深度学习实验六

组员 1 姓名：江经 学号：1190202328 班级：19030103 班
组员 2 姓名：朱健坤 学号：1190201924 班级：19030103 班

一. 选题说明及任务描述

本次深度学习实验六我们的选题是根据文字生成图片，即根据输入的一句话生成一张图片。我们了解到在该领域有一些经典网络结构：

1.条件生成对抗网络 CGAN，CGAN 是在 GAN 基础上做的一种改进，通过给原始 GAN 的生成器 Generator 和判别器 Discriminator 添加额外的条件信息（如类别标签或者其它的辅助信息），实现条件生成模型；

2.文本生成图像的开山之作 GAN-INT-CLS，该网络在对抗网络中加入了匹配感知鉴别器和流形插值学习。通过在判别器的输入增加真实图像和错误的文本描述，让判别器能够更好地学习文本描述和图片内容的对应关系。同时简单地在训练集文本的嵌入之间进行插值来生成大量额外的文本嵌入，增加了文本的变化，从而让生成器具有更强大的生成能力。

3.AttnGAN，Attentional Generative Ad-versarial Network(AttnGAN)是一种注意力驱动的多阶段的细粒度文本到图像生成器。同时网络还借助一个深层注意多模态相似模型(deep attentional multimodal similarity model)来训练该生成器。首次表明分层注意 GAN 能够自动选择单词级别的条件来生成图像的不同部分。

最终，我们选择了发表于 CVPR 的网络 DF-GAN 和 SSA-GAN 来完成我们的实验。DF-GAN 模型抛弃了以往的堆叠结构，只使用一个生成器、一个鉴别器、一个预训练过的文本编码器。SSA-GAN 在 DF-GAN 的模型基础上进行了改进，它把把 UPBlocks 改成了 SSACN Blocks，将文本编码器可以与图像生成器联合训练，更好地利用文本信息生成图像。

二. 数据集描述

1.CUB-200-2011 数据集

altech-UCSD Birds-200-2011 (CUB-200-2011) 是 CUB-200 dataset 的一个扩充版本,每个类的图像数量大约增加两倍和新的部位注释。CUB-200-201 包含 200 种鸟类，图片总数目为 11788 张，每张图片的标注信息有 15 个 Part Locations、312 个 Binary Attributes、1 个 Bounding Box。下载地址：
http://www.vision.caltech.edu/datasets/cub_200_2011/

本实验中，使用该数据集的图片作为训练集、测试集，而对于图片的描述（caption）来源于 https://drive.google.com/file/d/1O_LtUP9sch09QH3s_EBAgLEctBQ5JBSJ/view，每张图片有 10 个 caption，并且重新划分了 CUB-200-2011 的训练集和测试集，训练集有 8855 张图片，测试集有 2933 张图片。

2.预处理过程

2.1 图片部分

```
image_transform = transforms.Compose([
    transforms.Resize(int(imsize * 76 / 64)),
    transforms.RandomCrop(imsize),
    transforms.RandomHorizontalFlip()])
```

```
self.norm = transforms.Compose([
    transforms.ToTensor(),
    transforms.Normalize((0.5, 0.5, 0.5), (0.5, 0.5, 0.5))])
```

图 1 图片处理

图片处理部分采用了随机裁剪、随机水平翻转，并且图片 resize 为 256*256，将图片进行了归一化。

2.2 文字部分

先利用所有 caption 构造词典，词典共 5450 个词，之后读取 caption，将 caption 进行 lower()操作，并且将其 tokenize，将每个 token 都根据词典转换为 id。

