

文章编号: 1672-8785(2019)07-0026-09

基于迁移学习的红外图像 多目标检测技术

林鸿生¹ 刘文正² 汤永涛¹

(1. 海军士官学校, 安徽 蚌埠 233012;

2. 中国人民解放军 75833 部队, 广东 广州 510000)

摘 要: 针对用传统方法难以解决城市背景下红外图像多目标检测的问题, 采用迁移学习技术把深度学习中可见光域的目标检测框架迁移到红外域中。利用该方法建立的模型的小目标检测性能非常好, 在制作的测试集上平均精度 mAP (IoU=0.50) 为 0.858。还对训练数据与模型检测性能之间的关系进行了初步研究。制作了大数据量和小数据量 2 个训练集, 对模型进行训练, 然后在相同的测试集上进行测试。通过小数据量训练的模型在制作的测试集上的平均精度 mAP (IoU=0.50) 为 0.615。实验结果表明, 数据的多样性、数量、质量等都会影响模型的好坏。

关键词: 迁移学习; 深度学习; 红外图像; 目标检测; 多目标

中图分类号: TP39 **文献标志码:** A **DOI:** 10.3969/j.issn.1672-8785.2019.07.004

Multi-target Detection Technology in Infrared Image Based on Transfer Learning

LIN Hong-sheng¹, LIU Wen-zheng², TANG Yong-tao¹

(1. Naval Petty Officer School, Bengbu 233012, China;

2. Unit 75833 of PLA, Guangzhou 510000, China)

Abstract: In order to solving the problem of multi-target detection of infrared image in urban background by traditional methods, migration learning technology is used to migrate the target detection framework of visible light in deep learning to the infrared domain. A model is built by the method. The model's small target detection performance is very good, and the average precision mAP (IoU=0.50) of the test set is 0.858 on the produced test set. A preliminary study of the relationship between training data and model detection performance was also conducted. Two training sets of large data volume and small data volume were produced, the model was trained, and then tested on the same test set. The average precision mAP (IoU=0.50) of the small data set is 0.615. The experimental results show that the diversity, quantity and quality of the data will affect the quality of the model.

Key words: transfer learning; deep learning; infrared image; target detection; multi-target

收稿日期: 2019-06-12

作者简介: 林鸿生(1980-), 男, 广东揭阳人, 讲师, 主要从事光学图像处理方面的研究。

E-mail: linhs_lhs@163.com

0 引言

相比于可见光成像系统, 红外成像系统具有明显的优势: 环境适应性好(全天候工作或大雾等恶劣条件下也能工作)、隐蔽性好、识别伪装能力强等。因此, 红外图像目标检测技术受到各国研究人员的关注, 已成为国防科技研究的热门领域之一^[1], 逐渐从军事领域扩展到民用领域中的自动驾驶、安全监控、勘探检测、电力设备监测等方面^[2]。随着技术的发展, 出于现实需要, 安全监控系统开始大量普及使用红外摄像机。因此, 研究城市背景下红外图像的目标检测具有重要意义。

红外图像目标检测技术一直是研究的热点, 但针对城市背景的红外图像目标检测的研究较少^[3]。另外, 传统的红外图像目标检测方法大多数集中在小目标(点目标)上, 而且通常不对目标类别进行预测。

随着技术的不断发展, 深度学习在图像处理领域取得了引人注目的成功, 在目标检测、目标分割、物体识别等各个方面均显示出巨大的优势^[4]。更重要的是, 它能自动从数据中学习特征并进行分类, 不需要人工干预。在可见光领域, 深度学习得到了广泛应用, 但在红外领域用得比较少。由于红外成像与可见光成像具有一定的相似性, 为深度学习模型的迁移提供了良好的基础。

学术界提出了很多基于深度学习的图像目标检测方法, 如 Faster-RCNN^[5]、YOLO^[6]

等, 但绝大部分是针对可见光图像。文献[7]将深度学习用于红外船只的目标检测, 但是只把它作为分类识别的工具, 目标区域的提取还是使用传统方法。本文以城市为背景, 以行人和车辆为主要研究对象, 利用深度学习中的迁移学习方法将可见光图像目标检测方法引入到红外图像多目标检测中, 把特征提取、目标定位、目标分类统一到一个网络中, 实现端到端的目标检测。

1 相关概念

1.1 城市背景下红外图像的目标特点

可见光成像系统采集的是物体反射的可见光信息, 红外成像系统采集的是物体本身发出的热辐射信息。因此红外图像中物体的边缘、轮廓较模糊, 细节不够丰富。

同一场景的可见光图像和红外图像如图 1 所示。可以看到, 红外图像中人和汽车没有可用的纹理信息, 而且场景中有人、有车辆, 目标到成像系统的距离有远有近, 导致目标的图像形状不一样, 尺寸也不一样。

