

基于循环生成对抗网络的图像转换

古 彭*, 刘宝镇*, 戚馨文, 李怀诚#, 刘嘉帆

西藏大学, 西藏 拉萨

Email: #1296907025@qq.com

收稿日期: 2021年7月2日; 录用日期: 2021年7月28日; 发布日期: 2021年8月4日

摘 要

在无监督学习领域中生成对抗网络是近几年发展速度较快的一个研究方向, 其主要特征是用一种概率估计的方式去逼近未知分布的模型。运用这种模型可以避免复杂的运算和编码, 特别是在计算机视觉方面, 可以生成质量较好的图片。本文基于循环生成对抗网络, 训练一个风格迁移的神经网络, 通过输入一张采样图片转化为输出一张与采样图片不同风格的图片。为防止生成器学习到具有欺骗性的虚假数据, 本文采用加入一个新的生成器, 把第一个生成器的输出当作输入再次使用, 使生成器的输出和原图具有较高的相似性, 同时不丢失原图片的特征并且确保输出一个和原始输入相似的图片。实验仿真数据集选取真实校园作为场景, 在训练初期并不能较好的还原回原始图片, 这意味着生成器使用了虚假的输出结果, 当训练次数达到10,000次以上, 结果显示可以较好的还原回原始图片, 证明第一个生成器的输出保留了大量原始图片特征, 输出结果较为可靠。

关键词

无监督学习, 计算机视觉, 循环生成对抗网络

Image Conversion Based on Cyclic Generative Confrontation Network

Peng Gu*, Baozhen Liu*, Xinwen Qi, Huaicheng Li#, Jiafan Liu

Tibet University, Lhasa Tibet

Email: #1296907025@qq.com

Received: Jul. 2nd, 2021; accepted: Jul. 28th, 2021; published: Aug. 4th, 2021

*第一作者。

#通讯作者。

Abstract

In the field of unsupervised learning, generative confrontation network is a research direction that has developed rapidly in recent years. Its main feature is to use a probability estimation method to approximate an unknown distribution model. Using this model can avoid complex calculations and coding, especially in computer vision, can generate better quality pictures. This article is based on the cyclic generative confrontation network, training a style transfer neural network, inputting a picture and outputting a picture of its different styles. In order to prevent the generator from learning deceptive false data, this article adopts adding a new generator, and uses the output of the first generator as input again, so that the output of the generator has a higher similarity with the original image, while not losing the characteristics of the original picture and ensuring that a picture similar to the original input is output. The experimental simulation data set selects the real campus as the scene, and the original picture cannot be restored well in the initial training stage. This means that the generator uses false output results. When the number of training times reaches 10,000 or more, the results show that it can be restored well. Going back to the original picture, it is proved that the output of the first generator retains a large number of original picture features, and the output result is relatively reliable.

Keywords

Unsupervised Learning, Computer Vision, Cyclic Generative Confrontation Network

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 研究背景与内容

随着计算机图形、图像处理和计算机视觉领域的发展[1], 越来越多的问题都可以转换为输入图像和输出图像的形式。很多实际场景都可以用红绿蓝三原色, 梯度域, 边缘图来渲染。相比于自然语言自动翻译, 图像到图像的转换问题可以定义为: 在给定足够图像数据训练的条件下, 将一种存在的场景转换为具有多种可能性的场景。计算机视觉和计算机图形学面临着多对一和一对多的映射问题。本质上的任务是从一个像素点预测另一个像素点。

卷积神经网络已逐渐成为解决大多数图像预测问题的重要手段。卷积神经网络是一种通过学习使损失函数最小化的方法[2], 损失函数的目标是评价结果的质量。虽然计算机在学习过程中是自我学习的, 但它仍然需要投入大量的人力和计算能力来设计和计算损失函数。换句话说, 最重要的目标就是使卷积神经网络明确最小化的方向。然而在不成熟方面, 需要使用卷积神经网络来最小化预测图像与真实图像之间的欧氏距离, 最终这种方法会使结果模糊。原因在于欧式距离是通过将所有输出平均来最小化, 产生模糊结果显而易见。为了使卷积神经网络输出的结果更加真实, 损失函数的设计变得越来越重要。本文使用循环生成对抗网络以学校的教学楼图片为训练图片训练网络参数, 最终达到生成图片在整体视觉效果上和原图片相似, 但在细节方面有一些区别。

