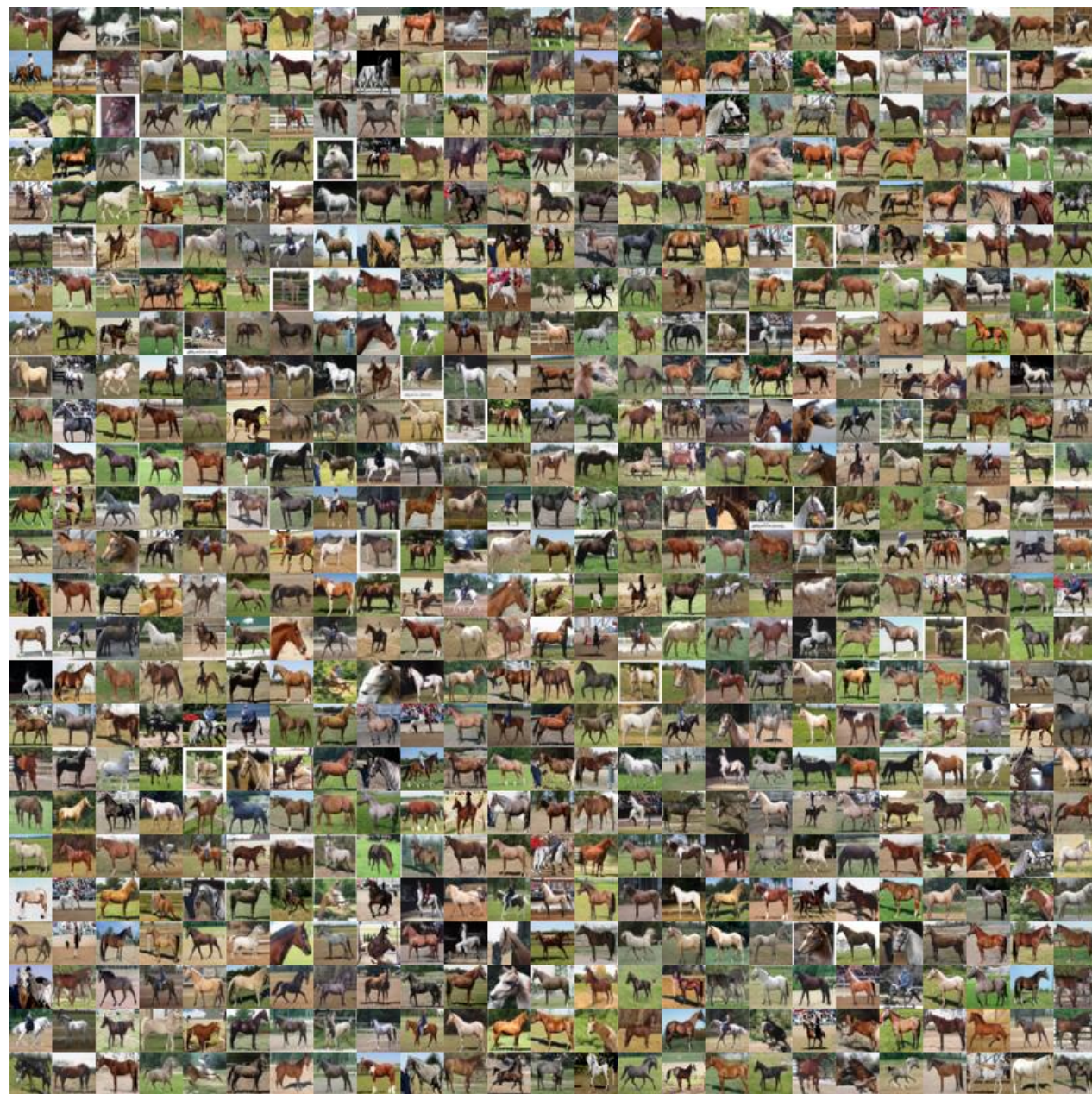
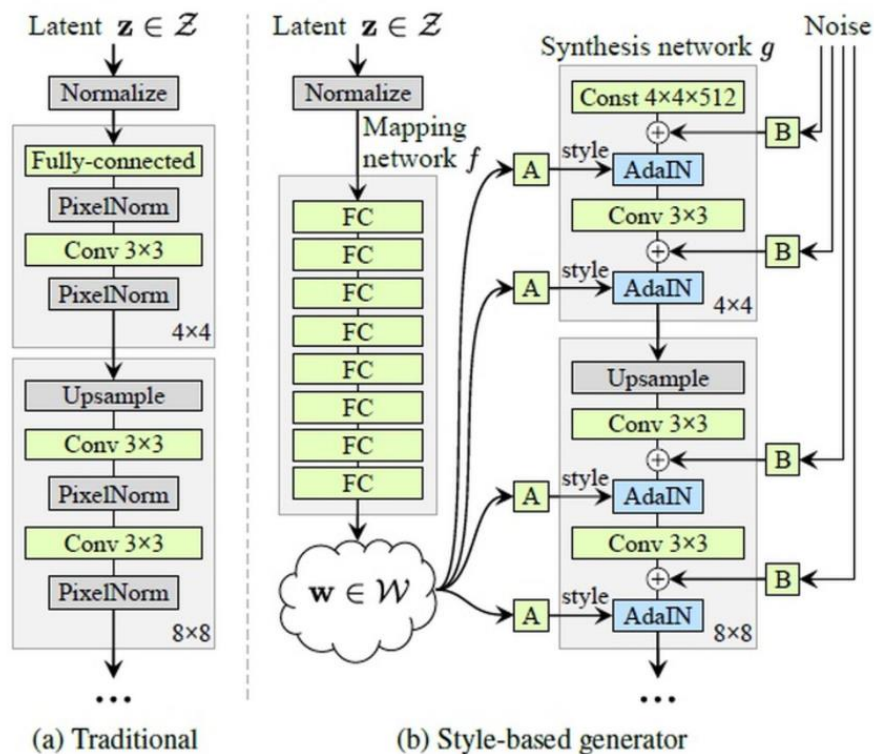


