A8 生成模型实验报告

一、任务目标

1、基本实验要求完成：

a. **任务1.请选择一个生成模型（例如VAE或GAN），以CIFAR-10为训练数据，训练一个生成器并从中生成新的图像。要求：代码需有注释，最终提交代码（jupyter notebook文件）和模型，并撰写报告，包含模型原理、实验设置、实验结果分析（例如损失图、生成图像的例子）等。**

2、进阶要求：

a. **在上述基础上，进一步复现条件生成网络（conditional VAE或GAN），并与第一题中方法比较生成图像的差异，最终提交代码、模型，并撰写报告。具体要求同上。**

b. **思考并定义一个图像生成问题，通过调研目前存在的方法和策略，设计相应的解决方法，并分析方法的可行性以及创新性。**

二、实验实现过程及结果：

1、实验平台概述：

本次实验采用 Python 完成。

依赖库有：

依赖库	版本
numpy	1.24.3
torch	1.13.0+cu116
matplotlib	3.6.0

2、实验及实现效果：

① 基本GAN模型实现

1. 模型实现原理

基本生成对抗网络（Generative Adversarial Network，GAN）是一种深度学习模型，由生成器（Generator）和判别器（Discriminator）两部分组成，它们通过对抗训练的方式互相竞争和提升。

GAN 的基本原理如下：

a. 生成器（Generator）接收一个随机噪声向量（通常服从均匀分布或高斯分布）作为输入，并输出一个与训练数据类似的样本。

b. 判别器（Discriminator）接收真实样本（来自训练数据）和生成器生成的样本作为输入，并尝试区分它们。它的目标是对真实样本给出高概率（接近1），对生成样本给出低概率（接近0）。

c. 在训练过程中，生成器和判别器相互博弈和对抗。生成器通过最小化判别器对其生成样本的概率来提高自己的性能，而判别器通过最大化对真实样本和生成样本的区分度来提高自己的性能。

d. 通过交替训练生成器和判别器，GAN 寻求达到一个动态平衡点，即生成器生成的样本越来越逼真，判别器无法区分真实样本和生成样本。

e. 在训练结束后，生成器可以被独立使用，接收随机噪声作为输入并生成逼真的样本。

生成对抗网络（GAN）的基本公式如下所示：

生成器（Generator）的目标是最小化生成样本与真实样本之间的差异：

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

其中：

- $\mathbf{x}$ 是真实样本，
- $\mathbf{z}$ 是从先验噪声分布 $p_{\mathbf{z}}(\mathbf{z})$ 中采样得到的随机噪声，
- $G(\mathbf{z})$ 是生成器的输出，表示生成的样本，
- $D(\mathbf{x})$ 和 $D(G(\mathbf{z}))$ 是判别器的输出，表示样本为真实样本的概率。

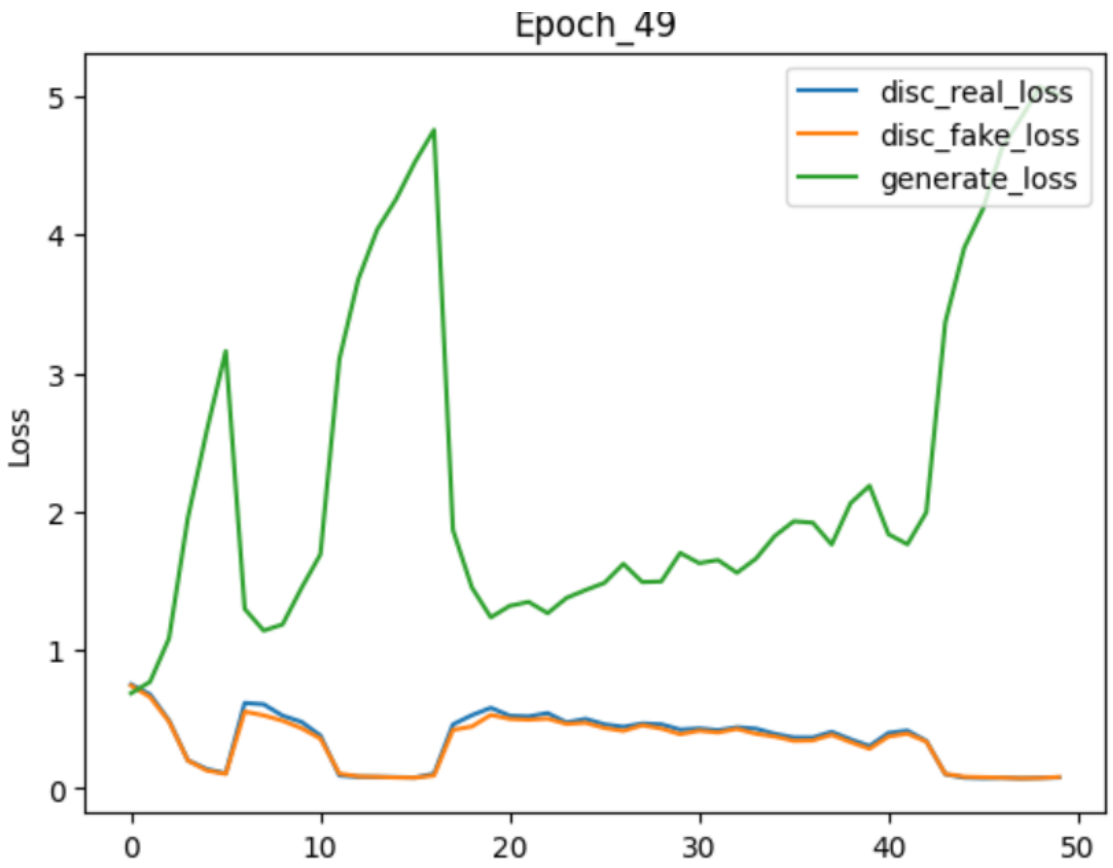
判别器（Discriminator）的目标是最大化对真实样本和生成样本的区分度：