2. GAN 网络的基本框架

生成对抗网络是通过对抗过程来估计生成模型的新框架, 并使用随机梯度下降算法实现优化。这避

免了反复应用马尔可夫链学习机制带来的分配函数计算[3], 降低了运算的复杂度。生成对抗网络需要同时训练两个模型: 生成模型 G 寻找合理生成分布和鉴别器模型 D 来估计训练数据不生成模型数据的概率[4]。生成器网络模型可以学习真实样本的数据分布, 生成新的数据样本。鉴别器网络模型是一种通用的二分类器。鉴别器网络模型的输入是真实数据和生成的样本数据, 然后判断数据的来源, 是真实数据还是虚假数据。为了能在博弈中胜出, 两个模型需不断提高自身的生成能力和判别能力。生成模型和判别模型都可以是类似深度神经网络的结构[5]。生成对抗网络 GAN 的优化的期望是能够达到“纳什均衡”的状态, 也就是说其优化过程是一个极大极小博弈(Minimax game)问题, 生成网络模型中的生成器可以很好地学习数据样本的分布, 从而生成所需的数据样本。生成对抗网络框架如图 1 所示:

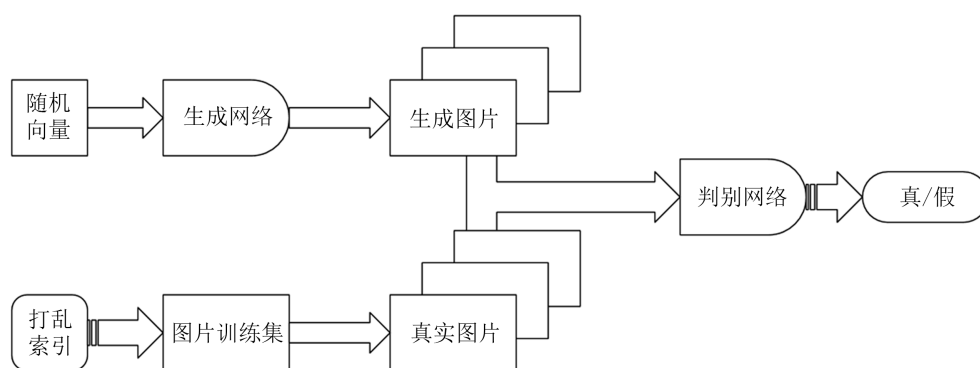


Figure 1. Generative confrontation network framework
图 1. 生成对抗网络框架

在训练过程中, 生成网络 G 的目的是尽可能多地生成真实的图像来欺骗和辨别网络 D 。 D 的目标是尽可能地将 G 的图像从真实图像中分离出来。目标函数是:

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

整个公式由两项组成。 X 是真实图像, Z 是输入到 G 网络的噪声, $G(Z)$ 是 G 网络生成的图像。 $D(x)$ 是 D 网络判断真实图片是否真实的概率(因为 x 是真实的, 所以对于 D , 值越接近 1 越好)。 $D(g(z))$ 是 D 网络判断 g 生成的图像是否真实的概率。 G 的目的: 如上所述, $D(G(z))$ 是 D 网络判断 G 生成的图像是否真实的概率。生成器希望 G 生成的图像“越接近真实越好”[6]。换句话说, 生成器希望 $D(G(z))$ 尽可能得大, 这时 $V(D, G)$ 会变小。因此目标函数最前面的记号是 $\min_G$ 。判别器的主要作用为: D 的判别效果越强, $D(x)$ 应该越大, $D(G(x))$ 应该越小。这时 $V(D, G)$ 会变大。因此式子对于 D 来说是求最大 $\max_D$ 。