三. 方案设计（包含损失函数和网络结构等）

1.DF-GAN

1.1 简介

从文本描述中合成高质量的真实图像是一项具有挑战性的任务。现有的文本到图像生成对抗性网络通常采用堆叠式架构作为主干，但仍然存在三个缺陷。首先，堆叠结构引入了不同图像尺度的生成器之间的纠缠。第二，现有研究倾向于在训练中修复额外的网络，以实现文本-图像语义一致性，但是这限制了这些网络的监控能力。第三，以往研究中广泛采用的基于跨模态注意的文本图像融合由于计算量大而局限于几种特殊的图像尺度。为此，模型提出了一种更简单但更有效的深度融合生成性对抗网络（DF-GAN）。具体来说，DF-GAN 提出： (i) 一种新的单级文本到图像主干，它直接合成高分辨率图像，而不同生成器之间没有纠缠； (ii) 一种由匹配软件梯度惩罚和单向输出组成的新的目标感知鉴别器，它在不引入额外网络的情况下增强了文本图像的语义一致性， (iii) 一种新的深文本图像融合块，它深化了融合过程，使文本和视觉特征完全融合。与目前最先进的方法相比，模型提出的 DFGAN 在合成真实感和文本匹配图像方面更简单但效率更高，并且在广泛使用的数据集上实现了更好的性能。

