

Self-Attention Generative Adversarial Networks

Han Zhang^{1,2} Ian Goodfellow³ Dimitris Metaxas¹ Augustus Odena²

Abstract

在本文中，我们提出了自注意生成对抗网络(SAGAN)，它允许对图像生成任务进行注意驱动的远程依赖建模。传统的卷积gan仅作为低分辨率特征图中空间局部点的函数生成高分辨率细节。在SAGAN中，可以使用来自所有特征位置的线索生成细节。此外，鉴别器还可以检查图像远处部分的高细节特征是否彼此一致。此外，最近的研究表明，发电机调节会影响GAN的性能。利用这一见解，我们将谱归一化应用于GAN生成器，并发现这改善了训练动态。提出的SAGAN比之前的工作¹表现得更好，在具有挑战性的ImageNet数据集上，将发布的最佳盗梦分数从36.8提高到52.52，并将frjs:1₆盗梦距离从27.62减少到18.65。注意层的可视化显示，生成器利用了与物体形状对应的邻域，而不是固定形状的局部区域。

1. 简介

图像合成是计算机视觉中的一个重要问题。随着生成对抗网络(GANs)的出现，在这个方向上已经取得了显著的进展(Goodfellow et al., 2014)，尽管仍然存在许多悬而未决的问题(Odena, 2019)。基于深度卷积网络(Radford et al., 2016; Karras et al., 2018; Zhang et al.)的gan尤其成功。然而，通过仔细检查从这些模型生成的样本，我们可以观察到，当在多类数据集(e.g., ImageNet (Russakovsky et al., 2015))上训练时，卷积gan (Odena et al., 2017; Miyato et al., 2018; Miyato & Koyama, 2018)在建模某些图像类时比其他图像类要困难得多。例如，虽然最先进的ImageNet GAN模型(Miyato & Koyama, 2018)擅长合成具有很少结构约束的图像类(e.g., 海洋，天空和景观类，它们更多地通过纹理而不是几何来区分)，但它无法捕获在某些类中一致出现的几何或结构模式(例如，狗通常用逼

¹Department of Computer Science, Rutgers University ²Google Research, Brain Team ³Work done while at Google Research. Correspondence to: Han Zhang <zhanghan@google.com>.

Proceedings of the 36th International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019. Copyright 2019 by the author(s).

¹Brock et al. (2018)，在很大程度上建立在这项工作的基础上，此后大大改善了这些结果。

真的皮毛纹理绘制，但没有明确定义的分开的脚)。一个可能的解释是，以前的模型严重依赖于卷积来模拟不同图像区域之间的依赖关系。由于卷积算子具有局部接受域，因此只能在经过几个卷积层后处理长范围依赖关系。由于各种原因，这可能会阻止学习长期依赖关系：小模型可能无法表示它们，优化算法可能难以发现仔细协调多层次以捕获这些依赖关系的参数值，并且这些参数化在统计上可能是脆弱的，并且在应用于以前未见过的输入时容易失败。增加卷积核的大小可以增加网络的表示能力，但这样做也会损失使用局部卷积结构所获得的计算和统计效率。另一方面，自关注(Cheng et al., 2016; Parikh et al., 2016; Vaswani et al., 2017)在建模长期依赖关系的能力与计算和统计效率之间表现出更好的平衡。自关注模块将某个位置的响应计算为所有位置特征的加权和，其中权重(或注意力向量)的计算成本很小。

在这项工作中，我们提出了自注意生成对抗网络(SAGANs)，它将自注意机制引入卷积gan中。自关注模块是卷积的补充，有助于对图像区域之间的长距离、多层次依赖关系进行建模。在自我关注的武装下，生成器可以绘制图像，其中每个位置的精细细节与图像远处部分的精细细节精心协调。此外，鉴别器还可以更准确地对全局图像结构施加复杂的几何约束。

除了自我关注之外，我们还结合了最近有关网络调节与GAN性能的见解。(Odena et al., 2018)的工作表明，条件良好的发电机往往表现更好。我们建议使用谱归一化技术对GAN发生器进行良好的调节，该技术以前仅应用于鉴别器(Miyato et al., 2018)。

我们在ImageNet数据集上进行了大量的实验，以验证所提出的自注意机制和稳定技术的有效性。SAGAN在图像合成方面显著优于先前的工作，将发布的最佳盗梦分数从36.8提高到52.52，并将frjs:1₆切盗梦距离从27.62减少到18.65。注意层的可视化显示，生成器利用了与物体形状对应的邻域，而不是固定形状的局部区域。我们的代码可在<https://github.com/brain-research/self-attention-gan>上获得。

2. 相关工作

生成对抗网络。gan在各种图像生成任务中取得了巨大的成功，包括图像到图像的翻译(Isola et al., 2017;