一、结果展示

| 实验编号 | 模型架构 | FID |
|------|------------|-----|
| 1 | GAN | 198 |
| 2 | AAE | 299 |
| 3 | VAE | 327 |
| 4 | DCGAN | 111 |
| 5 | WGAN | 241 |
| 6 | LSGAN | 172 |
| 7 | DRAGAN | 184 |
| 8 | BEGAN | 205 |
| 9 | styleGAN 3 | 100 |
| 10 | styleGANx1 | 22 |



二、算法介绍——StyleGAN

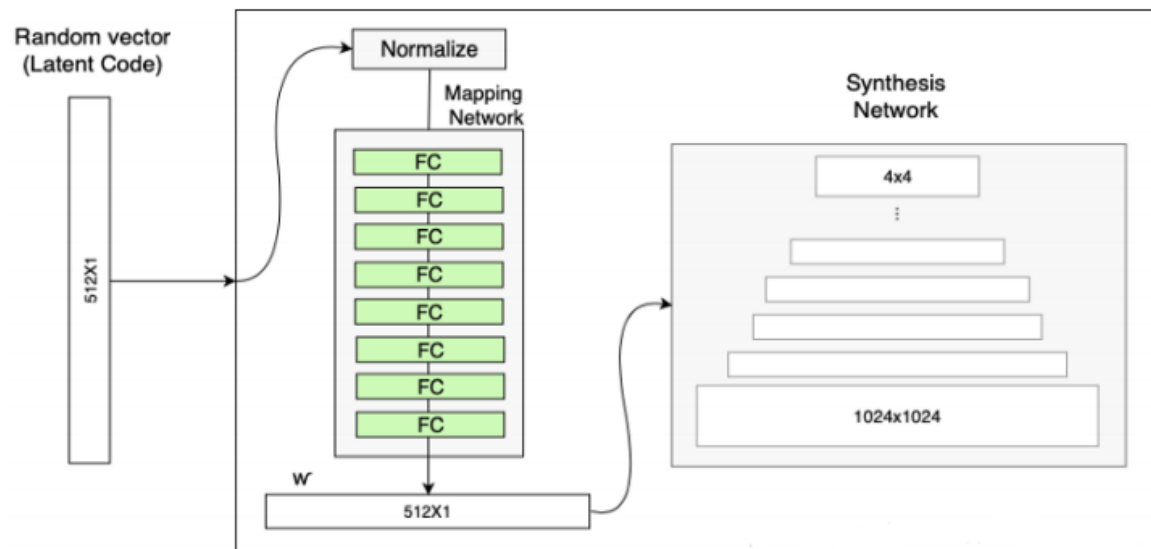


stylegan的模型架构

左图为传统的GAN，右图为StyleGAN。

StyleGAN主要分为两个子网络：

- **Mapping network :**
 - 用于将latent code z 转换为 w
- **Synthesis network :**
 - 用于生成图像