2.1. 条件生成对抗网络

传统的生成对抗网络是通过随机向量 Z 生成图像 Y 。但这样的初始化数据在训练过程中较为复杂, 有的时候训练找不到纳什均衡点, 最终导致训练不稳定。

条件生成对抗网络的损失函数和普通的生成对抗网络有一定的区别也有一些联系。条件生成对抗网络增加了有辅助特征, 并且多了 L1 损失函数, 这使得源域与目标域之间的图像距离尽可能的靠近。

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x, y} [\log D(x, y)] + \mathbb{E}_{x, z} [\log(1 - D(x, G(x, z)))] , \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D)$$

和 GAN 网络的目标函数类似, 但是 GAN 网络是根据随机噪声 z 生成图像, 而条件生成对抗网络是根据随机噪声 Z 和输入图像 X 去生成。此时公式意义有一定的调整: 首先整个式子由两项构成。 x 表示

源域图像, y 表示真实图片, z 表示输入生成器网络的噪声, 而 $G(x, z)$ 表示生成器网络根据源域图像和随机噪声生成的目标域图片。其次 $D(x, y)$ 表示判别器网络判断真实图片是否真实的概率[7], 真实图片用 y 表示, 生成器网络中加入输入噪声 z , 而 $G(x, z)$ 表示生成器网络根据源域图像和随机噪声生成的目标域图片。其次 $D(x, y)$ 表示判别器网络判断输入一个批次的图片真实性的概率, 其中 Y 是真实图片, 对于判别器得到的最终结果越靠近 1 说明图片效果较好。而 $D(x, G(x, z))$ 是判别器网络判断生成器生成的图片的是否真实的概率。再其次生成器网络的目的是 $D(x, G(x, z))$ 是判别器网络判断生成器生成的图片是否真实的概率, 生成器最终的方向是拟合真实图片的概率。也就是说, 生成器希望 $D(x, G(x, z))$ 尽可能得大, 这时 $1 - D(x, G(x, z))$ 会变小。因此式子的最前面的记号是 $\min_G$ 。判别器的目的: 判别器的能力越强, $D(x, y)$ 应该越大, $D(x, G(x, z))$ 应该越小。这时 $1 - D(x, G(x, z))$ 会变大。因此式子对于判别器来说是求最大 ($\max_D$)。条件生成对抗网络的生成器作用除了生成可以愚弄判别器的图像之外, 还需要尽量接近目标域的图像 Y :

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

关于损失函数的选取, 选用 L1 损失函数而不用 L2 损失函数从而保证更少的模糊。对于一维的数据, 最小化 L2 损失函数是获得算术平均数; 最小化 L1 值得到的是中位数。最终的损失函数为:

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

2.2. 循环生成对抗网络损失函数

循环对抗网络中有两个生成器, 分别为 D 和 F 表示, 生成器有 X 和 Y 两个定义域。生成器 G 用来基于 X 域的图像生成 Y 域的图像 ($X \rightarrow Y$), 生成器 F 用来基于 Y 域的图像生成 X 域的图像 ($Y \rightarrow X$), 这两个生成器的定位是相反的过程, 通过损失函数 $\mathcal{L}(G, F, D_X, D_Y)$ 进行约束。同时循环生成对抗网络中有 2 个判别器, 分别用 $D(X)$ 和 $D(Y)$ 表示, 它用于判断输入图像在 X 域或 Y 域中是真是假。因此, 循环生成对抗网络可视为两种对抗网络的融合。一种对抗网络由发生器 G 和鉴别器 $D(y)$ 组成, 实现了从 X 域到 y 域的图像生成和识别; 另一种对抗网络由生成器 F 和鉴别器 $D(x)$ 组成, 实现了从 y 域到 x 域的图像生成和识别。

