**第三次作业课题报告**

——自注意力改进的DCGAN网络

**第九组**

组员： 陈高欣 1611640212

组员： 刘 钰 1611640428

组员： 李怡琳 16116404061

组员： 詹璧瑞 1611640206

**摘要**

DCGAN（Deep Convolutional Generative Adversarial Networks，DCGAN）作为GAN生成对抗网络的变种之一，在许多图像生成的任务上都取得了不错的结果。本文简述了前人提出的自注意力机制，在深度卷积生成对抗网络的基础上添加了自注意力机制，并使用手写数字MNIST数据集进行实验。我们对比了改进前后生成图片的质量，发现当训练数据较少的时候，添加了自注意力机制的结果比原结果有所提升。最后对进一步改进GAN网络模型的方法进行探讨。我们讨论了应用在Transformer网络里的自注意力机制——多头注意力机制（Muti-Head Attention）,它是本文所实现的注意力机制的一种强有力的改进。

1. **概述：**

1.生成式对抗网络GAN( Generative Adversarial Networks)：

生成式对抗网络（GAN）中最主要的两个部分是判别模型D和生成模型 G生成模型G的作用是不断学习训练集中真实数据的概率分布，目标是将输入的随机噪声转化为可以以假乱真的图片（生成的图片与训练集中的图片越相似越好）。判别模型D的作用是判断输入的图片是否是真实的图片，目标是将生成模型G产生的“假”图片与训练集中的“真”图片分辨开。

GAN网络的优化目标函数定义如下：

优化判别器D： max V(D,G)=EX-Pdata(x)[log(D(x))] (1)

D

优化生成器： min V(D,G)=Ez-p(z)[log(1-D(G(z)))] (2)

G

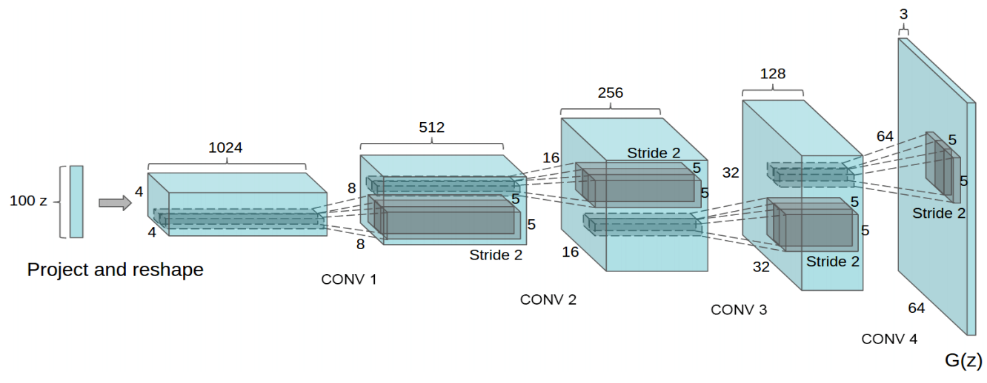
min max V(D,G)= EX-Pdata(x)[log(D(x))]+ Ez-p(z)[log(1-D(G(z)))] (3)

G D

2. 深度卷积对抗网络DCGAN

（Deep Convolutional Generative Adversarial Networks）：

深度卷积对抗网络（DCGAN）在GAN的基础上增加了深度卷积网络结构，其中的D、G的含义以及损失都和原始GAN中完全一致,但是它在D和G中采用了较为特殊的结构，以便对图片进行有效建模。



如图所示，生成器G的输入是一个100维的噪声向量。网络的第一层实际是一个全连接层 ，将 100 维的向量变成一个 4\*4\*1024 维的向量，从第二层开始，使用转置卷积做上采样，逐渐减少通道数，最后得到的输出64\*64\*3，即输出一个三通道的宽和高都为 64 的图像。判别器D的结构是一个卷积神经网络，输入的图片经过若干层卷积后得到一个特征，将特征输入Logistic函数得到为真的概率。

