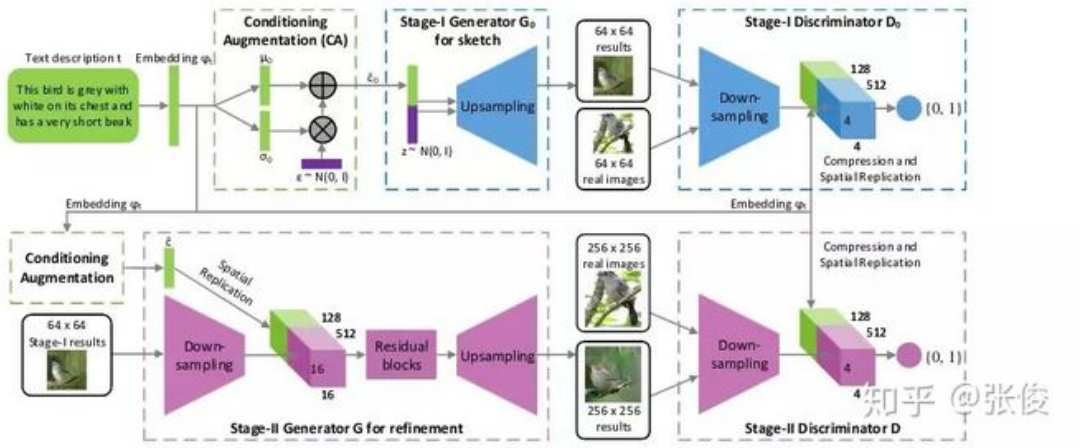
其他GAN

堆叠GAN的文本到图像：*StackGAN*

StackGAN [21] 作为一种迭代方法，只有两层生成器。第一个生成器接收输入 (z,c)，然后输出模糊图像，可以显示粗略的形状和对象的模糊细节，而第二个生成器采用 (z,c) 和前一个生成器生成的图像，然后输出更大的图像，可以得到更加真实的照片细节。

StackGAN [21] 建议使用两个不同的生成器进行文本到图像的合成，而不是只使用一个生成器。第一个生成器负责生成包含粗糙形状和颜色的对象的低分辨率图像，而第二个生成器获取第一个生成器的输出并生成具有更高分辨率和更清晰细节的图像，每个生成器都与其自己的判别器相关联。



StackGAN ++ [27] 建议使用更多对生成器和判别器而不是仅仅两个，为判别器增加无条件图像合成损失，并使用由均值平均损失计算的色彩一致性正则化项和真实和虚假图像之间的差异。

SGAN

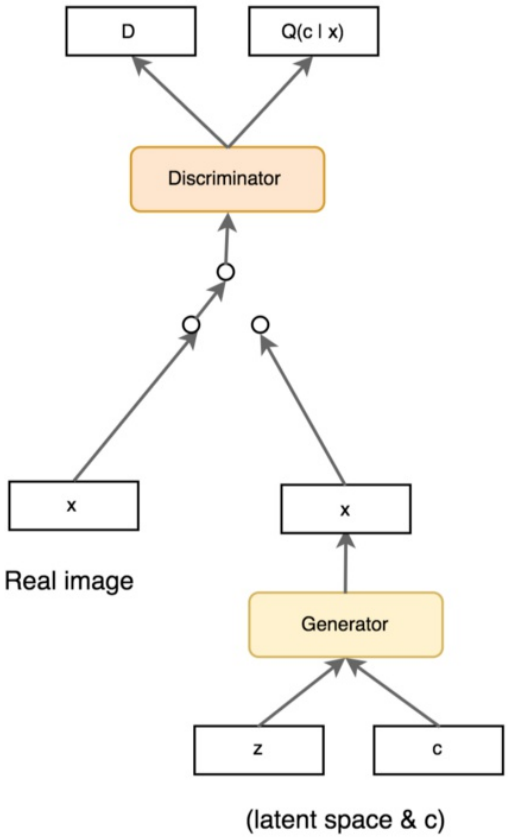
其堆叠生成器，其将较低级别的特征作为输入并输出较高级别的特征，而底部生成器将噪声矢量作为输入并且顶部生成器输出图像。