1.2 网络结构

1.2.1 整体结构

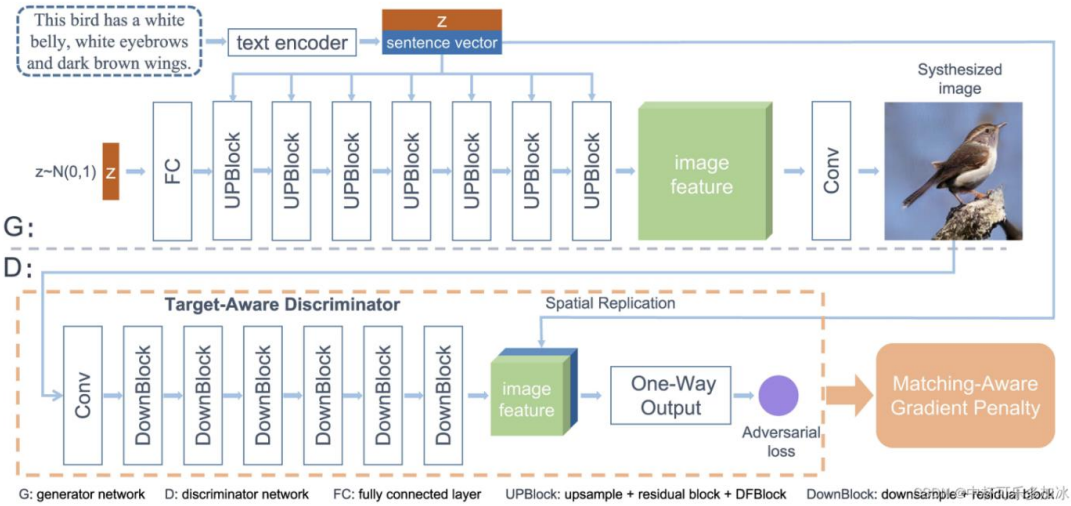


图 2-1 DF-GAN 框架

1.2.2 匹配感知梯度惩罚

作者设计的鉴别器叫 Target-Aware Discriminator，由匹配感知梯度惩罚（MA-GP）和

单向输出(One-Way Output)组成，主要作用就是促使生成器合成更真实更符合语义一致性的图像。

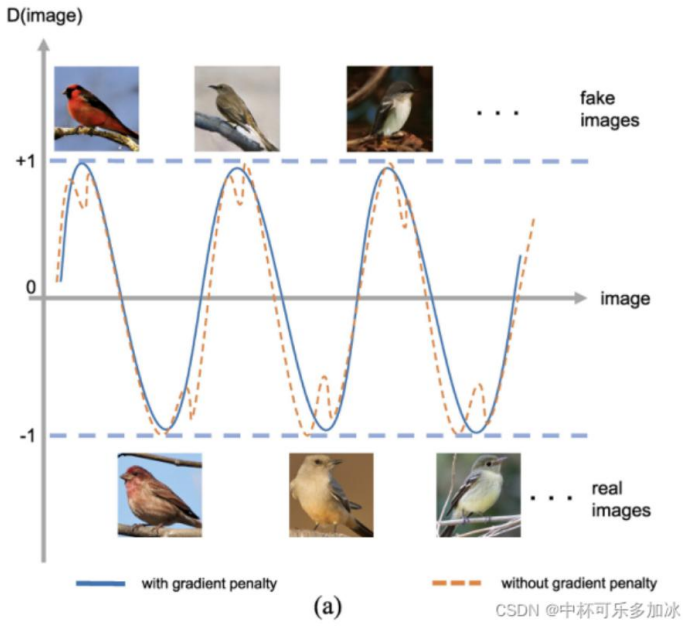


图 2-2 梯度惩罚

梯度惩罚（Gradient Penalty）是 WGAN-gp 曾提出的一种梯度变化方法，如上图所示，首先使用 hinge loss 将鉴别器损失范围限制在-1 和 1 之间，梯度越大，惩罚越大，即改变梯度的程度越大。损失函数的计算公式如下：

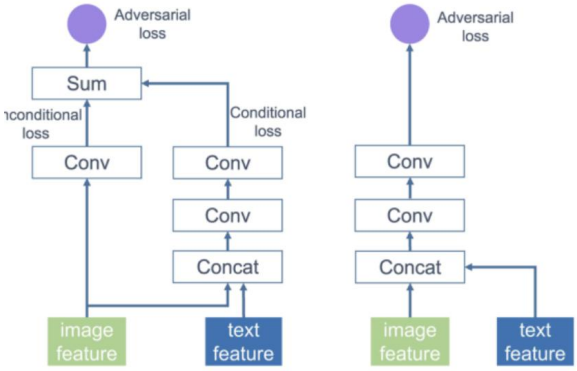
$$
\begin{aligned}
L_D = & -\mathbb{E}_{x\sim\mathbb{P}_r}[\min(0,-1+D(x,e))] \\
& -(1/2)\mathbb{E}_{G(z)\sim\mathbb{P}_g}[\min(0,-1-D(G(z),e))] \\
& -(1/2)\mathbb{E}_{x\sim\mathbb{P}_{mis}}[\min(0,-1-D(x,e))] \\
& +k\mathbb{E}_{x\sim\mathbb{P}_r}\left[(\|\nabla_x D(x,e)\|+\|\nabla_e D(x,e)\|)^p\right] \\
L_G = & -\mathbb{E}_{G(z)\sim\mathbb{P}_g}[D(G(z),e)]
\end{aligned}
$$

1.2.3 单向输出

在以往的 T2I 任务中，鉴别器以两种方式来进行判断，一是判断图像是真是假，二是将图像特征与句子向量连接起来，判断图像与文本是否语义一致（有条件损失）。这个被作者称 Two-Way Output。研究表明，这种 Two-Way Output 其实减慢了生成器的收敛速度。

条件损失给出的梯度α指向图像与匹配文本的方向，无条件损失的梯度β仅指向真实图像的方向，最终的梯度方向只是简单的求和α+β，并不像预期那样指向（真实，匹配）的方向，这样的过程会减慢图像与文本的一致性。

故作者提出了单向输出（One-Way Output），其鉴别器将图像特征和句子向量连接起来，然后通过两个卷积层仅输出一个对抗性损失。这样设计可以使单个梯度γ直接指向目标数据点（真实和匹配），从而优化和加速生成器的收敛。



(a) Two-Way Output (b) One-Way Output

gure 4. Comparison between Two-Way Output and our One-

图 2-3　单向输出和双向输出

上图表明了 Two-Way Output 和 One-Way Ouput 的区别，Two-Way Output 首先根据图像特征计算无条件损失，然后将图像特征连接文本特征再计算有条件损失，再将两个损失连接。而 One-Way Output 将图像特征与文本特征直接连接后经过两个卷积层直接计算总损失。

