

GAN中的图像到图像生成模型综述

周宇梁
21821351
浙江大学
yuliang.zhou@hotmail.com

摘要

图像到图像转换是计算机视觉中的一类重要问题，包括了内容的转换和风格的转换，而Conditional GAN则是解决这类问题常用的模型。本文按照训练数据分类，分别介绍了有监督的配对图像转换模型和无监督的非配对图像转换模型，并对他们的优缺点进行了阐述。

1. 引言

计算机视觉中的许多问题都包含了将图像从一个域转换到另一个域，包括超分辨率，图像上色，图像填充，风格转换，内容转换等等，这些模型通常采用了Conditional GAN [1]架构，根据训练集的不同，还可以分为有监督的配对图像转换和无监督的非配对图像转换。

使用GAN [2]进行训练也存在着一些缺点，比如训练不稳定，模式崩溃，生成细节缺失等问题，但这并不妨碍人们继续对GAN的探索，不断提升GAN的生成效果。

2. Conditional GAN

Conditional GAN是在原始GAN基础上的一个改进版本，分别在生成器和判别器中加入了条件信息来进行训练，网络结构如图 1 所示。

传统GAN的目标函数是：

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

Conditional GAN加入了条件y作为联合输入：

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

虽然改动非常简单，但是效果很好，非常具有启发性。模型可以通过修改条件内容来控制生成结果，在生成手写数字，识别图像内容方面都取得了不错的效果。

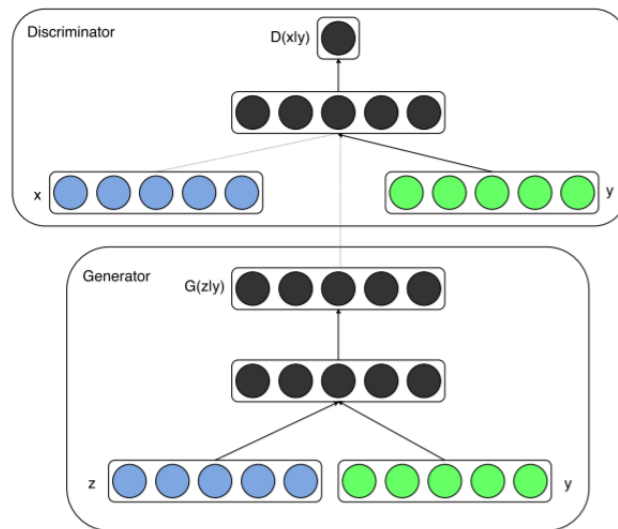


图 1: Conditional GAN结构

3. 有监督的配对图像转换

有监督的配对图像转换，即输入图像和生成图像为一一对应关系，由于手工标注太过复杂，通常采用边缘提取或滤波算法得到训练数据，并进行训练，所以输入图像和生成图像往往共享一些形状特征。而对于类别或风格差异较大的图像，如从苹果图像转换到橘子图像，或是将照片转换为油画，则很难找到配对图像。

3.1. pix2pix

pix2pix [3]采用了典型的Conditional GAN结构，并进行了一些改进，提高了生成效果。如图 2 所示，判别器将两张图像作为联合输入进行真假判别。

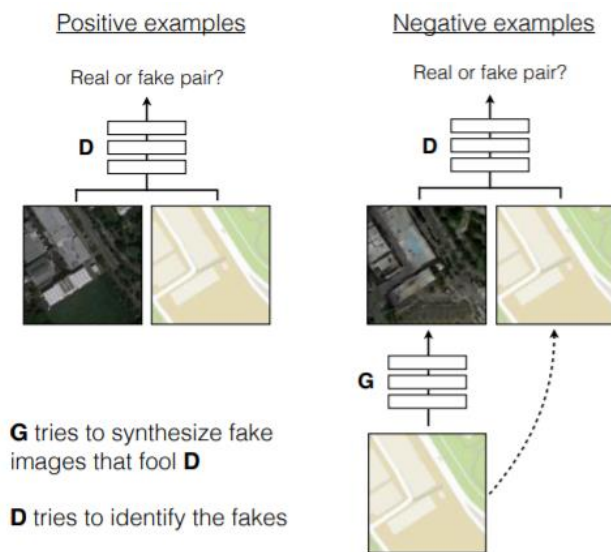


图 2: pix2pix的判别器和生成器示意图

在模型结构方面，pix2pix的生成器采用了带有skip的U-Net结构，将decoder和encoder连接起来，因为生成图像和输入图像通常会共享一些如边缘之类的低频特征，这样做是为了方便共享一些低频信息，如图3右侧所示。

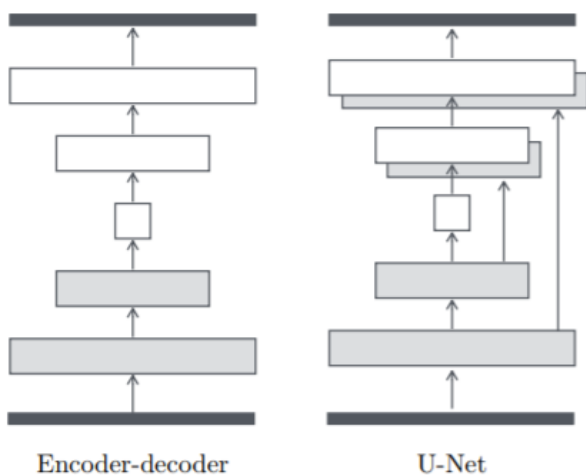


图 3: pix2pix的生成器结构

有研究发现对GAN目标函数加上一些传统的损失函数很有效，pix2pix采用了L1 distance而不是L2 distance来减少了模糊：

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

最终得到的目标函数：

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

此外，还采用了PatchGAN来判断图像是否为真，即将图像分成N×N块，在每一块上进行判别，并将结果进行平均得到最终输出。

3.2. PI-REC

PI-REC [4]是一种将边缘信息和色块图像转换为高清图像的模型，采用了残差块作为模型结构，在数据预处理和模型训练方面提出了一些创新。

提取边缘图像和色块需要一些参数，为了提升模型的鲁棒性和生成效果，分别采用了多种不同的参数组合提取得到训练图像，称之为超参数混淆（hyperparameter confusion），实际起到了数据增强的作用，图4对比了采用超参数混淆和不采用的效果。

