Text to image

Lin Shaochuan

SCUT.   Intelligent Information Fusion Research Group

The Objection:

Text $\xrightarrow{\text{embedding}}$ Vector $\xrightarrow{\text{Up-sampling}}$ Image

e.g.

一只黑色羽毛的鸟 $\longrightarrow$ [0.22,0.10,0.11...0.64] $\xrightarrow[\text{Up-sampling}]{\text{Reshape}}$
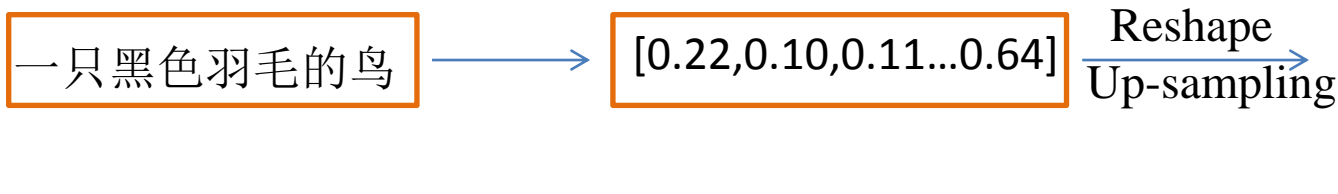
Main Challenge:

Multimodal!

Text description maps a number of images correctly

Generative Adversarial Networks (GAN)