对不同级别的特征使用单独的生成器的必要性是 SGAN 关联编码器，判别器和 Q 网络（用于预测 P(zi|hi) 的后验概率以进行熵最大化，其中 hi 每个生成器的第 i 层的输出特征），以约束和改善这些特征的质量。

InfoGAN

相比于有监督方法，无监督方法不使用任何标签信息。因此，无监督方法需要对隐空间进行解耦得到有意义的特征表示。

InfoGAN对把输入噪声分解为隐变量 z 和条件变量 c （训练时，条件变量c从均匀分布采样而来。），二者被一起送入生成器。在训练过程中通过最大化 c 和 $G(z,c)$ 的互信息 $I(c;G(z,c))$ 以实现变量解耦（ $I(c;G(z,c))$ 的互信息表示 c 里面关于 $G(z,c)$ 的信息有多少，如果最大化互信息 $I(c;G(z,c))$ ，也就是最大化生成结果和条件变量 c 的关联性）。模型结构和CGAN基本一致，除了Loss多了一项最大互信息。具体如下[10]:

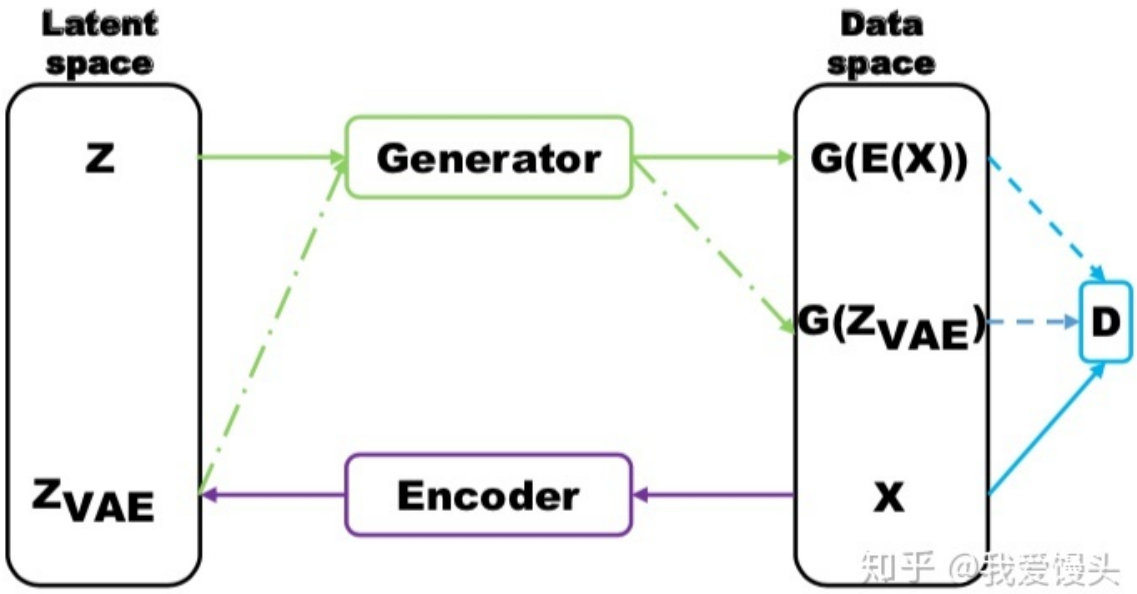


从上面分析可以看出，InfoGAN只是实现了信息的解耦，至于条件变量c每一个值的具体含义是什么，我们无法控制。于是ss-InfoGAN出现了，ss-InfoGAN采用半监督学习方法，把条件变量 c 分成两部分，$c = c_{ss} \bigcap c_{us}$ 。 c_{ss} *则利用标签像CGAN一样学习，* c_{us} 则像InfoGAN一样学习。