Mapping network详细结构图

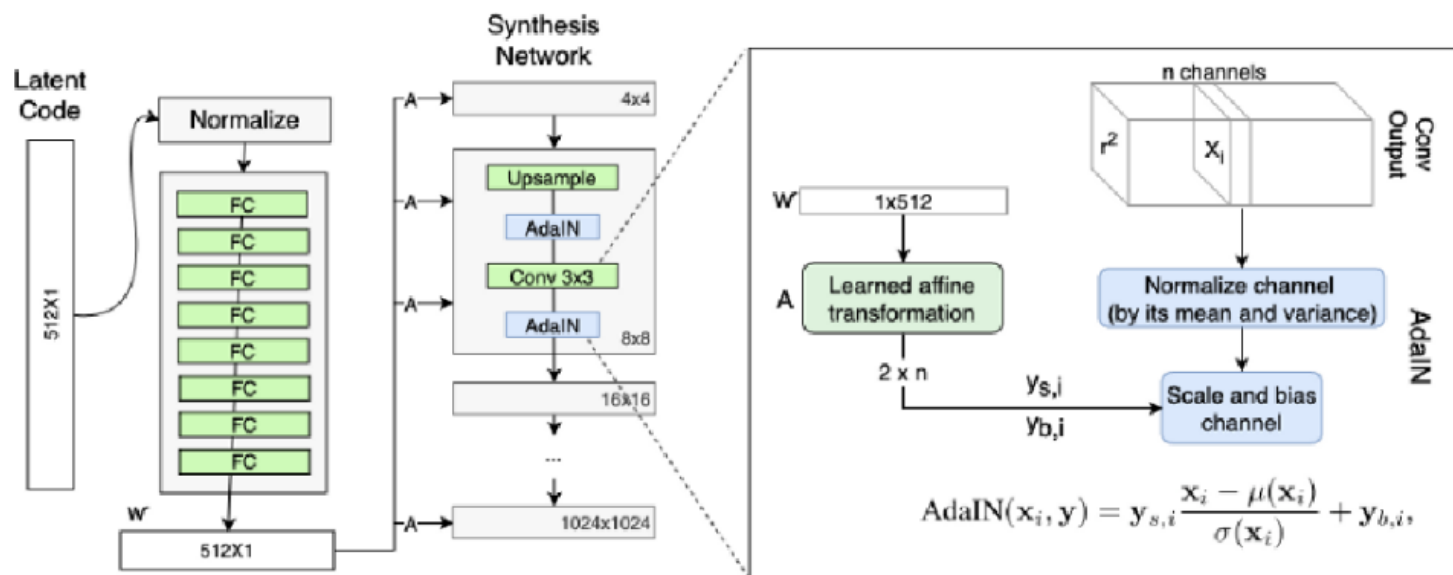
由8个全连接层 (fc) 组成，同时输出向量 w 和输入的 z 维度相同 (512*1)

Mapping Network作用：

z 是符合均匀分布或者高斯分布的随机变量，所以变量之间的耦合性比较大。将latent code z 进行解耦，才能更好地进行后续操作，来改变其不同特征。因此将latent code z 变成 w 。

将latent code转换得到 w 后，然后经过仿射变换生成 A ，分别送入Synthesis network的每一层网络，进行控制特征，Synthesis network的网络层有18层，所以通过 w 生成得到了18个控制向量，用于控制不同的视觉特征。

二、算法介绍——StyleGAN



Synthesis network详细结构图

在Synthesis network中，最初的输入向量shape为 $512 * 4 * 4$ ，最后输出为 $3 * 1024 * 1024$ 。

从 $4 * 4$ 到 $8 * 8 \dots (4, 8, 16, 32 \dots 1024)$ ，其中每一个又包含两个卷积层（一个用于upsample，一个用于特征学习），所以一共包含18层。

StyleGAN相较于传统GAN的改进

1、删除了传统输入

利用 $512 * 4 * 4$ 的输入代替传统初始输入

- 避免初始输入值取值不当而生成不正常图片
- 有助于减少特征纠缠

2、添加噪声(noise)