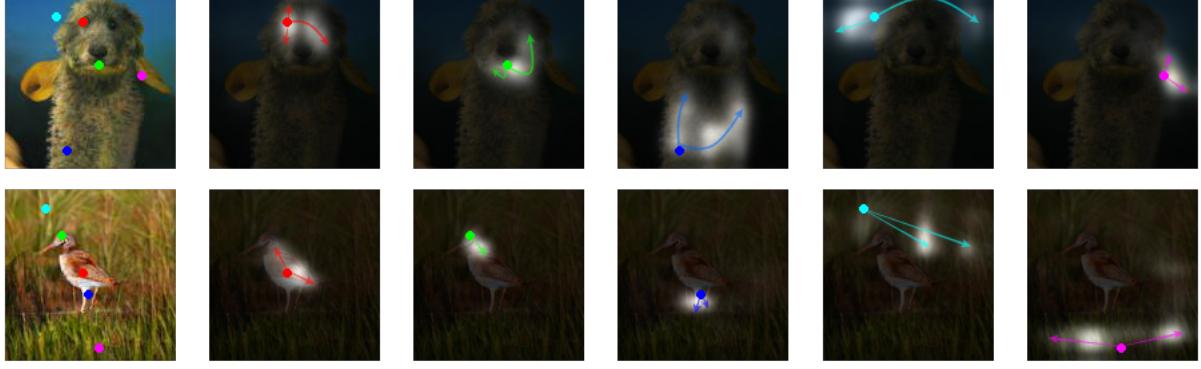


Figure 1. 提出的SAGAN通过利用图像远处部分的互补特征而不是固定形状的局部区域来生成一致的对象/场景来生成图像。在每一行中，第一张图像显示了五个具有代表性的查询位置，并用颜色编码的点表示。其他五张图片是这些查询位置的注意力图，用相应的颜色编码箭头总结了最受关注的区域。

Zhu et al., 2017; Taigman et al., 2017; Liu & Tuzel, 2016; Xue et al., 2018; Park et al., 2019)、图像超分辨率(Ledig et al., 2017; Sønderby et al., 2017)和文本到图像的合成(Reed et al., 2016b;a; Zhang et al., 2017; Hong et al., 2018)。尽管取得了这样的成功，但已知gan的训练是不稳定的，并且对超参数的选择很敏感。一些作品试图通过设计新的网络架构(Radford et al., 2016; Zhang et al., 2017; Karras et al., 2018; 2019)、修改学习目标和动态(Arjovsky et al., 2017; Salimans et al., 2018; Metz et al., 2017; Che et al., 2017; Zhao et al., 2017; Jolicoeur-Martineau, 2019)、添加正则化方法(Gulrajani et al., 2017; Miyato et al., 2018)和引入启发式技巧(Salimans et al., 2016; Odena et al., 2017; Azadi et al., 2018)来稳定GAN训练动态和提高样本多样性。最近，Miyato et al.(Miyato et al., 2018)提出限制鉴别器中权矩阵的谱范数，以约束鉴别器函数的Lipschitz常数。结合基于投影的鉴别器(Miyato & Koyama, 2018)，光谱归一化模型极大地改善了ImageNet上类条件图像的生成。

注意力模型。最近，注意力机制已经成为必须捕获全局依赖关系的模型的一个组成部分(Bahdanau et al., 2014; Xu et al., 2015; Yang et al., 2016; Gregor et al., 2015; Chen et al., 2018)。特别是，自我注意(Cheng et al., 2016; Parikh et al., 2016)，也称为内注意，通过关注同一序列中的所有位置来计算序列中某个位置的响应。Vaswani et al.(Vaswani et al., 2017)证明了机器翻译模型可以通过单独使用自注意模型获得最先进的结果。Parmar et al.(Parmar et al., 2018)提出了一个Image Transformer模型，将自关注添加到图像生成的自回归模型中。Wang et al.(Wang et al., 2018)将自注意形式化为非局部操作，以模拟视频序列中的时空依赖关系。尽管取得了这些进展，但在gan的背景下，自我注意尚未被探索。(AttnGAN (Xu et al., 2018)对输入序列中的词嵌入使用注意，但对内部模型状态不使用自注意)。SAGAN学会在图像的内部表示中有效地找到全局的、长期的依赖关系。

3. 自注意生成对抗网络

大多数基于gan的图像生成模型(Radford et al., 2016; Salimans et al., 2016; Karras et al., 2018)都是使用卷积层构建的。卷积处理局部邻域的信息，因此单独使用卷积层对图像中的远程依赖关系建模计算效率低下。在本节中，我们采用(Wang et al., 2018)的非局部模型将自关注引入GAN框架，使生成器和鉴别器都能够有效地建模广泛分离的空间区域之间的关系。我们将提出的方法称为自注意生成对抗网络(Self-Attention Generative Adversarial Networks, SAGAN)，因为它的自注意模块(见图2)。

首先将前一隐藏层 $\mathbf{x} \in \mathbb{R}^{C \times N}$ 的图像特征转换为两个特征空间 \mathbf{f}, \mathbf{g} 来计算注意力，其中 $\mathbf{f}(\mathbf{x}) = \mathbf{W}_f \mathbf{x}$, $\mathbf{g}(\mathbf{x}) = \mathbf{W}_g \mathbf{x}$

$$\beta_{j,i} = \frac{\exp(s_{ij})}{\sum_{i=1}^N \exp(s_{ij})}, \text{ where } s_{ij} = \mathbf{f}(\mathbf{x}_i)^T \mathbf{g}(\mathbf{x}_j), \quad (1)$$

$\beta_{j,i}$ 表示模型在合成 j^{th} 区域时对 i^{th} 位置的关注程度。其中 C 为通道数， N 为前一隐藏层特征的特征位置数。注意层的输出为 $\mathbf{o} = (\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_j, \dots, \mathbf{o}_N) \in \mathbb{R}^{C \times N}$ ，其中：

$$\mathbf{o}_j = \mathbf{v} \left(\sum_{i=1}^N \beta_{j,i} \mathbf{h}(\mathbf{x}_i) \right), \quad \mathbf{h}(\mathbf{x}_i) = \mathbf{W}_h \mathbf{x}_i, \quad \mathbf{v}(\mathbf{x}_i) = \mathbf{W}_v \mathbf{x}_i. \quad (2)$$

