

生成对抗网络改进与应用简介

条件GAN

深度卷积GAN

InfoGAN

图像超分辨

由文字生成图像

AI学习与实践平台



www.sigai.cn

生成对抗网络是一种通用的框架，自出现之后，出现了大量的改进方案与应用

典型的改进点

网络结构，使用不同的神经网络

目标函数，构造新的目标函数

实际应用

图像和其他数据生成

图像超分辨

由文字生成图像

条件生成对抗网络，简称CGAN

生成模型的输入数据除了随机噪声之外还有人工控制的变量，通过控制这个变量可以生成不同类型的数据

生成模型除了接受随机噪声作为输入之外还加上了一个输入变量 y ，判别模型的输入中也加上这个变量
变量 y 称为条件变量，它是类别标签或其他信息

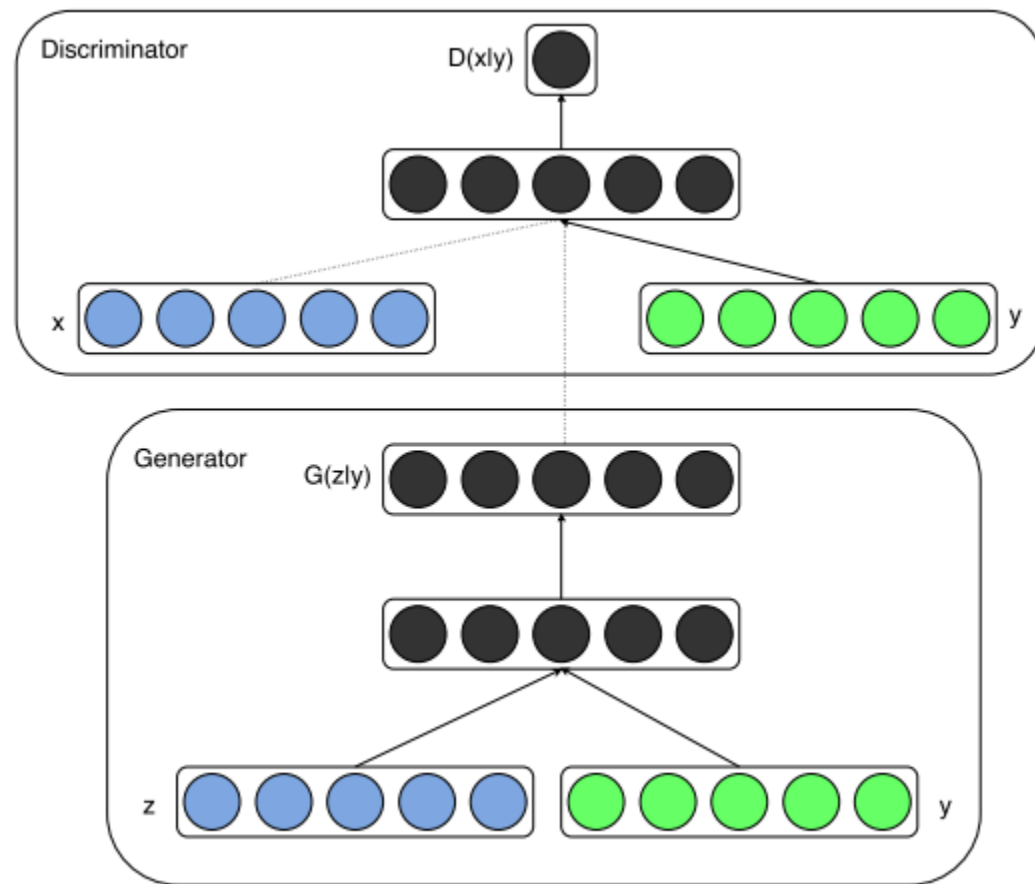
优化目标函数

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} (\log D(x|y)) + \mathbb{E}_{z \sim p_z(z)} (\log(1 - D(G(z|y))))$$

实现时，如果条件变量作为类别变量，则可以采用one-hot向量编码方式

在生成模型中向量 y 和随机噪声向量 z 串联输入到神经网络中。判别模型将 x 和 y 合并起来送入神经网络进行训练，
这里是一个多分类问题

Mirza M, Osindero S. Conditional Generative Adversarial Nets. Computer Science, 2672-2680, 2014