总的来说，DCGAN是GAN的一个变体，可以理解为GAN的反面，它的工作原理大致如下：输入一张图，经过卷积层提取它的一些特征信息，以大于一的步幅做卷积来代替池化，通过卷积层池化层，将信息逐层深入的去解析。由复杂的图像最终得到特征。

1. **自注意力机制：**

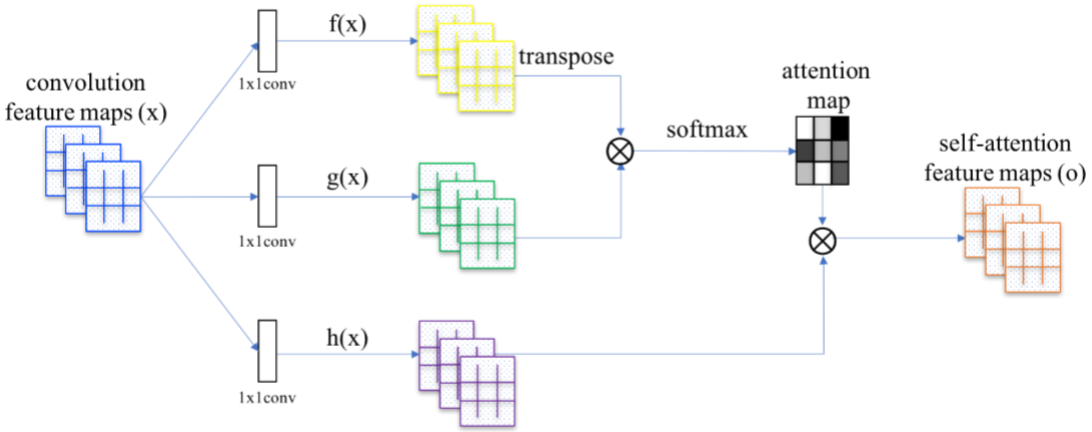
如图所示，f(x)，g(x)和h(x)都是对图片做线性变化；将f(x)的输出转置，并和g(x)的输出相乘，再经过softmax归一化得到一个attention map； 将得到的attention map和h(x)逐像素点相乘，得到自适应注意力的特征图。

(5)