在上述公式中， $\mathbf{W}_g \in \mathbb{R}^{\bar{C} \times C}$, $\mathbf{W}_f \in \mathbb{R}^{\bar{C} \times C}$, $\mathbf{W}_h \in \mathbb{R}^{\bar{C} \times C}$ ，和 $\mathbf{W}_v \in \mathbb{R}^{C \times \bar{C}}$ 是学习到的权重矩阵，它被实现为 1×1 个卷积。由于我们没有注意到当减少通道数时，性能会有明显的下降 \bar{C} 成为 C/k ，其中 $k = 1, 2, 4, 8$ 在ImageNet上训练了几次之后。为了内存效率，我们选择 $k = 8$ (i.e., $\bar{C} = C/8$ 在我们所有的实验中)。

此外，我们进一步将注意力层的输出乘以一个尺度参

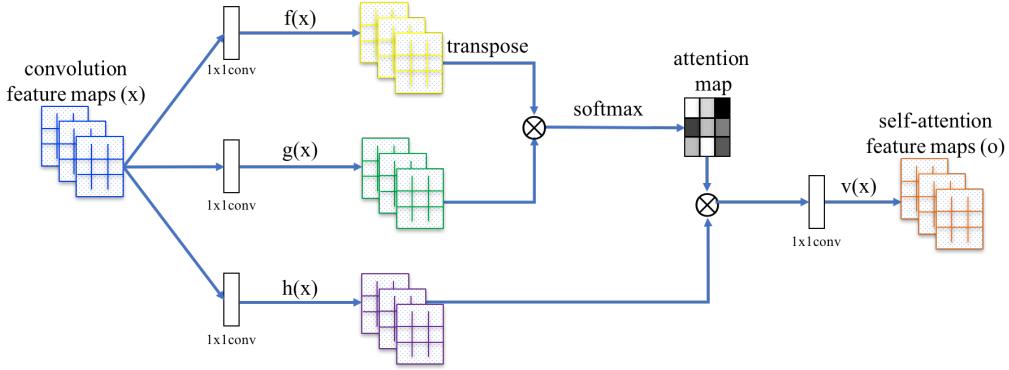


Figure 2. 为SAGAN提出的自我关注模块。 \otimes 表示矩阵乘法。对每一行执行softmax操作。

数，并将输入特征映射加回去。因此，最终输出为：

$$\mathbf{y}_i = \gamma \mathbf{o}_i + \mathbf{x}_i, \quad (3)$$

其中 γ 是一个可学习的标量，初始化为0。引入可学习的 γ 允许网络首先依赖本地邻居的线索——因为这更容易——然后逐渐学会给非本地证据分配更多的权重。我们这么做的直观原因很简单：我们想先学习简单的任务，然后逐步增加任务的复杂性。在SAGAN中，提出的注意力模块已应用于生成器和鉴别器，它们通过最小化对抗性损失的铰链版本以交替的方式进行训练(Lim & Ye, 2017; Tran et al., 2017; Miyato et al., 2018)。

$$\begin{aligned} L_D = & -\mathbb{E}_{(x,y) \sim p_{data}} [\min(0, -1 + D(x, y))] \\ & -\mathbb{E}_{z \sim p_z, y \sim p_{data}} [\min(0, -1 - D(G(z), y))], \quad (4) \\ L_G = & -\mathbb{E}_{z \sim p_z, y \sim p_{data}} D(G(z), y), \end{aligned}$$

4. 稳定gan训练的技术

我们还研究了两种技术来稳定gan在挑战性数据集上的训练。首先，我们在生成器和鉴别器中使用频谱归一化(Miyato et al., 2018)。其次，我们确认了双时间尺度更新规则(TTUR) (Heusel et al., 2017)是有效的，并且我们主张专门使用它来解决正则化鉴别器中的缓慢学习问题。

4.1. 生成器和鉴别器的谱归一化

Miyato et al.(Miyato et al., 2018)最初提出通过将谱归一化应用于鉴别器网络来稳定gan的训练。这样做通过限制每一层的谱范数来限制鉴别器的利普希茨常数。与其他归一化技术相比，谱归一化不需要额外的超参数调优(将所有权重层的谱范数设置为1在实践中始终表现良好)。此外，计算成本也相对较小。

我们认为生成器也可以从谱归一化中受益，基于最近的证据，生成器的调节是gan性能的重要原因(Odena et al., 2018)。在发生器中进行谱归一化可以防止参数幅度的上升，避免异常梯度。我们从经验上发现，生

成器和鉴别器的谱归一化使得每次生成器更新使用更少的鉴别器更新成为可能，从而显着降低了训练的计算成本。该方法具有更稳定的训练行为。

4.2. 生成器和鉴别器更新的学习率不平衡

在以前的工作中，鉴别器(Miyato et al., 2018; Gulrajani et al., 2017)的正则化通常会减慢gan的学习过程。在实践中，使用正则化鉴别器的方法通常需要在训练过程中每个生成器更新步骤需要多个(*e.g.*, 5)鉴别器更新步骤。另外，Heusel et al.(Heusel et al., 2017)主张对生成器和鉴别器使用单独的学习率(TTUR)。我们建议专门使用ttur来补偿正则化鉴别器中缓慢学习的问题，使每个生成器步骤使用更少的鉴别器步骤成为可能。使用这种方法，我们能够在相同的时钟时间内产生更好的结果。

