【实验题目】对抗生成网络-文字到图片的合成

Generative Adversarial Text to Image Synthesis

【背景】

GAN作为一个具有“无限”生成能力的模型，它的直接应用就是建模，生成与真实数据分布一致的数据样本，例如可以生成图像、视频等。基于文字描述生成图像在生活中应用场景广泛，比如在影视创作、场景重建中存在很大潜力。本实验进行的内容为利用GAN对抗生成网络实现文字描述到图片的合成。

对于文本生成图像这个任务，我们将这个任务分解成两个子问题，首先，获取重要细节的----文本特征表示。然后，用这个特征合成一张较好的图片。

【数据处理】

使用的是标准数据集Oxford-102 flower，包含 102 种花类的图像数据集（主要是一些英国常见的花类），每个类别包含 40—258 张图像。这些图像在比例、姿势以及光照方面有着丰富的变化。

由于GAN生成对抗网络在训练过程中需要的数据输入为图片对应的文字描述，因此在训练GAN之前要先对Oxford-102 flower数据集进行处理。

第一个子问题直接使用的数据集是Oxford-102 flower，使用RNN来做图片生成描述，可用来对文字进行判别和表示，由于它是根据图片的内容和他前一个词生成下一个词，是遵循链式规则的；

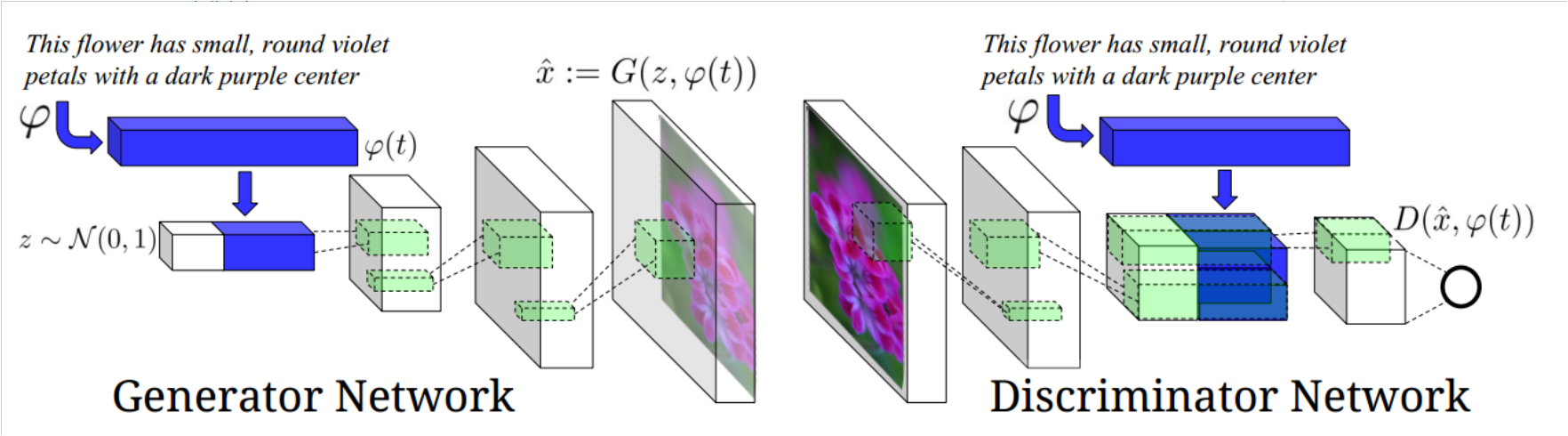
第二个子问题是使用GAN完成图片生成，模型输入为图片的文字描述。故模型使用的数据集为第一个问题的实验结果，即图片及其文字描述的数据集。

【算法原理】

介绍GAN生成对抗网络的算法原理：

1、网络架构

基于文本的DCGAN结构：



生成对抗网络基于博弈模型，其中生成器网络（generator network）必须与其对手判别器（discriminator）竞争。训练过程可以总体表示为：生成器试图最大化鉴别器将其输入错认为正确的概率，而鉴别器引导生成器生成更逼真的图像。

模型输入为文本描述向量（上图中深蓝色的长条φ），在生成网络和判别网络中都有输入。在生成器网络中，首先我们将之前得到的文本向量，用网络全连接层对文本向量压缩，最后得到128维向量，然后和一个从正态分布中得到的随机向量进行拼接，输入生成网络中进行图像生成。

在判别网络中，首先对输入做几个stride=2的卷积，当feature map的大小变为2x2时，则又一次对文本编码结果做一个全连接层，将全连接层的结果拼接到这个大小为2x2的feature map上。然后对拼接结果做一个1x1的卷积和2x2的卷积。

DCGAN对卷积神经网络的结构做了一些改变，以提高样本的质量和收敛的速度，这些改变有：取消所有pooling层。G生成器网络中使用转置卷积进行上采样，判别器网络中用加入stride的卷积代替pooling；在判别器和生成器中均使用批归一化batch normalization等。

2、loss函数

GANs在训练时需要同时运行两个优化算法，我们需要为discriminator和generator分别定义一个优化器，一个用来来最小化discriminator的损失，另一个用来最小化generator的损失。即loss = d\_loss + g\_loss。Loss都采用交叉熵计算得来。

①判别器Discriminator network：

对于辨别器discriminator，其损失等于真实图片和生成图片的损失之和，即 d\_loss = d\_loss\_real + d\_loss\_fake 。在 tensorflow 中可使用以下函数：

tf.nn.sigmoid\_cross\_entropy\_with\_logits(logits=logits, labels=labels)

在计算真实数据产生的损失d\_loss\_real时，我们希望辨别器discriminator输出1；而在计算生成器生成的 ‘假’ 数据所产生的损失d\_loss\_fake时，我们希望discriminator输出0。因此对于真实数据，我们使用label=1计算代价函数来训练判别器，对于生成数据产生的d\_loss\_fake使用label=0来计算代价函数。

②生成器 Generator network

生成器经训练尝试输出能使辨别器分配接近概率1的样本。

生成器的代价函数为

g\_loss=tf.reduce\_mean(tf.nn.sigmoid\_cross\_entropy\_with\_logits(logits=d\_logits\_fake, labels=tf.ones\_like(d\_logits\_fake)))

随着以上训练的进行，判别器会逐步增强自身的判别能力，而生成器生成越来越逼真的输出，以欺骗判别器。

3、与传统训练方法相比，GAN的优势和不足

①优势

GANs是一种以半监督方式训练分类器的方法，在你没有很多带标签的训练集的时候，可以不做任何修改的直接使用网络训练；

GAN不需要马尔科夫链，模型优化仅用到反向传播，优势在于采样概率高，生成数据更接近真实数据分布；

GAN能够解决数据不足的问题，例如多视角多任务的生成等问题；

②劣势

在训练过程中，训练GAN需要达到纳什均衡，比较难以训练，存在训练不同步及梯度消失的问题，如何定义训练策略及损失函数成为关键。其中，GAN训练不同步的原因在于更新生成器G的N次，更次判别器D的1次。此外，梯度消失的原因在于当真实样本和生成样本之间具有极小重叠甚至没有重叠时，其目标函数的散度是一个常数，导致优化目标不连续。

【实验细节】

//修改过程

//实验结果

【实验总结】

//实验结果分析

//心得体会