SCUT. Intelligent Information Fusion Research Group

Lin Shaochuan
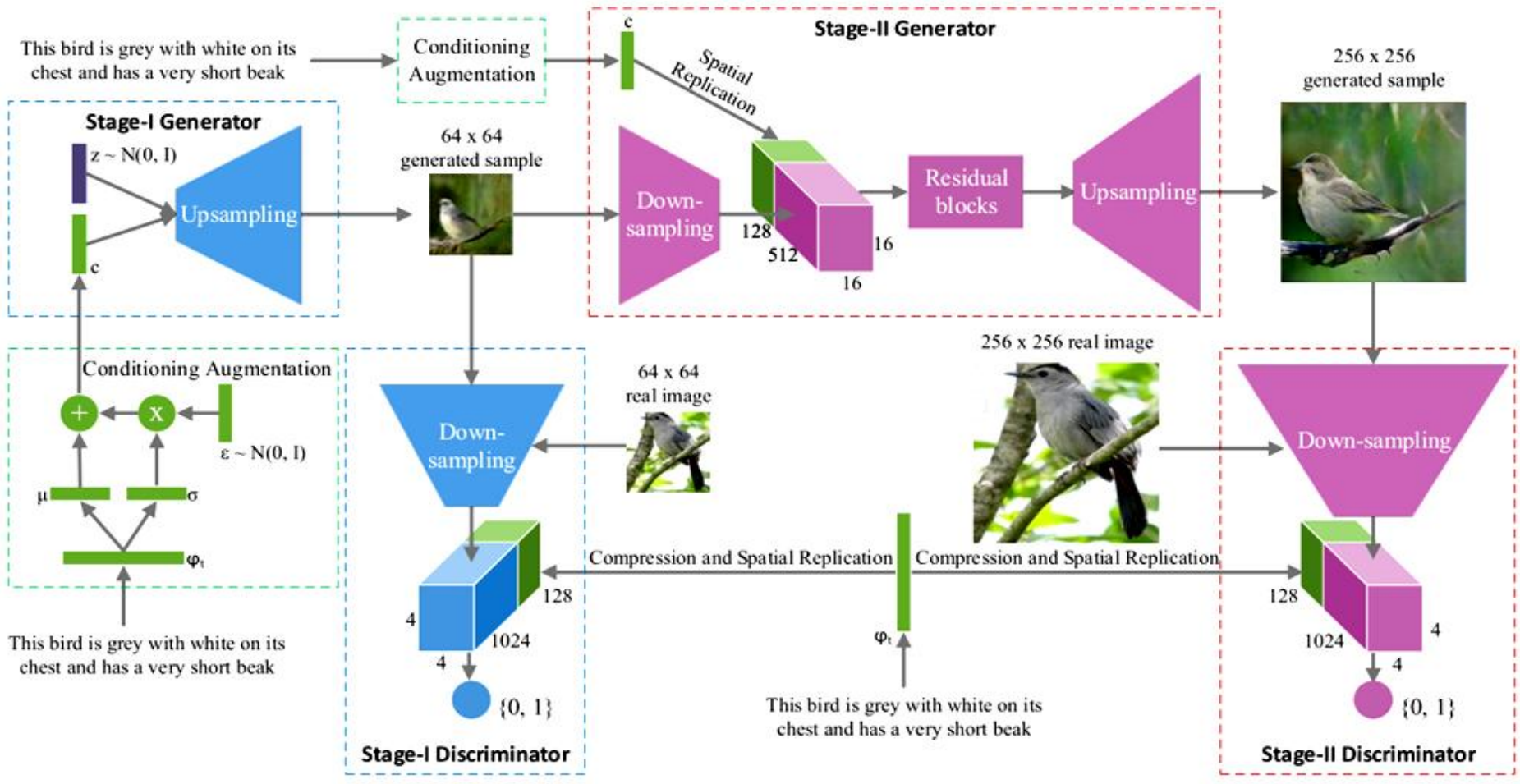
Preliminaries GAN：

生成器G与鉴别器D：

• 鉴别器D想要<u>鉴别能力最大化</u>，既对于真实图片的输入都想要尽可能输出1的概率，对于假的图片（生成器G生成的图片），想要尽可能输出0的概率。
（这里用1代表真实，0代表伪造。越接近1真实度越高，越接近0伪造度越高）

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[\log D\left(\boldsymbol{x}^{(i)}\right) + \log\left(1 - \boxed{D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)}\right) \right]$$

• 生成器G则是想要<u>生成能力尽可能好</u>。想生成的图片尽可能骗过鉴别器，既想让D(G生成的图片)输出概率尽可能接近1，也就是1-D(G生成的图片)尽可能地接近0。

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G\left(\boldsymbol{z}^{(i)}\right)\right)\right)$$
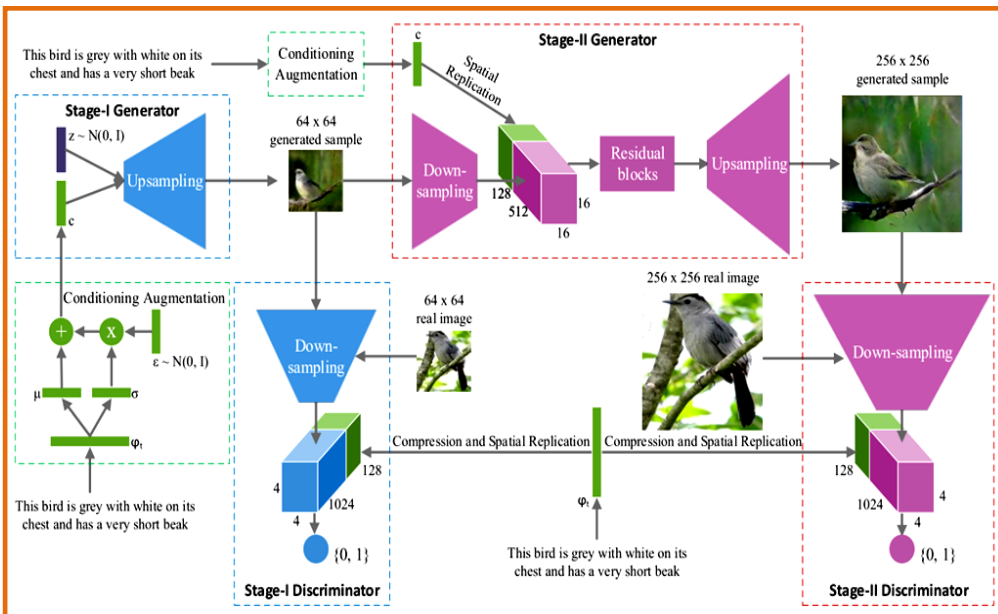
Model Architecture:

Reference : StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks . In ICCV 2017

Model Architecture:

-Stage-I GAN:
根据文本产生物体大致的形状和
基本的颜色，并且从随机噪声中产生背景。
从而生成低分辨率的图像
64*64

-Stage-II GAN:
从Stage-I生成的模型中读取生成的数据，
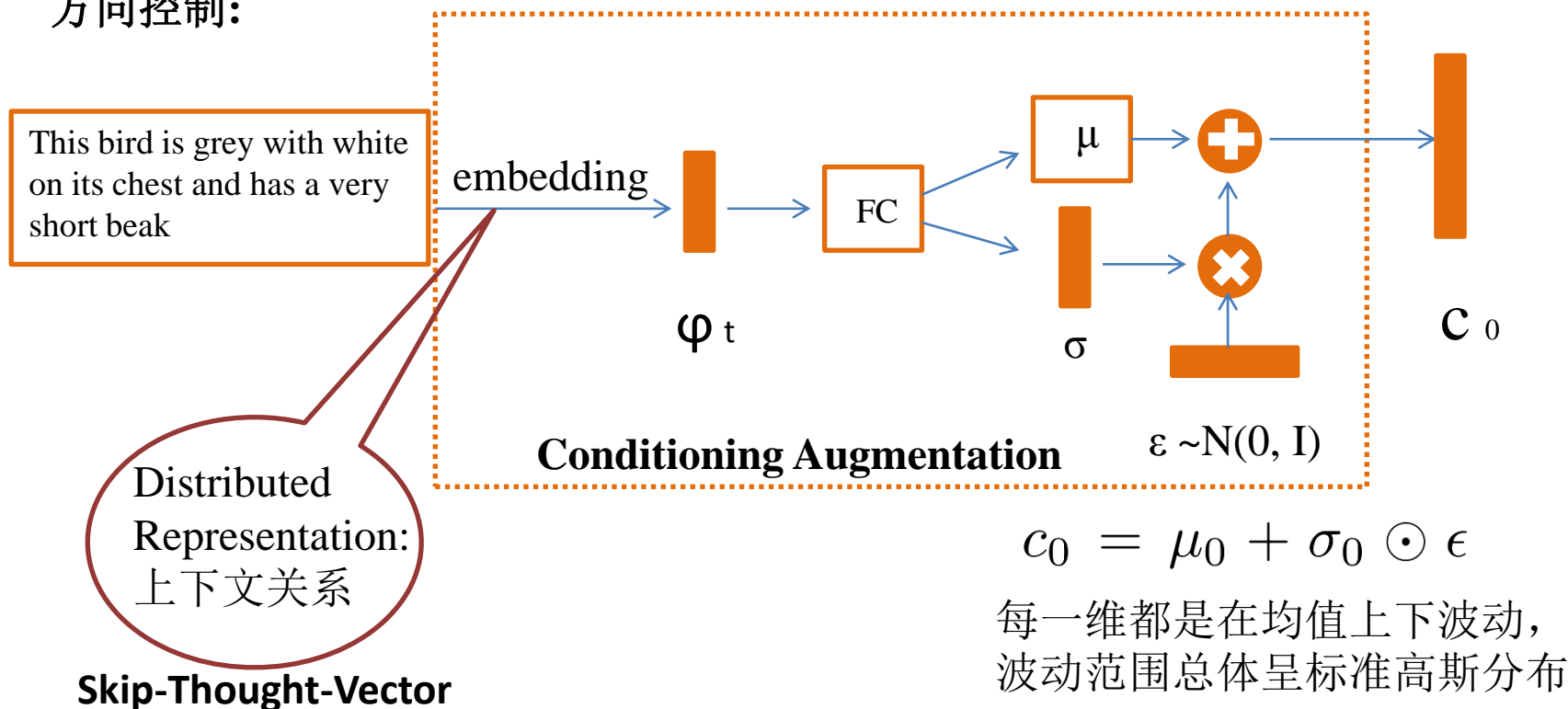并再次读取文本，生成细节更多，
与文本更加贴切的图像，且分辨率更大：
256*256

-亮点：
1.将生成任务分成两个阶段训练，每个阶段模型的重点都不同

2.由于之前训练好的向量化模型会**文字映射到高维（128维），**
而我们现在描述的文本数据较少，会导致映射到高维的向量比较离散不连续。
对映射到高维向量进行数据增强。