5. 实验

为了评估所提出的方法，我们在LSVRC2012 (ImageNet)数据集(Russakovsky et al., 2015)上进行了大量实验。首先，在5.1节中，我们提出了旨在评估两种提出的稳定gan训练技术的有效性的实验。接下来，本文将在5.2章节中对所提出的自我注意机制进行研究。最后，将我们的SAGAN与最先进的方法(Odena et al., 2017; Miyato & Koyama, 2018)在5.3节中的图像生成任务上进行比较。每个模型在4个gpu上训练大约2周，使用同步SGD(因为异步SGD有众所周知的困难-参见例如(Odena, 2016))。

评估指标。

我们选择Inception分数(IS)(Salimans et al., 2016)和frjs:1; Inception距离(FID) (Heusel et al., 2017)进行定量评价。虽然存在替代品(Zhou et al., 2019; Khrulkov & Oseledets, 2018; Olsson et al., 2018)，但它们并没有被广泛使用。盗梦空间得分(Salimans et al., 2016)计算条件分布和边际类分布之间的KL散度。盗梦空间分数越高，图像质量越好。我们包含了盗梦空间的配乐，因为它被广泛使用，因此可以将我们的结果与以前的工作

进行比较。然而，重要的是要理解盗梦空间评分有严重的局限性——它主要是为了确保模型生成的样本可以被自信地识别为属于特定类别，并且模型从许多类别中生成样本，不一定是为了评估细节的现实性或类内多样性。**FID**是一种更具原则性和综合性的度量，并且在评估生成样本的真实性和变化时已被证明与人类评估更加一致(Heusel et al., 2017)。**FID**在Inception-v3网络的特征空间中计算生成图像与真实图像之间的Wasserstein-2距离。除了在整个数据分布上计算的**FID** (*i.e.*, ImageNet中所有1000类图像)，我们还计算每个类中生成的图像和数据集图像之间的**FID**(称为**intra FID** (Miyato & Koyama, 2018))。**FID**值和**FID**内值越低，意味着合成数据和真实数据分布之间的距离越近。在我们所有的实验中，每个模型随机生成50k个样本来计算Inception分数、**FID**和内部**FID**。

网络结构和实现细节。

我们训练的所有SAGAN模型都设计为生成128张×128张图像。默认情况下，谱归一化(Miyato et al., 2018)用于生成器和鉴别器中的层。与(Miyato & Koyama, 2018)类似，SAGAN在生成器中使用条件批处理归一化，在鉴别器中使用投影。对于所有模型，我们使用Adam优化器(Kingma & Ba, 2015)与 $\beta_1 = 0$ 和 $\beta_2 = 0.9$ 进行训练。缺省情况下，鉴别器的学习率为0.0004，生成器的学习率为0.0001。

5.1. 评估所提出的稳定技术

在本节中，进行实验来评估所提出的稳定技术的有效性，*i.e.*，将频谱归一化(SN)应用于生成器并利用不平衡学习率(TTUR)。在图3中，我们的模型“G / D上的SN”和“G / D +TTUR上的SN”与基于最先进的图像生成方法(Miyato et al., 2018)实现的基线模型进行了比较。在该基线模型中，SN仅用于鉴别器。当我们对鉴别器(D)和生成器(G)进行1:1平衡更新训练时，训练变得非常不稳定，如图3最左边的子图所示。它在训练中很早就表现出模式崩溃。例如，图4的左上角子图展示了基线模型在第10k次迭代时随机生成的一些图像。虽然在原始论文(Miyato et al., 2018)中，通过对D和G使用5:1的不平衡更新大大减轻了这种不稳定的训练行为，但为了提高模型的收敛速度，需要能够使用1:1的平衡更新来稳定地训练。因此，使用我们提出的技术意味着该模型可以在给定相同的挂钟时间下产生更好的结果。鉴于此，不需要为生成器和鉴别器搜索合适的更新比率。如图3中间的子图所示，在生成器和鉴别器中同时添加SN，极大地稳定了我们的模型“G / D”上的SN，即使是在1:1平衡更新的情况下。然而，在训练过程中，样本质量并不是单调提高的。例如，由FID和IS测量的图像质量在第260k次迭代时开始下降。该模型在不同迭代下随机生成的示例图像如图4所示。当我们也应用不平衡学习率来训练鉴别器和生成器时，我们的模型“G / D +TTUR”上的SN在整个训练过程中生成的图像质量单调提高。如图3和4所示，在一百万次训练迭代期间，我们没有

观察到样本质量或FID或Inception得分有任何显著的下降。因此，定量结果和定性结果都证明了所提出的稳定技术对gan训练的有效性。他们还证明，这两种技术的效果至少部分是叠加的。在其余的实验中，所有模型都对生成器和鉴别器使用谱归一化，并使用不平衡学习率以1:1的更新训练生成器和鉴别器。