图 2 中, 本文用红色框把特殊小目标圈出来, 在图像下方放大显示, 用绿色框表示放大后的场景。

图 2(a)有目标被遮挡、多个目标靠在一起、目标与背景热辐射接近共 3 种常见情况的图像。图 2(b)中, 两个目标靠在一起, 但辐射量不一样, 显示特性不同; 另外 3 个小目标靠在一起, 还被电线杆部分遮挡。



图 1 可见光图像与红外图像



图 2 复杂情况下的红外图像

这些复杂情况下的目标检测问题采用传统的红外目标检测方法难以解决。因此,本文尝试引入深度学习来解决这个问题。

1.2 基于深度学习的目标检测框架

目前,在可见光图像目标检测方面,基于深度学习的目标检测框架主要有两类:两级式检测框架和单级式检测框架^[8]。

两级式检测框架包含一个用于区域建议的预处理步骤,即先选取候选区域,然后再进行分类。此类代表有 Faster RCNN、Mask RCNN^[9]等。

单级式检测框架采用回归的方法直接从特征层中回归出目标的位置以及种类,即采用网格化的方法或者不同尺度的卷积对特征层进行操作,直接得到目标的位置和分类信息。此类代表有 YOLO、SSD^[10]等。

一般来说,两级式检测框架的精度更高,单极式的速度更快。表 1 给出了 3 种方法在同一数据库下的训练结果。可以看到,在相同的训练条件下, Faster RCNN 的精度相对较高,所以本文利用 Faster RCNN 的预训练模型进行迁移训练。

表 1 MS COCO 下的训练结果

检测方法	Speed/ms	mAP
Faster RCNN	58	0.28
Mask RCNN	79	0.25
SSD	42	0.24

2 基于深度学习的红外图像多目标检测方法

2.1 迁移学习

文献 [11] 给出的迁移学习定义是:给定一个源域 D_s 和一个学习任务 T_s 、一个目标域 D_t 和一个学习任务 T_t , 迁移学习的目的是使用在 D_s 和 T_s 上的知识帮助提高在目标域 D_t 上的预测函数 $f_T(x)$ 的学习(见图 3)。

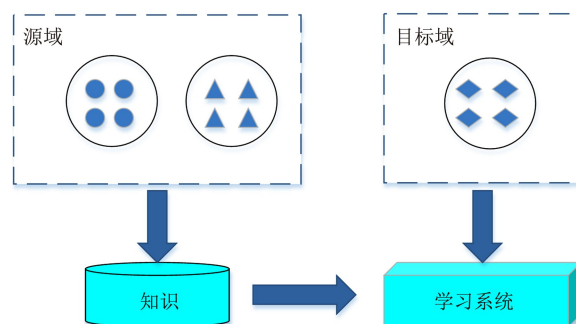


图 3 迁移学习示意图

具体到本文的目标检测任务,源域的任务定义为基于大型数据库 MS COCO 的目标检测,目标域的任务定义为基于小型红外数据库的目标检测。

传统的机器学习算法中的一个主要假设是,训练和测试数据必须在相同的特征空间中,并且具有相同的分布。而迁移学习则放宽了这一基本假设,训练和测试数据可能在不同的特征空间中,或者遵循不同的数据分布^[12]。红外成像与可见光成像都是通过光学系统收集

目标信息进行成像,具有一定的相似性。所以通过源任务获得的预训练模型来初始化目标任务中的网络参数,实现源域到目标域的特征信息迁移是可行的。

2.2 工作流程

本文的工作流程如图 4 所示。首先,取出预训练模型的网络参数,赋予新网络,进行初始化。然后,用红外图像训练数据训练新网络,得到训练模型,进行固化保存,用于推理。最后利用红外图像的测试集对推理模型进行测试,得到测试结果。

网络训练的过程中采用深度卷积神经网络提取图像的特征。

深度卷积神经网络在特征提取方面被证明是有效的^[13],如 VGGNet^[14]、ResNet^[15]、Inception^[16]等各种结构均在历年的图像分类比赛中获得不俗的成绩。