The Stage-I GAN
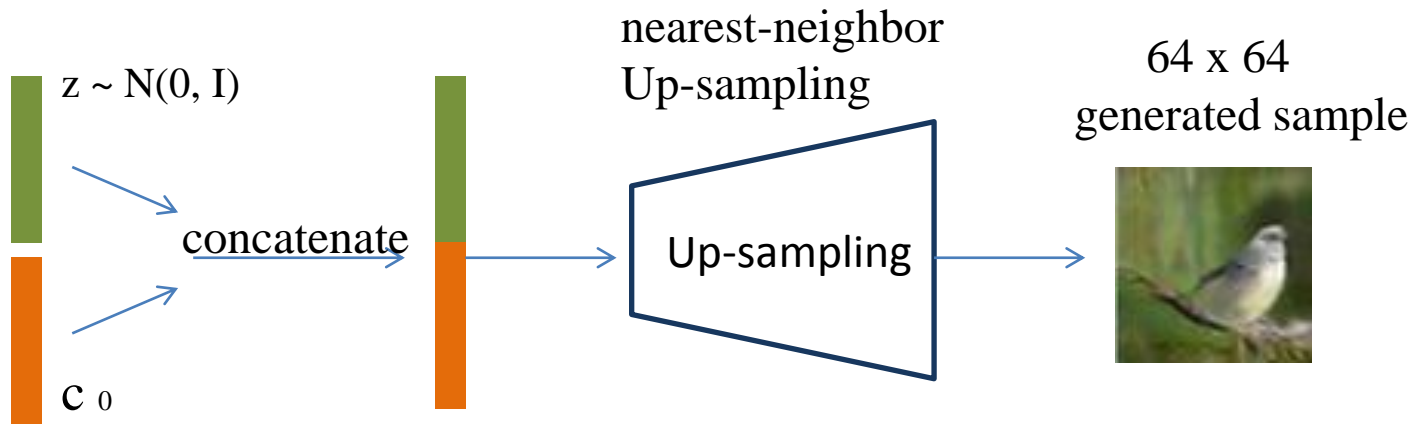
由于原始生成器G的输入一般为高斯噪声z，为了控制生成器的生成方向，我们加入条件变量c，使**生成器G**不仅能从z中生成比较随机的数据，还能朝c的方向生成。同时对于**辨别器D**来说，不仅要判别图像的真实性，还有判别方向的正确性。而这里的文本刚好可以来对原始生成器G和原始鉴别器进行方向控制。
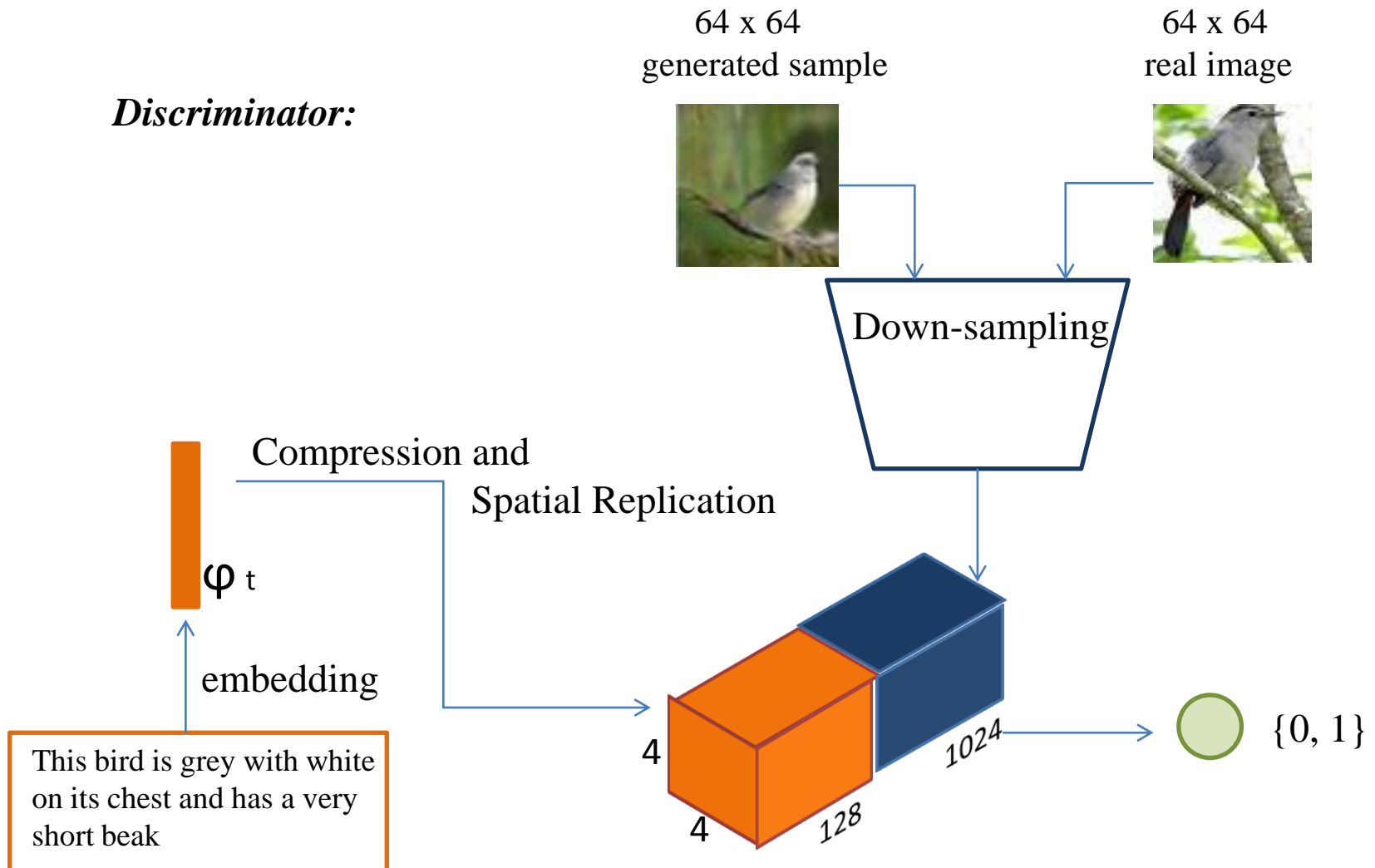
方向控制:

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

embedding

φ_t

FC

μ

σ

$\varepsilon \sim N(0, I)$

c_0

Conditioning Augmentation

Distributed Representation: 上下文关系

Skip-Thought-Vector

$$c_0 \ = \ \mu_0 \ + \ \sigma_0 \odot \epsilon$$

每一维都是在均值上下波动，波动范围总体呈标准高斯分布

The Stage-I GAN

Generator:

z ~ N(0, I)

c_0

concatenate

nearest-neighbor
Up-sampling

Up-sampling

64 x 64
generated sample

SCUT. Intelligent Information Fusion Research Group

Lin Shaochuan

The Stage-I GAN

Discriminator:

The Stage-I GAN

Maximize:

$$\mathcal{L}_{D_0} = \mathbb{E}_{(I_0,t)\sim p_{data}}[\log D_0(I_0, \varphi_t)] + \\ \mathbb{E}_{z\sim p_z, t\sim p_{data}}[\log(1 - D_0(G_0(z, c_0), \varphi_t))]$$

Minimize:

$$\mathcal{L}_{G_0} = \mathbb{E}_{z\sim p_z, t\sim p_{data}}[\log(1 - D_0(G_0(z, c_0), \varphi_t))] + \\ \lambda D_{KL}(\mathcal{N}(\mu_0(\varphi_t), \Sigma_0(\varphi_t)) \,\|\, \mathcal{N}(0, I)),$$

Notes: 实际上 tf.nn.sigmoid_cross_entropy_with_logits (logits ,label)

SCUT. Intelligent Information Fusion Research Group

Lin Shaochuan

The Stage-I GAN

row 0
although primarily black and white this bird has a yellow orange patch on his crown

row 1
the bird has a curved throat and a long black bill.

row 2
a bird with a large rounded beak, a dark nape, and a light cheek patch.