5.2. 自我注意机制。

为了探究所提出的自注意机制的效果，我们将自注意机制添加到生成器和鉴别器的不同阶段，建立了多个SAGAN模型。如表1所示，在中高层特征映射(*e.g.*、 $feat_{32}$ 和 $feat_{64}$)上具有自注意机制的SAGAN模型比在低层特征映射(*e.g.*、 $feat_8$ 和 $feat_{16}$)上具有自注意机制的SAGAN模型取得了更好的性能。例如，模型“SAGAN, $feats_8$ ”的**FID**由“SAGAN, $feat_{32}$ ”从22.98提高到18.28。原因是自关注接收到更多的证据，在更大的特征图上有更大的选择条件的自由(*i.e.*，它是对大型特征图的卷积的补充)，然而，在对小型(*e.g.*, 8 × 8)特征图建模依赖关系时，它的作用与局部卷积相似。结果表明，注意机制赋予了生成器和鉴别器更大的能力来直接对特征映射中的远程依赖关系进行建模。此外，我们的SAGAN与不加注意的基线模型(表1第二列)的比较进一步表明了所提出的自注意机制的有效性。

与参数数相同的残差块相比，自注意块也取得了更好的效果。例如，当我们把8张×8个特征图中的自注意块替换为残差块时，训练结果不稳定，导致性能显著下降(*e.g.*, FID从22.98增加到42.13)。即使在训练顺利进行的情况下，用剩余块替换自注意块仍然会导致FID和Inception得分更差。*(e.g.*, FID 18.28 vs. 特征图32 × 32中的27.33)。这个比较表明，使用SAGAN所带来的性能改进不仅仅是由于模型深度和容量的增加。

为了更好地理解在生成过程中所学到的内容，我们将SAGAN中生成器对不同图像的关注权重可视化。图5和图1显示了一些需要注意的示例图像。请参阅图5的标题，了解学习到的注意图的一些属性的描述。

5.3. 与最先进的比较

我们的SAGAN还与最先进的GAN模型(Odena et al., 2017; Miyato & Koyama, 2018)进行了比较，用于ImageNet上的类条件图像生成。如表2所示，我们提出的SAGAN达到了最好的Inception分数，内部**FID**和**FID**。提出的SAGAN显著提高了《盗梦空间》已发表的最佳分数，从36.8分提高到52.52分。SAGAN获得的较低**FID**(18.65)和**intra FID**(83.7)也表明SAGAN通过使用自关注模块来建模图像区域之间的远程依赖关系，可以更好地近似原始图像分布。

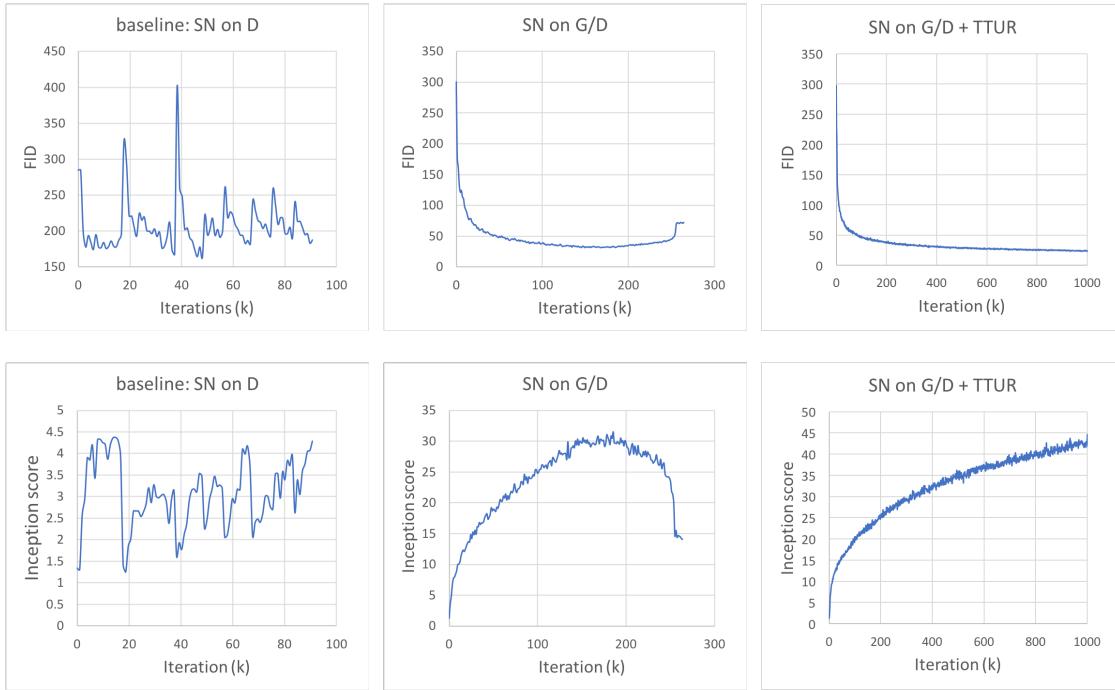


Figure 3. 基线模型和我们的模型的训练曲线与提出的稳定技术，“G / D上的SN”和双时间尺度学习率(TTUR)。所有模型都以1:1平衡更新G和D进行训练。

Model	no attention	SAGAN				Residual			
		$feat_8$	$feat_{16}$	$feat_{32}$	$feat_{64}$	$feat_8$	$feat_{16}$	$feat_{32}$	$feat_{64}$
FID	22.96	22.98	22.14	18.28	18.65	42.13	22.40	27.33	28.82
IS	42.87	43.15	45.94	51.43	52.52	23.17	44.49	38.50	38.96

