《人工神经网络》大作业开题报告

陶天骅
计算机科学与技术系
清华大学
tth17@mails.tsinghua.edu.cn

杨雅儒
计算机科学与技术系
清华大学
yangyr17@mails.tinghua.edu.cn

1 任务定义

本课题希望构建一个神经网络以及一些简单的界面，可以根据用户提供的一些特征的比例，自动生成一张风景图片。程序界面示意图如下。

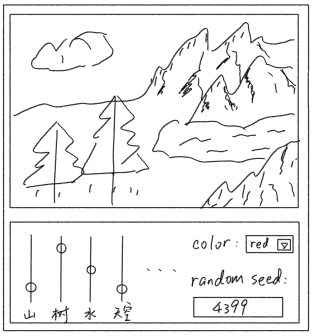

图 1: A Simple Sketch.

用户通过滑动滑条，确定例如树木、山、水、天空等要素在图片中的占比，并可以设定希望使用的主题颜色，程序便生成一张符合以上要求的风景图片。用户可以设置随机种子来获得不同的图片。

用数学语言形式化程序的任务即为：

令 pic 是一张图片，$label_{pic}$ 是图片的标签，值为 $real$ 或 $generated$ ，表示是真实的或者生成的。需要构建一个辨别器 D ，参数是 θ_D，$D(pic;\theta_D)$ 输出一个标签，表示辨别器认为 pic 是否是生成的。训练目标为：

$$\arg\max_{\theta_D}(Prob(D(pic;\theta_D) = label_{pic}))$$

即 D 能给出正确的标签。

再构建生成器 G，参数是 θ_G，对于一个特征向量 x，$G(x;\theta_G)$ 输出一张图片。训练目标为:

$$\arg\min_{\theta_G}(Prob(D(G(x;\theta_G);\theta_D) = generated))$$

即 G 的输出应足够逼真，以至于 D 不能分辨图片是否是生成的。

同时为了让 G 能正确地展现 x 的特征，还需要一个编码器 E，$E(pic)$ 输出图片对应的特征向量。E 的训练就是 AutoEncoder 的 $Encoder$ 的训练，而 $decoder$ 的部分就是前面的 G。当 E 是一个理想编码器的时候，G 的训练目标还要包括:

$$\arg\max_{\theta_G}(Prob(E(G(x;\theta_G)) = x))$$

即对 G 在 x 上生成的图片进行编码的话，还能得到 x，这表示生成的图片确实含有 x 的特征。

2 数据集

用于训练的数据集可以是从各大图片社交平台（如 Pinterest 、Flicker）上下载获得的风景图片。

有一些现存的数据集可以使用，但是需要剔除一些内容，包括 MIT 的 Computational Visual Cognition Laboratory，Github 上的 ml5-data-and-models，MIT Computer Science and Artificial Intelligence Laboratory 提供的 places 数据集等。

目前估计图片的大小为 256 x 256，数量在 3000 以上。

3 挑战和参考工作

3.1 挑战

- 确定使用哪些特征作为输入标签。
- 将特征和主题颜色向量化。
- 考虑到算力有限，可能无法生成分辨率较高的图片。
- 融合不同的神经网络架构，打造一个本程序专用的神经网络。

3.2 参考工作

- 参考工作 1.

  Nvidia 的 SPADE 项目构建了一个程序，可以根据景物轮廓绘出风景图像，与本项目有类似之处，但是它使用了不同的原理。Nvidia 的工作构造了一个称为 GauGAN 的网络，在 COCO-Stuff, Cityscapes 和 ADE20K 数据集上学习了图像语义分割，再由特定类别的图像语义生成图像。但是考虑到本项目不会涉及到非常多的语义分割的部分，且 SPADE 项目的规模要大的多，因而参考意义有限。

- 参考工作 2.

  Google 在《A deep-learning photographer capable of creating professional work》中的工作可以根据一张街景照片，生成一张专业级摄影照片。其中的一些技术可以参考，但是主要的流程和目的与本项目不相符。

4 研究计划

1. 收集训练数据

   计划使用手动（打包下载）或者自动（爬虫）的方法，在一些图片平台上获取与风景相关的图片数据。

2. 清洗数据

   对各种图片，删去一些不相关的内容，调整为统一大小，统一格式。

3. 构建基于 CNN、GAN、AutoEncoder 的神经网络

   网络主要含两部分，一部分是 GAN 的部分，目标是使图片显得真实。另一部分是 Encoder 的部分，目的是体现出特征向量中的特征。CNN 集成在两部分网络中间。

4. 构建 GUI

   可以使用 Python Tkinter 或者 Qt 等框架构建简单的 GUI 程序。

5. 训练

   目前的训练设备有 1 个 RTX 2080Ti GPU。在训练的时候不断调整网络架构和参数。

6. 测试

5 可行性

本项目使用到的主要方法为一些十分成熟的神经网络架构，主要难点在于将他们进行融合，并设计本项目专用的网络，因为既能保证可行性，又有创新点。在计算规模上，我们尽量将图片的大小控制在可以在单个 GPU 上训练的规模。

6 参考文献

[1] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. ”Semantic Image Synthesis with Spatially-Adaptive Normalization”, in CVPR, 2019.

[2] Fang, H., Zhang, M. (2017). Creatism: A deep-learning photographer capable of creating professional work. ArXiv, abs/1707.03491.