小特征是随机的，将这些小特征插入GAN图像的常用方法是在输入向量中添加noise；加入噪声后的生成图往往更加逼真与多样。

3、自适应实例归一化 (AdaIN)

基于AdaIN可以快速实现任意图像风格的转换。

4、样式混合 (通过混合正则化)

进一步明确风格控制，低分辨率的style控制姿态等，高分辨率的style控制背景色等。

5、感知路径长度 (Perceptual path length)

是一个指标，用于判断生成器是否选择了最近的路线。

6、线性可分性

训练特征，方便之后编辑，操纵图像特征。

二、算法介绍——StyleGAN 2

➤ 背景

- StyleGAN 1 可以通过分别修改网络中每个级别的输入来控制在该级别中所表示的视觉特征，而不会影响其它级别(通过非线性映射消除特征纠缠)。但是其有一个明显的缺陷——生成的图像有时包含斑点似的伪影(artifacts)，即StyleGAN中出现水滴问题，如下图所示：

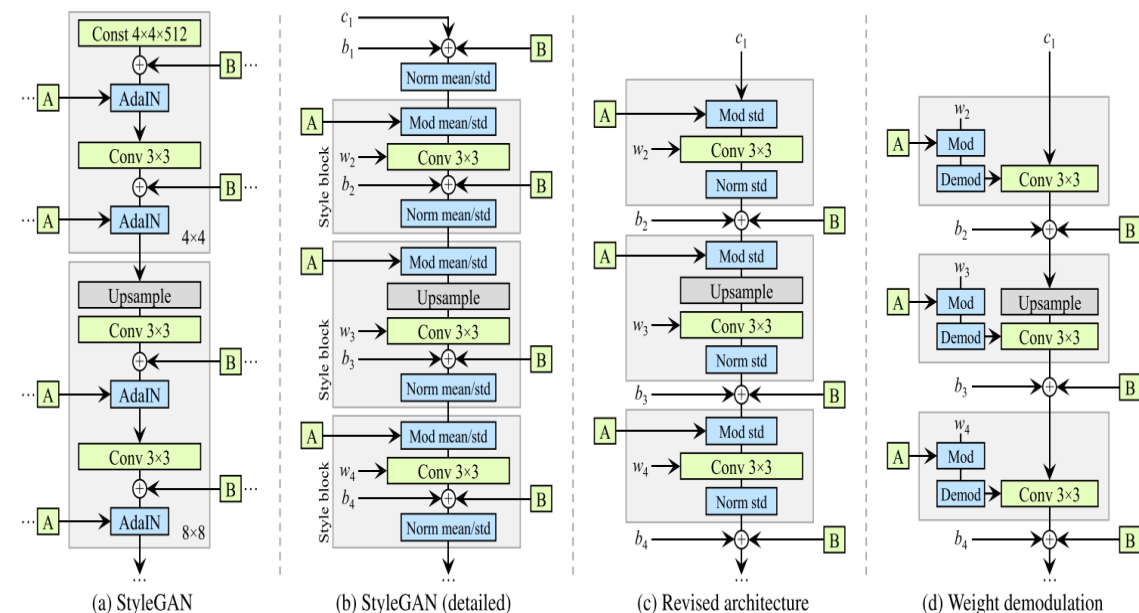


- Instance normalization导致 StyleGAN 图像中出现水滴状伪影。这些在生成的图像中并不总是很明显，但是在生成器网络内部，从 64x64 分辨率开始的所有特征图中总是存在这样的问题。这是一个困扰所有 StyleGAN 图像的系统性问题。

二、算法介绍——StyleGAN 2

➤ 改进

- 实验发现：水滴主要由于AdaIN层导致，但是除去AdaIN层，缺少style mixing模块，StyleGAN结构如下图a、b， StyleGAN结构进行修改得到StyleGAN 2结构如图c、d所示



$$\omega'_{ijk} = s_i \cdot \omega_{ijk} \quad (1)$$

$$\omega''_{ijk} = \omega'_{ijk} / \sqrt{\sum_{i,k} \omega'_{ijk}{}^2 + \varepsilon} \quad (2)$$

➤ 图c网络中的改进：

- 删除了最开始输入的噪声B和偏置b₁
- 仅调整方差，不调整均值
- 将b和B的添加移动到style的活动区域之外，即灰框之外

➤ 增加调节反调节机制（如图d）：

- 式1为图d中Mod模块，调整卷积权重
- 式2为图中Demod模块，得到新的卷积权重，与输入特征进行卷积计算
- i表示输入特征图序号i，j表示输出层序号j，k为卷积核
- 该机制将训练效率由37img/s提升至61img/s