Table 1. gan自注意与残差块的比较。这些块被添加到网络的不同层中。所有的模型都经过了一百万次迭代的训练，并且报告了最佳的初始分数(IS)和初始距离(FID)。 $feat_k$ 意味着给 $k \times k$ 个特征图添加自关注。

图6显示了ImageNet的代表性类的一些比较结果和生成的图像。我们观察到，我们的SAGAN在合成具有复杂几何或结构模式的图像类别(如金鱼和圣伯纳犬)方面比最先进的GAN模型(Miyato & Koyama, 2018)实现了更好的性能(*i.e.*, 更低的内部FID)。对于结构约束较少的类别(*e.g.*, 山谷, 石墙和珊瑚真菌, 它们更多地通过纹理而不是几何来区分), 我们的SAGAN与基线模型(Miyato & Koyama, 2018)相比显示出较少的优势。同样, 原因是SAGAN中的自关注是对卷积的补充, 用于捕获在几何或结构模式中持续发生的远程、全局级依赖关系, 但在为简单纹理建模依赖关系时扮演类似于局部卷积的角色。

练习。SAGAN在ImageNet上实现了最先进的类条件图像生成性能。

6. 结论

在本文中, 我们提出了自注意生成对抗网络(SAGANs), 它将自注意机制纳入到GAN框架中。自关注模块在远程依赖关系建模方面是有效的。此外, 我们还证明了将谱归一化应用于生成器可以稳定GAN训练, 并且TTUR可以加速正则化鉴别器的训

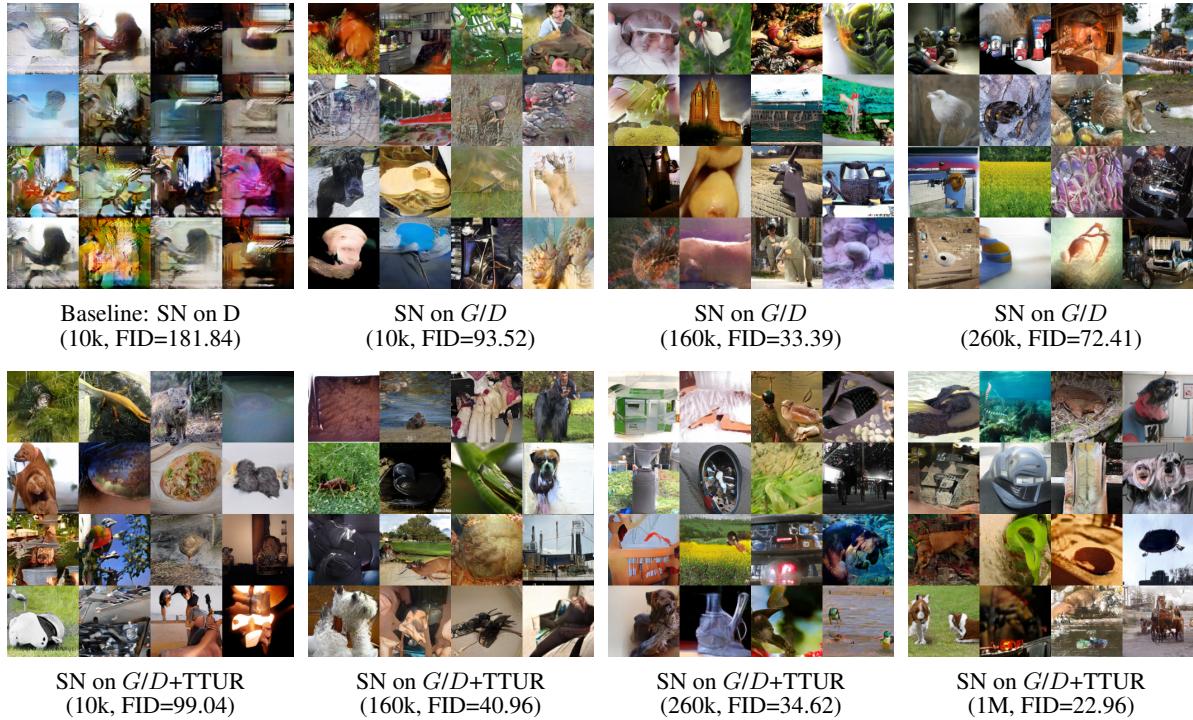


Figure 4. $128 \times$ 由基线模型和我们的模型“ G / D 上的SN”和“ G / D 上的SN +TTUR”随机生成的128个示例。

Model	Inception Score	Intra FID	FID
AC-GAN (Odena et al., 2017)	28.5	260.0	/
SNGAN-projection (Miyato & Koyama, 2018)	36.8	92.4	27.62*
SAGAN	52.52	83.7	18.65

Table 2. 提出的SAGAN与最先进的GAN模型(Odena et al., 2017; Miyato & Koyama, 2018)在ImageNet上进行类条件图像生成的比较。sngan投影的FID是根据官方公布的权重计算的。