**(6)**

**(7)**

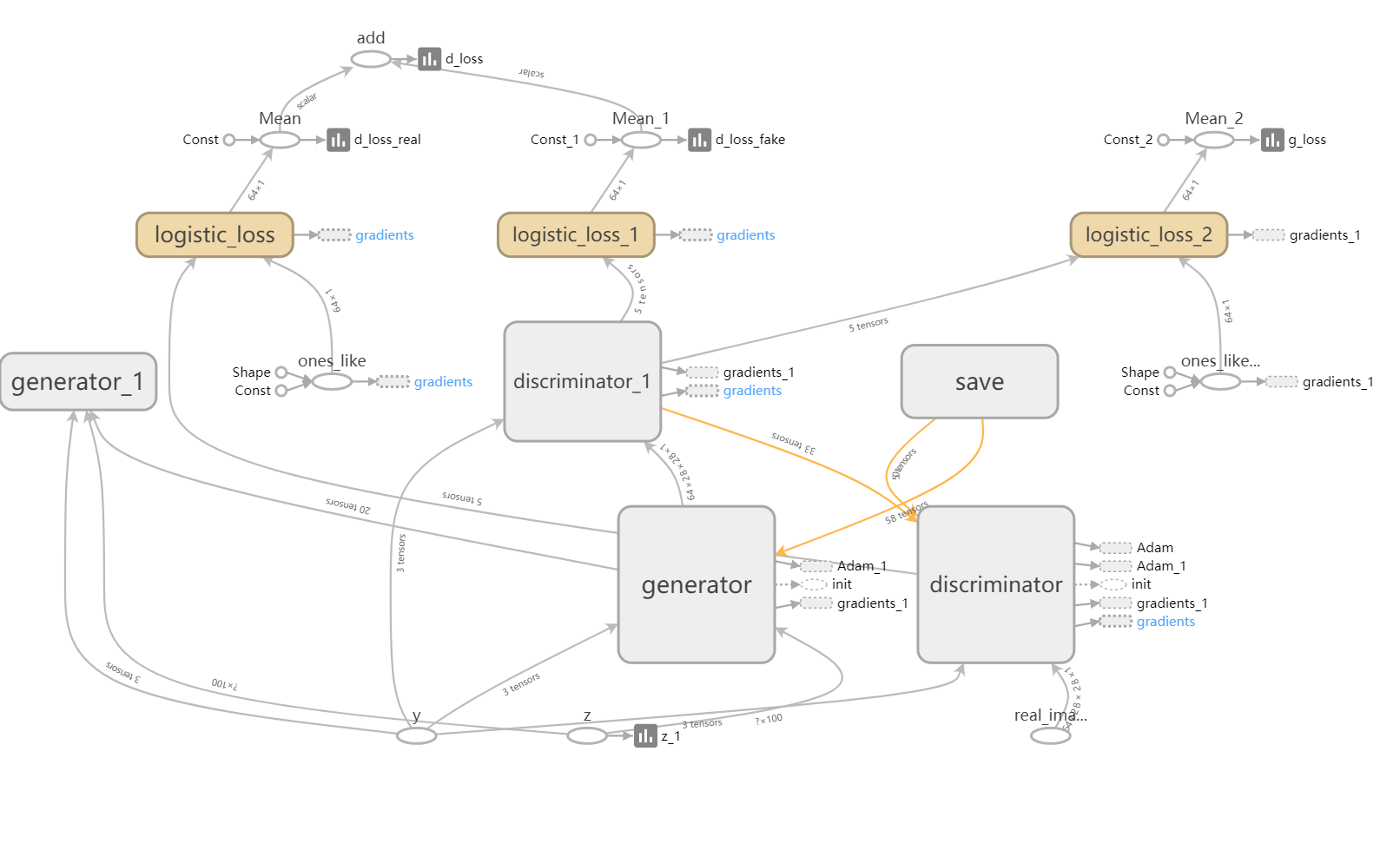
自注意力的网络结构图如下：



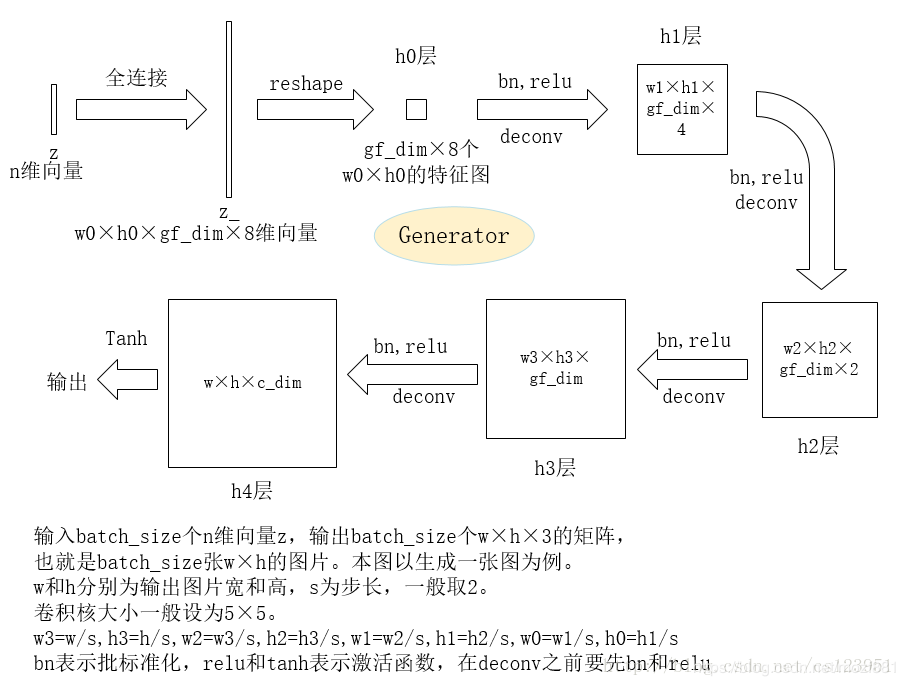
1. **搭建网络结构与实验：**

本次实验中，我们在DCGAN的基础上往生成器和判别器里都添加了自我注意力机制。