GAN与VAE的结合

GAN相比于VAE可以生成清晰的图像，但是却容易出现mode collapse问题。VAE由于鼓励重构所有样本，所以不会出现mode collapse问题。

一个典型结合二者的工作是VAEGAN，结构很像前文提及的MRGAN，具体如下：



上述模型的Loss包括三个部分，分别是判别器某一层特征的重构误差，VAE的Loss，GAN的Loss。

BEGAN

后面根据不同的任务，生成对抗网络架构也有更多的变化。在 17 年提出来的 BEGAN 中，它为判别器加上了一个自编码器。与一般的 GAN 相比，BEGAN 度量生成分布与自编码器损失函数的距离，而不是直接度量生成分布与真实图像分布之间的距离。通过这种修改，模型能帮助生成器生成自编码器容易重建的数据，因此早期训练更加高效。

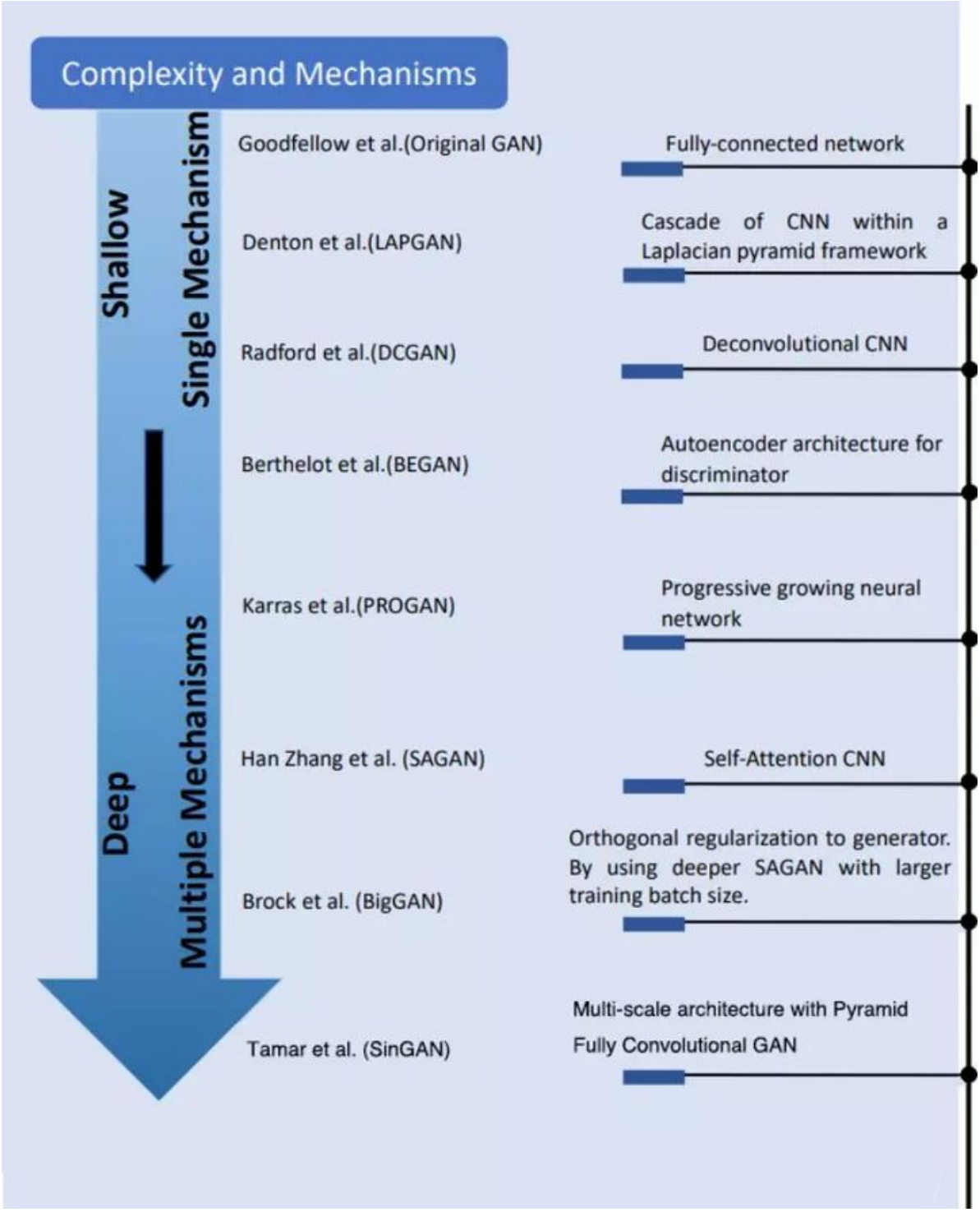
GAN的发展过程

- 出现在2014：Goodfellow et al. [27] proposed it in 2014

BigGANs[38]	2018	1. Generate high-resolution, diverse samples from complex datasets.	Resolution 512×512
StyleGAN[39]	2019	1. Enable intuitive, scale-specific control of the synthesis.	Resolution 1024x1024