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y, X, Y) + \mathcal{L}_{GAN}(F, D_X, Y, X) + \lambda \mathcal{L}_{cyc}(G, F)$$

循环生成对抗网络的整体优化目标包括三个部分: 第一部分 $\mathcal{L}_{GAN}(G, D_Y, X, Y)$ 表示生成器 G 和判别器 D_Y 的优化目标; 第二部分为了改善生成器 F 和判别器 D_X 采用函数 $\mathcal{L}_{GAN}(F, D_X, Y, X)$ 作为优化目标, 二部分的目的是生成对抗网络算法本身的优化目标。第三部分采用 $\mathcal{L}_{cyc}(G, F)$ 来提升从变换域再变换到原域时图像的一致性的能力。

$\mathcal{L}_{GAN}(G, D_Y, X, Y)$ 如公式所示, 采用一般生成对抗的优化目标, 优化目标的最大化采用判别器 D_Y , 生成器 G 要最小化公式中的后半部分。 $\mathcal{L}_{GAN}(G, D_Y, X, Y)$ 也是同理。

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

为了使训练过程更加稳定, 实际训练是基于最小二乘损失的。例如, 鉴别器 D 的优化目标变为最小化 $Ex \rightarrow p_{data}(y) [(D(y) - 1)^2] + Ex \rightarrow p_{data}(x) [D(G(x))^2]$

优化目标的第一部分是期望 $D(y)$ 尽可能接近 1, 也就是说, 当输入真实图像 y 时, 鉴别器 D 具有相

信输入是真实的最大概率。第二部分是期望 $D(g(x))$ 尽可能接近 0, 也就是说, 当输入假图像 $g(x)$ 时, 鉴别器 D 相信输入是真的概率最低。生成器 G 的优化目标变成最小化 $E \rightarrow p_{data}(x) \left[(D(G(x)) - 1)^2 \right]$, 这个优化目标是使 $D(G(x))$ 的期望值更接近 1, 而 $D(G(x))$ 越接近 1 表示判别器 D 越相信生成器 G 生成的图像 $G(x)$ 是真实图像。

$\mathcal{L}_{cyc}(G, F)$ 如下公式所示, 其中 $G(x)$ 表示从 X 域到 Y 域的变化, $F(G(X))$ 表示再从 Y 域变换到 X 域, 模型训练的目标当然是希望 $F(G(X))$ 尽可能和 X 接近, 因此用其距离来约束, 也就是 $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$, 将这个称之为正向循环一致性; 同理对于 $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$ 称之为后向循环一致性。

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

3. 实验设计

3.1. 实验设计

操作系统使用 Windows10 系统, 安装 Anaconda3 (64bit) 和 PyCharm2020 (64bit) 集成编译环境, 使用 python3.6.8 版本编译和 tensorflow2.0。需要的 package: matplotlib.pyplot、keras2.2.4、time、os、tensorflow.keras、numpy 等。

3.2. 网络结构

为了保证生成器生成的图片是基于输入图片, 而不是在发现判别器的“漏洞”, 循环生成对抗网络引入了与目标生成器和判别器对称的另一组生成器和判别器。如果目标网络寻找的是 x 到 y 的映射, 这组另外添加的网络就是寻找 y 到 x 的映射。如图 2 所示, 目标生成器的输出可以作为额外添加生成器的输入, 输入图片连续经过两个生成器, 最终结果与输入图片的差别可作为网络是否成功训练的依据。

