# 阅读报告

English 中文

## 一、论文标题：

Diffusion Models Beat GANs on Image Synthesis

## 二、作者单位、发表时间：

Prafulla Dhariwal OpenAI Alex Nichol OpenAI 2021

## 三、论文背景和问题：

GAN已经得到广泛充分的研究，但依然存在一些问题：多样性差（模式崩溃）；很难训练、很容易崩溃。扩散模型虽然比GAN能更好地覆盖样本分布，性质更为理想，但在一些困难的数据集上表现依然不如BigGAN-deep等。

本文假设这种差距来自于：1、GAN的架构已经经过大量改进；2、GAN用多样性换取了保真度。本文将设计一个用多样性换取保真度的方案，将这些好处带给扩散模型。

## 四、论文动机和贡献解读

作者旨在改进扩散模型。具体而言，作者通过一系列消融实验找到了更好的架构改进以提升FID性能（第三节）；描述了一种在采样过程中使用分类器的梯度来引导扩散模型的方法（第四节），通过调整分类器梯度的尺度这一超参即可在多样性和保真度之间进行权衡。

将改进后的模型和上采样堆栈对比，发现改进互补，结合后在ImageNet 256×256和512×512上获得最佳结果。

五、方案设计详细分析：

1、模型架构优化

本文使用的基础模型是其引文[24]中的UNet加一个单头全局注意力模块的架构。原文中提到，他们探索了如下架构变化：

1. 增加深度与宽度，保持模型尺寸相对恒定
2. 增加注意力头的数量
3. 在32×32、16×16和8×8分辨率下使用注意力，而不仅仅是在16×16分辨率下
4. 根据[60]，使用BigGAN[5]的残差块进行激活的上采样和下采样。
5. 根据[60,27,28]，用 缩放残差块

除了（5）其他提高性能，但（2）增加了时间成本所以不再考虑。

此外，作者发现更多head和每个head更少的通道可以提高FID。根据实验结果，最终使用64通道为默认值。

作者还尝试了自适应组归一化（AdaGN），即对于GroupNorm的结果缩放后加偏置。

总结：最终使用如下架构：可变宽度，每种分辨率尺度都有 2 个残差块；多头，每头 64 个通道，在 32、16 和 8 分辨率上进行注意力处理；使用 BigGAN 的残差块进行上采样和下采样，并采用自适应组归一化将时间步和类别嵌入注入到残差块中。

2、类别信息指引

（没看懂）

六、实验效果及其分析：

## 七、结论：

## 八：思考：

关于“类别信息指引”部分需要仔细研究