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

在训练过程中，通过交替优化生成器和判别器的目标函数，GAN 寻求一个平衡点，使得生成器可以生成逼真的样本，而判别器无法区分真实样本和生成样本。

注意：上述公式中的期望操作 $\mathbb{E}$ 表示对样本进行求和或积分，具体取决于样本是否是离散或连续的。在实际中，为了优化这些目标函数，通常使用随机梯度下降（SGD）或其变种算法来进行优化。

2. 实验设置
    a. 学习率设置的优化技巧
        i. 经调研，可采用差异化的学习率来进行控制判别器与生成器的训练进度，来均衡两方能力
        ii. 判别器学习率设置为2e-5,生成器学习率设置为5e-5(防止判别器训练过快，使生成器难以训练)
        iii. 使用AdamW优化器
    b. 批次大小设置为64
    c. 训练轮数设置为50
        i. 为平衡判别器相对生成器任务过于简单，我们让生成器每轮训练2次，判别器训练只训练1次
    d. 数据集设置：进行均值为0.5，方差为0.5的归一化操作
    e. 模型设置：
        i. 判别器：采用了Resnet18+全连接层+Sigmoid激活函数
        ii. 生成器：采用了多个反卷积层结合Batchnorm2d与ReLU激活函数的生成块，最后又使用了平均池化层+全连接层+Tanh激活函数
3. 实验结果分析
    a. 损失图：



disc_real_loss为判别器对真实图像的判别损失，disc_fake_loss为判别器对虚假生成图像的判别损失，generate_loss为生成器欺骗判别器损失。

可明显的观察到在总共50轮的训练过程中，生成器和判别器处于明确的对抗状态，即生成器与判别器损失的趋势是几乎相反的，这符合基本原理。

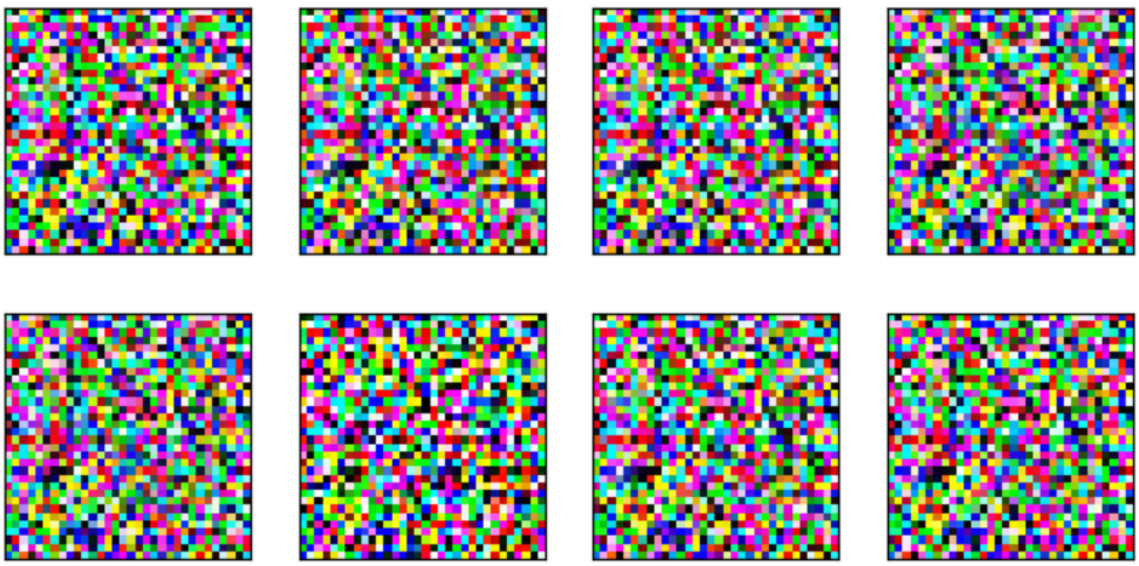
值得一提的是，在训练初期曾有短暂的处于纳什均衡状态（epoch=0），双方损失都为0.69，这一值恰好为ln(2)=ln(-0.5)=0.69，即判别概率都处于在0.5左右，达到平衡。