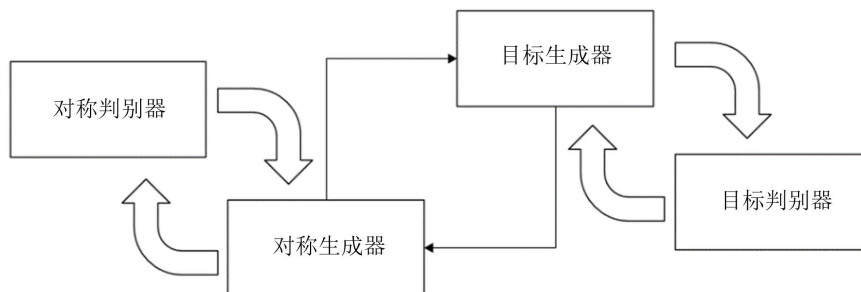


Figure 2. Cyclic generative confrontation network

图 2. 循环生成对抗网络

3.3. U-Net 为生成器主要的网络结构

生成器网络为 U 型结构, 整体可分为两个部分: 特征通道逐渐变小, 通道数逐渐变多的“下采样”过程; 特征通道逐渐变大, 通道数逐渐变少的“上采样”过程。上图的左半部分为生成器网络的整体结构。上图的右半部分为“上采样”过程中 Block 的具体实现方法。一个卷积层、一个归一化层和激励函数组成每个 Block 块。与“下采样”过程的区别是: “上采样”中残差和上一个 Block 的输出同时作为此 Block 的输入, 卷积层为反卷积; “下采样”中输出在传递到下一个 Block 的同时也作为残差传递到相应的 Block。如图 3 所示。

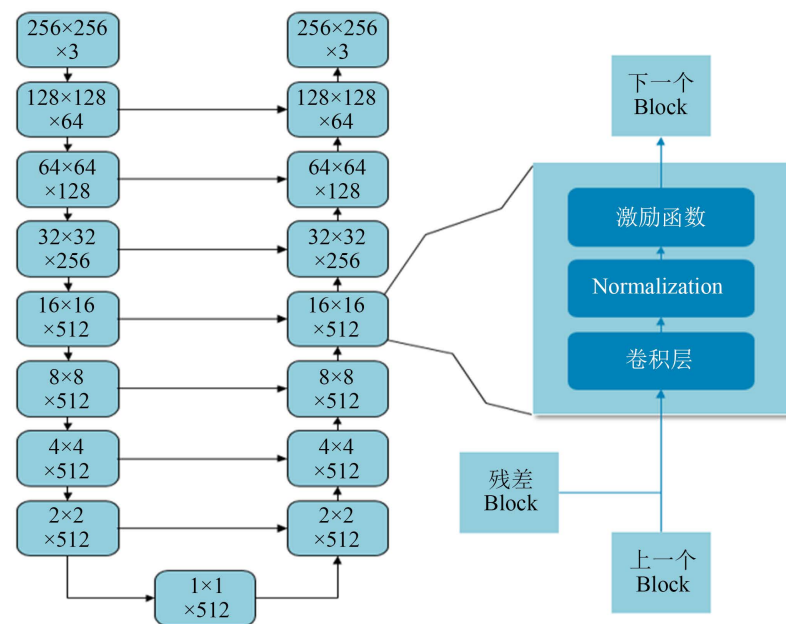


Figure 3. Generator network structure
图 3. 生成器网络结构

3.4. 判别器网络结构

判别器网络与传统的卷积神经网络很相似，区别主要是没有在最后添加全连接层。上图为判别器网络的结构。判别器网络可以看作另外的生成器，它的输入为生成器的输入和输出图像，或者是数据集中的两种对应图片，而它的输出是一张 30×30 的小图片，通过统计判别器生成的图片中的像素整体情况来判断输入图片间的关系。如图 4 所示：

