

小样本红外图像的样本扩增与目标检测算法

吴 晗, 张志龙[†], 李楚为, 李航宇

(国防科技大学 电子科学学院 自动目标识别重点实验室, 湖南 长沙 410073)

摘要: 深度卷积神经网络模型在很多公开的可见光目标检测数据集上表现优异, 但是在红外目标检测领域, 目标样本稀缺一直是制约检测识别精度的难题. 针对该问题, 本文提出了一种小样本红外图像的样本扩增与目标检测算法. 采用基于注意力机制的生成对抗网络进行红外样本扩增, 生成一系列保留原始可见光图像关键区域的红外连续图像, 并且使用空间注意力机制等方法进一步提升YOLOv3目标检测算法的识别精度. 在Grayscale-Thermal与OSU Color-Thermal红外-可见光数据集上的实验结果表明, 本文算法使用的红外样本扩增技术有效提升了深度神经网络对红外目标检测的精度, 与原始YOLOv3算法相比, 本文算法最高可提升近20%的平均精确率(mean average precision, mAP).

关键词: 红外图像; 目标检测; 稀缺样本; 生成对抗网络; 注意力机制; YOLOv3算法

引用格式: 吴晗, 张志龙, 李楚为, 等. 小样本红外图像的样本扩增与目标检测算法. 控制理论与应用, 2021, 38(9): 1477 – 1485

DOI: 10.7641/CTA.2021.10057

Infrared image sample amplification and object detection method with small samples

WU Han, ZHANG Zhi-long[†], LI Chu-wei, LI Hang-yu

(National Key Laboratory of Science and Technology on ATR, College of Electronic Science and Technology, National University of Defense Technology, Changsha Hunan 410073, China)

Abstract: The deep convolutional neural network model performs well on many public visible-lighted object detection datasets, but in the field of infrared object detection, the scarcity of object samples has always been a problem that plagues the increase in detection and recognition accuracy. In response to this problem, this paper proposes an algorithm for sample amplification and object detection of infrared images with small samples. The attention-based generative adversarial network was adopted to amplify the infrared samples to generate a series of infrared continuous images retaining the key areas of the original visible light image, and the spatial attention model and other methods were used to further improve the recognition accuracy of the YOLOv3 object detection algorithm. The experimental results on the OSU Color-Thermal infrared-visible light dataset and Grayscale-Thermal dataset show that the infrared data amplification technology of the proposed algorithm effectively improves the accuracy of the deep convolutional neural network for infrared object detection, and the mAP (mean average precision) of the proposed method is 20% higher than the original YOLOv3 algorithm.

Key words: infrared images; object detection; scarce samples; generative adversarial network; attention model; YOLOv3 algorithm

Citation: WU Han, ZHANG Zhilong, LI Chuwei, et al. Infrared image sample amplification and object detection method with small samples. *Control Theory & Applications*, 2021, 38(9): 1477 – 1485

1 引言

红外成像系统利用目标和环境间的红外辐射差异成像, 具有动态范围大、穿透性强、温度灵敏度高、可昼夜工作等优点. 因此利用红外成像设备侦察是当代

军事领域中感知敌方战场的重要手段, 并且随着我国军民融合政策的实施, 大量的红外成像系统被利用在社区防盗监控或者大型工厂、企业的设备安全监测等, 目前使用红外监测图像且结合目标检测的分析手段

收稿日期: 2021-01-16; 录用日期: 2021-04-23.

[†]通信作者. E-mail: zhangzhilong@nudt.edu.cn; Tel.: +86 15730307128.

本文责任编辑: 潘泉.

国家自然科学基金项目(61101185), 湖南省研究生科研创新项目(CX20200044)资助.

Supported by the National Natural Science Foundation of China (61101185) and the Postgraduate Scientific Research Innovation Project of Hunan Province (CX20200044).

在民用和军事领域的重要性日益提高。

在计算机视觉领域,基于传统手工特征设计的目标检测器在目标检测任务上展现出了较好的性能.在深度学习兴起之前,手工特征提取算法一直占据了重要地位,典型代表为:SIFT^[1]算法、HOG^[2]算法、neocognitron^[3]算法等.其中,Fukushima提出的neocognitron算法是一种具有生物视觉启发意义的平移不变性模型,这一模型引发了学者们对传统目标检测领域中手工提取特征等方法的反思.

随着深度学习技术的发展,LeCun证明随机梯度下降通过反向传播^[4]可以训练卷积神经网络,这一结论大大加速了卷积神经网络在目标检测领域上的应用.Ross Girshick的R-CNN模型^[5]第一次将卷积神经网络运用到目标检测领域,其两阶段进行检测识别的思想也深刻影响了后来的Fast R-CNN算法^[6]与Faster R-CNN算法^[7].针对两阶段目标检测算法存在的运行速度较慢和难以实现工程化等问题,YOLO算法^[8]和SSD算法^[9]先后被提出,它们将目标检测任务视为一个回归问题,在不影响检测精度的同时大幅度地提高了检测速度.

注意力机制是人类视觉系统固有的一种筛选图像信息和聚焦显著物体的信息处理机制^[10],深度学习中注意力机制借鉴了人类的视觉注意力思维方式.人类可以通过快速浏览整体图像,迅速找出需要重点关注的区域,从而抑制次要信息对整体的影响,极大地提高了视觉信息处理的效率与准确性.深度卷积神经网络在经过多层卷积处理后将产生大量的高维特征图,这些由各层网络生成的特征图包含输入图像的各种纹理、特征、形态等信息^[11].Jie Hu的SE模型^[12]将这些饱含信息的特征图在传递给更深层网络前赋予不同的权值.Sanghyun的CBAM模型^[13]通过并行的特征信息编码将特征图赋予相应的权值.上述模型都能在增加相应计算量的前提下,使最终深度网络模型的输出更多地受到包含关键区域特征图的影响.

