

基于生成对抗网络的红外图像数据增强

陈佛计^{1,2,3,4*}, 朱 枫^{1,2,3,4}, 吴清潇^{1,2,3,4}, 郝颖明^{1,2,3,4}, 王恩德^{1,2,3,4}

(1. 中国科学院沈阳自动化研究所, 沈阳 110016; 2. 中国科学院机器人与智能制造创新研究院, 沈阳 110016;

3. 中国科学院大学, 北京 100049; 4. 中国科学院光电信息处理重点实验室, 沈阳 110016)

(* 通信作者电子邮箱 1754208529@qq.com)

摘要:深度学习在视觉任务中的良好表现很大程度上依赖于海量的数据和计算力的提升,但是在很多实际项目中通常难以提供足够的数据来完成的任务。针对某些情况下红外图像少且难以获得的问题,提出一种基于彩色图像生成红外图像的方法来获取更多红外图像数据。首先,用现有的彩色图像和红外图像数据构建配对的数据集;然后,基于卷积神经网络、转置卷积神经网络构建生成对抗网络(GAN)模型的生成器和鉴别器;接着,基于配对的数据集来训练GAN模型,直到生成器和鉴别器之间达到纳什平衡状态;最后,用训练好的生成器将彩色图像从彩色域变换到红外域。基于定量评估标准对实验结果进行了评估,结果表明,所提方法可以生成高质量的红外图像,并且相较于在损失函数中不加正则化项,在损失函数中加入L1和L2正则化约束后,该方法的FID分数值平均分别降低了23.95和20.89。作为一种无监督的数据增强方法,该方法也可以被应用于其他缺少数据的目标识别、目标检测、数据不平衡等视觉任务中。

关键词:红外图像生成;生成对抗网络;图像转换;数据增强;生成图像质量评估

中图分类号:TP391.41 **文献标志码:**A

Infrared image data augmentation based on generative adversarial network

CHEN Fojij^{1,2,3,4*}, ZHU Feng^{1,2,3,4}, WU Qingxiao^{1,2,3,4}, HAO Yingming^{1,2,3,4}, WANG Ende^{1,2,3,4}

(1. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang Liaoning 110016, China;

2. Institutes for Robotics and Intelligent Manufacturing Innovation, Chinese Academy of Sciences, Shenyang Liaoning 110016, China;

3. University of Chinese Academy of Sciences, Beijing 100049, China;

4. Key Laboratory of Opto-Electronic Information Processing, Chinese Academy of Sciences, Shenyang Liaoning 110016, China)

Abstract: The great performance of deep learning in many visual tasks largely depends on the big data volume and the improvement of computing power. But in many practical projects, it is usually difficult to provide enough data to complete the task. Concerning the problem that the number of infrared images is small and the infrared images are hard to collect, a method to generate infrared images based on color images was proposed to obtain more infrared image data. Firstly, the existing color image and infrared image data were employed to construct the paired datasets. Secondly, the generator and the discriminator of Generative Adversarial Network (GAN) model were formed based on the convolutional neural network and the transposed convolutional neural network. Thirdly, the GAN model was trained based on the paired datasets until the Nash equilibrium between the generator and the discriminator was reached. Finally, the trained generator was used to transform the color image from the color field to the infrared field. The experimental results were evaluated based on quantitative evaluation metrics. The evaluation results show that the proposed method can generate high-quality infrared images. In addition, after the L1 or L2 regularization constraint was added to the loss function, the FID (Fréchet Inception Distance) score was respectively reduced by 23.95, 20.89 on average compared to the FID score of loss function not adding the constraint. As an unsupervised data augmentation method, the method can also be applied to many other visual tasks that lack train data, such as target recognition, target detection and data imbalance.