- Han Zhang, Ian Goodfellow, Dimitris Metaxas, Augustus Odena提出了自我注意生成对抗网络（SAGAN）(2018)： Self-Attention Generative Adversarial Networks
- DeepMind 带来的 BigGAN 创造性的将正交正则化的思想引入 GAN(2018)： Large Scale GAN Training for High Fidelity Natural Image Synthesis

SAGAN [14] 将谱归一化的思想用在判别器，限制判别器的能力。



SAGAN

SAGAN用自注意力机制构建生成器与判别器，能学到生成图像的全局依赖性关系。所有这些创新，都为更真实的图像生成打下了基础。

SAGAN [14] 将谱归一化的思想用在判别器，限制判别器的能力。

BigGAN

它在 SAGAN 的基础上证明，通过增加批量大小和模型复杂度，我们能极大地提升生成的图片质量。从 BigGAN 提出以来，我们看到的生成图片真的能欺骗人类的判断，

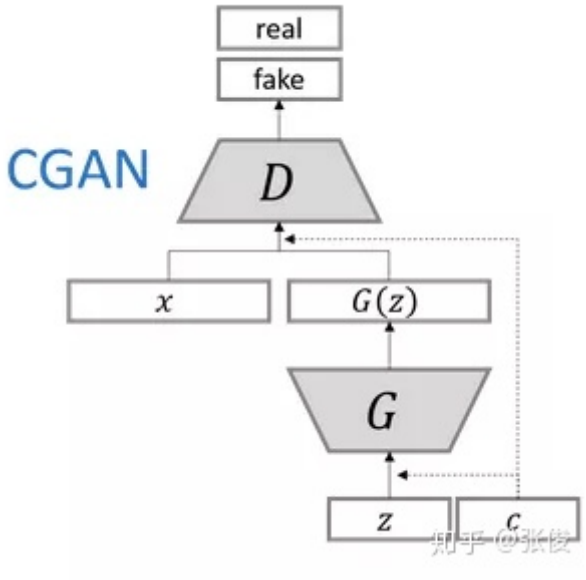
DeepMind 带来的 BigGAN 创造性的将正交正则化的思想引入 GAN(2018

SinGAN：

它提出了一种新的 Unconditional GAN。该模型能捕捉图像块（patch）的内部分布，从而生成高质量、多样化的样本，并承载与训练图像相同的视觉内容。

CGAN / 条件GAN

在原始 GAN 中，无法控制要生成的内容，因为输出仅依赖于随机噪声。我们可以将条件输入 c 添加到随机噪声 Z，以便生成的图像由 G(c,z) 定义。这就是 ***CGAN*** [6]，通常条件输入矢量 c 与噪声矢量 z 直接连接即可，并且将得到的矢量原样作为发生器的输入，就像它在原始 GAN 中一样。条件 c 可以是图像的类，对象的属性或嵌入想要生成的图像的文本描述，甚至是图片。



辅助分类器GAN (ACGAN)

为了提供更多的辅助信息并允许半监督学习，可以向判别器添加额外的辅助分类器，以便在原始任务以及附加任务上优化模型。这种方法的体系结构如下图所示，其中 C 是辅助分类器。

添加辅助分类器允许我们使用预先训练的模型（例如，在 ImageNet 上训练的图像分类器），并且在 ***ACGAN*** [7] 中的实验证明这种方法可以帮助生成更清晰的图像以及减轻模式崩溃问题。使用辅助分类器还可以应用在文本到图像合成和图像到图像的转换。