军事复杂战场环境下获取大量完备红外图像数据进行作战分析的难度较高,因此基于深度学习的红外目标检测算法的性能一直难以获得有效提升.近年来有些学者提出了一些具有启发意义的样本扩增技术,Alex提出的PCA Jittering算法^[14]采用对原始数据通道做PCA处理以及对其主成分添加高斯噪声的策略,使得训练得到的模型降低了1%的错误率.Zhang提出了一种融合现有样本、扩展训练分布的数据增强方法^[15],通过构建虚拟的训练样本,增强样本之间的线性表达,提升神经网络的泛化能力.上述方法只是缓解了深度学习模型过拟合的程度,并没有从根本上对红外数据本身进行有效增强,于是有些研究人员开始研究如何对已获得的红外数据进行扩增处理,试图通过扩增的红外样本来完善并重构原始红外环境状况.

本文提出了一种小样本条件下红外图像的样本扩增与目标检测算法.该算法通过生成对抗网络寻找可见光域与红外域的高维数据表征,并采用注意力机制将图像从可见光域映射至红外域,在保留关键区域的前提下实现完备的图像风格迁移.这一策略能够将样本数量充足、视角信息丰富的可见光图像转换为对应的红外图像,从而有效解决红外图像数据集样本稀缺的难题.最后用生成对抗网络生成的红外数据和原始红外样本作为训练集,训练改进后的YOLOv3模型,进一步地提高红外目标检测的精度.结合上述2种策略,本算法在小样本条件下的红外场景上实现了具有良好鲁棒性、准确的目标检测.

2 相关工作

生成对抗网络(generative adversarial network, GAN)^[16]旨在生成不存在于真实世界的数据.在传统图像仿真领域,图像的仿真需经过建模、添加纹理和光照、渲染等一系列复杂的步骤.GAN网络的出现简化了这一过程,其利用深度神经网络模型构建生成器和判别器,生成器负责生成伪样本,而判别器用于分辨样本的真伪,生成器和判别器之间采用对抗博弈的方法自动学习样本数据的真实分布,从而实现图像的样本扩增,GAN网络的结构示意图如图1所示.

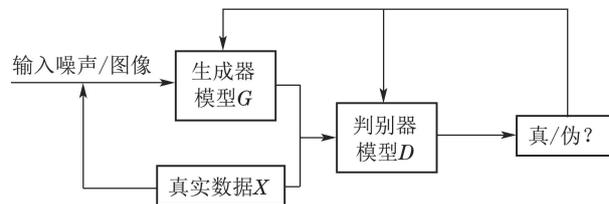


图1 GAN网络结构示意图

Fig. 1 Generative adversarial network structure diagram

在动态博弈的过程中,GAN网络的生成器和判别器取样于随机噪声.这样生成的伪样本状态并不可控,即生成的红外域样本图像的具体内容并不受用户控制.此外,由随机噪声生成的伪样本还存在图像边缘不清晰、细节模糊等问题.

为解决GAN网络生成图像状态不可控的问题,目前主流的方法是对输入样本进行约束性选择.以pix2pix^[17]算法为例,该算法有条件地使用用户输入样本,将成对的数据进行网络模型的训练,从而让模型寻找原始域和目标域之间的映射关系.然而,在复杂战场环境下,具有价值的红外样本数据的获取较为困难,因此使用大量成对的可见光原始域与红外目标域样本数据进行网络模型的训练这一设想是难以实现的.

当可见光和红外数据的样本数量不匹配时,pix2pix算法难以进行有效训练,因此有学者提出了CycleGAN算法^[18]来解决这一数据显示配对问题.然而CycleGAN网络在训练过程中容易产生模型崩溃的现