Key words: infrared image generation; Generative Adversarial Network (GAN); image transformation; data augmentation; quality evaluation of generative image

0 引言

近年来,基于深度学习的方法在很多领域取得了良好的

表现,但是很大程度上依赖于海量标数据。在某些条件下,标注红外数据很难获取,而且需要大量的人力成本。因此,针对红外数据少、数据难以获得的问题,基于现有的数据进行数据

收稿日期:2020-01-09;修回日期:2020-03-01;录用日期:2020-03-11。 基金项目:国家自然科学基金资助项目(U1713216)。

作者简介:陈佛计(1994—),男,山西忻州人,硕士,主要研究方向:图像生成、机器学习、机器人视觉;朱枫(1962—),男,辽宁沈阳人,研究员,博士生导师,博士,主要研究方向:机器人视觉;吴清潇(1978—),男,辽宁沈阳人,研究员,博士,主要研究方向:机器人视觉、机器视觉;郝颖明(1966—),女,辽宁沈阳人,研究员,博士,主要研究方向:视觉定位、图像优化、视觉检测、红外图像仿真、三维成像、数据处理;王恩德(1980—),男,辽宁沈阳人,研究员,博士,主要研究方向:图像目标识别、检测与跟踪、微弱信号检测。

增强是一种比较好的、获取更多数据的方法。目前,数据增强的方法有两种:基于有监督的方式和基于无监督的方式。有监督方式的数据增强是采用预设的规则,在已有的数据上进行数据的扩增,包括几何变换类和颜色变换类等方法。但是,基于有监督方式进行数据增强并没有对数据集进行实质性的改变;而无监督数据增强是通过模型学习数据所服从的分布,随机生成与样本集分布一致的数据,该方法可以使得数据集覆盖更多的模式,更有利于提高模型的性能。生成模型是无监督学习任务中最为关键的技术,目前比较常用模型有基于有向图模型的赫姆霍兹机(Helmholtz Machines)^[1]、深度信念网络(Deep Belief Network, DBN)^[2]、变分自动编码器(Variational Auto-Encoder, VAE)^[3];基于无向图模型的受限玻尔兹曼机(Restricted Boltzmann Machine, RBM)^[4]、深度玻尔兹曼机(Deep Boltzmann Machine, DBM)^[5]、自回归模型(AutoRegressive model, AR)^[6]以及基于零和博弈的生成对抗网络(Generative Adversarial Network, GAN)^[7]等。AR和VAE是通过一种显示方式对数据进行建模,但是AR是按照像素点去生成图像,导致计算成本太高,而且在并行性上受到限制,在处理大分辨率图像的时候有一定的困难。VAE虽然在图像生成上是可并行的,但是其生成的图像比较模糊。相较于显示建模方式,隐式建模方法GAN不仅避免了难解的推断,而且还可以生成高质量的图像;并且由于其拟合高维数据分布的能力,以及在图像生成方面的优良表现,GAN是生成模型中目前较好的一种方法。

生成红外图像模型的总体结构如图1所示,GAN由生成器和鉴别器两个部分组成。生成器的目的是生成和样本数据所服从分布一致的新样本来骗过鉴别器;而鉴别器的目的是鉴别输入的图像是否来自于真实数据分布。通过对抗训练,生成器和鉴别器的性能不断提升,最终达到纳什平衡状态。林懿伦等^[8]对生成对抗网络的基本思想以及其常见的模型架构和训练方式进行了详细的介绍,并且对其一些应用场景进行了介绍。曹仰杰等^[9]对生成对抗网络的优点和缺点进行了总结和归纳,并且对其在高质量图像生成和图像翻译等领域研究现状和未来的发展方向进行了阐述。因此,本文的目标是基于生成对抗网络的方法生成更多的类似于真实红外图像样本的数据。最近有很多工作基于GAN来作图像生成,一种方式是基于服从某一分布的随机向量生成图像,该方法的基本思想是用一个多层神经网络实现一个非线性的映射,该映射可以将采样的随机向量映射为目标域中的图像。基于该思想的模型有标准GAN、集成高斯混合模型的条件生成对抗网络^[10]、信息最大化生成对抗网络(Info-GAN)^[11]、自注意力生成对抗网络(Self-Attention Generative Adversarial Network, SAGAN)^[12]等。另一种方式是图像到图像的转换,该方法的基本思想是用神经网络学习域之间的映射关系,基于该映射可以将图像从一个域转换到目标域,而且变换前后的图像在内容上是一致的,基于该思想的模型有图像翻译生成对抗网络(Pix2Pix)^[13]、循环一致性对抗网络(Cycle-GAN)^[14]、辅助分类器生成对抗网络(Auxiliary Classifier Generative Adversarial Network, ACGAN)^[15]、星型生成对抗网络(Star-GAN)^[16]等。在实际项目中,有些场景和目标的红外图像很难获取,而且数据量少,但是可以获取目标和场景大量彩色图像,为了得到大量红外图像数据,因此将彩色图像转换为红外图像是进行数据增强的一种很好的思路。在对抗网络之前主要是基于仿真的技术将图像从彩色域转换到红外域,例如:许洪等^[17]研究的红外多光谱图像的仿真生成;陈珊等^[18]进行的基于可见光图像的红外场景仿真。但是仿真的方法是一个复杂的过程,需要大量计算和建立复杂的模型。而生成对抗网络可以基于神经网络强大的拟合能力学习一个端到端的映射,基于该映射直接将图像从彩色域转换到红外域,从而避免仿真方法那些复杂的计算。因此,针对红外数据少的问题,基于GAN将彩色图像转换成红外图像是一种解决该问题很好的方法。