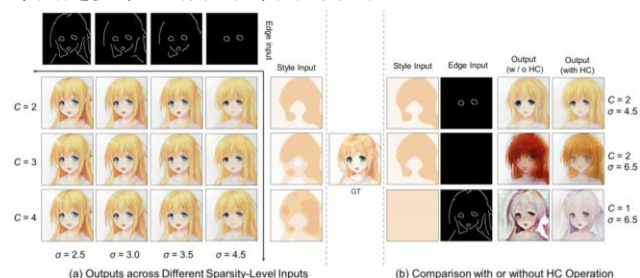


图 4: 超参数混淆的效果

绘画的时候，我们一般会先勾勒出轮廓，然后再一步步细化、上色，最终完成画作。受绘画过程的启发，PI-REC的训练过程分为3个阶段：模拟阶段，生成阶段和微调阶段。三个阶段训练同一个生成器和判别器，从而加速了训练过程。模拟阶段输入边缘图像和被部分遮挡（遮挡面积小于70%）的图像进行训练，模型收敛后进入生成阶段，使用边缘图像和颜色域图像进行训练，再次收敛后进入微调阶段，使用边缘图像和第二阶段生成器生成的图像进行训练，得到最终的模型，训练过程如图5所示。

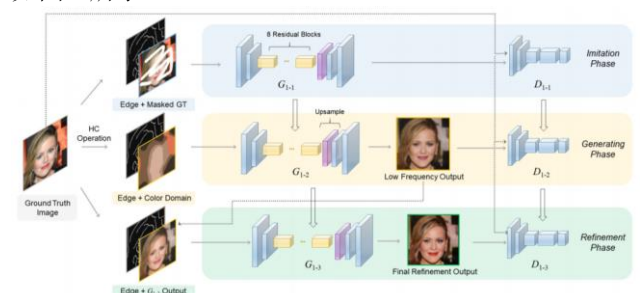


图 5: PI-REC的训练过程

4. 无监督的非配对图像转换

与配对图像转换相比，无监督的非配对图像训练对数据的要求更少，训练数据也更容易获取。通常非配对图像对两个不同类的图像进行训练，可以是风格不同，也可以是内容不同，对图像的限制较少。

4.1. CycleGAN

CycleGAN [5]与DualGAN [6]和DiscoGAN [7]都是同一时期独立发表的三篇论文，想法却惊人的相似，提出了循环损失函数这一概念，即衡量图像从源域X转换到目标域Y，再转换回源域X后的损失，如图6所示，这样做可以防止模式坍塌，即所有图像都转换为同一张图像，保证了两个域图像之间的一一对应关系。

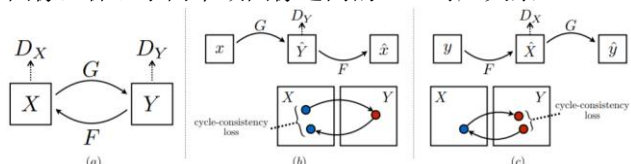


图6: CycleGAN的循环一致性损失函数

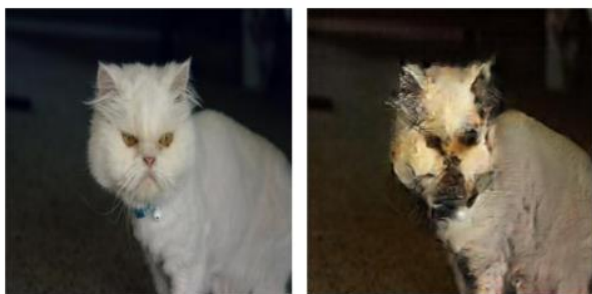
模型的损失函数由GAN目标函数和循环一致性损失函数组成，其中GAN目标函数包含了两部分，即从源域X到目标域Y的生成器G，和从目标域Y到源域X的生成器F：

$$\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)} [\log D_Y(y)] + \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log(1 - D_Y(G(x)))]$$

循环一致性损失函数同样也包含了两部分：

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

CycleGAN的原理很简单，效果也很好，但是仍存在一些局限，尤其是在几何变换方面。对于形状不同的对象，如图7所示的狗→猫的任务中，猫的颜色和纹理发生了变化，而轮廓则没有发生改变。这些不仅仅是CycleGAN所面临的问题，同样也是大部分图像到图像转换任务所面临的。



cat → dog

图7: 从猫转换到狗

- [3] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1125-1134.
- [4] You S, You N, Pan M. PI-REC: Progressive Image Reconstruction Network With Edge and Color Domain[J]. arXiv preprint arXiv:1903.10146, 2019.
- [5] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2223-2232.
- [6] Yi Z, Zhang H, Tan P, et al. Dualgan: Unsupervised dual learning for image-to-image translation[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2849-2857.
- [7] Kim T, Cha M, Kim H, et al. Learning to discover cross-domain relations with generative adversarial networks[C]//Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017: 1857-1865.

Reference

- [1] Mirza M, Osindero S. Conditional generative adversarial nets[J]. arXiv preprint arXiv:1411.1784, 2014.
- [2] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in neural information processing systems. 2014: 2672-2680.