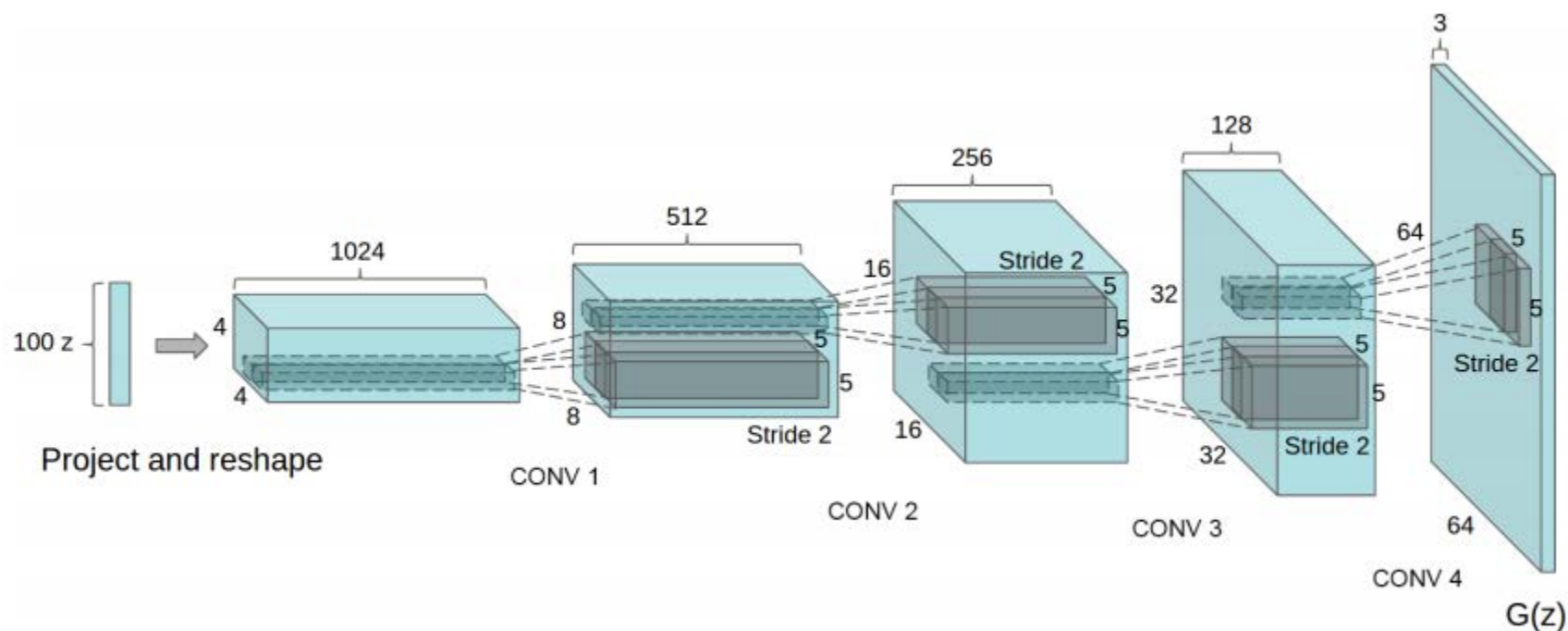
深度卷积生成对抗网络，简称DCGAN

用生成对抗网络生成尺寸较大的图像时会有模糊的问题，其中一个原因是生成模型的网络过于简单

DCGAN用深度卷积神经网络作为生成模型，以随机噪声向量作为输入，输入向量通过一个反卷积网络映射为二维的输出图像

生成网络的输入是均匀分布随机向量，被神经网络映射成三维图像，接下来是反卷积层，最后输出RGB图像，这个网络没有池化层和全连接层

判别网络结构和生成网络相反



Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434, 2015

InfoGAN

引入了隐变量以对生成的样本加入语义控制信息
系统由3个网络组成，包括1个生成网络两个判别网络

生成网络的映射为

$$\mathbf{x} = G(\mathbf{z}, \mathbf{c})$$

第一个判别网络的映射为

$$y_1 = D_1(\mathbf{x})$$

它用于区分样本是真实的还是生成器生成的
第二个判别网络的映射为

$$y_2 = D_2(\mathbf{x})$$

它用于判断样本的类别。除了最后一层之外，两个判别网络共享参数

Chen X, Duan Y, Houthoofd R, et al. InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets[J]. arXiv preprint arXiv:1606.03657, 2016