row 3
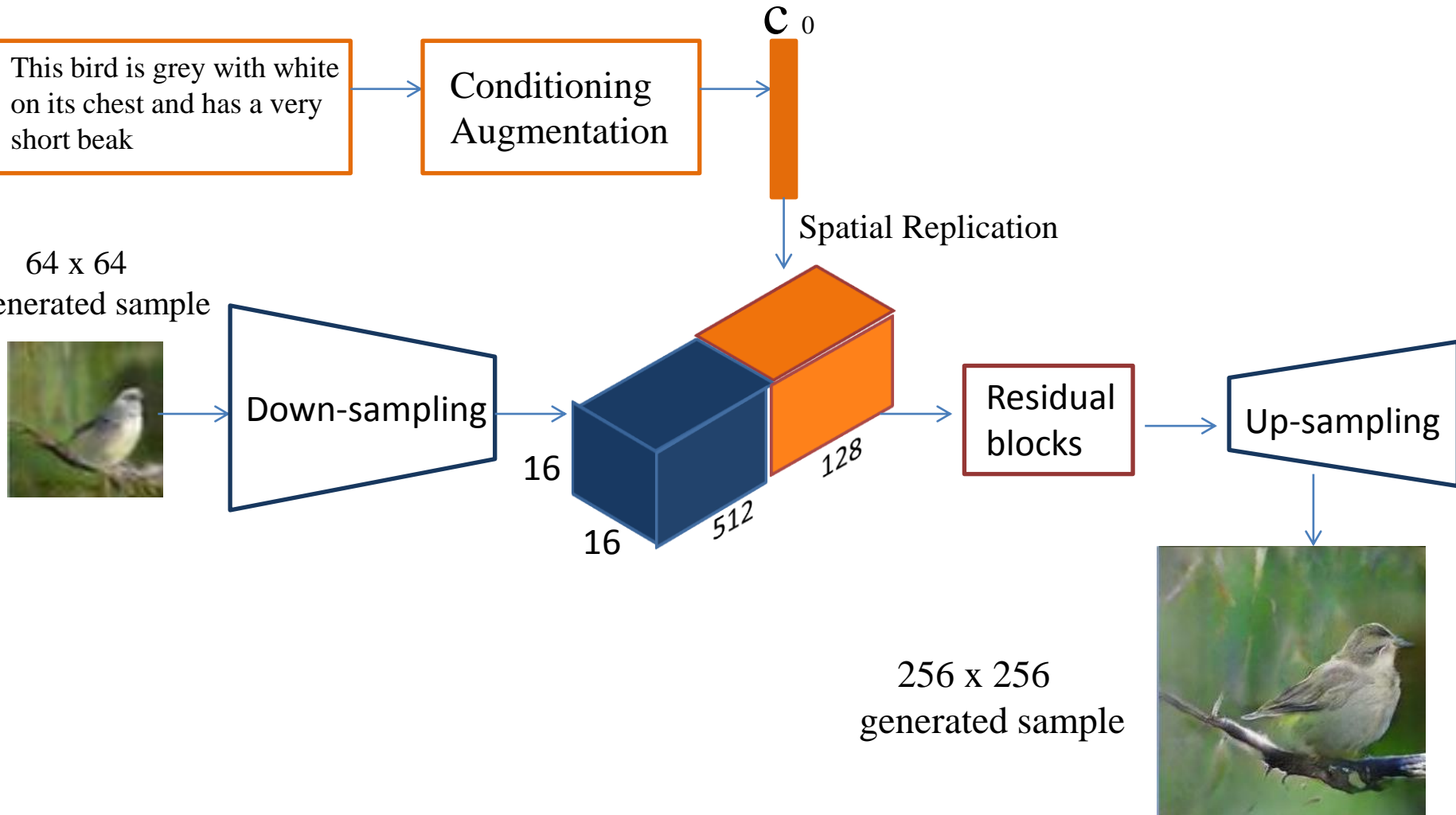this little bird has a blue crown, yellow throat and chest, small gray bill, greenish yellowish primaries.
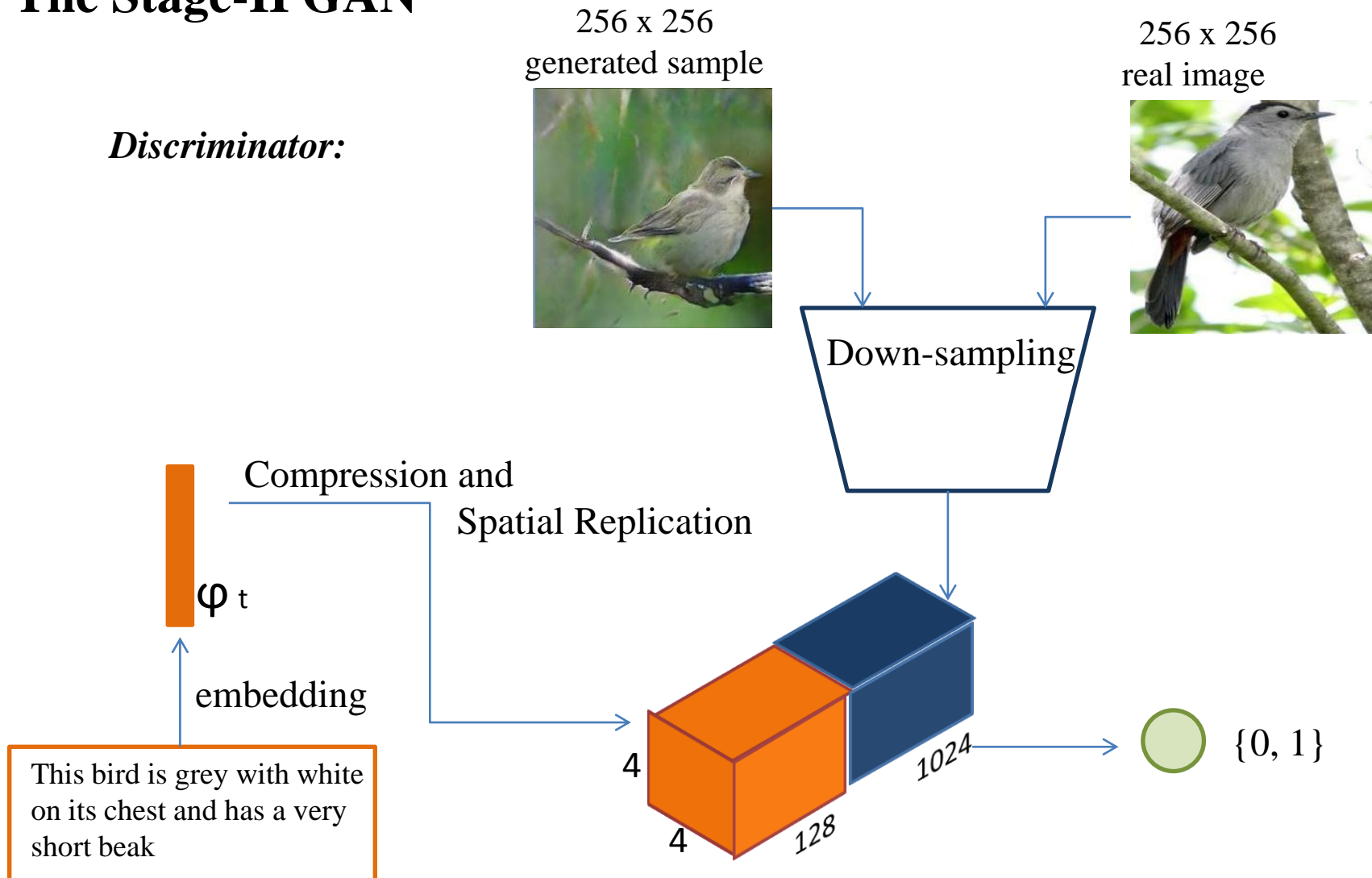

Real

SCUT. Intelligent Information Fusion Research Group

Lin Shaochuan

The Stage-II GAN

Generator:

This bird is grey with white on its chest and has a very short beak → Conditioning Augmentation → C_0 → Spatial Replication

64 x 64 generated sample → Down-sampling → 16, 16, 512, 128 → Residual blocks → Up-sampling → 256 x 256 generated sample

The Stage-II GAN

Discriminator:

256 x 256
generated sample

256 x 256
real image

Down-sampling

Compression and
Spatial Replication

ϕ t

embedding

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

4

4

128

1024

{0, 1}

The Stage-II GAN

Maximize:

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

Minimize:

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

The StackGAN Result on CUB birds dataset:

Warning:
Inception Score 判别模型生成情况，但是无法判别与文字的匹配程度

SCUT. Intelligent Information Fusion Research Group

Lin Shaochuan

The StackGAN Result on CUB birds dataset:

The bird is completly red → The bird is completly yellow

This bird is completely red with black wings and pointy beak → this small blue bird has a short pointy beak and brown on its wings

This bird is completely red with black wings and pointy beak → The bird has a yellow breast with grey features and a small beak

SCUT. Intelligent Information Fusion Research Group

Lin Shaochuan