象,即原始域输入图像都映射到相同的目标域输出图像,并且优化无法取得有效的进展。

针对上述GAN网络中存在的问题,本文算法在CycleGAN算法的基础上进行改进,利用其未成对数

据定义网络的输入和输出策略(如图2所示),并使用注意力机制区分源域和目标域,帮助生成对抗网络模型集中转换关键区域,从而有效缓解高维数据跨域重复映射问题。

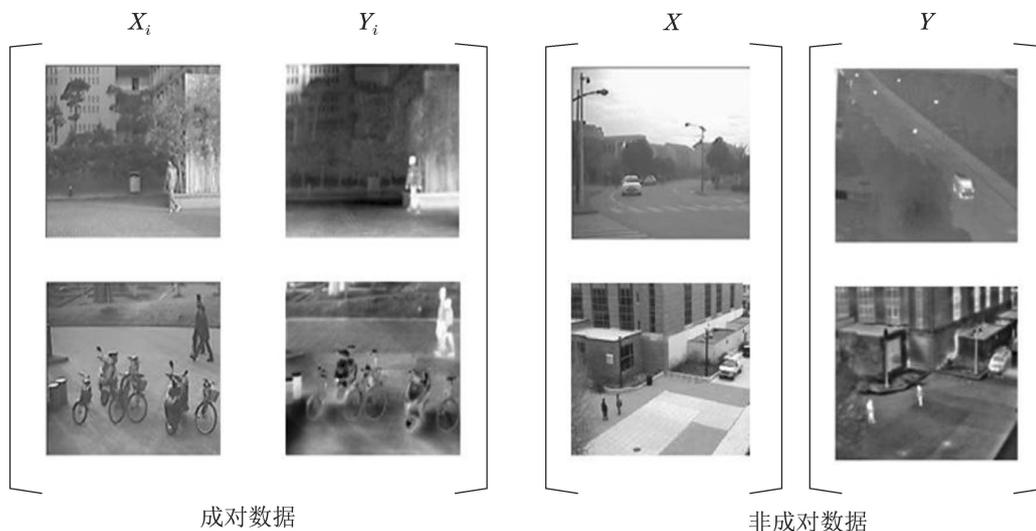


图2 由成对或未成对数据定义网络输入输出示意图

Fig. 2 Network input and output diagrams are defined by paired or unpaired data

3 基于生成对抗网络的样本扩增

3.1 网络架构

本文中生成对抗网络采用未成对样本数据训练网络模型,需给网络增加限制条件使得网络学习高维数据间的映射关系.为了将可见光域的图像转换为红外域的图像,本文采用2个判别器和2个生成器构成循环的网络结构,2个生成器 G_1 和 G_2 用于对可见光域和红外域的图像进行相互转换,2个判别器 D_1 和 D_2 分别辨别红外域和可见光红外域图像的真伪。

如图3所示,可见光域图像 x 经过生成器 G_1 后生成一幅红外域上的伪样本图像 $G_1(x)$,本文利用判别器 D_1 来辨别真伪后得出判别器 D_1 的损失函数 L_{D_1} ,接着将这个伪样本再次经过生成器 G_2 生成一幅可见光域上的重构图像 $G_2(G_1(x))$,然后将重构图像 $G_2(G_1(x))$ 同原始输入图像 x 进行对比分析得出重构损失 L_1 .既然网络构成一个环形结构,那么输入图像若为红外域图像 y ,也将经过上述类似的过程得到判别器损失函数和重构损失 L_2 ,具体流程如图4所示.网络模型在训练的过程中不断优化目标损失函数,从而学习到高维数据的特征分布,并构建出可见光域至红外域的映射关系。

$$L_{D_1} = E_{y:\text{data}(y)}[\log D_1(y)] + E_{x:\text{data}(x)} \cdot [\log(1 - D_1(G_1(x)))], \quad (1)$$

$$L_{D_2} = E_{x:\text{data}(x)}[\log D_2(x)] + E_{y:\text{data}(y)} \cdot [\log(1 - D_2(G_2(y)))], \quad (2)$$

$$L_1 = E_{x:\text{data}(x)}[\|G_2(G_1(x)) - x\|_1], \quad (3)$$

$$L_2 = E_{y:\text{data}(y)}[\|G_1(G_2(y)) - y\|_1]. \quad (4)$$

通过将2个单向的生成对抗网络耦合成一个环状网络结构,可以构建可见光域至红外域的映射关系.加入重构损失函数后,生成对抗网络在训练的过程中给生成伪样本的生成器添加了强制性约束,即真实样本与通过连续的生成器所生成的伪样本间的差异应尽可能的小.这可以在一定程度上缓解生成对抗网络中普遍存在的模式崩溃问题,即来自一个域的多个模式的数据映射到另一个域的单个模式,从而实现可见光域图像至红外域图像的风格迁移。

3.2 生成器网络结构设计

本文算法将注意力机制模型添加至生成对抗网络的生成器网络结构中.如图5所示,将可见光域图像输入生成器下采样后,将经过多个残差块处理后得到的特征图送入注意力机制模型.模型为包含不同区域关键信息的特征图赋予相应的权值,最后经残差块处理后,上采样得到红外域生成图像。

生成器中注意力机制模型的结构图如图6所示,模型将输入的特征图经过全局平均池化和全局最大池化操作得到依托通道数的特征向量.将特征向量经过全连接层、拼接处理后送入分类器做源域和目标域的分类,反向传播学习到特征图对应的权值 w_n , n 为特征图对应的通道数.最后, w_n 和特征图进行位乘得到经过注意力机制增强的特征图,使生成器着重于可见

光域图像的关键区域信息以实现红外风格迁移.

4 YOLOv3算法的网络改进

4.1 骨干网络的注意力机制改进

YOLOv3算法能够取得比其他单阶段目标检测算法更好的检测精度,其关键在于YOLOv3的基础网络骨干(backbone): DarkNet53. 本文采用注意力机制模型改进DarkNet53,使其在通道和局部空间维度上对特征图进行基于注意力的加权,使得模型能更加关注输入图像的关键区域.

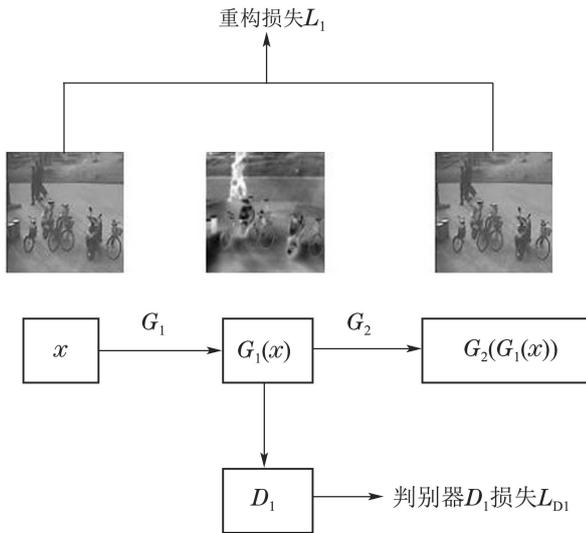


图3 输入为可见光域样本的单网络结构示意图
Fig. 3 The input is the schematic diagram of the single network structure of the visible sample