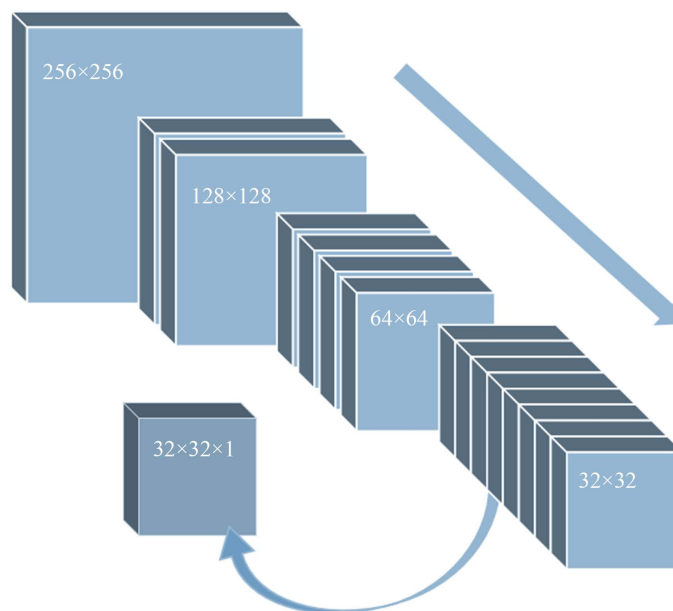


Figure 4. Discriminator network framework
图 4. 判别器网络框架

4. 结果分析

第一组实验如图 5、图 6 所示, 输入都是标准图像, 左侧为输入, 右侧为输出。观察第一组的实验结果可以发现, 由简笔画生成的建筑图片色调上有明显的差别, 建筑的结构只在图片的边缘处出现错位等情况; 而建筑图片生成的简笔画总体上具有建筑的结构, 但其中会出现一些没有意义的色块。

第二组实验如图 7 所示, 输入标准的建筑图像, 得到输出的简笔画图像后, 再将该简笔画图像输入, 最后得到生成的建筑图像。通过研究第二组的实验结果可以发现, 生成对抗网络从简笔画图像中还原出了过于完整的细节, 而在简笔画图像中不能看到这部分的细节。这是循环生成对抗网络中的经典问题, 由于在损失中加入了生成图像与原图像的区别, 生成对抗网络会将信息隐藏在生成的简笔画图像中, 在最终的生成结果中将这信息还原出来。