* 整体的网络结构

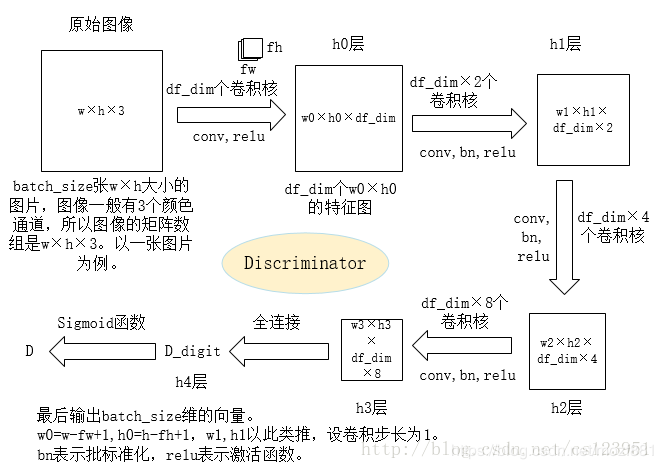


* 生成器的网络结构



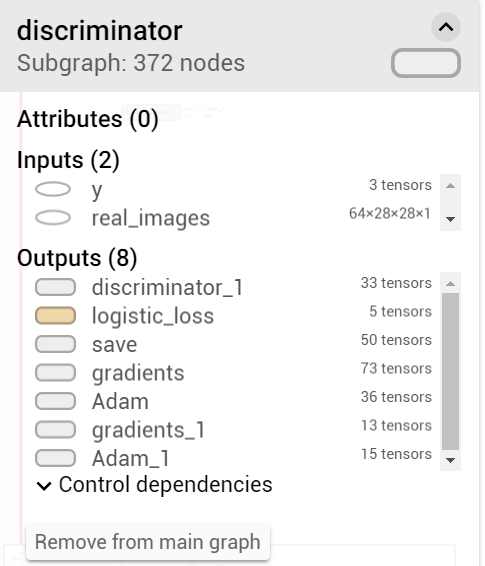
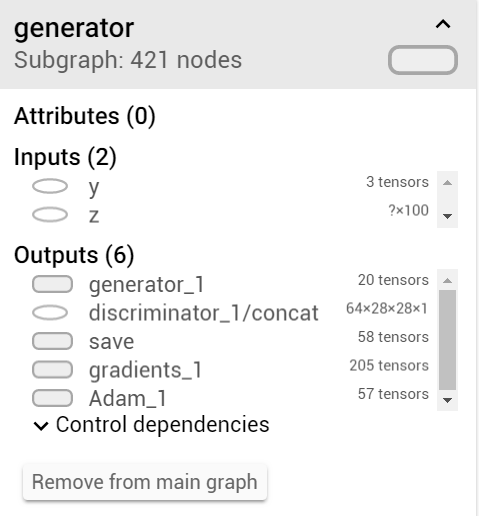
自我注意力机制