二、算法介绍——StyleGAN 3

➤ 背景

- 目前生成器网络架构是卷积+非线性+上采样的结果，而这样的架构无法做到很好的等变性
- StyleGAN 2 存在图像坐标与特征粘连的问题
- StyleGAN 2 无法满足旋转一致性问题
- 改进后的 StyleGAN 3 实现了真正的图像平移、旋转等不变性

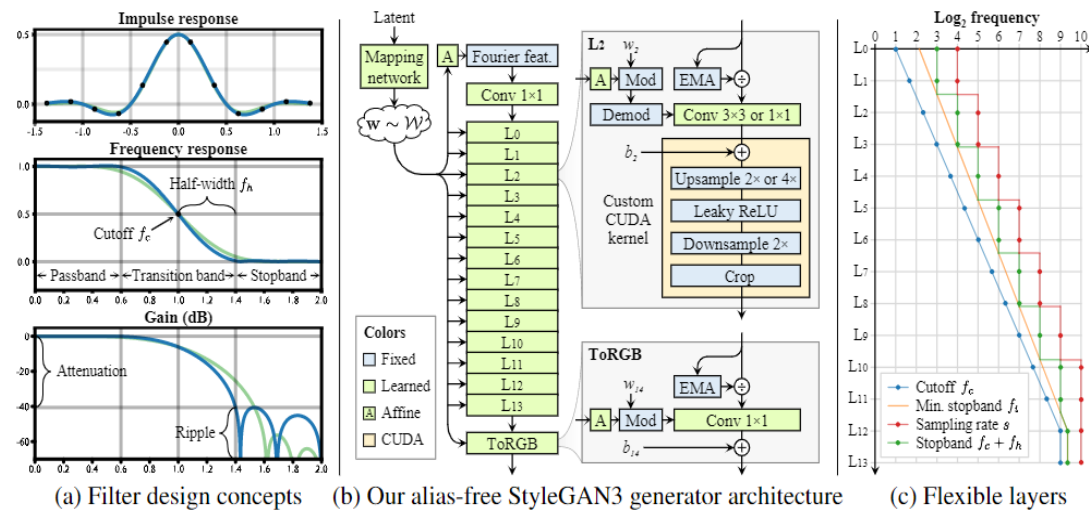


二、算法介绍——StyleGAN 3

➤ 针对StyleGAN2的改进之处

- 提出傅里叶特征以及基线模型简化;
- 引入了一个固定大小的boundaries来近似, 每一层操作后再crop到这个扩展的空间;
- 改进了非线性滤波, 并将整个unsample-LReLU-downsample过程写到一个自定义的CUDA内核中;
- 为了抑制aliasing, 可以简单地将截止频率降低;
- 加入一个放射层, 通过输出全局平移和旋转参数输入傅里叶特征, 改善FID;
- 提出跳频在最低分辨率层中较高, 在最高分辨率层中较低;

- 为了得到旋转不变性网络, 做出以下两个改进
 - ✓ 将所有层的卷积核大小从3*3替换为1*1
 - ✓ 通过将feature map的数量增加一倍, 用来弥补减少的特征容量