之后判别器能力一直强于判别器，判别器损失有很强的振荡现象，且判别器越来越强，生成器损失也不断提高。

但生成器损失也不能完全说明生成器能力强弱。

2. 中间生成结果：

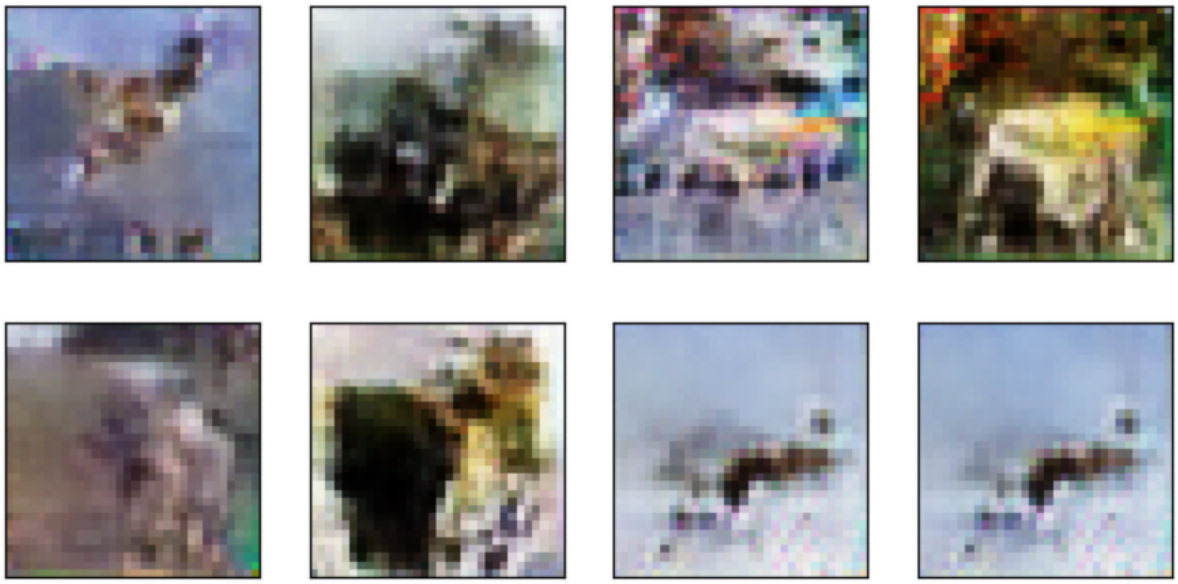
Epoch 0：



Epoch 15：

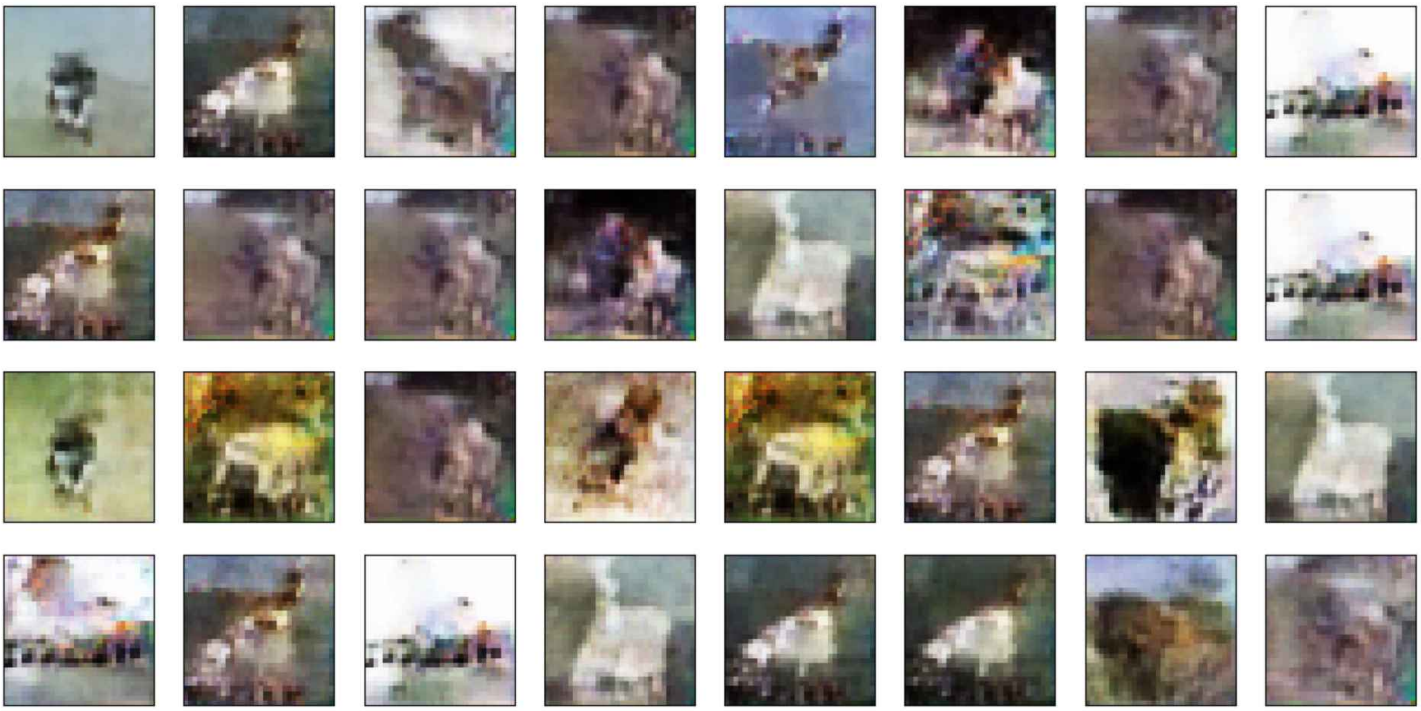
Epoch 35：

Epoch 50：



模型一开始生成较无意义的噪声图像，后逐渐开始生成贴近CIFAR10数据集的内容，在这一过程中，发现了GAN模型的一经典训练难题——模式崩溃，即GAN模型开始偏向于只生成一类图像以欺骗判别器，使生成多样性减少。如Epoch50时右下角两幅图像几乎一致。

1. 较优结果展示：



我们保存了当前认为的最优模型Epoch 50来生成32张测试结果，在一些图片中能明显观察到些动物特征，船体特征，但模式崩溃问题依旧很明显。

② 条件GAN模型实现

1. 模型实现原理

   条件生成对抗网络（Conditional Generative Adversarial Network，Conditional GAN）是生成对抗网络（GAN）的一个变体，它在生成器（Generator）和判别器（Discriminator）之间引入了条件信息。

   条件GAN 的基本原理如下：

   a. 生成器接收两个输入：一个是随机噪声向量（通常服从均匀分布或高斯分布），另一个是条件向量（包含额外的类别、属性或上下文等信息）。生成器的目标是生成与给定条件相匹配的逼真样本。

   b. 判别器接收真实样本（来自训练数据）和生成器生成的样本，同时还接收与生成样本相对应的条件向量。判别器的目标是区分真实样本和生成样本，并判断它们是否与给定条件匹配。

   c. 在训练过程中，生成器和判别器相互博弈和对抗。生成器通过最小化判别器对其生成样本的概率来提高自己的性能，而判别器通过最大化对真实样本和生成样本的区分度来提高自己的性能。