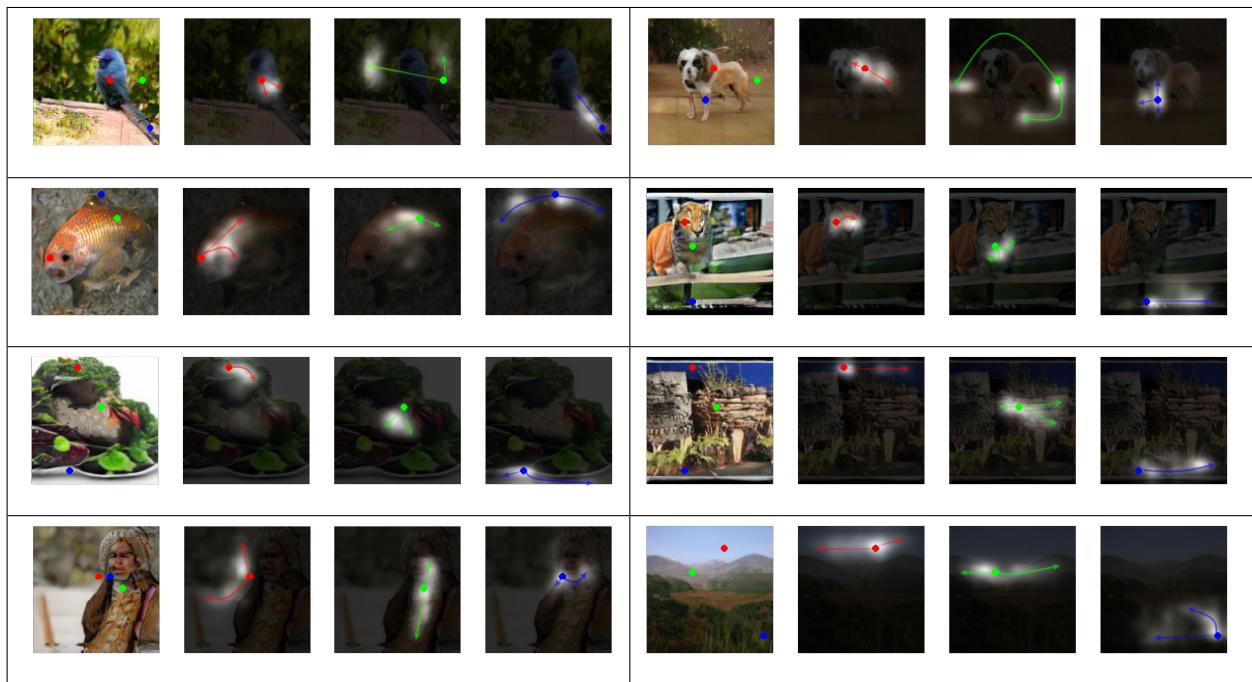


Figure 5. 注意图的可视化。这些图像是由SAGAN生成的。我们将使用注意力的最后一个生成器层的注意力图可视化，因为这一层最接近输出像素，并且最直接地投射到像素空间并进行解释。在每个单元格中，第一张图像显示了三个具有代表性的查询位置，并用颜色编码的点表示。其他三张图片是这些查询位置的注意力图，用相应的颜色编码箭头总结了最受关注的区域。我们观察到，网络学习根据颜色和纹理的相似性来分配注意力，而不仅仅是空间邻接(见左上角的单元格)。我们还发现，尽管一些查询点在空间位置上非常接近，但它们的注意地图可能非常不同，如左下角的单元格所示。如图右上单元格所示，SAGAN能够画出两条腿分开的狗。蓝色的查询点表示注意有助于得到正确的关节区域结构。关于习得的注意图属性的更多讨论，请参阅本文。



Figure 6. SAGAN为不同的类生成128x128的示例图像。每行显示一个类的示例。在最左边的一栏中，列出了我们的SAGAN的内部FID(左)和最先进的方法(Miyato & Koyama, 2018)(右)。

致谢

我们感谢Surya Bhupatiraju对本文草稿的反馈。我们还感谢David Berthelot和Tom B. Brown在实现细节方面提供的帮助。最后，我们感谢Jakob Uszkoreit、Tao Xu和Ashish Vaswani的有益讨论。

References

- Arjovsky, M., Chintala, S., and Bottou, L. Wasserstein GAN. *arXiv:1701.07875*, 2017.
- Azadi, S., Olsson, C., Darrell, T., Goodfellow, I., and Odena, A. Discriminator rejection sampling. *arXiv preprint arXiv:1810.06758*, 2018.
- Bahdanau, D., Cho, K., and Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv:1409.0473*, 2014.
- Brock, A., Donahue, J., and Simonyan, K. Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2018.
- Che, T., Li, Y., Jacob, A. P., Bengio, Y., and Li, W. Mode regularized generative adversarial networks. In *ICLR*, 2017.
- Chen, X., Mishra, N., Rohaninejad, M., and Abbeel, P. Pixelsnail: An improved autoregressive generative model. In *ICML*, 2018.
- Cheng, J., Dong, L., and Lapata, M. Long short-term memory-networks for machine reading. In *EMNLP*, 2016.
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. C., and Bengio, Y. Generative adversarial nets. In *NIPS*, 2014.
- Gregor, K., Danihelka, I., Graves, A., Rezende, D. J., and Wierstra, D. DRAW: A recurrent neural network for image generation. In *ICML*, 2015.
- Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C. Improved training of wasserstein GANs. In *NIPS*, 2017.
- Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. GANs trained by a two time-scale update rule converge to a local nash equilibrium. In *NIPS*, pp. 6629–6640, 2017.
- Hong, S., Yang, D., Choi, J., and Lee, H. Inferring semantic layout for hierarchical text-to-image synthesis. In *CVPR*, 2018.
- Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. Image-to-image translation with conditional adversarial networks. In *CVPR*, 2017.
- Jolicoeur-Martineau, A. The relativistic discriminator: a key element missing from standard GAN. In *ICLR*, 2019.
- Karras, T., Aila, T., Laine, S., and Lehtinen, J. Progressive growing of GANs for improved quality, stability, and variation. In *ICLR*, 2018.
- Karras, T., Laine, S., and Aila, T. A style-based generator architecture for generative adversarial networks. In *CVPR*, 2019.
- Khrulkov, V. and Oseledets, I. Geometry score: A method for comparing generative adversarial networks. *arXiv preprint arXiv:1802.02664*, 2018.
- Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. In *ICLR*, 2015.
- Ledig, C., Theis, L., Huszar, F., Caballero, J., Aitken, A., Tejani, A., Totz, J., Wang, Z., and Shi, W. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, 2017.