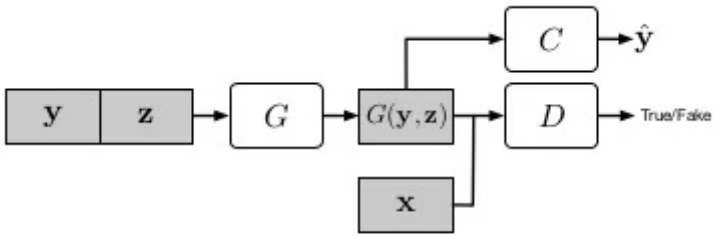
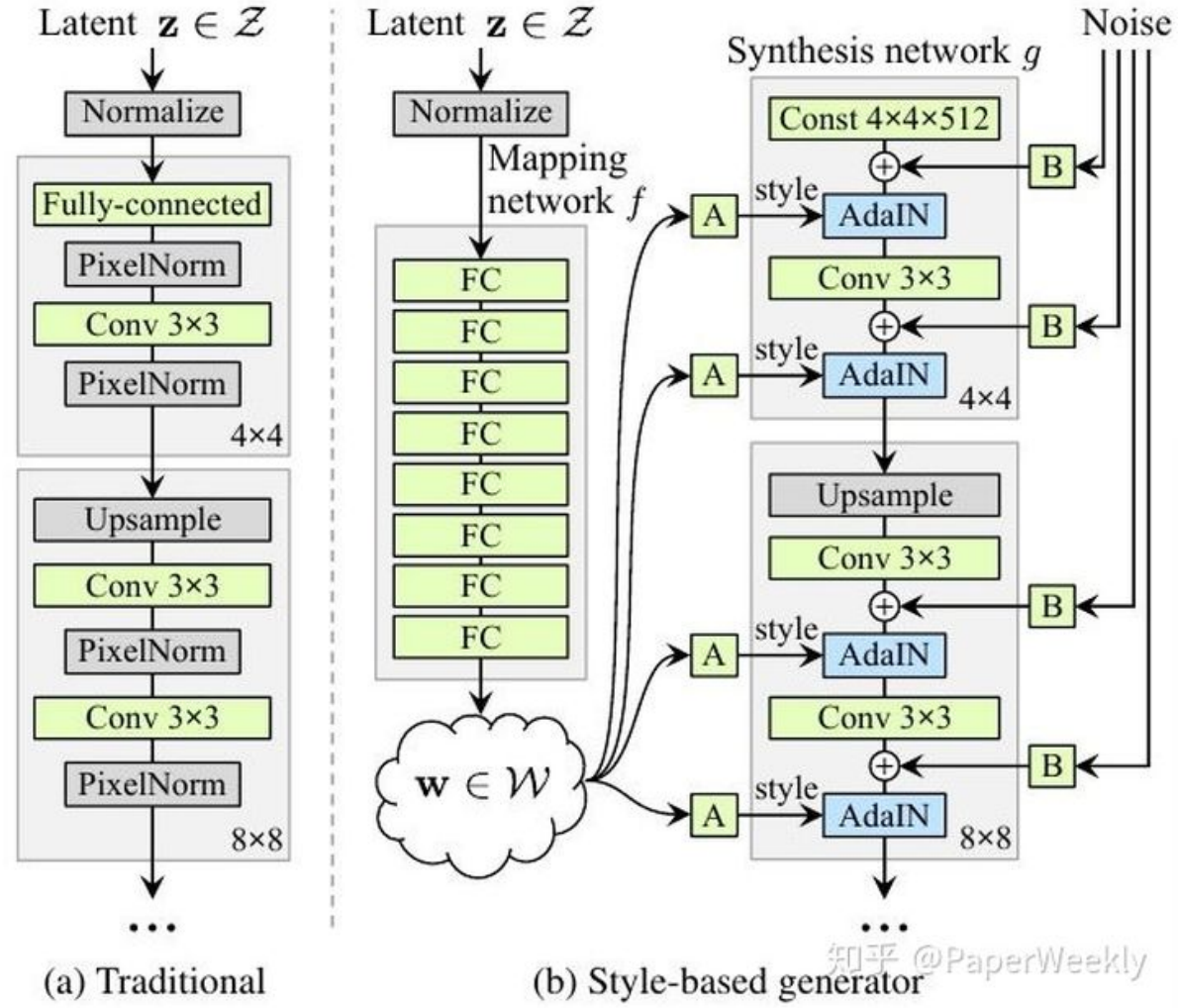