   d. 通过交替训练生成器和判别器，条件GAN 寻求达到一个动态平衡点，即生成器生成的样本在给定条件下越来越逼真，判别器无法准确区分真实样本和生成样本。

   e. 在训练结束后，生成器可以接收随机噪声和特定条件，生成与条件匹配的逼真样本。

条件GAN 的优点在于可以控制生成样本的特定属性或类别，并生成符合特定条件的样本。它在许多任务中具有广泛应用，例如图像到图像的转换、图像修复、图像生成等。

条件GAN 的目标函数可以表示为以下形式：

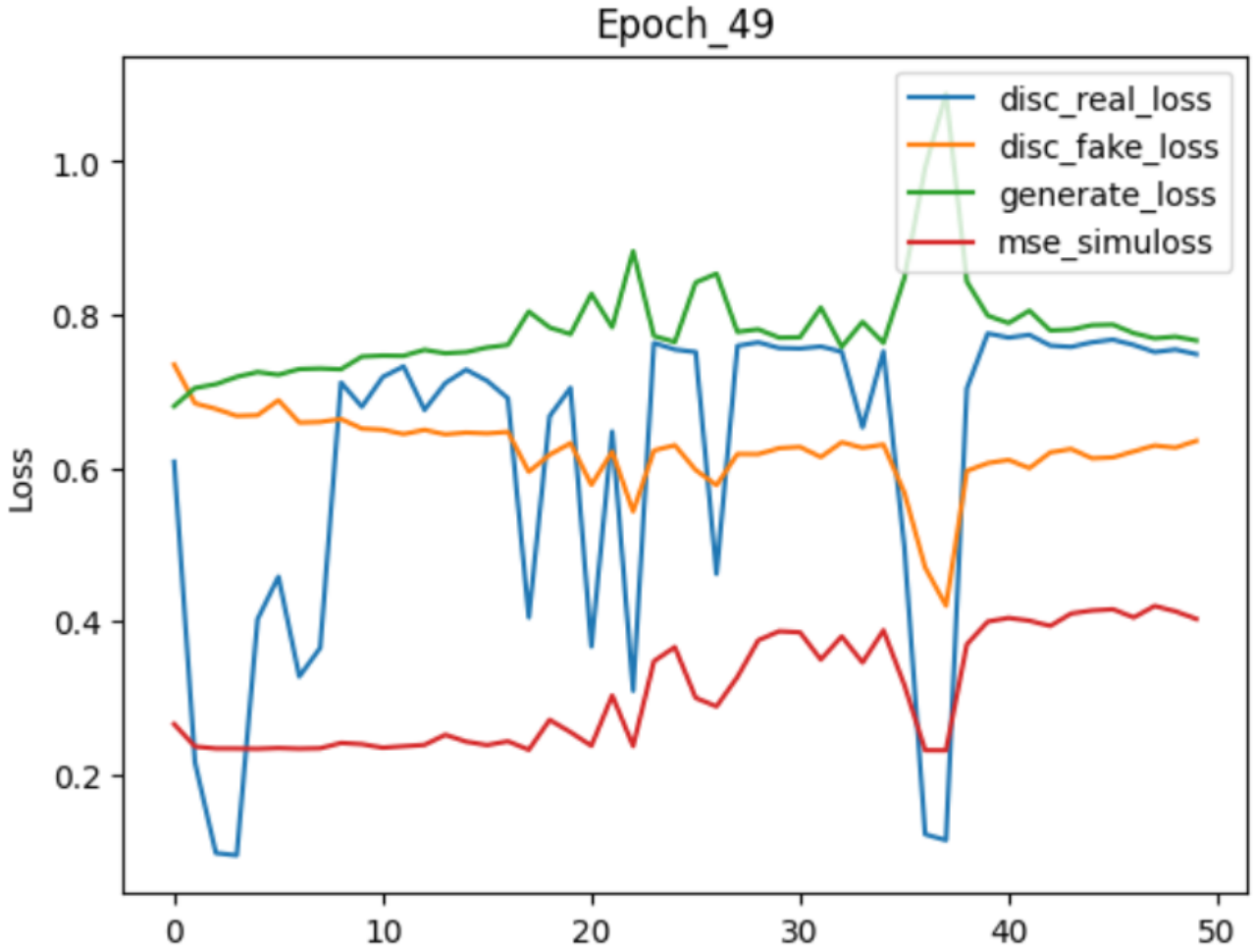
$$\min_G \max_D V(D,G) = \mathbb{E}_{\mathbf{x}\sim p_{\text{data}}(\mathbf{x})}[\log(D(\mathbf{x}|\mathbf{y}))] + \mathbb{E}_{\mathbf{z}\sim p_{\mathbf{z}}(\mathbf{z})}[\log(1-D(G(\mathbf{z}|\mathbf{y})))],$$

其中：

- $\mathbf{x}$ 是真实样本，
- $\mathbf{y}$ 是条件向量，
- $\mathbf{z}$ 是随机噪声向量，
- $D(\mathbf{x}|\mathbf{y})$ 和 $D(G(\mathbf{z}|\mathbf{y}))$ 是判别器的输出，表示样本为真实样本的概率。