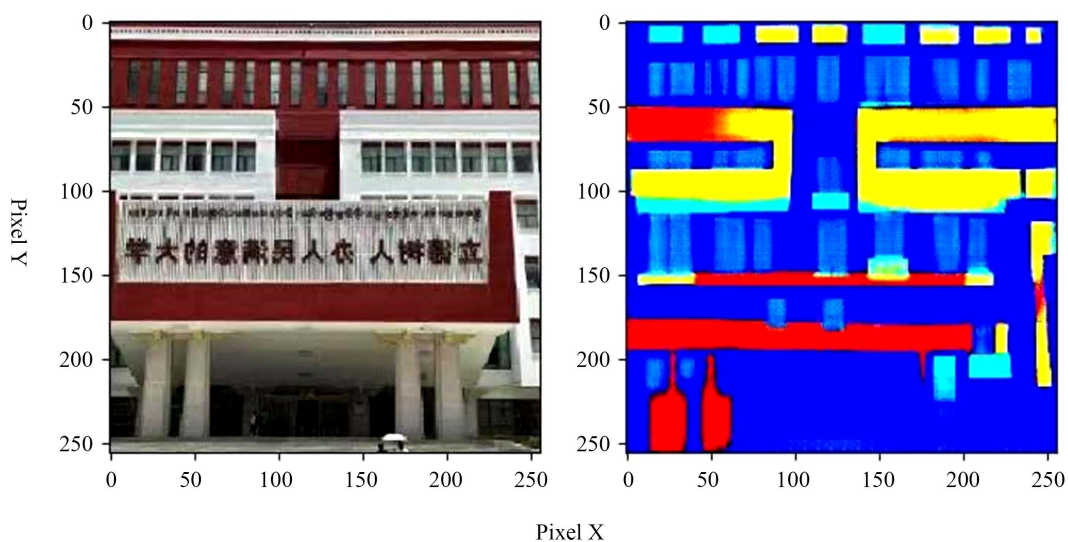


Figure 5. Building a standard drawing to stick figure
图 5. 建筑物一标准图到简笔画

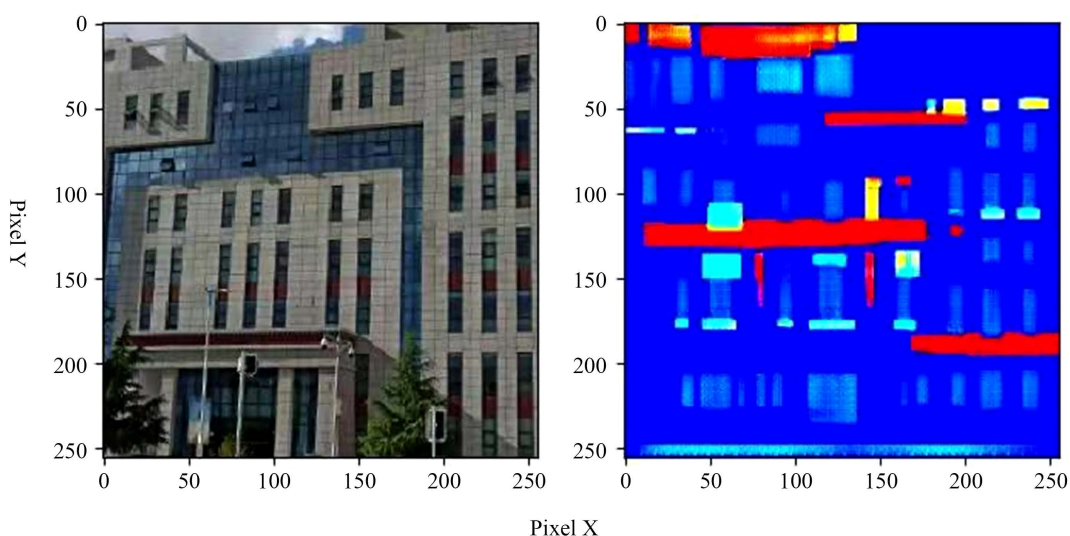


Figure 6. The second standard drawing of the building to the stick figure
图 6. 建筑物二标准图到简笔画