文献[17-19]提出,可采用可视化的方法

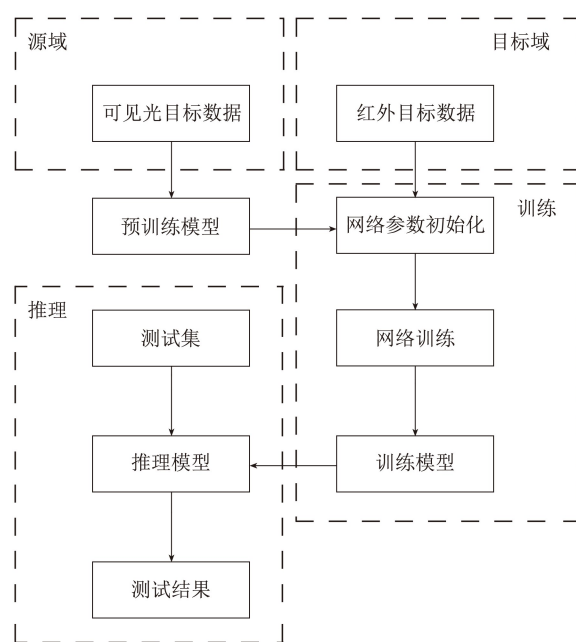


图 4 工作流程

对卷积网络提取的特征图进行叠加、上采样和反卷积,然后以热力图的形式显示网络提



(a) 红外图像1特征图的可视化



(b) 红外图像2特征图的可视化

图 5 可视化结果: 左边为红外图像, 右边为特征可视化结果

取的是哪部分的特征。

本文利用深度卷积神经网络对红外图像提取特征,然后用热力图的形式查看特征提取的准确性。结果如图 5 所示,红色区域代表卷积网络重点提取的区域。结果表明,用卷积网络提取红外图像的特征基本是有效的。

训练网络的结构采用经典的 Faster RCNN 模型结构,如图 6 所示。

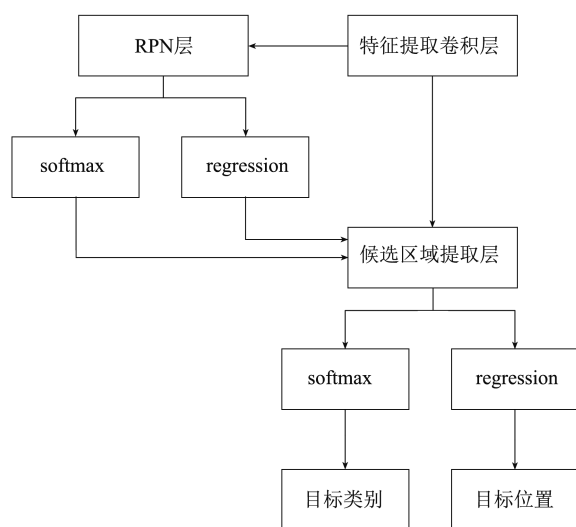


图 6 训练网络的结构

RPN 层用于选择候选目标区域。该层通过 softmax 判断候选区域属于前景或者背景,同时利用回归方法获得目标的位置。

候选区域提取层收集输入的特征数据和 RPN 层送过来的候选目标信息,综合这些信息后提取候选区域的特征数据;送入分类层,判定目标类别;再送入回归层,计算精确的目标位置。

训练过程中分类损失函数使用的是交叉熵损失,而计算回归损失函数使用的是 Smooth_L1_loss。

3 实验结果

3.1 实验细节

3.1.1 训练环境

算法基于 tensorflow 框架实现,在 Ubuntu 中用 Python 进行实验。训练平台采用 CPU 为 Intel i7-7700K、GPU 为 Nvidia GTX 1080、

内存为 16 G 的台式机。训练过程中采用的 GPU 加速。

3.1.2 训练参数

预训练模型的参数用 Faster RCNN 模型在 MS COCO 数据集下训练得到的结果进行初始化;优化算法采用可以加速 SGD 并且抑制振荡的 Momentum 算法,动量取值 0.9;初始学习率为 0.00001;Batchsize 设置为 1;周期数设置为 20000。

3.1.3 数据集

菲力尔(FLIR)公司提供的红外图像数据集包含了很多红外小目标,可以用来训练和测试小目标检测算法。该数据集分为训练集和验证集两个部分,共 14000 张图像,其中带注释标签的有 9214 张。该数据集中,人的数量为 28151 个,汽车的数量为 46692 辆。

3.1.4 训练样本

为了生成训练数据,我们对数据集进行了整理,剔除无效数据,同时选择一定数量的小目标作为训练数据的一部分(本文图像的大小为 640×512 。按照红外小目标定义^[20],目标的纵向尺寸和横向尺寸的乘积在 393 个像素以下为小目标)。为了观察训练数据量大小对模型的影响,本文进行了两次训练,一次训练样本比较小(下称模型 A),一次训练样本比较大(下称模型 B)。模型 A 的训练样本中包含人的数量为 352,汽车数量为 1034,两个类别的数量差别较大。模型 B 的训练样本中包含人的数量为 17335,汽车的数量为 19419,两个类别的数量差别较小。