通过引入条件向量，生成器和判别器可以利用额外的信息来生成和判别样本。训练过程中，根据给定的条件向量，生成器生成样本，判别器判断生成样本是否与条件匹配，并提供反馈信号以更新生成器和判别器的参数。

2. 实验设置
   a. 学习率
      i. 判别器学习率设置为5e-5,生成器学习率设置为5e-5
      ii. 使用AdamW优化器
   b. 批次大小设置为64
   c. 训练轮数设置为50
      i. 为平衡判别器相对生成器任务过于简单，我们让生成器每轮训练2次，判别器训练只训练1次
   d. 数据集设置：进行均值为0.5，方差为0.5的归一化操作
   e. 模型设置：
      i. 判别器：采用了Resnet18+全连接层+Sigmoid激活函数
      ii. 生成器：采用了多个反卷积层结合Batchnorm2d与ReLU激活函数的生成块，最后又使用了平均池化层+全连接层+Tanh激活函数
      iii. 与上一个任务不同的是，我们针对CIFAR10的分类标签条件，设计了一个判别器和生成器各自拥有的全连接的Embedding层对条件与噪声进行升维处理，之后将其拼接输入判别器或生成器。
3. 实验结果分析
   a. 损失图：



disc_real_loss为判别器对真实图像的判别损失，disc_fake_loss为判别器对虚假生成图像的判别损失，generate_loss为生成器欺骗判别器损失，新增mse_simuloss用于衡量模型生成的图像和标准图像的相似度。

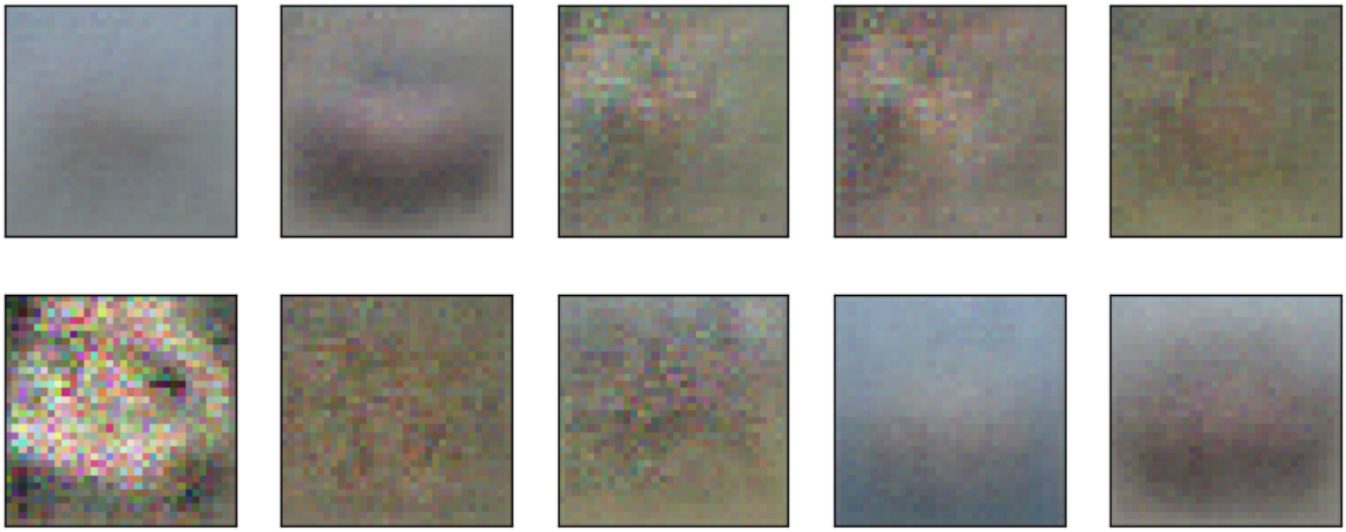
在这一任务中，我进行了trade-off的尝试，在生成器的总损失中添加了相似度损失用于加快模型训练。但在损失曲线中并不能过多体系其相似能力，可能是由于CIFAR-10数据集每一类内容差距较大。