损失函数通过互信息构造

互信息是信息论中的一个概念，用来衡量两个随机变量的相互依赖程度

对于两个离散型随机变量，互信息定义为

$$I(x, y) = \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$

对于两个连续型随机变量，互信息定义为

$$I(x, y) = \int_x \int_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy$$

如果两个随机变量相互独立，互信息为0。如果它们存在确定的、可逆的函数关系，则二者的互信息有最大值

基于互信息和隐含变量，InfoGAN构造的损失函数为

$$\min_G \max_D V_I(D, G) = V(D, G) - \lambda I(\mathbf{c}; G(\mathbf{z}; \mathbf{c}))$$

函数的第一项为生成对抗网络的标准目标函数。函数的第二项可以看做是正则化项，用互信息进行惩罚

图像超分辨

由低分辨率的图像得到高分辨率的图像

超分辨生成对抗网络，简称SRGAN

使用了一种新的损失函数，由对抗损失和内容损失两部分构成。第一部分损失和标准生成对抗框架相同，通过一个判别模型，让生成网络生成的超分辨图像和真实高分辨率图像尽可能接近

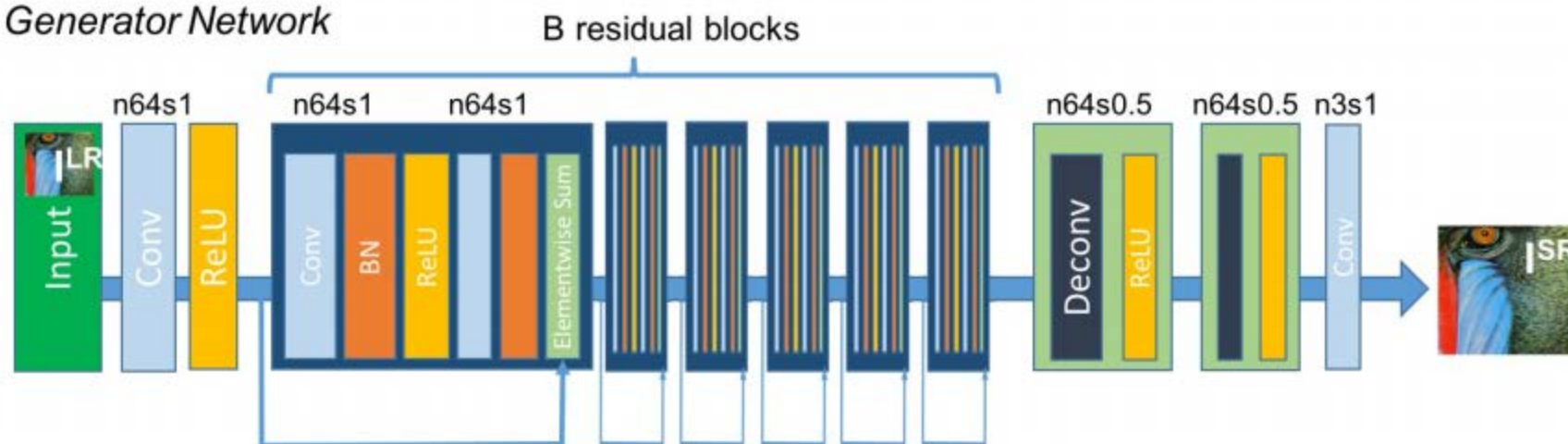
生成器卷积网络采用深度残差网络；判别模型也是一个层次很深的卷积网络，用于区分一张图像是真实的高分辨率图像还是由生成器网络生成的图像

假设低分辨率图像为 \mathbf{I}^{LR} ，目标是根据它估计出高分辨率的图像 \mathbf{I}^{SR} ，，在训练样本中与低分辨率图像相对应的真实高分辨率图像为 \mathbf{I}^{HR}