Fig. 2. Architecture of GAN with auxiliary classifier, where $\mathbf{y}$ is the conditional input label and C is the classifier that takes the synthetic image $G(\mathbf{y}, \mathbf{z})$ as input and predict its label $\hat{\mathbf{y}}$

StyleGAN

StyleGAN 已经创建了一匹极其真实的「假人类」。但 StyleGAN 在论文中表明，它并没有构建新方法以稳定 GAN 的训练，也没有引入新的架构，它只是引入了一种生成效果非常好的 GAN。

StyleGAN提出了一个新的生成器框架，号称能够控制所生成图像的高层级属性，如发型、雀斑等；并且生成的图像在一些评价标准上得分更好，作为无监督学习的一种GAN，它能够从噪声生成高分辨率的高清图像。具体的算法框架原理图如下所示：



StyleGAN中生成的图像清晰，细节丰富，当然训练的成本也是很高的。目标函数为

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G)$$

Pix2Pix [29] 提出将 CGAN 的损失与 L1 正则化损失相结合，使得生成器不仅被训练以欺骗判别器而且还生成尽可能接近真实标注的图像，使用 L1 而不是 L2 的原因是 L1 产生较少的模糊图像。

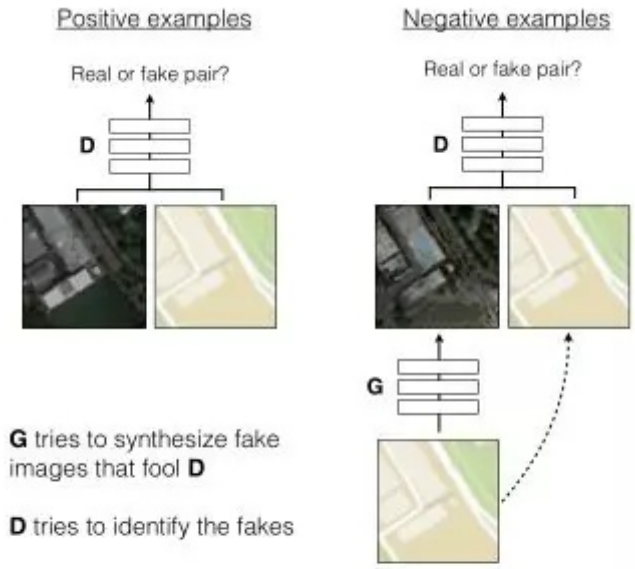


Figure 2: Training a conditional GAN to predict aerial photos from maps. The discriminator, D, learns to classify between real and synthesized pairs. The generator learns to fool the discriminator. Unlike an unconditional GAN, both the generator and discriminator observe an input image.

有条件的 GAN 损失定义为：

$$\mathcal{L}_{CGAN}(G,D) = \mathbb{E}_{x,y\sim p_{data}(x,y)}[logD(x,y)] + \mathbb{E}_{x\sim p_{data}z\sim p_z(z)}[log(1-D(x,G(x,z)))]$$