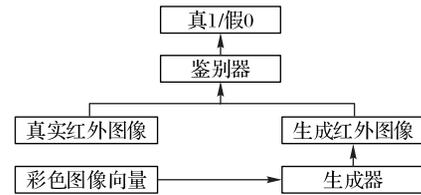


图1 生成红外图像模型的总体结构

Fig. 1 Overall structure of infrared image generation model

本文将重点关注如何基于彩色图像来生成红外图像,而为深度学习模型提供更多的训练数据。首先,基于编码器-解码器的神经网络架构来构建生成对抗网络中的生成器,它通过解码从编码器得到的隐变量来生成红外图像,而鉴别器是通过一个卷积神经网络来实现。其次,基于Wasserstein距离来度量生成分布和真实分布之间的距离,并且基于此构建模型的损失函数。为了更好地约束模型生成更高质量的图像,在损失函数中还加入了L1正则项。最后,基于由彩色图像和红外图像构成数据集对模型进行训练,并且验证了将彩色图像转换成红外图像方法的有效性。

1 相关工作

基于深度神经网络的生成对抗网络是生成模型中生成图像质量最好的一种方法,按照生成器输入量的不同,目前基于生成对抗网络做图像生成主要分为基于随机向量生成图像和基于图像转换生成图像两种方法,下面分别对其介绍。

1) 基于随机向量生成图像。标准GAN通过Kullback-Leibler散度对真实样本数据服从的分布和生成样本数据服从的分布之间的相似性进行度量,但是由于Kullback-Leibler散度固有的缺陷,导致模型会出现梯度消失和模式崩溃的问题,最终使得生成图像的质量不是很理想。因此瓦瑟斯坦生成对抗网络(Wasserstein Generative Adversarial Network, WGAN)^[19]采用瓦瑟斯坦距离来对真实分布和生成分布之间的距离进行度量,该方式很好地避免了梯度消失的问题。在标准GAN中,输入向量通常会被生成器进行过度耦合的处理,导致隐向量中每一个维度的量没有足够明确的语义信息。因此,Info-GAN^[11]将输入向量分解为隐变量和条件变量两个部分,并且将其一起送入生成器,在训练过程中通过加入互信息正则化的约束项来实现变量之间的解耦,从而使得输入向量中的某些维度可解释性。在构造生成对抗网络生成器的时候,小的卷积核使得网络难以发现图像中的依赖关系,大的卷积核会导致丧失计算上的效率。因此,SAGAN^[12]通过引入自注意力机制来捕捉数据或特征的内部相关性,并且全局信息也会被很好地利用去生成质量更好的图像。基于SAGAN^[12],大规模生成对抗网络(BIG-GAN)^[20]通过增加Batch的大小,模型的容量以及共享嵌入核正交正则化等技巧进一步改善了生成图像的质量和多样性。

2) 基于图像转换生成图像。计算机视觉和图像处理中的很多问题都可以被看作是图像转换的问题。图像转换的问题经常被很多模型看作是像素的分类或者是回归问题,这些模型以输出的每一个像素条件独立于输入图像中的所有像素为条件,逐个生成图像的每一个像素,但是这些模型生成的图像一般都比较模糊。Pix2Pix^[13]模型通过学习一个条件生成模型,在对抗损失函数和重构损失项的约束下,基于条件输入图像来生成相应的图像。在Pix2Pix^[13]模型的基础上,Pix2pixHD^[21]基于实例分割图像,使用多尺度的生成器和鉴别器来生成高分辨率的图像。Cycle-GAN^[14]通过在损失函数中加入循环一致性损失,基于不成对的数据来生成图像。Star-GAN^[16]实现了基于一个生成器来进行多域之间图像的转换,该模型以图像和目标域的分类标签作为输入,将输入图像转换由类标签指明的域。为了增加生成图像的多样性,