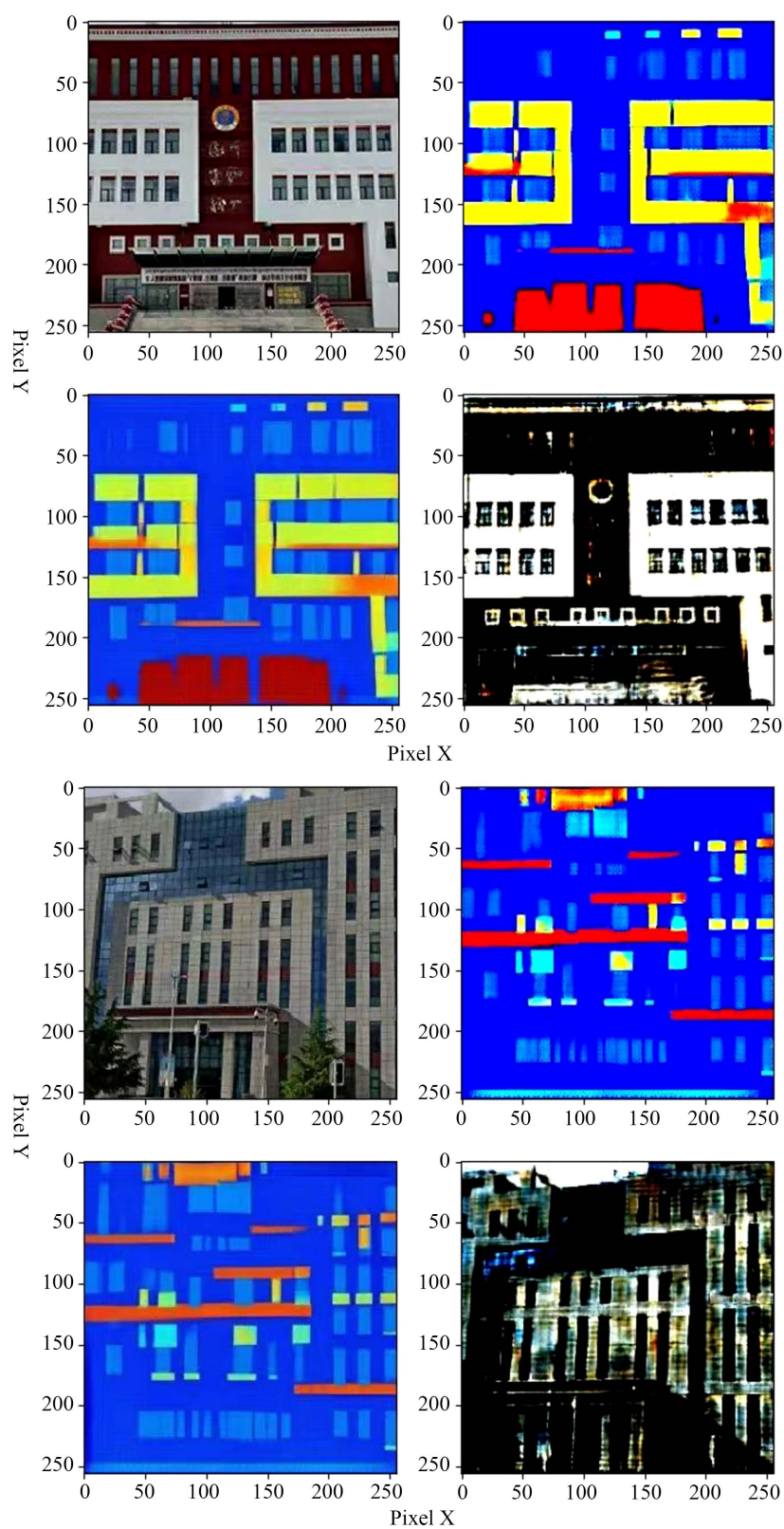


Figure 7. Simple strokes to generate pictures

图 7. 简笔画到生成图片

5. 展望

针对图片生成图片的研究, 本文运用循环生成对抗网络, 无须使用成对的图片进行训练, 可以将图片从一个领域变换到另一个领域, 训练学习两种映射的生成网络。循环生成对抗网络存在一种情况是训练稳定后某些部分会被隐藏起来, 由于服务器问题本次实验只运行了 5000 代, 在清晰度方面还存在一定问题。未来研究方向可以结合 styleGAN 网络框架, 可以有效的调整细节方面的需求, 改善生成图片的视觉效果和解码能力。

参考文献

- [1] 朱秀昌, 唐贵进. 生成对抗网络图像处理综述[J]. 南京邮电大学学报(自然科学版), 2019, 39(3): 1-12.
- [2] 黄宇, 韩璞, 王东风, 张婧. 基于 BP 神经网络整定的 PID 控制在过热汽温系统中的应用[J]. 仪器仪表学报, 2006(S3): 1980-1981.
- [3] 冯杰, 班彪华. 生成对抗网络模型的基本介绍和应用综述[J]. 现代计算机(专业版), 2019(4): 34-39.
- [4] 卢庆林, 叶伟, 李国靖. 基于 DCGAN 的 SAR 虚假目标图像仿真[J]. 电子信息对抗技术, 2020, 35(2): 57-61+65.
- [5] 时澄, 潘斌, 郭小明, 李芹芹, 张露月, 钟凡. 生成式对抗网络在图像补全中的应用[J]. 计算机科学与探索, 2019, 13(8): 1402-1410.
- [6] 朱良宽, 晏铭, 黄建平. 一种新型卷积神经网络植物叶片识别方法[J]. 东北林业大学学报, 2020, 48(4): 50-53.
- [7] 毕松, 刁奇, 孙贵宾, 韩存武. 交通场景物体检测模型研究[J]. 计算机仿真, 2018, 35(10): 193-197.