**利用更少的标签生成高保真图像**

**摘要**

深度生成模型将成为现代机器学习的奠基石（cornerstone）。最近在条件生成对抗网络（conditional generative adversarial networks）方面的研究说明通过自然图像学习复杂的、高维分布特征是有可能的。虽然最新的模型能够生成高保真、多样性的高分辨率自然图像，但是这要依赖于大量的带标签数据。在本文中，我们说明怎样从最近关于自监督和半监督学习的研究中获得超过SOTA（states-of-the-art）的表现效果，无论实在无监督的ImageNet合成方面还是在特殊的数据集合成方面。特别的，与现在表现优异的条件模型Big-GAN相比较，在ImageNet数据集上，本文所提到的方法仅用10%的标签球可以达到同样的效果，如果标签数据达到20%，本文所提到的方法将表现更佳。

**1.介绍**

由于可以学习到复杂的高维分布特征，例如：自然图像、视频、语音等分布特征，深度生成模型获得了广泛的关注。最近的研究进展主要通过扩大模型、模型的结构修改以及正则化技术。

高保真自然图像的生成（通常在ImageNet上训练）依赖于访问大量带标签的数据。这并不奇怪，因为标签将丰富的边信息引入到训练过程中，有效地将极具挑战性的图像生成任务划分为语义上有意义的子任务。

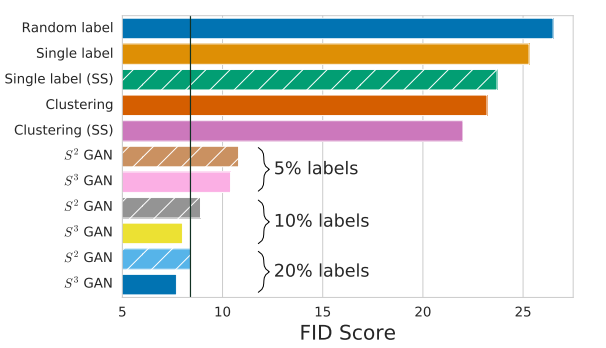


图1。基线的FID和提出的方法。垂直线表示使用所有带标签数据的基线（BigGAN）。所提出的方法（S3GAN）在只使用10%的标记数据的情况下，其表现能够与之相匹配，在使用20%带标签样本的请扩下，其性能更优。

然而，这种对大量标记数据的依赖与大多数数据没有标记的事实是不一致的，并且标记本身往往代价高昂且容易出错。尽管最近在无监督图像生成方面取得了一些进展，但条件模型和无监督模型在样本质量方面的差距仍然很大。

在这项工作中，使用生成性对抗网络（GANs），我们采取了一个重要的步骤，以缩小有条件和无监督生成的高保真图像之间的差距。我们利用两个简单而强大的**概念**：

（i） 自监督学习：通过自监督学习方法在训练数据上训练获得语义特征提取器，然后利用得到的特征表示来指导GAN的训练过程。

（ii）半监督学习：整个训练集的标签可以从小部分带标签的图像中推断出来，并且推断出来的标签可以用作GAN训练的条件信息。

我们在这项工作中的**贡献**

1.提出并研究各种方法来减少或完全忽略自然图像生成任务中的真值标签信息，

2.在ImageNet上实现无监督生成新的SOTA，仅使用10%的标签在128x128 ImageNet上获得SOTA，并仅使用20%的标签（通过FID测量）达到新的SOTA，以及

3.开源所有用于实验的代码github.com/google/compare\_gan。

**2.背景和研究现状**

**IMAGENET上的高保真GANs** 除了BIGGAN（Brock，2019）之外，只有少数先前的方法能够将GANs缩放到IMAGENET，其中大多数方法依赖于使用标签的类条件生成器（class-conditional generation）。最早的尝试之一是带有辅助分类器（ACGANs）（Odena，2017年）的GANs，它将一个带有潜在代码的one-hot encoded标签信息输入生成器，并为鉴别器配备一个辅助头，除了预测输入是真是假之外，还预测图像的类别。最近的方法依赖于鉴别器中的标签预测层，本质上产生每类真/假分类器（Miyato&Koyama，2018）和生成器中的自我关注机制（Zhang等人，2018）。这两种方法都使用调制批量归一化（modulated batch normalization）向生成器提供标签信息（De Vries等人，2017）。在无监督方面，Chen（2019b）表明附加在鉴别器上的辅助旋转损失函数对训练有稳定作用。最后，适当的梯度正则化可以在不使用标签的情况下将MMD-GANs扩展到ImageNet（Arbel等人，2018）。

**半监督GANs** 最近的一些工作利用GANs进行分类器的半监督学习。Salimans等人（2016）和Odena（2016）训练一个鉴别器，将其输入分类为K+1类：K图像类用于真实图像，一类用于生成图像。同样，Springenberg（2016）将标准的GAN目标扩展到了K类。Li（2017）等人也考虑了这种方法，使用分离的判别器和分类器模型。其他方法包括推断模型，来预测缺失的标签（Deng等人，2017）或利用联合分布（标签和数据）匹配半监督学习（Gan等人，2017）。我们强调，这项工作的重点是从几个标签中训练分类器，而不是使用几个标签来提高生成的模型的质量。据我们所知，Li（2017）；邓（2017年）；Sricharan等人（2017）提到可以通过部分标签信息提高样品质量，所有这些都只考虑来自有脸领域的低分辨率数据集。