- Lim, J. H. and Ye, J. C. Geometric GAN. *arXiv:1705.02894*, 2017.
- Liu, M. and Tuzel, O. Coupled generative adversarial networks. In *NIPS*, 2016.
- Metz, L., Poole, B., Pfau, D., and Sohl-Dickstein, J. Unrolled generative adversarial networks. In *ICLR*, 2017.
- Miyato, T. and Koyama, M. cGANs with projection discriminator. In *ICLR*, 2018.
- Miyato, T., Kataoka, T., Koyama, M., and Yoshida, Y. Spectral normalization for generative adversarial networks. In *ICLR*, 2018.
- Odena, A. Faster asynchronous sgd. *arXiv preprint arXiv:1601.04033*, 2016.
- Odena, A. Open questions about generative adversarial networks. *Distill*, 2019. doi: 10.23915/distill.00018. <https://distill.pub/2019/gan-open-problems>.
- Odena, A., Olah, C., and Shlens, J. Conditional image synthesis with auxiliary classifier GANs. In *ICML*, 2017.
- Odena, A., Buckman, J., Olsson, C., Brown, T. B., Olah, C., Raffel, C., and Goodfellow, I. Is generator conditioning causally related to GAN performance? In *ICML*, 2018.
- Olsson, C., Bhupatiraju, S., Brown, T., Odena, A., and Goodfellow, I. Skill rating for generative models. *arXiv preprint arXiv:1808.04888*, 2018.
- Parikh, A. P., Täckström, O., Das, D., and Uszkoreit, J. A decomposable attention model for natural language inference. In *EMNLP*, 2016.
- Park, T., Liu, M., Wang, T., and Zhu, J. Semantic image synthesis with spatially-adaptive normalization. In *CVPR*, 2019.
- Parmar, N., Vaswani, A., Uszkoreit, J., Łukasz Kaiser, Shazeer, N., and Ku, A. Image transformer. *arXiv:1802.05751*, 2018.
- Radford, A., Metz, L., and Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. In *ICLR*, 2016.
- Reed, S., Akata, Z., Mohan, S., Tenka, S., Schiele, B., and Lee, H. Learning what and where to draw. In *NIPS*, 2016a.
- Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., and Lee, H. Generative adversarial text-to-image synthesis. In *ICML*, 2016b.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. ImageNet large scale visual recognition challenge. *IJCV*, 2015.
- Salimans, T., Goodfellow, I. J., Zaremba, W., Cheung, V., Radford, A., and Chen, X. Improved techniques for training GANs. In *NIPS*, 2016.
- Salimans, T., Zhang, H., Radford, A., and Metaxas, D. N. Improving GANs using optimal transport. In *ICLR*, 2018.
- Sønderby, C. K., Caballero, J., Theis, L., Shi, W., and Huszar, F. Amortised map inference for image super-resolution. In *ICLR*, 2017.
- Taigman, Y., Polyak, A., and Wolf, L. Unsupervised cross-domain image generation. In *ICLR*, 2017.
- Tran, D., Ranganath, R., and Blei, D. M. Deep and hierarchical implicit models. *arXiv:1702.08896*, 2017.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. Attention is all you need. *arXiv:1706.03762*, 2017.
- Wang, X., Girshick, R., Gupta, A., and He, K. Non-local neural networks. In *CVPR*, 2018.
- Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A. C., Salakhutdinov, R., Zemel, R. S., and Bengio, Y. Show, attend and tell: Neural image caption generation with visual attention. In *ICML*, 2015.
- Xu, T., Zhang, P., Huang, Q., Zhang, H., Gan, Z., Huang, X., and He, X. AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks. In *CVPR*, 2018.
- Xue, Y., Xu, T., Zhang, H., Long, L. R., and Huang, X. SegAN: Adversarial network with multi-scale L1 loss for medical image segmentation. *Neuroinformatics*, pp. 1–10, 2018.
- Yang, Z., He, X., Gao, J., Deng, L., and Smola, A. J. Stacked attention networks for image question answering. In *CVPR*, 2016.
- Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., and Metaxas, D. N. StackGAN++: Realistic image synthesis with stacked generative adversarial networks. *TPAMI*.
- Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., and Metaxas, D. StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks. In *ICCV*, 2017.

- Zhao, J., Mathieu, M., and LeCun, Y. Energy-based generative adversarial network. In *ICLR*, 2017.
- Zhou, S., Gordon, M., Krishna, R., Narcomey, A., Morina, D., and Bernstein, M. S. HYPE: human eye perceptual evaluation of generative models. *CoRR*, abs/1904.01121, 2019. URL <http://arxiv.org/abs/1904.01121>.
- Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017.