{N}(0, I)),$$



# Photo-realistic image generation





# Experiments and Comparison

Text description	This bird is blue with white and has a very short beak	This bird has wings that are brown and has a yellow belly	A white bird with a black crown and yellow beak	This bird is white, black, and brown in color, with a brown beak	The bird has small beak, with reddish brown crown and gray belly	This is a small, black bird with a white breast and white on the wingbars.	This bird is white black and yellow in color, with a short black beak
Stage-I images							
Stage-II images							

## Experiments and Comparison

This flower has overlapping pink pointed petals surrounding a ring of short yellow filaments

Stage-I



Stage-II





## Experiments and Comparison











# Experiments and Comparison




















# Experiments and Comparison

Text description	This bird is red and brown in color, with a stubby beak	The bird is short and stubby with yellow on its body	A bird with a medium orange bill white body gray wings and webbed feet	This small black bird has a short, slightly curved bill and long legs	A small bird with varying shades of brown with white under the eyes	A small yellow bird with a black crown and a short black pointed beak	This small bird has a white breast, light grey head, and black wings and tail
64x64 GAN-INT-CLS							
128x128 GAWWN							
256x256 StackGAN							



# Experiments and Comparison

Text description	This flower has a lot of small purple petals in a dome-like configuration	This flower is pink, white, and yellow in color, and has petals that are striped	This flower has petals that are dark pink with white edges and pink stamen	This flower is white and yellow in color, with petals that are wavy and smooth	A picture of a very clean living room	A group of people on skis stand in the snow	Eggs fruit candy nuts and meat served on white dish	A street sign on a stoplight pole in the middle of a day
64x64 GAN-INT-CLS								
256x256 StackGAN								

# Experiments and Comparison

This bird is white with some black on its head and wings, and has a long orange beak

This bird has a yellow belly and tarsus, grey back, wings, and brown throat, nape with a black face

This flower has overlapping pink pointed petals surrounding a ring of short yellow filaments

(a) StackGAN Stage-I  
64x64 images



(b) StackGAN Stage-II  
256x256 images



(c) Vanilla GAN  
256x256 images





# Experiments and Comparison

A small bird with a black head and wings and features grey wings

This bird is completely red with black wings and pointy beak

256x256  
Stage-I GAN  
without CA



256x256  
Stage-I GAN  
with CA



256x256  
StackGAN  
with CA,  
Text twice



끝