* 判别器的网络结构

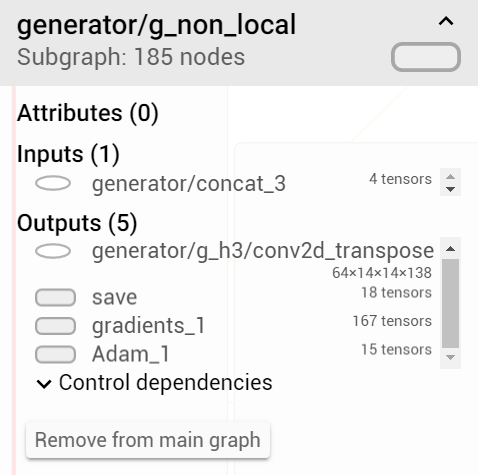
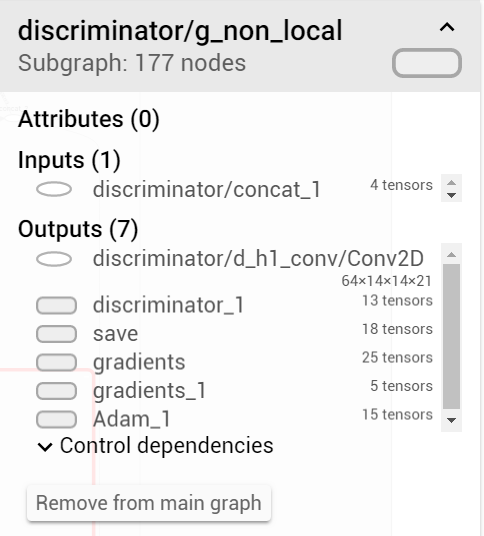


自我注意力机制

* 生成器与判别器的参数



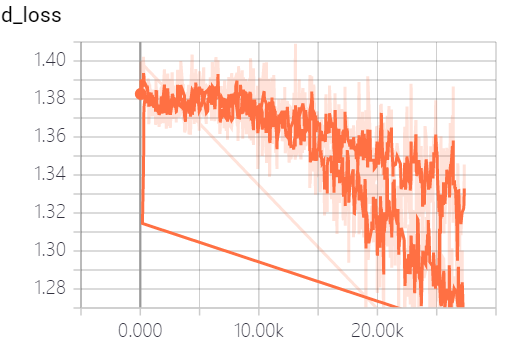
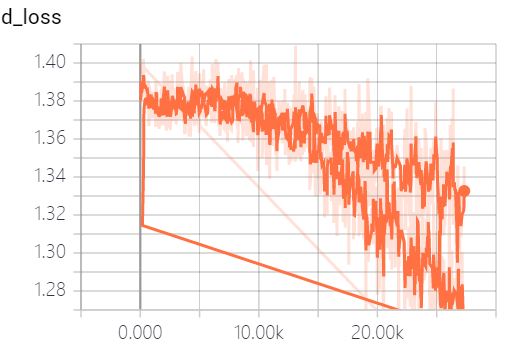
* 生成器与判别器中自注意力模型的参数

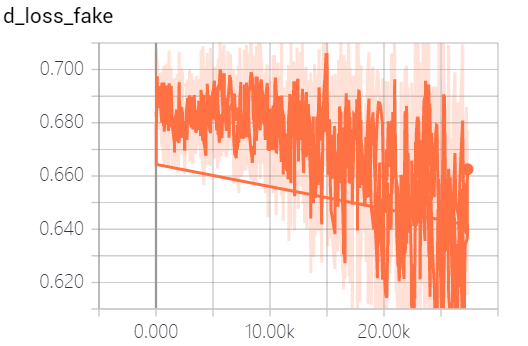
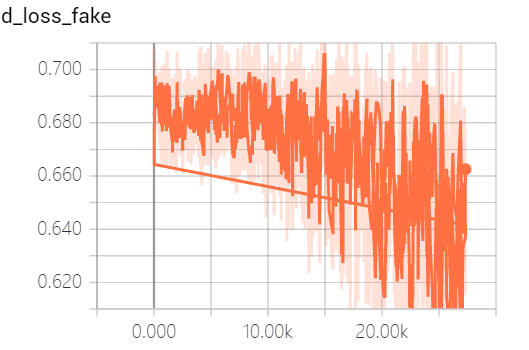
 