4. 中间生成结果：

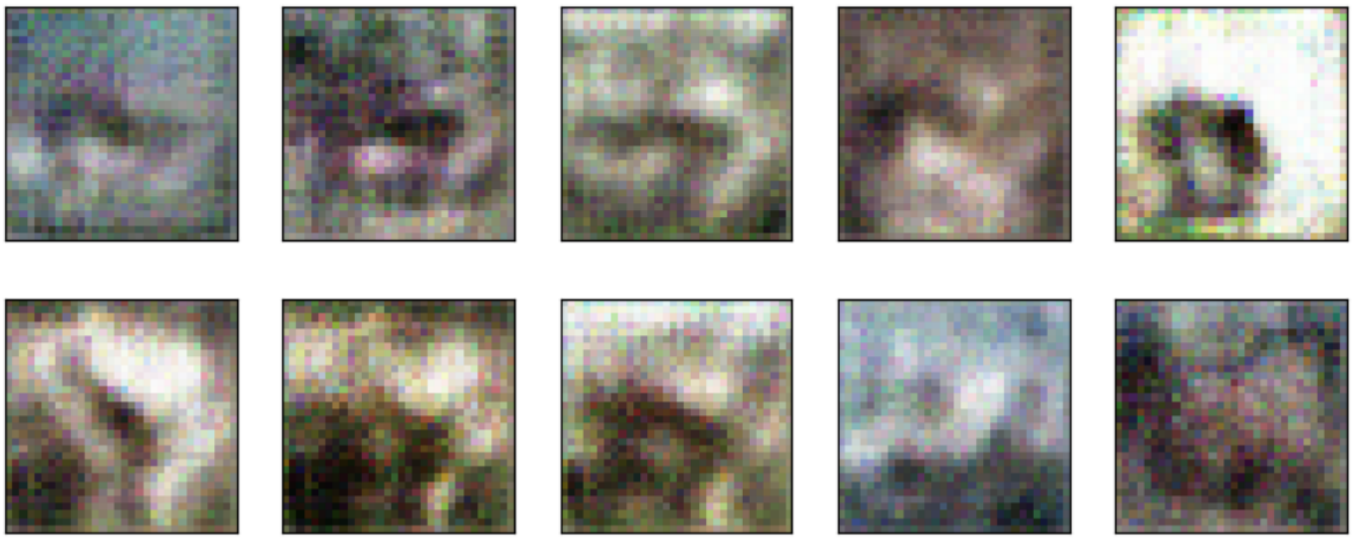
（每处展示都是不同的类别，即每个图象给予了不同的标签生成条件）

Epoch 0：

Epoch 15：



Epoch 35：



Epoch 50：

与之前不同的是，模型的多样化性能因为加入了条件标签作为监督，最终的模型训练结果变得多样化了起来，没有见到大量重复样本。但是可能由于需要进行标签语义理解，模型的训练速度变得慢了很多，且不是很稳定。可能跟模型的结构设计有关。

> 💡 在如上两个任务中，其实进行了多次的重复实验，发现在不固定随机数种子的情况下，GAN模型训练并不稳定可复现，已将多次ipynb中实验结果保存为html文件，将一并呈现在附件中。

③ 新的图像生成问题

基于小规模数据集利用生成模型生成高精度逼真训练数据以扩充训练图像

一、现存方法

1. 数据扩充方法：
    - 传统数据扩充：包括图像翻转、旋转、缩放、平移、裁剪等基本几何变换，以及亮度、对比度、饱和度等颜色变换。这些方法可以通过应用不同的变换参数来生成更多的训练样本，但可能无法捕捉数据集的真实分布或引入新的多样性。
    - 数据合成：通过合成图像元素或背景，将它们组合在一起生成新的图像样本。例如，可以使用图像合成软件或混合模型将不同的对象放置在不同的场景中，生成逼真的合成图像。
2. 生成模型方法：

- Variational Autoencoder (VAE)：VAE是一种生成模型，通过学习输入数据的潜在分布，可以生成新的样本。通过训练VAE模型，可以生成具有多样性的合成图像样本，以扩充小规模数据集，但VAE方法其能力一般较差。

- Generative Adversarial Networks (GANs)：GANs是一种生成模型，由生成器和判别器组成。生成器试图生成逼真的图像样本，而判别器则试图区分真实图像和生成的图像。通过对抗训练的过程，GANs可以生成高质量、多样性的合成图像，但GAN存在难以训练这一问题有待解决。

3. 迁移学习：

- 利用预训练模型：使用在大规模数据集上训练的预训练模型，如ImageNet上的预训练模型，将其迁移到小规模数据集的目标任务中。通过微调或特征提取的方式，可以利用预训练模型的泛化能力来提升小规模数据集上模型的性能。

4. 半监督学习：

- 利用有标签和无标签数据：如果在小规模数据集中有一部分数据是有标签的，可以使用半监督学习方法，结合有标签和无标签数据进行模型训练。半监督学习方法可以利用无标签数据的信息来提升模型的性能。

