

文章编号: 2095-4980(2021)04-0724-05

基于生成对抗网络的图像转换技术

李国威, 石志广, 张 焱

(国防科技大学 电子科学学院, 湖南 长沙 410000)

摘要: 针对不同谱段图像获取代价不同的问题, 提出一种基于生成对抗网络的图像转换方法。转换过程以肉眼可分辨范围内图像轮廓不变为出发点。首先, 通过成对的训练数据对生成器和判别器进行交替训练, 不断对损失函数进行优化, 直到模型达到纳什平衡。然后用测试数据对上述训练好的模型进行检测, 查看转换效果, 并从主观观察和客观上计算平均绝对误差和均方误差角度评价转换效果。通过上述过程最终实现不同谱段图像之间的转换。其中, 生成器借鉴U-Net架构; 判别器采用传统卷积神经网络架构; 损失函数方面增加L1损失来保证图像转换前后高、低频特征的完整性。以红外图像与可见光图像之间的转换为例进行实验, 结果表明, 通过本文设计的生成对抗网络, 可以较好地实现红外图像与可见光图像之间的转换。

关键词: 生成对抗网络; 图像转换; pix2pix; 红外图像

中图分类号: TN219; TP391.41

文献标志码: A

doi: 10.11805/TKYDA2019426

Image transformation technology based on generative adversarial networks

LI Guowei, SHI Zhiguang, ZHANG Yan

(College of Electrical Science and Technology, National University of Defense Technology, Changsha Hunan 410000, China)

Abstract: An image conversion method based on generative adversarial networks is proposed in order to solve the problem of different image acquisition costs in different spectral segments. In the conversion process, the image outline does not change into the starting point in the range that can be distinguished by the naked eye. Firstly, the generator and discriminator are trained alternately through pairs of training data, and the loss function is optimized until the Nash equilibrium of the model is reached. Then the test data are utilized to detect the trained model, to check the conversion effect, and to evaluate the conversion effect from the subjective observation and objective calculation of the average absolute error and mean square error. Through the above process, the conversion between different spectral images is finally realized. Among them, the generator learns from U-Net architecture; the traditional convolution neural network architecture is used by the discriminator; and L1 loss function is increased to ensure the integrity of high and low frequency features before and after image conversion. In this paper, the conversion between infrared image and visible image is taken as an example to carry out the experiment. The results show that the conversion between infrared image and visible image can be well realized through the generative adversarial networks designed in this paper.

Keywords: generative adversarial networks; image conversion; pix2pix; infrared image

对于图像转换, 可把它理解成将一张输入图像经过特定的变换得到相应的输出图像的过程。传统意义上的图像转换可追溯到 Hertzman 使用图像类比方法将纹理特征转移到目标图像上^[1], 实现低级图像特征的转移。目前, 实现可见光图像与红外图像的转换通常是利用反演的方式, 也就是通过寻找同一目标可见光图像与红外图像的映射关系, 得到同样条件下的反演关系^[2]。同时, 有人提出了利用地物光谱仪采集目标的光谱反射率, 并利用经过定标后的 CMOS 相机采集目标的可见光与红外图像, 而后结合反射率和相机标定结果建立可见光与红外图像灰度值的映射关系的方法^[3]。对于红外图像细节的描述上往往通过图像变换算法将红外数据映射到模拟图像^[4]。从上述方法可以看出要想实现转换, 都是需要针对不同的模型建立相应的映射关系。

在生成对抗网络出现之后, 图像转换任务便可以通过一个统一的框架来解决, 即 pix2pix^[5]。本文便是基于生成对抗网络来实现可见光图像与红外图像的转换。在此过程中, 通过生成器和判别器之间博弈式的训练方法不断提高各自的能力, 即生成器生成的图像越来越真实, 判别器判别图像的能力越来越精确, 这样极大地提高了训练效率以及生成图像的质量, 取得了良好的转换效果, 所付出的代价就是需要大量的成对数据进行训练。相较于原始的图像转换, 生成式对抗网络(Generative Adversarial Networks, GAN)的强大之处就在于可以自动地学习真实样本的分布以及自动学习损失函数, 而不是定义一个模型让数据去学习。实验表明: 基于生成式对抗网络的可见光图像与红外图像转换可以取得比较好的结果。