通过结合 MA-GP 和单向输出，目标感知鉴别器可以引导生成器合成更多真实和文本匹配的图像。

1.2.4 生成器

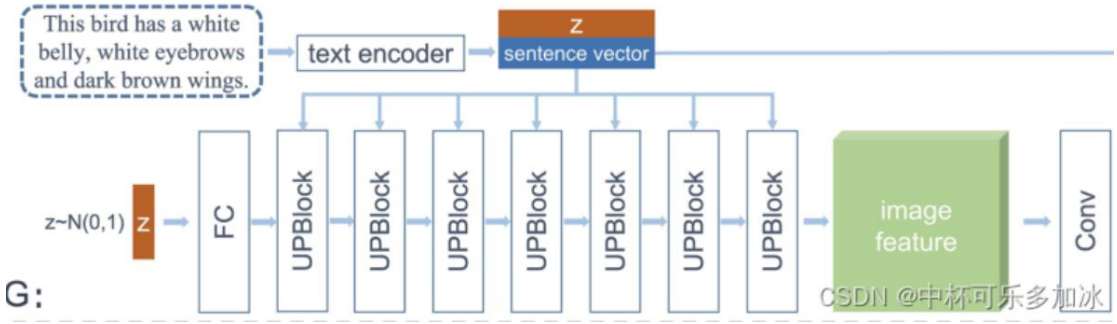


图 2-4 生成器

生成器由七个 UPBlocks 组成，UPBlocks 包括上采样、残差块和 DFBlock，DFBlock 是作者提出的一种深度文本图像融合块，其在融合块中叠加了多个仿射变换和 ReLU 层。

2.SSA-GAN

2.1 简介

Semantic-Spatial Aware GAN（SSA-GAN）提出了一种新的语义空间感知 GAN 框架，文章发表于 2021 年 10 月。文本到图像生成（T2I）模型旨在生成语义上与文本描述一致的照片逼真图像。在生成性对抗网络（GAN）最新进展的基础上，现有的 T2I 模型取得了巨大进展。然而，仔细检查它们生成的图像会发现两个主要局限性：（1）条件批量归一化方法平等适用于整个图像特征映射，忽略了局部语义；（2）文本编码器在训练过程中是固定的，它应该与图像生成器一起训练，以学习更好的文本表示，从而生成图像。为了解决这些局限性，SSA-GAN 提出了一种新的语义空间感知 GAN 框架，该框架以端到端的方式进行训练，以便文本编码器能够利用更好的文本信息。具体来说，SSA-GAN 介绍了一种新的语义空间感知卷积网络，该网络（1）学习以文本为条件的语义自适应仿射变换，以有效地融

合文本特征和图像特征；（2）以弱监督的方式学习掩码映射，该方法依赖于当前的文本-图像融合过程，以在空间上指导变换。在具有挑战性的 COCO 和 CUB bird 数据集上进行的实验表明，SSA-GAN 的方法在视觉保真度和与输入文本描述的一致性方面优于最近的最新方法。本实验在 CUB bird 上对 SSA-GAN 进行了复现。

2.2 网络结构

2.2.1 整体结构

SSA-GAN 的框架如下：

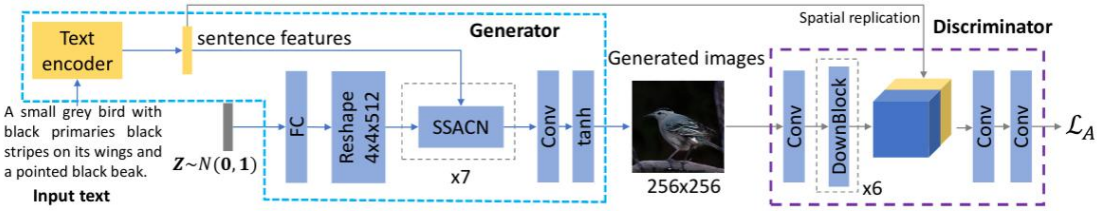


Figure 2: A schematic of our framework SSA-GAN. It has one generator-discriminator pair. The generator mainly consists of 7 proposed SSACN blocks which fuse text and image features through the image generation process and guarantee the semantic text-image consistency. The gray lines indicate the data streams only for training.