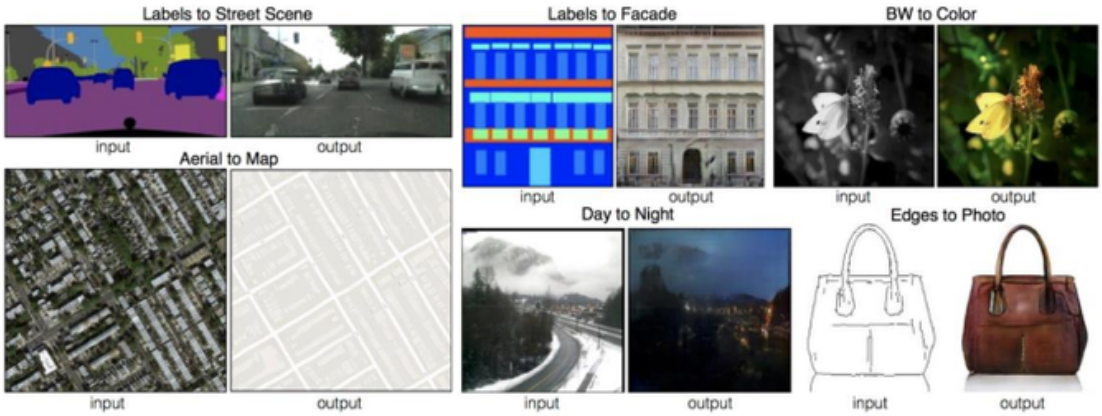
约束自相似性的 L1 损失定义为：

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x\sim p_{data}z\sim p_z(z)}[\|y - G(x,z)\|_1]$$

总的损失为：

$$G^*, D^* = arg\min_G \max_D \mathcal{L}_{CGAN} + \lambda\mathcal{L}_{L1}(G)$$

其中 λ 是一个超参数来平衡两个损失项，Pix2Pix 的生成器结构基于 UNet ，它属于编码器 - 解码器框架，但增加了从编码器到解码器的跳过连接，以便绕过共享诸如对象边缘之类的低级信息的瓶颈。

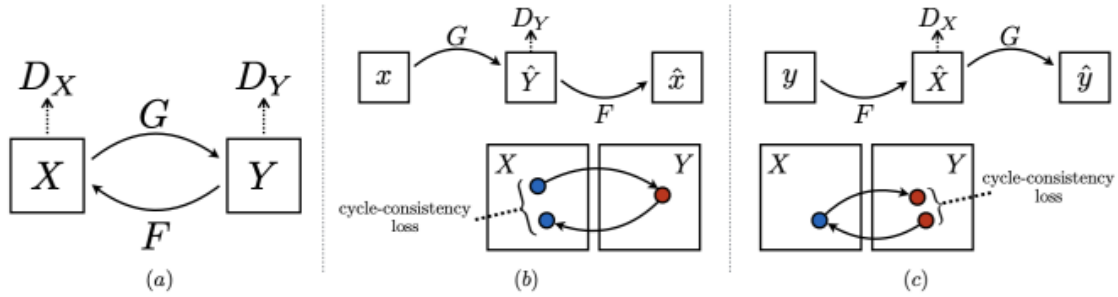


（2）无监督的图像到图像转换（unpaired image translation）

虽然有监督下图像转换可以得到很好的效果，但需要的条件信息以及paired image成为其很大的限制。但如果用无监督学习，学习到的网络可能会把相同的输入映射成不同的输出，这就意味着，我们输入任意 x_i 并不能得到想要的输出 y_i 。

CycleGAN [24]、DualGAN [25] 和DiscoGAN [26]突破了这个限制，这几项工作都提出了一致/重构损失（consistent loss），采取了一个直观的思想：即生成的图像再用逆映射生成回去应该与输入的图像尽可能接近。在转换中使用两个生成器和两个判别器，两个生成器 G_{XY} 和 G_{YX} 进行相反的转换，试图在转换周期后保留输入图像。