1 生成模型与判别模型

1.1 生成模型

概率生成模型, 简称为生成模型(Generative Model), 是概率统计中常见的一种模型^[5], 主要是对联合概率进行建模, 从统计学的角度描述数据的分布, 应用十分广泛, 例如朴素贝叶斯模型。在生成式对抗网络中, 生成模型是一个神经网络, 通过学习训练集中数据的真实分布, 希望通过输入随机噪声来生成以假乱真的图像。对于复杂的模型, 例如高维随机变量, 直接建模会比较复杂, 这时可以利用深层神经网络可以近似任意函数的特点进行建模。

1.2 生成器结构设计

本文主要使用了基于条件生成对抗网络(Conditional Generative Adversarial Network, cGAN)结构的 pix2pix 模型, 从原始的无监督学习变成了有监督学习。在生成器结构的设计上主要考虑 2 个问题: a) 希望转换前后在肉眼可分辨的前提下, 图像的轮廓信息可以被完整地保留下来, 也就是学习到图像的低频特征; b) 希望可以将图像的细节描述清楚, 即学习到图像的高频特征。上述这 2 个问题的实现对于普通的 Encoder-Decoder 模型^[6]来讲是一个矛盾的过程, 想要学习图像的高频特征就需要增加卷积层数, 但是增加卷积层数又会导致图像低频信息的损失, 这时 U-Net 结构就很好地解决了这一问题^[7]。相较于传统的 Encoder-Decoder 模型, U-Net 结构增加了特征图像和反卷积后同样大小的图像按通道拼接即跳跃连接这一部分, 这样通过共享低频图像信息来减轻网络的负担, 而不用每一层都承载、保存、传播所有的信息。原始 U-Net 结构^[8]见图 1, 本文在原始结构基础上增加下采样层数以更好地提取图像高频特征。

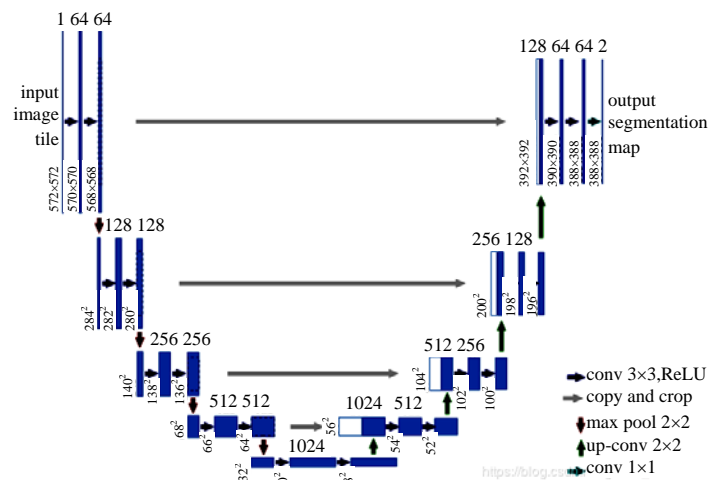


Fig.1 Structure diagram of original U-Net
图 1 原始 U-Net 结构图

1.3 判别器结构

对于 pix2pix GAN, 判别器的设计类似于 cGAN, 不能只接收一张图像, 来判别接收图像的真假, 因为要保证生成的图像和原始的图像的对应关系, 这样最终的结果就是对两张图像得出一个概率值。pix2pix 的原始论文中, 判别器 D 是通过 Patch-D 来实现的^[8]。在此之前都是对整张图像输出一个是否为真的概率, 论文中提到的 Patch 就是先把图像分成 $N \times N$ 的小块, 然后计算每一小块的概率, 最后将这些概率求平均值作为整体的输出, 这样每一次运算量就会减小, 从而可以加快训练的速度和收敛速度。在实践中发现直接对整张图像进行判别的计算速度不会产生很大的影响, 就计算总量而言, 无论是 Patch-D 还是采用普通的卷积神经网络, 总量是一样的。所以对于比较小的图像而言, 效果影响不是很大。所以本文采用普通的神经网络, 经过四层卷积和 Leaky_ReLU 激活后得到判别器输出。