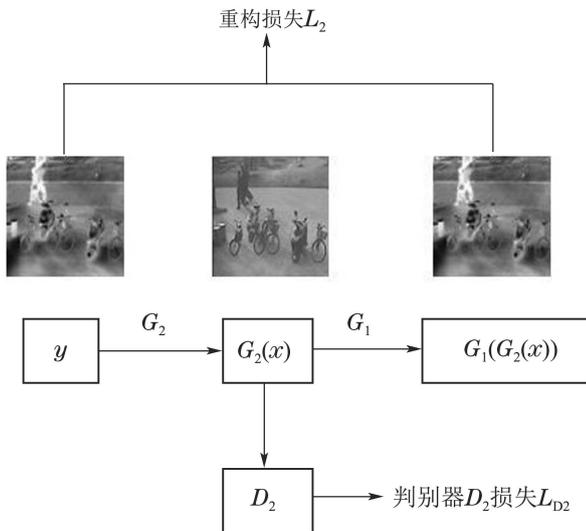


图4 输入为红外域样本的单网络结构示意图
Fig. 4 The input is the schematic diagram of the single network structure of the infrared sample

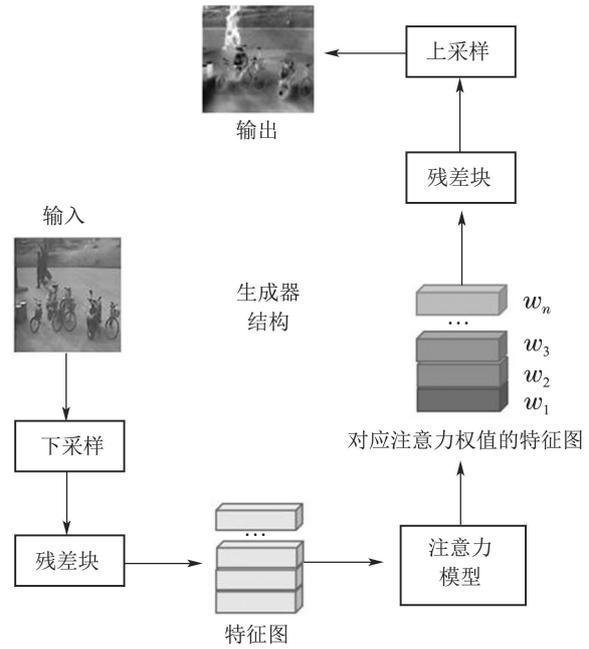


图5 生成器网络的结构示意图
Fig. 5 Schematic diagram of generator network

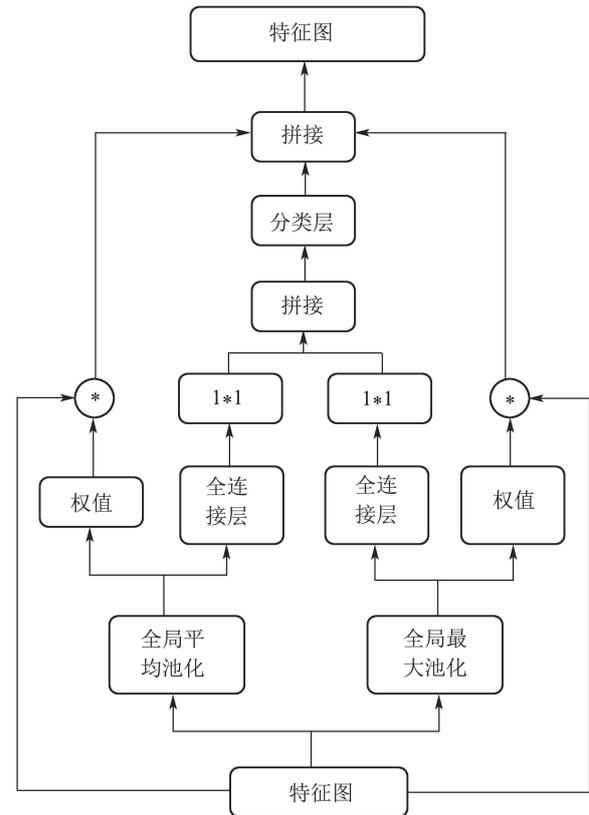


图6 生成器的注意力模型结构示意图
Fig. 6 Schematic of the attention model for the generator

如图7所示,本文算法将DarkNet53通过残差块卷积操作得到的特征图输入到通道注意力模型中,经过全局平均池化操作后得到当前特征图的全局压缩特征向量,通过激活函数Relu后再接入由1*1卷积

层构建的瓶颈结构(bottleneck). 如式(5)所示, w_c 为经上述池化与卷积操作后, 输入特征图的各个通道所对应的权重回归值; c 为特征图所对应的通道. w_c 经 Sigmoid 层归一化后得 weight 与原始输入的特征图进行位乘, 其中特征图中包含语义特征、目标边缘等信息更丰富的通道(weight 值越大)得到更高程度的响应, 最终将加权得到的特征图输入到空间注意力机制模型中.