图 3-1 SSA-GAN 框架

整体来看，SSA-GAN 与 DF-GAN 十分相似，也是单级主干结构。但 SSA-GAN 将 DF-GAN 的 UPBlocks 改成了 SSACN Blocks。SSA-GAN 包括一个文本编码器（Text encoder），一个生成器（Generator），一个鉴别器（Discriminator）。

生成器部分：首先输入网络一个随机整体噪声，经过 FC 层和一次 Reshape 后，连接七个 SSACN 层；而同时，输入 caption，经过 Text encoder 编码获得 caption 的特征，将其送入 SSACN 中；之后 text 和 noise 的融合向量在经过 Conv 层和 tanh 激活之后，生成图片；

鉴别器部分：生成的图片随后送入输入鉴别器进行鉴别，图片经过 Conv 和多个 DownBlock 加上多个 Conv 后进行判别。需要注意的是，在 SSA-GAN 中，文本编码器不固定参数，其也是生成器的一部分。

2.2.2 Text encoder

Text encoder 由 Bi-LSTM 组成，通过最小化深度注意多模态相似模型（DAMSM）损失，使用真实图像-文本对进行预训练得到。其中，深度注意多模态相似度模型（DAMSM）用于在单词级和句子级测量图像-文本相似度，以计算图像生成的细粒度损失。其将给定的文本描述编码为具有 256 维的句子特征向量（取自 LSTM 的最后隐藏状态）和 256 维的单词特征（取自 LSTM 每个时间步的隐藏状态，最多 18 个单词）。

2.2.3 Semantic-Spatial Aware Convolutional Network（SSACN block）

每个 SSACN 块由一个 upsample block（上采样块）、一个 mask predictor（掩码预测器）、一个 Semantic-Spatial Condition Batch Normalization (SSCBN，语义空间条件批量归一化和一个 residual block（残差块）组成。如图 3-2。

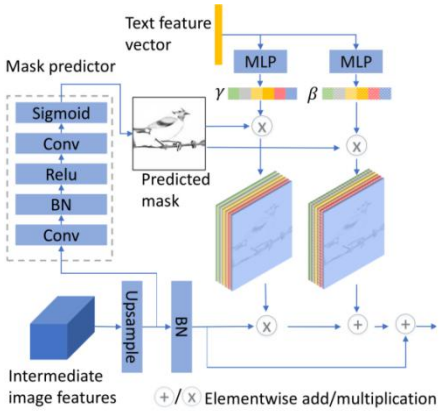


图 3-2 SSACN block 结构

(1) upsample block。其通过双线性插值将图像特征图的宽度和高度加倍。

(2) mask predictor。其将上采样的图像特征图作为输入，预测其 mask 图。mask 图中，每个像素的值在 0 到 1 之间，每个像素的值决定在该位置上进行仿射变换的程度。mask 图基于当前生成的图像特征图，因此，它直观地指示了当前图像特征图的哪些部分仍需要根据文本信息进行增强，以便于改进后的图像特征图与给定文本语义更一致。mask predictor 与整个网络一起训练，但 mask predictor 没有特定的损失函数来指导其学习过程，也没有额外的 mask annotation，唯一的监督来自鉴别器给出的损失，因此，这是一个弱监督的过程。

(3) Semantic-Spatial Condition Batch Normalization (SSCBN）。SSCBN 根据文本特征向量学习语义感知仿射参数。根据当前文本图像融合过程（即最后一个 SSCBN 块的输出），预测 mask map。这种仿射变换有效而深入地融合了文本和图像特征，并使得图像特征与文本语义一致；原始 BN 首先将 mini-batch 的数据标准化为每个特征通道均是零平均值和标准方差：

$$\hat{x}_{nchw} = \frac{x_{nchw} - \mu_c(x)}{\sigma_c(x)},$$

$$\mu_c(x) = \frac{1}{NHW}\Sigma_{n,h,w}x_{nchw},$$

$$\sigma_c(x) = \sqrt{\frac{1}{NHW}\Sigma_{n,h,w}(x_{nchw} - \mu_c)^2 + \epsilon},$$