二、算法介绍——StyleGAN XL

➤ 背景

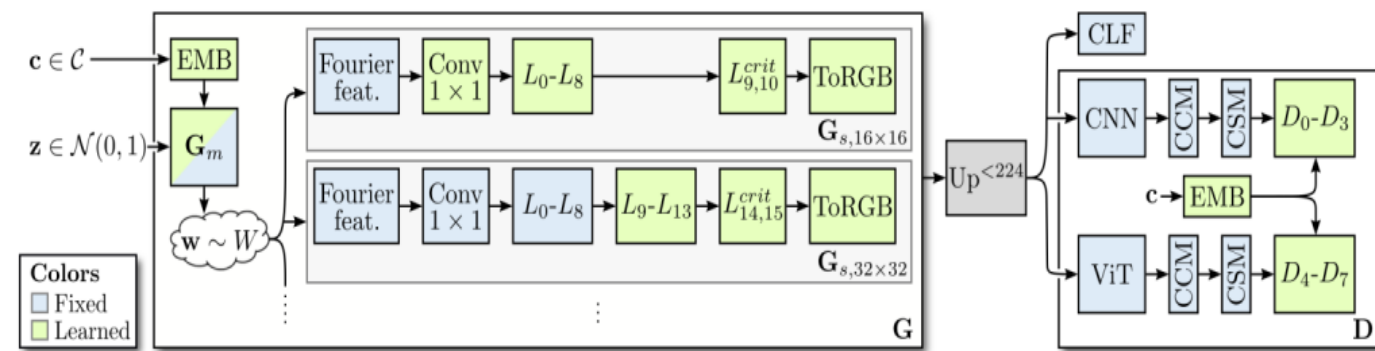
- 尽管StyleGAN 3相较StyleGAN 2有显著优化，但仍不能适用于大型非结构化数据集（如ImageNet）；
- 当需要更大的模型时或扩展到更高的分辨率时，StyleGAN3的计算成本过高；
- 通过将生成和实际的样本投射到一个固定的、预训练的特征空间，Projected GAN取得了较好的表现成果。



二、算法介绍——StyleGAN XL

➤ 针对StyleGAN3的改进之处

- 修改了生成器及其正则化损失，调整了潜在空间以适应 Projected GAN (Config-B) 和类条件设置 (Config-C);
- 重新讨论了渐进式增长，以提高训练速度和性能 (Config-D);
- 研究了用于 Projected GAN 训练的特征网络，以找到一个非常适合的配置 (Config-E);
- 提出了分类器引导，以便 GAN 通过一个预训练的分类器 (Config-F) 提供类信息。



StyleGAN XL 训练模型

| Configuration | | FID ↓ | IS ↑ |
|---------------|------------------------------|-------|-------|
| A | StyleGAN3 | 53.57 | 15.30 |
| B | + Projected GAN & small z | 22.98 | 57.62 |
| C | + Pretrained embeddings | 20.91 | 35.79 |
| D | + Progressive growing | 19.51 | 35.74 |
| E | + ViT & CNN as $F_{1,2}$ | 12.43 | 56.72 |
| F | + CLF guidance (StyleGAN-XL) | 12.24 | 86.21 |

StyleGAN XL 消融实验

三、对比分析

| 实验编号 | 模型架构 | 分析 |
|------|------------|---|
| 1 | GAN | 零和博弈存在“梯度消失”的问题，FID 分数在本次作业中表现中规中矩 |
| 2 | AAE | 使用对抗训练的方式来匹配 $q(z)$ 与 $p(z)$ ；但使得图像模糊，训练效果不佳 |
| 3 | VAE | 使用 KL 散度惩罚来利用 $q(z)$ ；同样只能获得模糊化图片，FID 较低 |
| 4 | DCGAN | 几乎完全使用了卷积层代替全链接层，判别器几乎是和生成器对称的，极大的提升了 GAN 训练的稳定性以及生成结果质量 是本次任务中除 sota 外表现最佳的改进方法，FID 分数及图片清晰度都大大提升 |
| 5 | WGAN | 用基于 Wasserstein 距离的代价函数取代基于 KL 散度和 JS 散度的代价函数，结果反而变差； 推测是由于训练困难，没有收敛；可以将 weight clipping 改为 gradient penalty |
| 6 | LSGAN | 传统 GAN 中, D 网络和 G 网络都是用简单的交叉熵 loss 做更新, LSGAN 则用最小二乘 (Least Squares) Loss 做更新 更严格地惩罚远离数据集的离群 Fake sample, 使得生成图片更接近真实数据 (同时图像也更清晰)FID 有所上升 |
| 7 | DRAGAN | 本质上也是一种梯度裁剪，其判别器和生成器的价值函数类似 WGAN-GP，避开局部最优解 获得更稳定的 GAN 训练，结果比 GAN,WGAN 都有提升，与预期保持一致 |
| 8 | BEGAN | 让生成图像的重构误差分布逼近真实图像的重构误差分布取代传统的 GAN 让生成图像的分布逼近真实图像的分布的做法 结果不降反升，猜测原因是 BEGAN 解决了 GAN 控制生成样本多样性难的问题，但在 Score-based 的任务表现不佳 |
| 9 | styleGAN 3 | 改进内容前面已经介绍过了，FID 对比原始 GAN 近乎达到一倍的提升 |
| 10 | styleGANxl | 是当前任务的 State-of-the-Art |