$$\text{weight} = \text{Sigmoid}(w_c) = \frac{1}{1 + e^{-w_c}}. \quad (5)$$

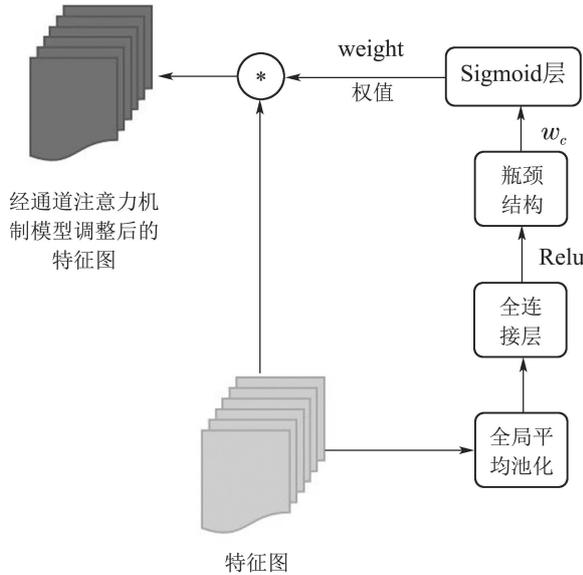


图 7 通道注意力模型示意图

Fig. 7 Schematic diagram of channel attention model

空间注意力模型以通道注意力模型的输出为输入, 如图8所示. 在完成基于通道并行的全局最大池化和全局平均池化操作后, 模型将两个操作产生的特征向量拼接, 输入到后续卷积层进行处理, 如式(6)所示, $w_{(i,j)}$ 为经上述并行编码等操作后, 输入特征图的各个区域所对应的权重回归值; (i, j) 为特征图所对应的空间区域. 经 Sigmoid 层归一化后输出空间注意力特征 w . 最后, 将 w 与模块的输入进行位乘, 得到最终基于通道与空间注意力机制的特征图.

$$w = \text{Sigmoid}(w_{(i,j)}) = \frac{1}{1 + e^{-w_{(i,j)}}}. \quad (6)$$

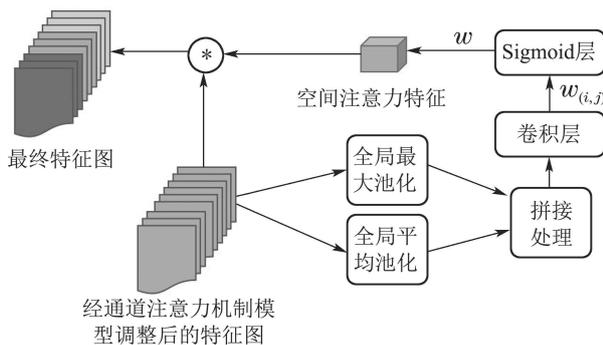


图 8 空间注意力模型示意图

Fig. 8 Schematic diagram of spatial attention model

基于注意力机制改进的YOLOv3算法在后续进行边界回归任务的过程中, 会更多地关注输入图像中包含目标的关键区域, 即经过上述注意力网络模型的池化与卷积等操作后, 特征图中对应语义信息丰富的区域得到更高程度的响应, 从而使得模型在训练的过程中对于包含目标的关键图像区域具有选择性, 最终提高神经网络对于目标检测任务的泛化能力.

不同于仅对特征图通道作加权处理的注意力机制模型, 添加空间注意力机制后, 模型将降低背景信息对目标检测任务的干扰, 从而提高目标检测模型的鲁棒性和检测精度.

4.2 学习率设置

在神经网络训练的过程中, 通常采用梯度下降的方式来优化目标损失函数, 而越接近损失函数的全局最优解时, 其学习率应设置为较小值, 从而使得深度网络损失函数尽可能收敛至全局最优点. 但是随着学习率衰减, 深度网络的训练容易陷入局部最优困境, 这将导致最终模型泛化能力的不足. 而余弦退火(cosine annealing)^[19]通过余弦函数来降低学习率, 并且在模拟退火与热重启的过程中可以使得学习率有一个不断变化的趋势, 这种学习率下降模式能够有效避免模型陷入局部最优解.

本文采用余弦退火算法来优化YOLOv3网络训练过程中学习率的衰减方式, 即学习率 α 在单个退火过程中余弦衰减, T_{\max} 次迭代后进行热重启, 以此循环直至网络训练完成. 新的学习率衰减函数如式(7)所示, α_{\max} 和 α_{\min} 分别表示学习率的最大值和最小值, T_{\max} 为单个退火过程中算法需要执行的总步数, T_{cur} 为单个退火过程中已经迭代的步数.

$$\alpha = \alpha_{\min} + \frac{1}{2}(\alpha_{\max} - \alpha_{\min})(1 + \cos(\frac{T_{\text{cur}}}{T_{\max}}\pi)). \quad (7)$$

5 实验结果与分析

本文中实验所采用的硬件平台为: Intel i5-9400 CPU@2.90GHz, NVIDIA 2080TI显卡、32GB内存. 操作系统为Ubuntu 16.04, 深度学习框架为Pytorch, 软件环境为CUDA 10.0, CUDNN 7.4, 训练过程中采用余弦退火算法优化的学习率设置为

$$\begin{cases} T_{\max} = \text{epoch} * \text{num_train} / \text{batch_size}, \\ \alpha_{\max} = 0.001, \alpha_{\min} = 0.0001, \end{cases} \quad (8)$$

其中: epoch 设置为200, batch_size 设置为8, num_train 为训练集图像总数量.

5.1 样本扩增实验结果分析

本部分实验在公开的 Grayscale-Thermal 数据集^[20]与 OSU Color-Thermal 红外-可见光数据集^[21]上展开, 利用2个数据集集中的1000对(尽可能地多)可见光-红外域图像数据进行生成对抗网络的训练, 其

中1000对可见光-红外域图像数据采用图2中未成对数据的匹配方式. 训练迭代10000次的训练结果如图9所示, 由图9可以看出, 生成对抗网络生成的行人和车辆序列红外样本数据保留了可见光域行人和车辆图像序列的连续性.