而 CBN 对通道进行了仿射变换，对每个 channel 都有两个参数开控制，而这两个参数都是学习得到：

$$\tilde{x}_{nchw} = \gamma_c\hat{x}_{nchw} + \beta_c,$$

也即是：

$$\tilde{x}_{nchw} = \gamma(con)\hat{x}_{nchw} + \beta(con).$$

在此基础上，为了能够更好地融合图像与文本的特征，这两个参数都是从文本向量$\bar{e}$得到，如：

$$\gamma_c = P_\gamma(\bar{e}), \quad \beta_c = P_\beta(\bar{e})$$

其中P_γ、P_β是两个 MLP，用于通过文本向量$\bar{e}$学习这两个参数。

如果不添加更多的空间信息，则上一步的 CBN 将在空间上均匀地进行图像特征映射，而 SSCBN 希望只调整特征图中与文本有关的部分。SSCBN 的实现如下：

$$\tilde{x}_{nchw} = m_{i,(h,w)}(\gamma_c(\bar{e})\hat{x}_{nchw} + \beta_c(\bar{e})).$$

其中$m_{i,(h,w)}$为 mask predictor 预测的掩码，用掩码来表示各个像素需要变换的权重。且在 SSCBN 是以文本向量（句子级）为条件的，与词级相比，文本向量（句子级）所需的计算量要少得多。

(4) residual block。其用于保持图像特征的主要内容，以防止与文本无关的部分发生变化，并防止图像信息被文本信息覆盖掉。

2.3 损失函数

2.3.1 鉴别器损失函数

MA-GP loss（Matching-Aware zero-centered Gradient Penalty）

$$\begin{aligned}\mathcal{L}_{adv}^{D} =& E_{x\backsim p_{data}}[max(0, 1 - D(x, s))] \\ &+ \frac{1}{2}E_{x\backsim p_G}[max(0, 1 + D(\hat{x}, s))] \\ &+ \frac{1}{2}E_{x\backsim p_{data}}[max(0, 1 + D(x, \hat{s}))] \\ &+ \lambda_{MA}E_{x\backsim p_{data}}[(\|\nabla_x D(x, s)\|_2 \\ &+ \|\nabla_s D(x, s)\|_2)^p],\end{aligned}$$

其中 s 是给定的文本描述，$\hat{s}$是不匹配的文本描述，x 是 s 对应的真实图像，$\hat{x}$是生成的图像。λ_{MA}、p都是 MA-GP 的超参数。

2.3.2 生成器损失

生成器损失由两部分组成，分别是对抗损失和 DAMSM 损失（单词级细粒度图像文本匹配）

$$\mathcal{L}_G = \mathcal{L}_{adv}^{G} + \lambda_{DA}\mathcal{L}_{DAMSM}$$
$$\mathcal{L}_{adv}^{G} = -E_{x\backsim p_G}[D(\hat{x}, s)],$$

四. 实验过程

1.实验环境

操作系统：Linux　CentOS 3.10.0-1062.el7.x86_64

torch 版本：torch 1.11.0

cuda 版本：11.3

显卡：Nvidia RTX A100 * 2

IDE：PyCharm 2020.2.5 (Professional Edition)

2.实验过程

2.1 DF-GAN

由于计算资源有限，我们并没有从头开始训练 DF-GAN，而是接着他提供的已经训练了 600 个 epoch 的预训练模型继续训练了 50 个 epoch。由于 tensorboard 的文件有 1.8G，如有需要可前往[https://pan.baidu.com/s/1xThAbE1VnuzNytpp-ioM0Q 提取码: f1eo]进行下载。

2.2 SSA-GAN

由于时间以及计算资源有限，SSA-GAN 的训练使用了论文作者提供的训练了 550 个 epoch 的预训练模型。我们在此基础上继续训练了 50 个 epoch。由于 log 文件过大，如有需要请前往[https://pan.baidu.com/s/1xThAbE1VnuzNytpp-ioM0Q 提取码: f1eo]进行下载。

五. 实验结果与分析

1.评价指标

1.1 IS(Inception Score)

IS 使用预训练的 inception v3 网络来计算类条件分布（生成图像）和类边缘分布（真实图像）之间的 KL 散度。较大的 IS 表示生成的图像质量较高，且每个图像的类别易被分类器判别。其公式如下：

$$IS(G) = \exp(\mathbb{E}_{\mathbf{x}\sim p_g} D_{KL}(p(y|\mathbf{x})||p(y)))$$

其中 x 为生成的图像，y 为其类别标签。

1.2 FID(Fréchet Inception Distance)

FID 计算生成图像和真实图像的特征分布之间的 Fréchet 距离，其中的特征提取自预训练的 inception v3 网络。较低的 FID 表示生成的图像更真实。