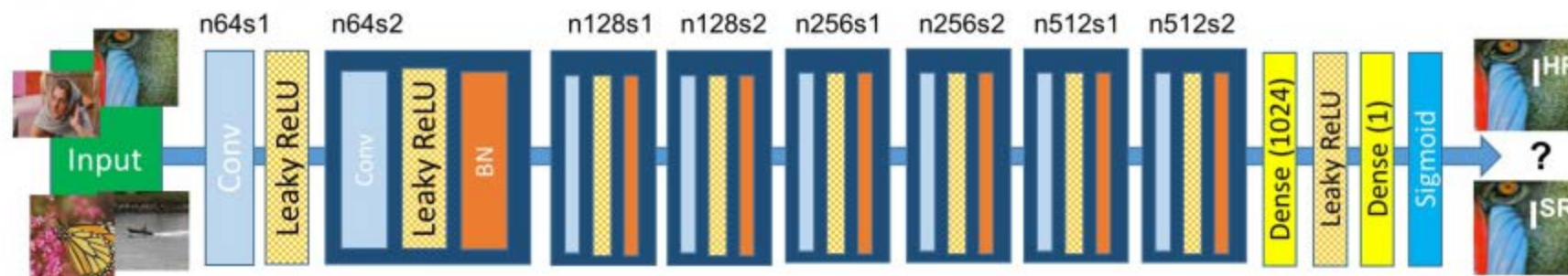
在训练时，低分辨率图像通过对高分辨率图像进行高斯平滑滤波然后下采样得到

Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, Wenzhe Shi. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. Computer Vision and Pattern Recognition, 2016

Generator Network



Discriminator Network



图像超分辨由生成器卷积网络 G_{θ_G} 完成，给定训练样本集 $(\mathbf{I}_n^{HR}, \mathbf{I}_n^{LR})$ ，生成器网络的训练目标是求解如下最优化问题

$$\min_{\theta_G} \frac{1}{N} \sum_{n=1}^N l^{SR} \left(G_{\theta_G} \left(\mathbf{I}_n^{LR} \right), \mathbf{I}_n^{HR} \right)$$

损失函数由多个部分加权和组成。。第一部分为内容损失，为逐像素的均方误差损失函数，称为VGG损失，定义为

$$l_{VGG/ij}^{SR} = \frac{1}{W_{ij} H_{ij}} \sum_{x=1}^{W_{ij}} \sum_{y=1}^{H_{ij}} \left(\phi_{ij} \left(\mathbf{I}^{HR} \right)_{xy} - \phi_{ij} \left(G_{\theta_G} \left(\mathbf{I}^{LR} \right) \right)_{xy} \right)^2$$

第二部分为对抗损失，它用判别模型输出的概率值构造

$$l_{Gen}^{SR} = \sum_{n=1}^N -\log D_{\theta_D} \left(G_{\theta_G} \left(\mathbf{I}^{LR} \right) \right)$$

训练目标是如下最优化问题

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{\mathbf{I}^{HR} \sim p_{train}(\mathbf{I}^{HR})} \left(\log D_{\theta_D} \left(\mathbf{I}^{HR} \right) \right) + \\ \mathbb{E}_{\mathbf{I}^{LR} \sim p_G(\mathbf{I}^{LR})} \left(\log \left(1 - D_{\theta_D} \left(G_{\theta_G} \left(\mathbf{I}^{LR} \right) \right) \right) \right)$$

由文字生成图像

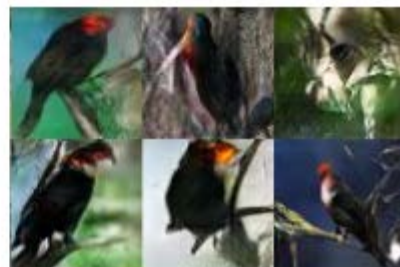
采用编码器-解码器思想

第一步是将一段文字转换成向量表示，即文字的语义信息。第二步在第一步生成的文本向量基础上训练一个生成对抗网

this small bird has a pink breast and crown, and black primaries and secondaries.



this magnificent fellow is almost all black with a red crest, and white cheek patch.