ACGAN^[15]在网络架构中加入辅助分来器,来约束模型生成更多种类的图像。

2 图像生成方法

基于GAN进行图像转换主要包含目标函数的设计、生成器架构的设计、鉴别器架构的设计以及训练算法的设计四个部分,下面对每一个部分进行详细介绍。

2.1 目标函数

GAN最终目的是最小化生成数据分布 P_G 和真实数据分布 P_{Data} 之间的距离,而度量该距离的方式有两种,分别是 f 散度和积分概率度量(Integral Probability Metric, IPM)。相较于 f 散度,IPM度量标准不会受到数据高维度的影响,并且即使是两个分布之间的支撑集没有相应的交集时,该方式也不会出现梯度消失的问题。因此,采用IPM度量标准中的Wasserstein距离来对 P_G 和 P_{Data} 之间的差异进行度量,并且该距离被定义为如式(1)所示:

$$W(P_G, P_{Data}) = \inf_{\gamma \in \Pi(P_G, P_{Data})} E_{(x,y) \sim \gamma} [\|x - y\|] \quad (1)$$

式中: $\Pi(P_G, P_{Data})$ 是由生成数据分布 P_G 和真实数据分布 P_{Data} 组成的联合分布的集合; $\|x - y\|$ 从联合分布 γ 中采样的真实样本 y 和生成样本 x 的距离; $E_{(x,y)} = [\|x - y\|]$ 表示在联合分布 γ 下真实样本 y 和生成样本 x 之间距离的期望值,而在所有的联合分布中该期望的下界就是 P_G 和 P_{Data} 之间的Wasserstein距离。因为直接对式(1)求解比较困难,因此利用Kantorovich-Rubinstein duality的技巧将式(1)转换成式(2)的形式:

$$W(P_G, P_{Data}) = \sup_{\|f\|_k \leq 1} E_{x \sim P_{Data}}(f(x)) - E_{x \sim P_G}(f(x)) \quad (2)$$

式(2)表示评价函数 f 在满足 k 利普希茨连续性约束的条件下, $f(x) - E_{x \sim P_G}(f(x))$ 的上确界就等价于 P_G 和 P_{Data} 之间的Wasserstein距离。并且式(2)中的 f 函数可以用一个 w 参数化的,最后一层神经网络不用非线性激活函数的多层神经网络 f_w 来实现,实际上就是对应对抗网络中的鉴别器。通过约束神经网络的参数 w 不超过某个范围的条件下, P_G 和 P_{Data} 间的Wasserstein距离可以表示成如下形式:

$$d_{wasserstein} = E_{x \sim P_{Data}}[f(x)] - E_{x \sim P_G}[f(x)] \quad (3)$$

由于生成器的目标是最小化 P_G 和 P_{Data} 之间的瓦瑟斯坦距离,而鉴别器的目标是最大化 P_G 和 P_{Data} 之间的瓦瑟斯坦距离。因此,该模型的损失函数可以被设计成式(4)和式(5)所示的形式。

$$\text{生成器 Loss:} \\ \text{Loss}_G = -E_{x \sim P_{Data}}[f_w(G(x'))] + \lambda L_1 \quad (4)$$

$$\text{鉴别器 Loss:} \\ \text{Loss}_D = E_{x \sim P_G}[f_w(y)] - E_{x \sim P_{Data}}[f_w(G(x'))] + \lambda L_1 \quad (5)$$

由于基于成对的数据来对模型进行训练,为了更好地保持输入图像和输出图像之间内容上的一致性,在生成器的损失函数部分加入了由生成样本和真实样本差的1模实现的正则化项,可以表示成如下形式:

$$L_1 = \|y - G(x')\|_1 \quad (6)$$

其中: y 代表真实的样本数据; $G(x')$ 表示生成的样本数据; x' 表示生成器的输入彩色样本数据。

2.2 生成器网络

生成器的目的是基于输入的彩色图像生成服从于真实数据分布 P_{Data} 的红外图像。因为生成器的输入和输出在内容上、物体的位置上是一致的,仅仅在表面上是不一样的,所以两者的高维特征是一致的。基于这样的构想,采用Encoder-Decoder^[22]的神经网络架构来实现生成器。Encoder和Decoder都是基于卷积块(卷积运算-批量正则化-ReLu激活函数)和转置卷积块(转置卷积运算-批量正则化-Leak-ReLu