2.性能指标评价与分析

本实验模型在 Pytorch 中实现，参考了论文作者的代码实现。

DF-GAN 的 batch_size 为 24，使用 1 块 Nvidia A100 GPU（40GB 显存）。训练中使用了β1=0.0 和β2=0.9 的 Adam 优化器。生成器和鉴别器的学习速率分别设置为 0.0001 和 0.0004。DF-GAN 模型在 CUB 数据集上训练了 650 个 epoch，选取最佳的模型（600epoch）进行评测。

SSA-GAN 的 batch_size 为 24，使用 2 块 Nvidia A100 GPU（40GB 显存）。训练中使用了β1=0.0 和β2=0.9 的 Adam 优化器。生成器和鉴别器的学习速率分别设置为 0.0001 和 0.0004。超参数 p、λMA 和λDA 分别设置为 6、2、0.1。SSA-GAN 模型在 CUB 数据集上训练了 600 个 epoch，选取最佳的模型（590epoch）进行评测。

为了评估 IS，本实验的模型均从测试数据集中随机选择文本描述，生成分辨率为 256×256 的 30k 张图片。为了评估 FID，本实验的模型均从测试数据集中随机选择文本描述，生成分辨率为 256×256 的 3k 张图片（为了能和测试集中的约 3k 图像数量保持一致）。实验结果如下：

表 DF-GAN 与 SSA-GAN 实验结果

模型	$IS\uparrow$	$FID\downarrow$	训练时间
DF-GAN	4.49 ± 0.06	20.96	**较短**

SSA-GAN	$\mathbf{4.91 \pm 0.08}$	**16.01**	较长

观察可知 SSA-GAN 的 IS 指标与 FID 指标均优于 DF-GAN，但缺点是 SSA-GAN 需要训练的模型更多，模型参数量更大，训练时间更长。

3.生成图像主观评价与分析

生成结果如下：

3.1 A small bird with an orange bill and grey crown and breast.

（一种有橙色喙、灰色冠和胸部的小鸟。）

GT　　DF-GAN　　SSA-GAN

观察可知，DF-GAN 与 SSA-GAN 都描述到了“orange bill and grey crown and breast”；但是 DF-GAN 的喙带有白色和黑色、不是较为纯净的橙色。且 DF-GAN 与 SSA-GAN 都没有特别地体现出“small bird”。

3.2 The bird has a bright red eye, a gray bill and a white neck.

（这只鸟有一只鲜红色的眼睛，灰色的喙和白色的脖子。）

GT　　DF-GAN　　SSA-GAN

观察可知，DF-GAN 与 SSA-GAN 都描述到了“a bright red eye, a gray bill and a white neck”；但是 DF-GAN 的喙带有一些白色，且其眼睛的鲜红色也不明显。

3.3 This bird has a long pointed beak with a wide wingspan.

（这种鸟有一个长而尖的喙，翼展很宽。）

GT　　DF-GAN　　SSA-GAN

观察可知，DF-GAN 与 SSA-GAN 都描述到了“long pointed beak with a wide

wingspan”；但是 DF-GAN 的头部出现了重叠，而 SSA-GAN 的头部没有出现在图片中。

3.4 A small bird with a black bill and a fuzzy white crown nape throat and breast.（一种有黑色喙和绒毛状白色冠的颈部和胸部的小鸟）

GT　　DF-GAN　　SSA-GAN

观察可知，DF-GAN 与 SSA-GAN 都描述到了“A small bird”、“a fuzzy white crown nape throat and breast”；但 DF-GAN 没有描述到“a black bill”，而 SSA-GAN 描述到了。

