Repurposing GANs for One-shot Semantic Part Segmentation

这篇文章是在计算机视觉的中应用 GAN 网络的又一个典型案例。Generative Adversarial Nets（下面简称 GAN） 是 Ian J. Goodfellow 在 2014 年发表一篇关于对抗生成网络的论文，是对抗生成网络的开山之作，在机器学习领域产生了很大的影响，同时也吸引更多人投身在 GAN 这个领域中，引发了新的有关对抗生成网络的热潮。GAN 也是机器学习中的一个非常典型的代表，如今已经衍生基于 GAN 的更多模型，如 CGAN（Conditional GAN）、InfoGAN（Information GAN）等模型。

GAN 的主要应用目标：生成式任务（生成、重建、超分辨率、风格迁移、补全、上采样等）

GAN 的核心思想：让生成器 G(generator)和判别器 D(Discriminator)的进行一次次迭代，最终获得一个性能较好的生成器或者判别器

生成器：生成网络，通过输入生成的数据，如图像、音频等

判别器：二分类网络，将生成器生成的数据作为负样本，真实数据作为正样本

因此，GAN 的最终目的是通过训练这两个模型，让生成器掌握真实数据集的分布的情况，也就是提取出原数据分布的情况，最终让这些信息存储在模型中，并且能够产生与原数据分布相同的假数据。

首先，会先定义一个生成器模型 $G(z;\theta_g)$，在自己的分布 p_g 上生成一个假的数据。为了让生成器 G 学到数据 *data* x 的分布情况，就先确定一个先验的输入噪声变量 $\mathrm{p}_z(z)$；然后用 $G(z;\theta_g)$ 代表到真实数据空间的一个映射函数，即通过一个输入 $\mathrm{p}_z(z)$ 获得一个假的数据。（其中 θ_g 是 G 模型中的参数，一般使用深度神经网络等形式进行训练）

同时，也定义一个判别器 $D(x;\theta_\mathrm{d})$，输出一个标量。$D(x)$ 代表了 x 来自于真实数据分布的可能性，而不是来自假数据分布 p_g。

即 GAN 的核心为如下：

用一个价值函数来评估对目标分布的掌握情况

$$V(D,G) = E_{x\sim p_{data}(x)}[\log\ D(x)] + E_{z\sim p_z(z)}[\log(\ 1 - D(G(z)))]$$

因此，我们训练判别器D。在给定模型G不变的情况下，通过G生成数据产生负样本p_g（即判定概率为0），并结合真实图像P_{data}作为正样本（即判定概率为1）。通过这样的一组样本$(x_i, p(x \in data))$来训练D。通过反向传播以及其他方式更新D中的参数θ_d，让对真数据的打分尽可能高，对假数据的打分尽可能低，最终找到合适的参数θ_d使得价值函数$V(D,G)$的取值最大。

同理，在训练生成器G时。给定模型D不变的情况下，以使得D对G生成的数据样本的评分尽可能高，最终接近G对于正样本的打分，即让价值函数$V(D,G)$的取值最小。

G和D的训练过程交替进行，这个对抗的过程使得G生成的数据越来越逼真，即越来越贴近真实的分布情况；同时D“打假”的能力也越来越强。

最终体现在数学公式的形式如下：

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

前面部分是对正样本的预测值的期望平均，后半部分是对负样本的否定期望平均，都是对D的性能评估。

上面是GAN的开山之作的核心思想，并且看过几篇GAN在人工智能安全领域的应用，如对抗样本生成、成员性质推理攻击方面。GAN在此类重建和生成任务中的性能确实是很强大。

这篇文章利用GAN模型在生成图形方面的卓越性能，借用三个网络，GAN网络生成同分布的图像，然后利用卷积网络CNN或多层感知网络MLP对GAN生成的图像带有注释进行分割，最后使用GAN的图像和分割结果对自动单镜头分割网络进行训练。最终实现了在成本较低的条件下，对少照片甚至单照片的情况，实现了较好的部分语义分割。

语义分割在计算机视觉中的应用是非常重要，尤其是在图像的初步处理中，获得下一步要处理的部分图像。而且，在日常应用的领域中，少图像甚

至是单图像的情况也是很常见，例如，在刷脸门禁中，正常照片的采集为了获取有足够信息的部分也为了减小后续处理中模型的开销，会对采集的照片进行初步分割，选取里面有用的部分。而且采集的照片不会与模型训练中几百上千张图像的训练集一样，受限于终端设备和服务器性能的限制，单人采集的图像仅有几张，甚至在极端情况下只有一张，这也是这篇文章和其他类似文章的立足点。