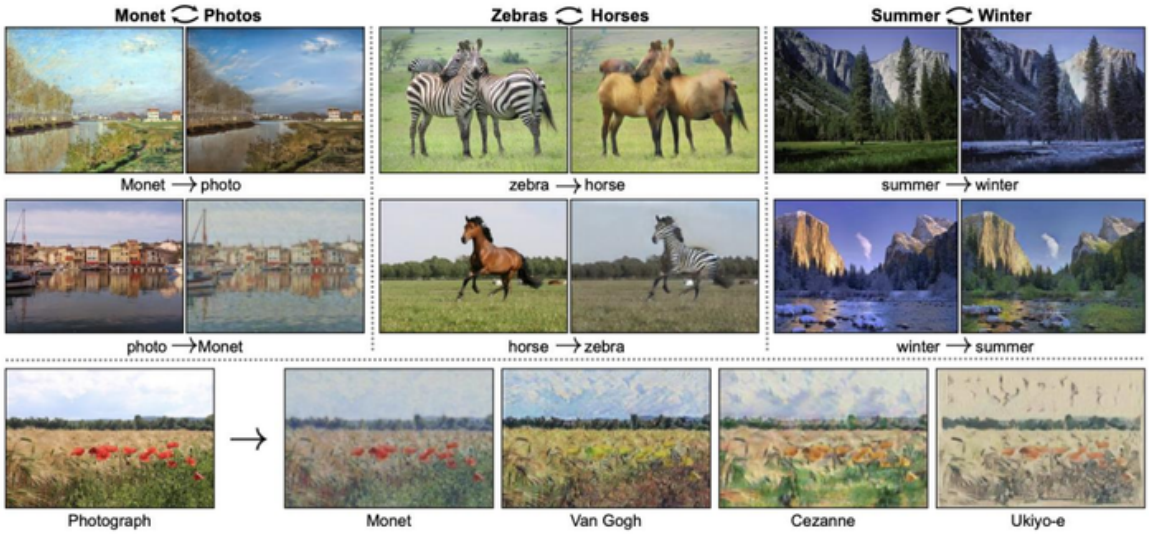
以CycleGAN为例，在CycleGAN中，有两个生成器， G_{XY} 用于将图像从域X传输到Y，G_{XY} 用于执行相反的转换。此外，还有两个判别器D_X 和 D_Y 判断图像是否属于该域。



图九：cycleGAN结构

其Consistent loss由L1进行描述：

$$\begin{aligned}\mathcal{L}_{\text{cyc}}(G, F) = & \mathbb{E}_{x \sim p_{\text{ data }}(x)}\left[\|F(G(x))-x\|_1\right] \\ & +\mathbb{E}_{y \sim p_{\text{ data }}(y)}\left[\|G(F(y))-y\|_1\right]\end{aligned}$$



图十：CycleGAN的生成效果

当然，尽管CycleGAN 和DualGAN具有相同的模型结构，但它们对生成器使用不同的实现。CycleGAN使用卷积架构的生成器结构，而DualGAN遵循U-Net结构。在实践中可以根据不同的需求更换生成器和判别器的实现结构。

OPEN QUESTIONS

such as **mode collapse**, **unstable training problem**, and **vanishing gradient problem**

In addition, GANs are also faced with the problem of **non-convergence** and **sensitivity to hyperparameters**.

On the other hand, **how to evaluate the quality of generated images** still lacks effective means.

At present, these problems remain to be solved, which need continuous research and efforts from the researchers, and many improved GAN-variants have emerged, including Least Square GAN (LSGAN) [101], Wasserstein GAN (WGAN) [102], WGAN-GP [103], and Spectral Normalized GANs (SNGAN) [104]. These models not only greatly improved the quality and the stability of GANs, but also make it easy to converge and aim to solve the problem of unstable training

However, the problem of collapse during training and the mode collapse have not been completely solved due to the high dimensional characteristics of image data. By using maximum likelihood pre-training, with the help of adversarial fine-tuning is now an effective solution to deal with mode collapse.

Many earlier studies on image synthesis based on GANs only used subjective visual assessments. Although it is very hard to quantify the quality of generated images, some studies of evaluating the GANs have begun to appear. For example, the Inception score (IS) [108] and Fréchet Inception Distance (FID) [109] are the most widely adopted evaluation metrics for quantitatively evaluating generated images

Future

1. video processing, the research on video using GANs is limited.
2. the generation and synthesis of 3D models, such as 3D colorization [124], 3D face reconstruction [125], [126], 3D character animation [127], and 3D textured object generation
3. reduce the use of data
4. data augmentation数据增强, due to its ability to synthesize high-quality images, especially in areas with data paucity, such as medical image analysis生成训练集，特别是对于数据珍贵难以获得的如医学领域
5. 模块化、游戏领域

[]https://sinpycn.github.io/2017/04/29/GAN-Tutorial-How-do-generative-models-work.html

[]https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9043519&tag=1

[]https://zhuanlan.zhihu.com/p/261871560

https://zhuanlan.zhihu.com/p/89887433