**自监督学习** 自监督学习方法采用无标签辅助任务来学习数据的语义特征表示。该方法成功地应用于不同的数据模式，如图像（Doersch等人，2015；Caron等人，2018）、视频（Agrawal等人，2015；Lee等人，2017）和机器人（Jang等人，2018；Pinto&Gupta，2016）。IMAGENET上当前最先进的方法是由Gidaris等人（2018）提出的，建议将预测被旋转只有的训练图像的旋转角度作为辅助任务。这种简单的自监督方法产生了对下游图像分类任务有用的表示。其他形式的自我监督方法包括预测给定图像的不相交图像块的相对位置（Doersch等人，2015；Mundhenk等人，2018）或估计规则网格上随机交换图像块的排列（Noroozi&Favaro，2016）。Kolesnikov（2019年）等人对现代神经结构的自监督学习进行了研究。



图2。顶行：128x128的样本，来自当前最先进的全监督模型BIGGAN。最下面一行：样本来自提到的的S3GAN，它在FID和IS方面的表现与BIGGAN相媲美，并且仅使用10%的真值标签。

**3. 减少对标记数据的需求**

简言之，我们将不向鉴别器提供真实图像的手动注释的真值标签，而是提供推断的真值标签。为了获得这些标签，我们将利用最近在自监督和半监督学习方面的进展。在详细介绍这些方法之前，我们首先讨论如何在最先进的GANs中使用标签信息。下面的论述假设熟悉GAN框架的基础知识（Goodfellow等人，2014）。

**合并标签** 根据Miyato&Koyama（2018）的建议，我们采用线性预测层将标签信息提供给鉴别器。为了使论述内容完整，我们将简要地回顾一下主要观点。在“香草”（无条件）GAN中，鉴别器D学习预测其输入x是真实的还是由生成器G生成的。我们将鉴别器分解为学习的鉴别器表示，，它被输入到线性分类器中，即鉴别器由()给出。在预测鉴别器中，学习与表示维数相同的每一类的**映射**。然后，对于给定的图像，标记输入x，y是真实的还是生成的决定于两个部分：（a）表示本身是否与真实数据一致，以及（b）表示是否与来自y类的真实数据一致。更正式地说，鉴别器采用

的形式，其中是应用于特征向量的线性预测层，一个热编码标签y作为输入。对于生成器，标签信息y通过类条件批次归一化合并（Dumoulin等人，2017；De Vries等人，2017）。带有预测鉴别器的条件GAN如图3所示。我们分别在第3.1节和第3.2节中描述了用于推断GAN训练标签的预训练和协同训练方法。

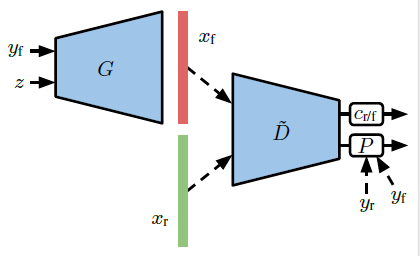


图3。带预测鉴别器的条件GAN。鉴别器通过将无条件分类器和通过预测层P实现的类条件分类器相结合，尝试从表示D预测真实图像xr（带标签yr）还是生成图像xf（带标签yf）是否在其输入处。这种形式的条件处理在BIGGAN中使用。向外的箭头输入损失函数。

**3.1 预训练的方法**

**基于无监督聚类的方法** 我们首先使用最先进的自监督方法学习真实训练数据的表示（Gidaris 等人，2018；Kolesnikov等人，2019），在此表示上执行聚类，并将聚类标记用作标签的替换。继Gidaris（2018）等人之后我们通过最小化以下自监督损失函数来学习特征抽取器F（通常是卷积神经网络）

, （1）

其中R是4个旋转度的集合，是图像x旋转r度，是预测旋转度r的线性分类器。在学习了特征提取器F之后，我们对训练图像的表示应用了小批量k均值聚类（Sculley，2010）。 最后，给定聚类标记函数，我们使用铰链损失函数训练GAN，或者最小化鉴别器损失函数和生成器损失函数，即

,

式中，是先验分布，其中和是训练集上聚类标签的经验分布。我们称这种方法为**聚类（CLUSTERING）**，并在图4中加以说明。

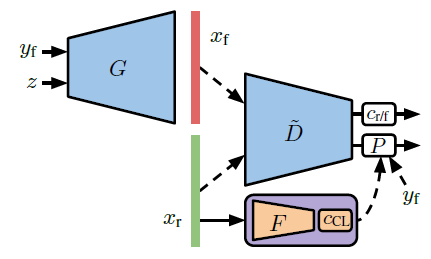


图4。 聚类：一种基于对通过解决自监督任务获得的表示进行聚类的无监督方法。F对应于通过自监督学习的特征提取器，是聚类标记函数。在预训练步骤学习了真实训练图像的F和后，我们通过推断标签为来进行条件GAN训练。