六. 方案评价

1.DF-GAN

该论文创新如下：

提出了一种新的单级文本到图像主干，可以直接合成高分辨率图像，而不需要不同生成器之间的纠缠。

提出了一种由匹配感知梯度惩罚（MA-GP）和单向输出组成的目标感知鉴别器。它在不引入额外网络的情况下显著增强了文本图像的语义一致性。

提出了一种新的深度文本图像融合块（DFBlock），它能更有效、更深入地融合文本和视觉特征。

但是 DF-GAN 也存在一些不足：

模型抛弃了 AttnGAN 以来提出的单词级信息，只引入了句子级的文本信息，这限制了细粒度视觉特征合成的能力

模型使用的 text encoder 仍然是 AttnGAN 中的 encoder。若引入预先训练过的大型语言模型来提供额外的知识可能会进一步提高性能。

Table 2. The performance of different components of our model on the test set of CUB.

Architecture	IS ↑	FID ↓	SC ↑
Baseline	3.96	51.34	-
OS-B	4.11	43.45	1.46
OS-B w/ DAMSM	4.28	36.72	1.79
OS-B w/ MA-GP	4.46	32.52	3.55
OS-B w/ MA-GP w/ OW-O	**4.57**	**23.16**	**4.61**

图 4 DF-GAN 论文中提供的消融实验结果

2.SSA-GAN

该论文提出了一种新的用于 T2I 生成的语义空间感知 GAN（SSA-GAN）框架，主要是

在生成器上做的工作，创新如下：

提出语义空间感知卷积网络（SSACN）模块，通过基于当前生成的图像特征图预测掩码映射草图，这种掩码图不仅可以指导生成器后续在何处添加文本信息，还起到了权重作用即决定要在某个部分上加强文本信息的程度。

提出新的仿射参数计算方法，将掩码图添加到 SSCBN 中作为权重，然后从编码的文本向量中学习仿射参数，再进行批量归一化。

而提供其提供的消融实验结果：

ID	Components		IS ↑	FID ↓
	SSACN	DAMSM		
0	-	-	4.86 ± 0.04	19.24
1	✓	-	4.97 ± 0.09	18.54
2	✓	✓	5.07 ± 0.04	**15.61**
3	✓	✓(fine-tune)	**5.17 ± 0.08**	16.58

图 5 SSA-GAN 论文中提供的消融实验结果

ID3 是指将文本编码器也加入训练进行微调，可以看到虽然 IS 有提高，但是 FID 却有所降低，分析的原因可能是微调文本编码器有助于文本图像融合，提高文本图像的一致性，从而提高 IS 分数，但文本与图像一致的同时导致图像多样性下降，所以 FID 会变差。且实验过程中发现 SSA-GAN 训练较慢，训练时长明显多余 DF-GAN，同等 batchsize 下占用显存也明显比 DF-GAN 大，训练 SSA-GAN 对计算设备要求较高。

七. 成员分工

1.江经

完成 SSA-GAN 的训练，对 IS、FID 指标进行评估，撰写实验报告

2.朱健坤

完成 DF-GAN 的训练，对 IS、FID 指标进行评估，撰写实验报告

参考文献

[1]H. Lee, U. Ullah, J. -S. Lee, B. Jeong and H. -C. Choi, "A Brief Survey of text driven image generation and maniulation," 2021 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia), 2021, pp. 1-4, doi: 10.1109/ICCE-Asia53811.2021.9641929.

[2]Hu K, Liao W, Yang M Y, et al. Text to image generation with semantic-spatial aware gan[J]. arXiv preprint arXiv:2104.00567, 2021.

[3]Tao M, Tang H, Wu S, et al. Df-gan: Deep fusion generative adversarial networks for text-to-image synthesis[J]. arXiv preprint arXiv:2008.05865, 2020.

附录

1.DF-GAN、SSA-GAN 最佳模型、tensorboard 训练的 log 在百度网盘：

链接: https://pan.baidu.com/s/1xThAbE1VnuzNytpp-ioM0Q 提取码: f1eo

2.IS、FID 评价引导请见与本报告同目录下的 SSA-GAN/README.md（IS 评价在 SSA-GAN/IS.py；FID 评价在 TTUR/fid.py）