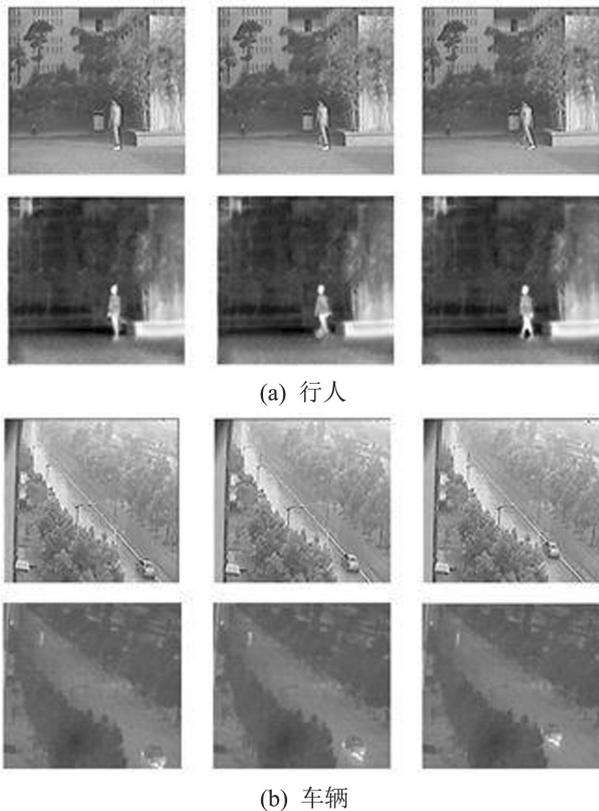


图9 生成对抗网络样本扩增结果

Fig. 9 The amplification result of the generative adversarial network sample

本文将有无注意力机制模型改进的生成对抗网络样本扩增结果进行对比分析, 如图10所示. 实验结果表明: 经注意力机制模型改进的生成对抗网络在对含多目标、小目标的可见光图像进行红外风格迁移时, 能够更好地关注可见光域图像中包含目标的关键区域, 减少数据信息的丢失.

5.2 红外-可见光数据集目标检测实验结果分析

本节实验中所使用的训练集与测试集图像均来源于Grayscale-Thermal与OSU Color-Thermal红外-可见光数据集, 数据集中主要包含行人和车辆目标. 训练集中, 红外域图像2000幅, 可见光域图像2000幅(其中1000幅可见光域图像与1000幅红外图像一一配对).

在本节实验中, 本文首先利用第5.1节中训练好的CycleGAN网络与改进后的生成对抗网络对未配对的可见光图像进行红外风格迁移, 得到2组1000幅扩增后的红外样本. 在此基础上, 展开后续对照试验: 1) 用2000幅可见光图像对原始YOLOv3进行训练;

2) 用2000幅真实红外域图像对原始YOLOv3进行训练; 3) 用2000幅真实红外域图像对改进的YOLOv3进行训练; 4) 用1000幅未配对的真实红外域图像对原始YOLOv3进行训练; 5) 用1000幅未配对的真实红外域图像对改进后的YOLOv3进行训练; 6) 用1000幅未配对的真实红外域图像和1000幅由CycleGAN网络迁移得到的红外域图像作为训练集, 对原始YOLOv3进行训练; 7) 用1000幅未配对的真实红外域图像和1000幅由CycleGAN网络迁移得到的红外域图像作为训练集, 对改进后的YOLOv3进行训练; 8) 用1000幅未配对的真实红外域图像和1000幅由改进的生成对抗网络迁移得到的红外域图像作为训练集, 对原始YOLOv3进行训练; 9) 用1000幅未配对的真实红外域图像和1000幅由改进的生成对抗网络迁移得到的红外域图像作为训练集, 对改进后的YOLOv3进行训练; 10) 用2000幅真实红外域图像和1000幅由改进的生成对抗网络迁移得到的红外域图像作为训练集, 对改进后的YOLOv3进行训练.

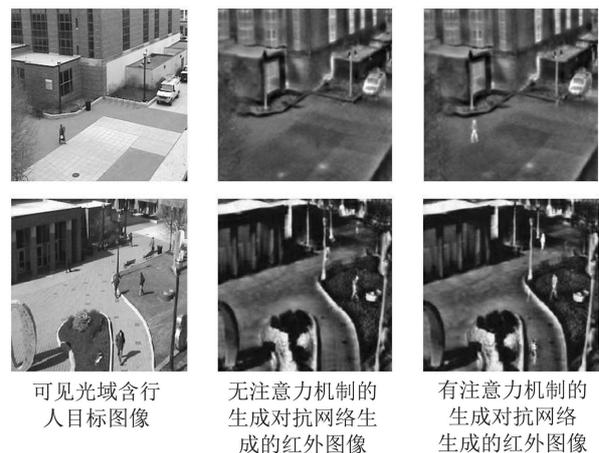


图10 有无注意力机制的GAN网络样本扩增结果对比

Fig. 10 Comparison of amplification results of GAN network samples with or without attention model

表1与表2展示了在测试集上(测试集的划分: 红外图像500幅; 可见光图像500幅)不同的图像域、不同数量的训练集、不同的网络结构对目标检测识别性能(平均精准度(average precision, AP); 所有类别下AP的均值(mean average precision, mAP))与检测速率(FPS: 每秒检测图像数)的影响.

表3展示了用2000幅红外域图像作为训练集对现有的一些小样本目标检测算法进行训练, 最终在测试集上(测试集的划分: 红外图像500幅)对目标检测识别性能与检测速率的对比分析.