2 损失函数

损失函数就是用来描述预测值和实际数据的差距大小,并且提供一个优化的方向。在本文中,损失函数可以看作 2 大部分:判别器网络的损失和生成器网络的损失。对于判别器网络,由于是对 2 张图像得出一个结果,也就是判别器每次输入 2 张图像得出一个概率,所以一方面对于输入的真实成对图像,判别器需要将它判为 1;另一方面对于生成的图像和原始图像,判别器需要将它判为 0。有了上面这个想法,判别器网络的损失函数就很容易定义了,同样是两部分:将真实图像误判为虚假图像所引入的损失以及将生成的虚假图像误判为真实图像所引入的损失。将二者相加就是最终判别器网络的损失。

对于生成器网络,总体上是希望生成器生成的图像可以骗过判别器,被判为 1,所以生成网络的损失函数一部分就是生成图像和 1 之间的损失。此外为了保证输入图像和输出图像的对应关系,在这里引入生成图像和真实图像之间的损失,关于这个损失函数,可以选择的有很多,考虑到本文增加了卷积层的数目,如果只依赖于生成器结构中的按通道拼接操作来实现图像低频特征的保留,也可以完成,但在这里引入 L1 损失使生成器进一步学习到图像低频特征。

总的损失函数为:

$$G^* = \arg \min_G \max_D L_{\text{cGAN}}(G, D) + \lambda L1_loss$$

$$L_{\text{cGAN}}(G, D) = E_{x,y} [\log D(x, y)] + E_{x,z} \{\log [1 - D(x, G(x, z))]\}$$

$$L1_loss = E_{x,y \sim p_{\text{data}}(x,y), z \sim p_z(z)} [\|y - G(x, z)\|]$$
(1)

式中: G^* 为总的损失函数; $L_{\text{cGAN}}(G, D)$ 为条件生成对抗网络损失函数; G 为生成器; D 为判别器; $L1_loss$ 为 L1 损失函数; $p_{\text{data}}(x, y)$ 为真实数据分布; λ 的数值没有硬性规定,只是为了加权 L1 损失使得生成器的两部分损失函数可以相互权衡而不至于某一个发挥全部作用。

3 参数更新

将生成器生成的图像与真实图像一同输入到判别器后,会得到一个概率输出,然后在反向传播过程中,通过梯度下降法不断对生成器和判别器中的卷积核的权重进行更新,不断增强生成器生成图像的真实度和判别器判别的准确度。

4 实验结果

本实验所用数据集是严格意义上红外图像与可见光图像对应的数据,其中既有变换的场景又有运动的小目标。训练用数据共 588 张,测试数据 13 张,图像大小为 256×256 ,共进行 4 000 次训练。训练用数据见图 2。