二、新解决方法：利用Diffusion生成模型进行数据集扩充

1. 问题定义：

在许多计算机视觉任务中，如目标检测和图像分类，模型的泛化能力往往依赖于大规模、多样化的训练数据。然而，对于一些特定领域或任务而言，获得大规模数据集可能是困难的。本次研究的目标是利用现有的小规模数据集，通过生成模型的方法进行数据扩充，进一步训练检测模型以提升其泛化能力。

通过对之前现存方法的调研可知，迁移学习与半监督学习都需要大量的其他数据集进行长时间预训练来提升模型性能，而数据扩充方法存在着无法捕捉真实分布等问题，传统生成模型又存在能力一般或难以训练的问题。

而近年来较为火热的Diffusion模型，凭借其生成质量高，相对于GAN训练又较为稳定，容易的优点迅速在AIGC产业中进行运用。故我们希望使用Diffusion方法来生成逼真的图像样本。

2. 解决方法设计：

基于调研，我们可以设计以下解决方法：

- 数据集准备和预处理：选择现有的小规模数据集，并进行预处理操作，如图像归一化、裁剪和增强等，以提高数据质量和可用性。
- Diffusion生成模型训练：使用Diffusion方法对小规模数据集进行训练，学习生成逼真的图像样本。通过逐步模糊和还原的过程，生成模型可以学习到数据集的分布，并生成具有多样性和高质量的合成图像样本。
- 数据扩充：利用训练好的Diffusion生成模型，生成额外的图像样本，并将其与原始数据集进行合并，形成扩充后的数据集。通过这种方式，可以增加数据集的规模和多样性，为后续的模型训练提供更多的训练样本。
- 检测模型训练：使用扩充后的数据集，训练目标检测模型（或其他视觉任务模型）。通过在扩充数据集上进行训练，模型可以更好地捕捉数据集的真实分布，从而提升其泛化能力。

3. 方法可行性和创新性分析：

- 可行性：利用生成模型进行数据扩充已经在许多任务中取得了成功。如在：“A Survey on GAN-Based Data Augmentation for Hand Pose Estimation Problem”一文中就使用了GAN的方法来扩充手部姿态数据集进行训练，获得了良好的效果。Diffusion作为一种生成模型训练方法，已经在图像生成任务中展现出了良好的效果。现有的小规模数据集和Diffusion方法提供了实施所述方法所需的基础。因此，本方法具有较高的可行性。
- 创新性：本方法的创新之处在于将Diffusion方法应用于小规模数据集的扩充，并结合检测模型训练，以提高其泛化能力。通过使用生成模型生成逼真的图像样本，并将其与原始数据集进行合并，可以增加数据集的多样性和规模。这种结合生成模型和任务模型训练的方法，在小规模数据集上提升模型性能的研究相对较少，因此具有一定的创新性。

4. 实验计划：

- 数据集准备：选择适合的小规模数据集，并进行预处理和增强操作，以提高数据质量和可用性。
- Diffusion生成模型训练：使用选定的小规模数据集，按照Diffusion方法进行生成模型的训练。调整模型的超参数和训练策略，以获得高质量的生成图像样本。
- 数据扩充：利用训练好的Diffusion模型，生成额外的图像样本，并将其与原始数据集进行合并，形成扩充后的数据集。
- 检测模型训练：使用扩充后的数据集，训练目标检测模型。选择适当的检测模型架构，并进行训练和调优，以提升模型在测试集上的性能。
- 实验评估：通过在验证集和测试集上进行评估，比较使用扩充数据集训练的检测模型与仅使用原始数据集训练的模型的性能差异。评估指标可以包括准确率、召回率、精确度等。

5. 预期结果和影响：

- 预期结果：通过使用Diffusion方法进行数据扩充，并结合检测模型训练，预计可以提升模型的泛化能力。扩充后的数据集可以更好地捕捉数据集的真实分布，从而改善模型在新样本上的表现。
- 影响：本方法的成功应用可以为小规模数据集的利用提供新的思路和解决方案。通过数据扩充和生成模型的结合，可以克服数据集规模有限的问题，提高模型的泛化能力，对于相关领域的研究和实践具有积极的影响。

6. 结论：

基于Diffusion的小规模数据集扩充与检测模型训练是一种有潜力的方法，可以通过生成逼真的图像样本扩充数据集，并结合检测模型的训练，提升模型的泛化能力。本报告提出了解决该问题的方法设计和实验计划，预计可以在小规模数据集上取得较好的实验结果，并对相关领域的研究和实践产生积极的影响。