实验结果表明: 1) 在红外稀缺样本条件下, 本文所提出的生成对抗网络进行红外样本扩增方法能有效地提升目标检测精度; 2) 经本文改进的YOLOv3算法仍可在原有的检测精度基础上提升近2%.

表 1 各算法的环境与检测速率

Table 1 The environment and detection rate of each algorithm

算法	图像域	训练集数量	FPS
原始YOLOv3	可见光	2000	31
原始YOLOv3	红外	2000	31
改进的YOLOv3	红外	2000	29
原始YOLOv3	红外	1000	31
改进的YOLOv3	红外	1000	29
CycleGAN样本扩增+原始YOLOv3	红外	1000+1000	31
CycleGAN样本扩增+改进YOLOv3	红外	1000+1000	29
生成对抗网络样本扩增+原始YOLOv3	红外	1000+1000	31
生成对抗网络样本扩增+改进YOLOv3	红外	1000+1000	29
生成对抗网络样本扩增+改进YOLOv3	红外	2000+1000	29

表 2 在测试集上各算法中行人和车辆检测的AP与mAP

Table 2 AP and mAP of person and car detection are implemented in each algorithm on the test set

算法	行人AP/%	车辆AP/%	mAP/%
原始YOLOv3	92.84	89.02	90.93
原始YOLOv3	91.25	87.13	89.19
改进的YOLOv3	93.73	88.31	91.02
原始YOLOv3	73.27	65.54	69.41
改进的YOLOv3	76.33	66.27	71.30
CycleGAN样本扩增+原始YOLOv3	83.41	76.85	80.13
CycleGAN样本扩增+改进YOLOv3	85.82	79.94	82.88
生成对抗网络样本扩增+原始YOLOv3	87.09	89.45	88.27
生成对抗网络样本扩增+改进YOLOv3	91.15	90.18	90.67
生成对抗网络样本扩增+改进YOLOv3	95.11	90.39	92.75

表 3 在测试集上各小样本目标检测算法指标对比

Table 3 The indexes of each object detection algorithm with small samples are compared on the test set

算法	Backbone	mAP/%	FPS
本算法	DarkNet53-Attention-FPN	92.75	29
LSTD ^[22]	Vgg16/ResNet-101	88.64	18
RepMet ^[23]	InceptionV3-FPN	93.12	21
Meta-SSD ^[24]	Vgg16	71.89	26

图11展示了本文算法在测试集上的目标检测结果. 可以看出, 在复杂场景下, 当红外域样本中出现多个目标且目标尺寸较小时, 本文算法仍具有良好的检测效果.



图 11 本文算法在测试集上的目标检测结果

Fig. 11 The algorithm of this paper on the test set of the object detection results

5.3 通用数据集目标检测实验结果分析

为了验证本文所改进的目标检测算法的通用性, 本文在公开的Microsoft Common Objects in Context (COCO)数据集^[25]与PASCAL-VOC2012数据集^[26]上展开目标检测精度评估实验. 在实验过程中所使用的COCO数据集包含80个目标类别, 82783张图像作为训练集, 20504张图像作为验证集, 10775张图像作为测试集, 平均每幅图像有5个标签信息. 所使用的VOC数据集包含20个目标类别, 总共包含5717张图像, 平均每幅图像有2个标签信息. 在实验过程中, VOC数据集划分为: 4002张图像作为训练集, 572张图像作为验证集, 1143张图像作为测试集.

IOU(交并比)衡量的是2个区域的重叠程度, 是2个区域重叠部分面积占二者总面积的比例, 常用于目标检测数据集中测量真实框与预测框之间的相关度, 相关度越高, 该值越大.

表4与表5分别展示了在COCO数据集与VOC数据集上, 本文改进后的YOLOv3算法与当前主流目标检测算法的目标检测精度与速度对比(在COCO数据集上测试IOU=0.75下的mAP; 在VOC数据集测试IOU=0.5下的mAP).

实验结果表明: 1) 本文改进后的YOLOv3算法与当前主流目标检测算法相比, 具有较好的目标检测性能; 2) 本文改进后的YOLOv3算法可在提升目标检测精度的同时保证检测速率, 基本可以满足工程应用前景下嵌入式设备部署所需求的高效性与实时性.

6 结论

本文提出了一种小样本红外图像的样本扩增与目标检测算法. 该算法将生成对抗网络应用于红外图像的样本扩增, 并与改进的YOLOv3网络相结合, 有效地提高了红外目标检测的性能. 实验结果表明, 该算

法可以在保留可见光域图像关键区域的前提下实现红外风格迁移,在Grayscale-Thermal与OSU Color-Thermal红外-可见光数据集上的检测精度比原始YOLOv3算法提升了近20%。下一步将继续优化生成对抗网络模型的结构,以生成分辨率更高、细节更丰富的红外图像。

表4 在COCO数据集上各目标检测算法指标对比

Table 4 The indexes of each object detection algorithm are compared on COCO data set

算法	Backbone	mAP/%	FPS
本算法	DarkNet53-Attention-FPN	40.1	29
YOLOv3	DarkNet53	38.4	31
SSD300	Vgg16	29.6	26
Faster-RCNN	ResNet-101-FPN	36.9	7
FCOS	ResNet-101-FPN	41.8	10

表5 在VOC数据集上各目标检测算法指标对比

Table 5 The indexes of each object detection algorithm are compared on VOC data set