激活函数)来实现。Encoder的作用是将输入的彩色图像映射成一个高维特征向量,而Decoder的作用是将这个高维特征向量解码成和输入图像内容一致的红外图像。由于生成器输入和输出的许多低层信息是一致的,为了更好地利用输入图像的这些低层信息,在Encoder和Decoder之间加入了跳跃连接来更好地共享这些信息。生成器的网络架构如图2所示,具体说明如表1所示。

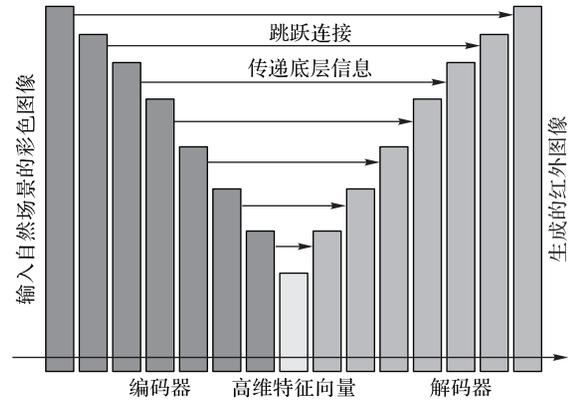


图2 生成器的网络架构

Fig. 2 Network architecture of generator

表1 生成器的编解码神经网络结构

Tab. 1 Encoder-decoder neural network structure of generator

层数	生成器结构
1	卷积块:4×4卷积核,64通道,步长为2
2	卷积块:4×4卷积核,128通道,步长为2
3	卷积块:4×4卷积核,256通道,步长为2
4-8	卷积块:4×4卷积核,512通道,步长为2
9-12	转置卷积块:4×4卷积核,512通道,步长为2
13	转置卷积块:4×4卷积核,256通道,步长为2
14	转置卷积块:4×4卷积核,128通道,步长为2
15	转置卷积块:4×4卷积核,64通道,步长为2
16	卷积块:4×4卷积核,3通道,步长为2

2.3 鉴别器网络

鉴别器的目的是区分真实红外图像样本和生成红外图像样本,其作用等价于一个二分类的问题,因此,可以基于一个多层的卷积神经网络来实现鉴别器。为了更好地捕捉图像中的高频信息,将图像分成 N 小块,然后让鉴别器鉴别小块图像是来自于真实数据分布 P_{Data} 还是来自于生成数据分布 p_G ,最后将鉴别器对 N 小块图像的平均输出作为最终鉴别器的输出。鉴别器的网络结构如图3和表2说明所示。



图3 鉴别器的网络架构

Fig. 3 Architecture of discriminator

表2 鉴别器的二分类神经网络结构

Tab. 2 Two-classification neural network structure of discriminator

层数	鉴别器结构
1	卷积块:4×4卷积核,64通道,步长为2
2	卷积块:4×4卷积核,128通道,步长为2
3	卷积块:4×4卷积核,256通道,步长为2
4	卷积块:4×4卷积核,512通道,步长为1
5	卷积层:4×4卷积核,1通道,步长为1

2.4 训练算法

在训练过程中为了更好地减小生成样本和真实样本之间

的差距,必须先拥有一个优良的鉴别器,因此,在训练对抗网络的过程中更新鉴别器参数 k 次,才更新生成器参数1次。同时,使用批量正则化都的技巧来解决初始化差的问题,在一定程度上缓解生成器过拟合,并且防止生成器将所有的样本都收敛到样本空间中的一点。该节对训练模型的算法进行详细的说明,其伪代码形式如算法1所示。

算法1 红外图像生成模型训练算法。

参数设置:初始化生成器的参数 θ_G 和鉴别器的参数 θ_D ,迭代次数 N 和步数 k 是在训练鉴别器的时候会用到的超参数。

For Number in N :

For Steps in k :

1) 从真实红外图像分布 P_{Data} 中采样 m 样本 x ;

2) 从生成数据分布 P_G 中采样 m 样本 y ;

3) 计算鉴别器的损失函数项和正则化项;

$$V = \frac{1}{m} \sum_{i=1}^m [\lg(D(Y)) - \lg(D(G(X)))] + \lambda L_1$$

4) 通过 V 的梯度来更新鉴别器的参数 θ_D ;

$$\theta_D \leftarrow \theta_D + \eta \nabla V(\theta_D)$$

End For

5) 从彩色图像集中采样 m 样本;