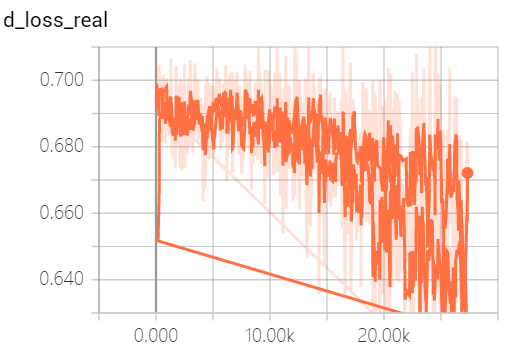
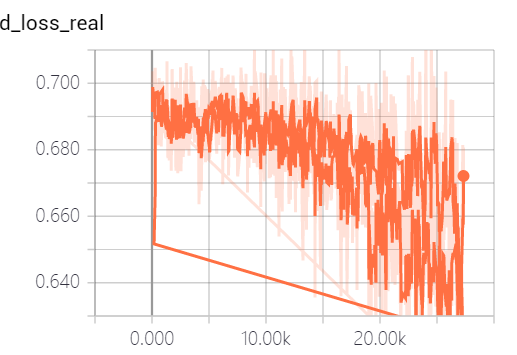
1. **实验结果即分析：**

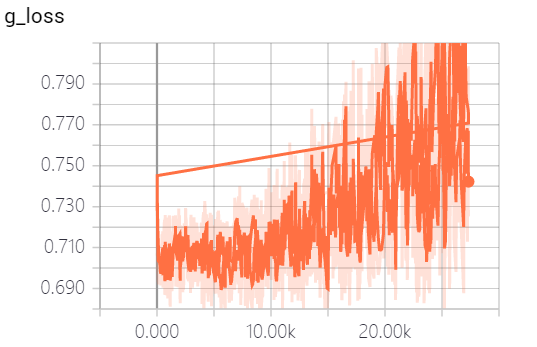
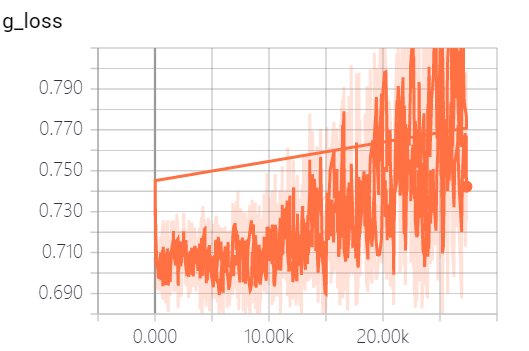
**5.1 损失函数对比**

**SADCGAN DCGAN**





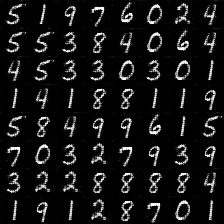
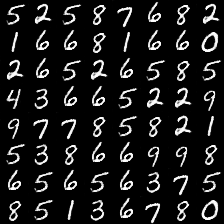




**5.2实验结果对比**

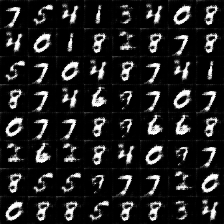
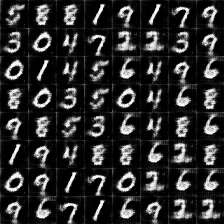
**(1)大数据集**

**DCGAN SADCGAN**



**（2）小数据集**

**DCGAN SADCGAN**

**5.3结果分析**

在全部数据集上训练后，通过对比可以发现两者都能很好地生成清晰且容易辨别的图片，而且损失函数也没有明显的变化，因此自注意力机制并没有对该数据集产生更好的效果。

但我们尝试着在小数据集上进行训练，得到了新的结果，自注意力机制在该数据集上的表现更优秀，SADCGAN已经全部都能看出来数字信息，而DCGAN却仍然有些数字无法分辨。

通过我们的讨论，觉得导致这种结果的原因是MNIST数据集的全局性比较低，而且该数据集足够大，导致DCGAN也能得出很好的结果，所以没有发挥出自注意力的作用，我们未来的工作将是在其他更注重全局逻辑的数据集上进行尝试。