**半监督学习方法** 半监督方法是一个活跃的研究领域，已经提出了大量的算法，我们遵循Beyer等人（2019年）并简单地扩展前段所述的带有半监督损失函数的自监督方法。这确保了两种方法在模型容量和计算成本方面具有可比性。假设我们为训练数据的子集提供了标签，我们试图通过自监督学习一个好的特征表示，同时训练一个好的线性分类器在获得的表示上（使用提供的标签）。更正式地说，我们最小化损失函数

（2）

其中，和分别是预测旋转角度r和标签y的线性分类器，并且平衡损失项。（2）中的第一项对应于（1）中的自监督损失，第二项对应于（半监督）交叉熵损失。在训练过程中，后一个期望值被标记训练样本子集上的经验平均值所代替，而前一个期望值被设置为整个训练集上的经验平均值（本文遵循此惯例）。在我们获得F和后，我们继续进行GAN训练，将真实图像标记为。特别地，我们可以选择最小化与聚类相同的生成器和鉴别器损失函数，除了我们使用通过最小化（2）获得的和F：

,

其中，和是统一的范畴。这种方法使用缩写S2GAN。

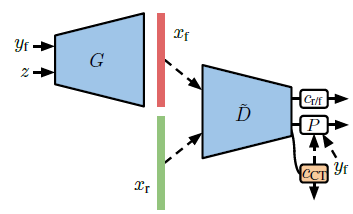


图5。S2GAN-CO：在GAN训练过程中，我们学习了一种基于鉴别器表示的辅助分类器cCT，它基于标记的真实样本，来预测无标记样本的标签。这避免了在GAN训练之前训练特征提取器F和分类器cS2L，如S2GAN中所述。

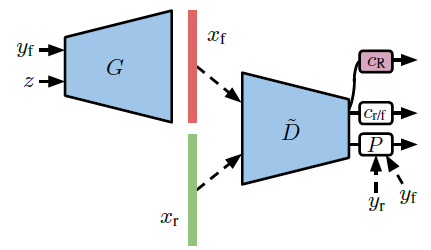


图6。利用旋转预测进行GAN训练的自监督。除了预测输入处的图像是真实的还是生成的之外，还通过辅助线性分类器训练鉴别器来预测旋转的真实图像和伪图像的旋转角度。Chen（2019b）等人成功地应用了这种方法稳定GAN训练。在这里，我们将它与我们预训练和共同训练的方法相结合，用预测的标签代替真实标签。

**3.2 联合训练法**

基于迁移学习的方法的主要缺点是需要通过自监督训练特征抽取器F和学习标签的推理机制（线性分类器或聚类）。在接下来的内容中，我们详细介绍了避免这两步过程的协同训练方法，并学习在GAN训练期间推断标签信息。

**无监督方法** 我们考虑两种方法。第一种方法是，我们通过简单地用相同的标签标记所有真实的和生成的样本并从鉴别器中移除预测层来完全移除标签，即我们设。这种方法使用缩写**单标签(SINGLE LABEL)**。对于第二种方法，我们将随机标签分配给（未标记的）真实图像。虽然真实图像的标签没有向鉴别器提供任何有用的信号，但采样标签可以通过提供不同于z的统计的附加随机性，以及由于在类条件批量归一化(class-conditional BatchNorm)中映射矩阵而产生的附加可训练参数，潜在地帮助生成器。此外，假数据的标签可以在向鉴别器提供假图像的旁侧信息时促进鉴别器。我们称这种方法为随机标签（RANDOM LABEL）。

**半监督方法** 在GAN训练过程中，如果实际数据子集有标签，我们直接根据鉴别器的特征表示训练辅助线性分类器，并利用它对未标记的真实图像进行标签预测。在这种情况下，鉴别器损失函数的形式是

, （3）

其中，第一项对应于（k%）标记的真实图像上的标准条件训练，第二项对应于标记的真实图像上辅助分类器的交叉熵损失函数（权重），第三项对应于未标记真实图像标签由预测的无监督鉴别器损失函数，最后一项是生成数据的标准条件鉴别器损失函数。我们用缩写S2GAN-CO来表示这种方法。 如图5所示。

**3.3 GAN训练中的自监督**

到目前为止，我们利用自监督来设计好的特征表示，或者学习半监督模型（参见第3.1节）。然而，考虑到鉴别器本身只是一个分类器，或许可以通过旋转预测增加一个辅助任务，即自监督，而从中受益。这种方法已经在（Chen等人，2019b）中进行了探索，观察到它可以稳定GAN训练。在这里，我们希望结合第3.1节和第3.2节中介绍的方法来评估其影响。为此，类似于F 在（1）和（2）中的训练，我们在鉴别器特征表示上训练一个额外的线性分类器，以预测被旋转的真实图像和被旋转的假图像的旋转角度。添加到判别器和生成器损失函数上的损失项分别是

(4)

和  
 （5）

其中，是用来平衡损失项的权重。这种方法如图6所示。