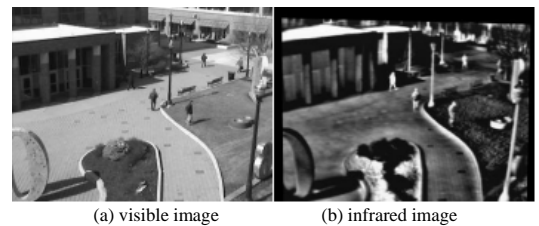
第一次训练结果见图 3,可以看出在第一次学习中并没有学习到什么东西。

第 1 000 次 epoch 训练结果见图 4。从结果可以看出,随着训练次数的增加,生成器生成图像的能力在逐渐增强,对于输入图像的整体结构已经描述得比较清楚。

中间的训练过程不再一一阐述,直接观察最后一次训练结果,结果见图 5。对比图 5(b)和图 5(c),主观上已经分辨不出差别,可以认为生成器结构已经训练完毕。

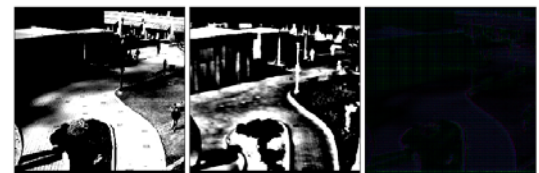
对于上述训练好的模型,用测试集进行测试,来检验训练效果,测试结果见图 6。

对于图像转换前后效果的评价方式有很多,本文主要从主观和客观的角度出发进行评判。主观上通过观察转换后的图像或者观察残差图的方式直接评价,客观上通过平均绝对误差以



(a) visible image (b) infrared image
Fig.2 Example of dataset image

图 2 数据集图像举例



(a) visible image (b) infrared image (c) generated image
Fig.3 First epoch

图 3 第 1 次 epoch



(a) visible image (b) infrared image (c) generated image
Fig.4 1 000th epoch

图 4 第 1 000 次 epoch



(a) visible image (b) infrared image (c) generated image
Fig.5 4 000th epoch

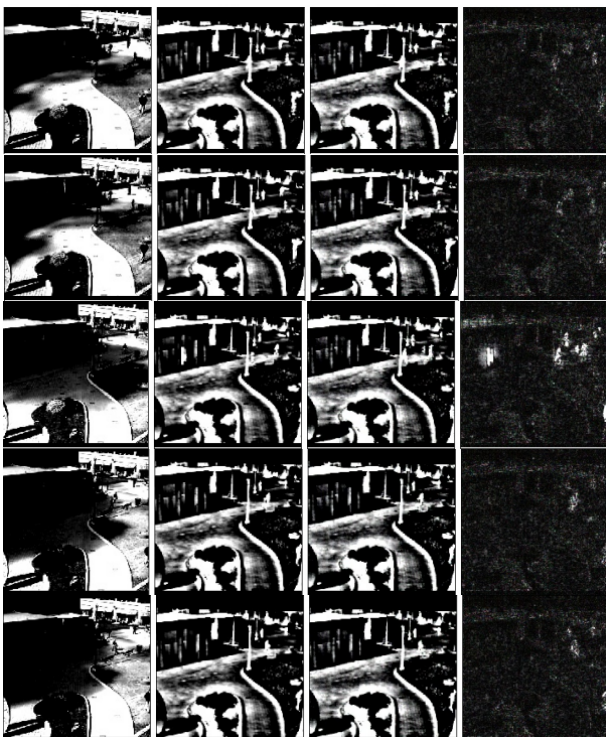
图 5 第 4 000 次 epoch

及均方误差来评判。从测试结果图可看出，生成器模型的训练比较成功，如果不看残差图已经分辨不出真实图像与生成图像；从表 1 可以看出，转换后的图像与原始图像差值已经很小。图 7 为文献[2]所述方法图像转换结果图。该算法整体思想是先将目标分类，而后根据不同的波段及天气状况选用不同的反演方法，最后将图像平滑得到输出。相较于传统方法，本方法优势在于不需要对不同的转换问题选取不同的反演方法，都可用一个统一的架构实现图像转换。同时在保证转换质量的前提下本方法更加灵活，复杂环境也可完成转换任务。

表 1 误差统计表

Table 1 Error statistic

number	mean absolute error	mean square error
1	0.084 007	0.013 699
2	0.085 620	0.014 569
3	0.123 252	0.075 154
4	0.088 778	0.017 115
5	0.086 825	0.015 204
mean	0.093 696 4	0.027 148 2



(a) visible image (b) infrared image (c) generated image (d) residual diagram

Fig.6 Test result diagram

图 6 测试结果图



(a) original image (b) infrared image (c) image after inversion

Fig.7 Using gray level transformation to realize image conversion

图 7 利用灰度映射法实现图像转换

5 结论

本文通过 GAN 的衍生模型 pix2pix GAN 实现了可见光图像与红外图像的转换。通过增加生成器中的下采样层来增加图像细节的描述，引入 batch normalization 和 dropout 加速收敛并且有效地防止了过拟合。实验结果表明，仿真结果和真实图像之间的差别已经很小，有较高的实用性和可行性。此外，本文所提的转换方法不止局限于可见光图像与红外图像之间的转换，也可以应用到其他波段图像之间的转换。

参考文献：

- [1] HERTZMAN A,JACOBS C E,OLIVER N,et al. Image analogies[C]// Proceedings of the 28th annual conference on computer graphics and interactive techniques. NewYork,US:ACM, 2001:327-340.
- [2] 武国军,白廷柱,白获宁. 基于可见光图像的红外图像反演研究[J]. 红外技术, 2011,33(10):574-579. (WU Guojun, BAI Tingzhu,BAI Funing. Research on infrared image inversion based on visible image[J]. Infrared Technique, 2011, 33(10):574-579.)
- [3] 周强,白廷柱,刘明奇,等. 基于可见光图像的近红外场景仿真[J]. 红外技术, 2015,37(1):11-15. (ZHOU Qiang,BAI Tingzhu,LIU Mingqi,et al. Simulation of near infrared scene based on visible image[J]. Infrared Technique, 2015,37(1): 11-15.)

(下转第 732 页)