the flower has petals that are bright pinkish purple with white stigma



this white and yellow flower have thin white petals and a round yellow stamen



S Reed, Zeynep Akata, Xincheng Yan, Lajanugen Logeswaran, Bernt Schiele, Honglak Lee. Generative Adversarial Text to Image Synthesis. international conference on machine learning, 2016

通过一个深度卷积网络和一个循环神经网络为文本产生向量输出，这个输出和图像对应

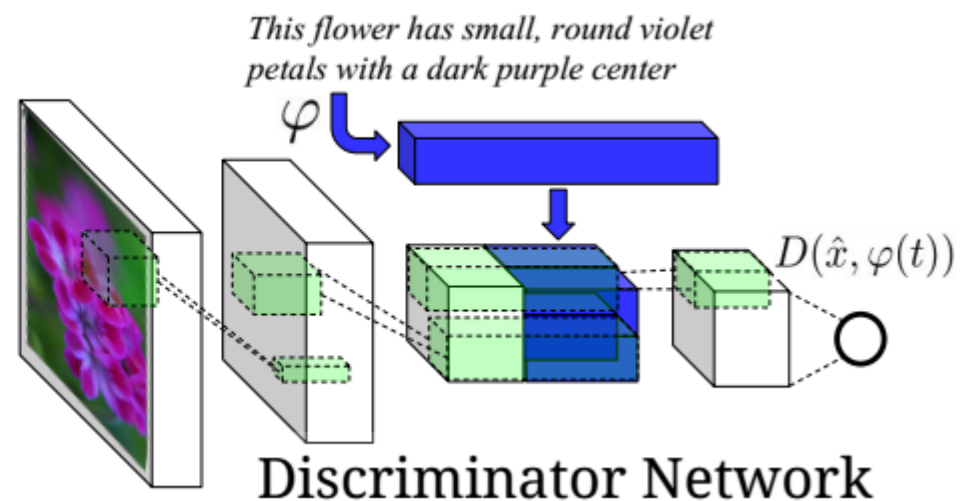
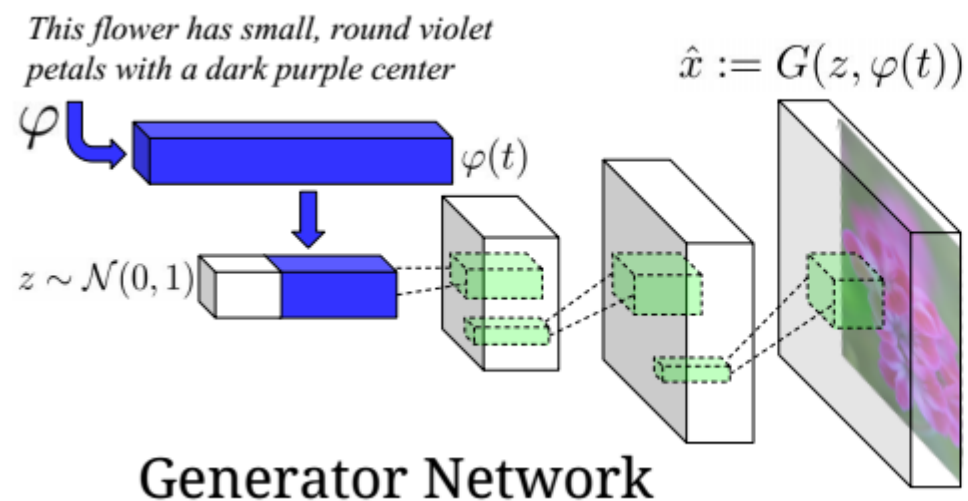
生成模型是一个深度卷积网络，负责生成图像。生成器网络实现的映射为

$$\mathbb{R}^Z \times \mathbb{R}^T \rightarrow \mathbb{R}^D$$

生成网络接受随机噪声向量和文字的特征向量作为输入，输出指定大小的图像
判别器实现的映射为

$$\mathbb{R}^D \times \mathbb{R}^T \rightarrow \{0,1\}$$

其中T为文字描述向量化后的向量维数，Z为随机噪声的维数，D是生成的图像的维数



AI学习与实践平台



www.sigai.cn