算法	Backbone	mAP/%	FPS
本算法	DarkNet53-Attention-FPN	79.3	29
YOLOv2	DarkNet53	71.8	27
YOLOv3	DarkNet53	77.3	31
SSD300	Vgg16	69.5	26
SSD512	ResNet-101-SSD	74.2	16
Faster-RCNN	ResNet-101-FPN	76.7	7
Mask-RCNN	ResNet-151-FPN	77.4	6
FCOS	ResNet-101-FPN	80.5	10

参考文献:

- [1] LOWE D. Distinctive image features from scaleinvariant keypoints. *International Journal of Computer Vision*, 2004, 60(2): 91 – 110.
- [2] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. San Diego, USA: IEEE, 2005: 886 – 893.
- [3] FUKUSHIMA K. Neocognitron: a self-organizingneural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 1980, 36(4): 193 – 202.
- [4] LECUN Y, BOSER B, DENKER J, et al. Backpropagation applied to handwritten zip code recognition. *Neural Computing*, 1989, 4: 541 – 551.
- [5] ROSS G, JEFF D, TREVOR D. Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, USA: IEEE, 2014: 580 – 587.
- [6] ROSS G. FSAT R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*. Santiago: IEEE, 2015: 1440 – 1448.
- [7] REN S Q, HE K M, ROSS G, et al. Faster R-CNN: towards real-time object detection with region proposal networks. *The Neural Information Processing Systems Conference*. Montreal, Canada: NIPS, 2015.
- [8] JOSEPH R, SANTOSH D, ROSS G, et al. You only look once: unified, real time object detection. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA: IEEE, 2016: 779 – 788.
- [9] LIU W, DRAGOMIR A, SUMITRA E, et al. SSD: single shot multi-box detector. *European Conference on Computer Vision*. Amsterdam, Netherland: ECCV, 2016.
- [10] LI Chuwei, ZHANG Zhilong, YANG Weiping. Salient object detection method by combining boolean map and grayscale rarity. *Journal of Image and Graphics*, 2020, 25(2): 267 – 281.
(李楚为, 张志龙, 杨卫平. 结合布尔图和灰度稀缺性的小目标显著性检测. *中国图象图形学报*, 2020, 25(2): 267 – 281.)
- [11] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks. *European Conference on Computer Vision*. Zurich, Switzerland: ECCV, 2014.
- [12] HU J, SHEN L, SUN G. Squeeze-and-excitation networks. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE, 2018: 7132 – 7141.
- [13] SANGHYUN W, JONGCHAN P, JOON Y L, et al. CBAM: convolutional block attention module. *Proceedings of the IEEE International Conference on Computer Vision*. USA: IEEE, 2018: 3 – 19.
- [14] ALEX K, ILYA S, GEOFFREY E H. ImageNet classification with deep convolutional neural networks. *The Neural Information Processing Systems Conference*. Lake Tahoe: NIPS, 2012.
- [15] ZHANG H Y, MOUSTAPHA C, YANN N. *Mixup: beyond empirical risk minimization*. URL <https://arxiv.org/pdf/1710.09412.pdf>, 2018.
- [16] IAN G, JEAN P A, MEHDI M, et al. Generative adversarial nets. *The Neural Information Processing Systems Conference*. Montreal: NIPS, 2014.
- [17] PHILLIP I, ZHU J Y, ZHOU T H, et al. Image-to-image translation with conditional adversarial networks. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Hawaii: IEEE, 2017: 1125 – 1134.
- [18] ZHU J Y, TAESUNG P, PHILLIP I, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE International Conference on Computer Vision*. Shanghai: IEEE, 2017: 2223 – 2232.
- [19] ILYA L, FRANK H. Decoupled weight decay regularization. *International Conference on Learning Representations*. New Orleans: IEEE, 2019.
- [20] LI C L, WANG X, ZHANG L, et al. Weighted low rank decomposition for robust grayscale-thermal foreground detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 2017, 27: 725 – 738.
- [21] DAVIS J, SHARMA V. Background-subtraction using contour-based fusion of thermal and visible imagery. *Computer Vision and Image Understanding*, 2007, 106: 162 – 182.
- [22] CHEN H, WANG Y L, WANG G Y, et al. LSTD: a low-shot transfer detector for object detection. *Thirty-Second AAAI Conference on Artificial Intelligence*. New Orleans: AAAI, 2018.

- [23] LEONID K, JOSEPH S, SIVAN H, et al. RepMet: representative-based metric learning for classification and few-shot object detection. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Los Angeles: IEEE, 2019: 5197 – 5206.
- [24] FU K, ZHANG T F, YAN M L, et al. Meta-SSD: towards fast adaptation for few-shot object detection with meta-learning. *IEEE Access*, 2019, 7: 77597 – 77606.
- [25] TSUNG Y L, MICHAEL M, SERGE B, et al. Microsoft COCO: common objects in context. *European Conference on Computer Vision*. Zurich: ECCV, 2014: 740 – 755.
- [26] EVERIN G M, VAN G L, WILLIAMS C K, et al. The pascal visual object class (VOC) challenge. *International Journal of Computer Vision* volume, 2010: 303 – 338.

作者简介:

吴 晗 硕士研究生, 主要研究方向为目标检测、计算机视觉等, E-mail: 296420575@qq.com;

张志龙 研究员, 主要研究方向为图像处理、自动目标识别等, E-mail: zhangzhilong@nudt.edu.cn;

李楚为 博士研究生, 主要研究方向为图像与视频处理、深度学习等, E-mail: iammsili@qq.com;

李航宇 硕士研究生, 主要研究方向为目标位姿估计、计算机视觉等, E-mail: 798269415@qq.com.