6) 计算生成器的损失函数项;

$$\tilde{V} = \frac{1}{m} \sum_{i=1}^m [-\lg(D(G(X)))] + \lambda L_1$$

7) 通过 \tilde{V} 的梯度来更新生成器参数 θ_G ;

$$\theta_G \leftarrow \theta_G + \eta \nabla \tilde{V}(\theta_G)$$

End For

3 实验和结果分析

为了验证该模型的有效性和泛化性能,在多个数据集上对该模型进行了测试。其中一个数据集RGB-INR是由成对彩色图像和近红外图像组成的针对自然场景的数据集。该数据集包含9个类别477张图像^[23],自然场景类别包括乡村、田地、森林、室内、山地、建筑物、街道、城市以及水等。另一个数据集VEDAI^[24]也是基于成对彩色图像和红外图像组成的空对地拍摄的场景图像,该数据集中的目标除了有尺度大小变化外,还包含多方位、光照、阴影和阻挡的变化。同时,该数据集对同一场景采集了不同波段和分辨率的图像。

3.1 分析目标函数

在实验中为了探索不同正则化对生成图像质量的影响。分别设计了不带正则化、带有L1正则化以及带有L2正则化的三种不同损失函数。如图4所示,图中右边的三列分别对应三种损失函数下生成器生成的红外图像,通过对比可以看出,当目标函数没有正则化的时候,生成的图像比较模糊。当带有正则化时,生成器可以很好地捕捉图像中的细节信息,并且生成图像的质量比较高。

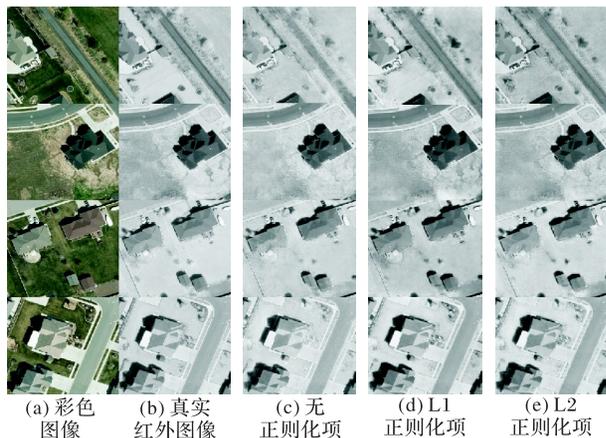


图4 不同损失函数下的实验结果

Fig. 4 Experimental results under different loss functions

3.2 定量评估

对生成图像的质量进行评估是一个比较难的问题。传统的像素均方误差评估标准无法对实验结果的联合统计进行评估,因此没法对损失函数想要捕捉的结构进行合理的测量。所以新的方法FID(Fréchet Inception Distance)^[25]被用于对生成图像的质量和多样性进行评估。该方法的基本思想基于Inception网络^[26]的卷积特征层将真实数据分布 P_G 和生成数据分布 P_{Data} 建模为一个均值为 μ_x, μ_g ,方差为 Σ_x, Σ_g 的多元高斯分布。基于这些信息,FID这种度量方式可以被表示成如下形式:

$$FID(X, G) = \left\| \mu_x - \mu_g \right\|_2^2 + \text{Tr}(\Sigma_x + \Sigma_g - 2(\Sigma_x \Sigma_g)^{\frac{1}{2}}) \quad (7)$$

该度量方式的思想与人类的判断相似,该指标的值越小,表示生成图像的质量和多样性越好。同时该度量方式对噪声不是很敏感,并且可以很好地反映模式崩溃的问题。基于该评估方式对生成图像的定量评估结果如表3所示。

表3 对实验结果的FID定量评估

Tab. 3 Quantitative evaluation of FID on experimental results

数据集	无正则化项	L1正则化项	L2正则化项
RGB-INR	122.743	94.912	97.837
VEDAI	119.531	99.462	102.654

通过上述定量评估指标的比较,可以发现加入正则化使得FID的数值降低,也就意味着模型在正则化的约束下,生成图像的质量有所提高,并且L1正则化的作用比L2正则化更加有效。