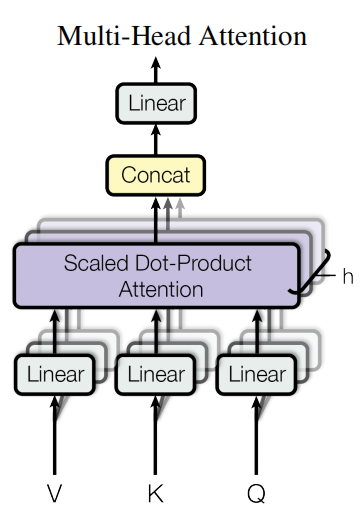
综上，便是SAGAN提出来的用自注意力机制去解决全局信息获取的问题，它既在每一层都考虑了全局信息，也没有引入过多的参数量，在提高感受和减小参数量之间找到了一个很好的平衡。因此，如果我们的生成任务是全局相关性比较高的图片，就可以考虑使用SAGAN。

1. **改进模型的设想：**

由卷积网络我们得知，每个卷积层都有若干个卷积核，每个卷积核负责一种特征的提取，最后将所有的特征组合起来。那么，能否像卷积核一样，引入多个上述的注意力机制，把每个注意力都组合到一起，更大地提升网络的性能呢？事实上，已经有这样的网络出现，它就是Transformer，可译为变形金刚。

Transformer是完全基于注意力，不需要卷积和循环等操作的网络架构。在这里我们主要讨论其对注意力进行了改进的部分，以做出日后完善GAN网络的设想。

网络中有一个结构叫Muti-head Attetion，这里译为多头注意力，它的作用是将多个注意力机制得到的结果连接在一起，最后经过一个线性变换得到最终的输出，其结构如下图：



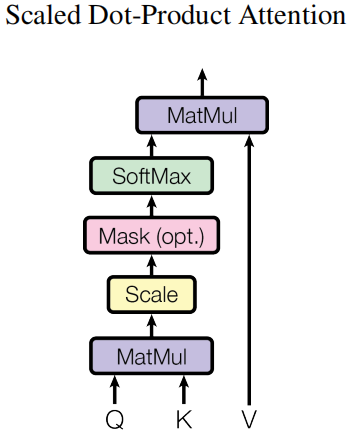
在自然语言处理（NLP）中，传进去的是三个不同的向量，而在图像处理中，传进去的可以是三张相同的图像，就像上文展示的自注意力结构图一样。此外，多头注意力中连接的每个注意力机制与本次作业使用到的注意力机制有轻微的不同，叫比例点积注意力机制（Scaled Dot-Product Attention），它的不同之处是，在对矩阵相乘后的结果进行Softmax运算前，对每个分量乘上一个比例因子，其中d为输入向量的维度，计算公式以及结构图如下：

Attention(x)=softmax（）h(xi)

其中f,g,h是对向量x做的线性变换，之所以加上比例因子，是为了防止当维度过大的时候，Softmax的梯度过小。多头注意力的计算方法：

MutiHead(x)=Concat(head1,…headh)WO

其中headi由Attention（xi）计算得来。



因此，如果要改进模型的话，我们就将同一层卷积得到的所有卷积核计算出的特征图进行如上的多头自注意力机制，每一个注意力传进的都是3张相同完整的图片，最后得到的结果维度肯定很大，但我们通过线性变换可以将它重新变回一张图片的大小，也就实现了所有比例点积注意力的整合。

**参考资料**

[1]《21个项目玩转深度学习——基于TensorFlow的实践详解》.何之源.中国工信出版集团、电子工业出版社联合出版

[2]Han Zhang and Ian Goodfellow and Dimitris Metaxas and Augustus Odena.Self-Attention Generative Adversarial Networks

[3]Niki Parmar and Ashish Vaswani and Jakob Uszkoreit.Image Transformer

**小组分工：**

四位成员分别阅读苑老师给定的四篇论文，讨论与分析网络结构

其中，陈高欣、詹璧瑞负责搭配环境，编写代码

刘钰、李怡琳负责调试、优化代码

最终的报告由刘钰、李怡琳编写

由陈高欣整理所有的运行结果和对代码做最后的完善

由詹璧瑞在课堂展示作业成果