3.3 与基于红外仿真方法生成图像的比较

基于仿真技术生成的红外图像,如图5所示,通过与相应场景真实的红外图像比较,发现基于红外仿真技术生成的图像丢失了场景中海水表面的一部分信息,而且场景中舰艇目标的形状相较于真实红外图像中舰艇目标的形状发生了改变。而基于生成对抗网络生成的红外图像,如图5所示,很好地保留了原始图像中各种目标的形状、纹理等信息,同时生成的红外图像和原始红外图像比较接近。但是,基于对抗网络生成的红外图像会存在局部模式崩溃的现象,而基于仿真方法生成的红外图像不会出现模式崩溃的现象。

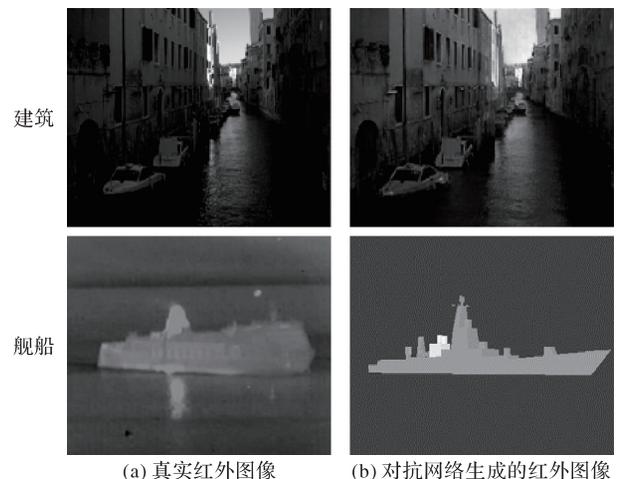


图5 真实红外图像和对抗网络生成的红外图像以及仿真技术生成的红外图像

Fig. 5 Real infrared images, infrared images generated by adversarial neural network and infrared images generated by simulation technology

4 结语

针对实际项目中红外图像少的问题,本文提出了一种基

于生成对抗网络将彩色图像转变为红外图像的方法,从而为模型的训练提供更多可用的训练样本。通过成对的数据训练该模型,使得基于编码器-解码器神经网络架构的生成器具有在高维空间中彩色图像映射为红外图像的能力,而且通过实验证明,该方法可以生成高质量的红外图像。目前该方法只是针对分辨率比较小的图像,实际中可能要求更大分辨率的图像,因此后续将继续改进该方法,进一步对生成红外图像有更加深入的研究。同时,在模型的训练过程中生成器生成的图像存在模式崩溃的现象,而且目前对生成对抗网络生成图像的评估还没有一个统一的标准,因此,希望研究者在将来的研究中对上述两个问题有更加深入的研究。

参考文献 (References)

- [1] DAYAN P. Helmholtz machines and wake-sleep learning [M]// ARBIB M A. Handbook of Brain Theory and Neural Network. Cambridge: MIT Press, 2000:522-525.
- [2] HINTON G E. Deep belief networks [J]. Scholarpedia, 2009, 4(5): No. 5947.
- [3] KINGMA D P, WELING M. Auto-encoding variational Bayes [EB/OL]. [2019-09-22]. <https://arxiv.org/pdf/1312.6114.pdf>.
- [4] SALAKHUTDINOV R, MNH A, HINTON G. Restricted Boltzmann machines for collaborative filtering [C]// Proceedings of the 24th International Conference on Machine Learning. New York: ACM, 2007: 791-798.
- [5] SALAKHUTDINOV R, HINTON G. Deep Boltzmann machines [C]// Proceedings of the 2009 Artificial Intelligence and Statistics. Cambridge: JMLR. org, 2009: 448-455.
- [6] VAN DEN OORD A, KALCHBRENNER N, KAVUKCUOGLU K. Pixel recurrent neural networks [EB/OL]. [2019-09-22]. <https://arxiv.org/pdf/1601.06759.pdf>.
- [7] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]// Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2014: 2672-2680.
- [8] 林懿伦,戴星原,李力,等. 人工智能研究的新前线:生成式对抗网络[J]. 自动化学报, 2018, 44(5): 775-792. (LIN Y L, DAI X Y, LI L, et al. The new frontier of AI research: generative adversarial networks[J]. Acta Automatica Sinica, 2018, 44(5): 775-792.)
- [9] 曹仰杰,贾丽丽,陈永霞,等. 生成式对抗网络及其计算机视觉应用研究综述[J]. 中国图象图形学报, 2018, 23(10): 1433-1449. (CAO Y J, JIA L L, CHEN Y X, et al. Review of computer vision based on generative adversarial networks[J]. Journal of Image and Graphics, 2018, 23(10):1433-1449.)
- [10] 陈文兵,管正雄,陈允杰. 基于条件生成式对抗网络的数据增强方法[J]. 计算机应用, 2018, 38(11):3305-3311. (CHEN W B, GUAN Z X, CHEN Y J. Data augmentation method based on conditional generative adversarial net model [J]. Journal of Computer Applications, 2018, 38(11):3305-3311.)
- [11] CHEN X, DUAN Y, HOUTHOOFT R, et al. InfoGAN: interpretable representation learning by information maximizing generative adversarial nets [C]// Proceedings of the 30th International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc., 2016: 2172-2180.
- [12] ZHANG H, GOODFELLOW I, METAXAS D, et al. Self-attention generative adversarial networks [EB/OL]. [2019-09-22]. <https://arxiv.org/pdf/1805.08318.pdf>.
- [13] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 5967-5976.
- [14] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 2242-2251.
- [15] ODENA A, OLAH C, SHLENS J. Conditional image synthesis with auxiliary classifier GANs [C]// Proceedings of the 34th International Conference on Machine Learning. New York: JMLR. org, 2017: 2642-2651.
- [16] CHOI Y, CHOI M, KIM M, et al. StarGAN: unified generative adversarial networks for multi-domain image-to-image translation [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8789-8797.
- [17] 许洪,王向军,刘峰,等. 基于可见光光谱图像的红外多光谱图像仿真生成[J]. 红外与激光工程, 2009, 38(2):200-204. (XU H, WANG X J, LIU F, et al. Infrared multispectral image simulation based on spectral images in visible bands [J]. Infrared and Laser Engineering, 2009, 38(2):200-204.)
- [18] 陈珊,孙继银. 基于可见光图像的红外场景仿真[J]. 红外与激光工程, 2009, 38(1):23-26, 30. (CHEN S, SUN J Y. IR scene simulation based on visual image [J]. Infrared and Laser Engineering, 2009, 38(1):23-26, 30.)
- [19] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein GAN [EB/OL]. [2019-09-22]. <https://arxiv.org/pdf/1701.07875.pdf>.
- [20] BROCK A, DONAHUE J, SIMONYAN K. Large scale GAN training for high fidelity natural image synthesis [EB/OL]. [2019-09-22]. <https://arxiv.org/pdf/1809.11096.pdf>.
- [21] WANG T C, LIU M Y, ZHU J Y, et al. High-resolution image synthesis and semantic manipulation with conditional GANs [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8798-8807.
- [22] NG A. Sparse autoencoder [EB/OL]. [2019-10-20]. https://web.stanford.edu/class/cs294a/sparseAutoencoder_2011new.pdf.
- [23] CHOE G, KIM S H, IM S, et al. RANUS: RGB and NIR urban scene dataset for deep scene parsing [J]. IEEE Robotics and Automation Letters, 2018, 3(3): 1808-1815.
- [24] SAKLA W, KONJEVOD G, MUNDHENK T N. Deep multi-modal vehicle detection in aerial ISR imagery [C]// Proceedings of the 2017 IEEE Winter Conference on Applications of Computer Vision. Piscataway: IEEE, 2017: 916-923.
- [25] HEUSEL M, RAMSAUER H, UNTERTHINER T, et al. GANs trained by a two time-scale update rule converge to a local Nash equilibrium [C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc., 2017: 6626-6637.
- [26] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4: inception-ResNet and the impact of residual connections on learning [C]// Proceedings of the 31st AAAI conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2017: 4278-4284.

This work is partially supported by the National Natural Science Foundation of China (U1713216).

CHEN Foji, born in 1994, M. S.. His research interests include image generation, machine learning, robot vision.

ZHU Feng, born in 1962, Ph. D., research fellow. His research interests include robot vision.

WU Qingxiao, born in 1978, Ph. D., research fellow. His research interests include robot vision, machine vision.

HAO Yingming, born in 1966, Ph. D., research fellow. Her research interests include visual localization, image optimization, vision inspection, infrared image simulation, 3D imaging, data processing.

WANG Ende, born in 1980, Ph. D., research fellow. His research interests include image target recognition, detection and tracking, weak signal detection.