3.1.5 测试样本

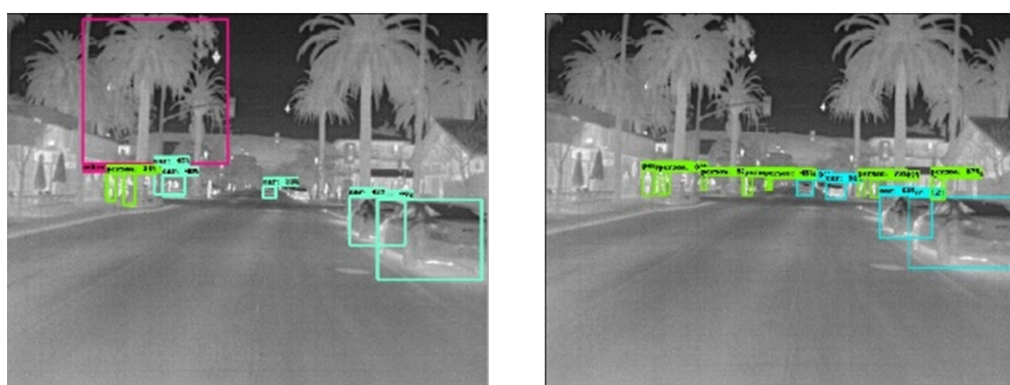
为了检测算法的有效性,我们从验证集中取出包含小目标较多、背景比较复杂的 22 张典型图片,作为测试集。

3.2 实验结果及分析

将 Faster RCNN 在 MS COCO 数据集下训练得到的模型、经过训练样本训练得到的模型 A 和模型 B 用测试集进行目标检测,部分检测结果如图 8 和图 9 所示。检测结果显示了目标



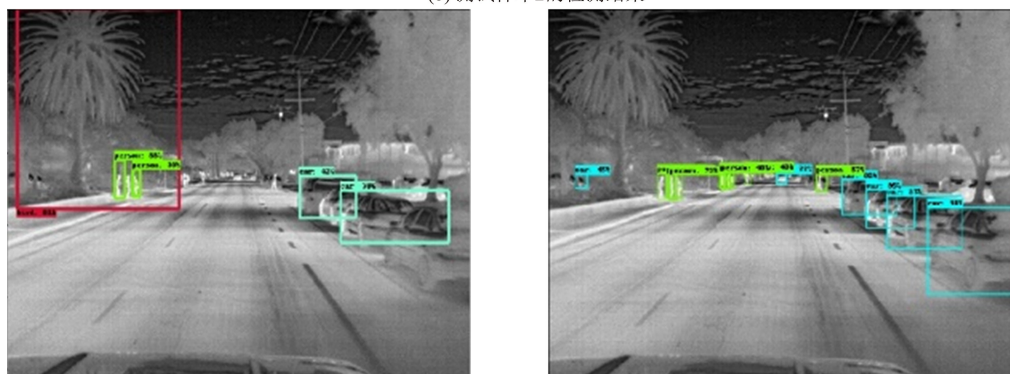
图 7 测试样本的示意图



(a) 测试样本1的检测结果



(b) 测试样本2的检测结果



(c) 测试样本3的检测结果

图 8 测试样本检测结果: 左边为 Faster RCNN 预训练模型检测结果, 右边为模型 B 检测结果

的检测框和检测概率。将检测框的显示阈值设置为 0.5。检测概率低于 0.5 的将不予显示。绿色的表示检测到的人,蓝色表示检测到车,其他颜色表示检测到其他物体。

从图 8 中可以看出,直接使用预训练模型进行检测,有一定的效果,说明红外成像与可见光成像确实有一定的相似性。但漏警和虚警比较严重,如测试样本 1 和 3 中把树识别为雨伞,测试样本 1、2、3 中还有很多的人和车辆

没有检测到,特别是比较小的目标。

相比而言,模型 B 的检测能力非常好,小目标的检测能力更强。如测试样本 1 中,模型 B 基本解决了预训练模型的虚警和漏警问题。

从图 9 中可以看出,模型 A 的检测能力相比预训练模型也有显著提升,但也还会出现较大的虚警和漏警问题。相比于模型 B 而言,其小目标检测能力更差。测试过程中,我们还发现模型 A 会出现检测错误的情况,把人的



(a) 测试样本1的检测结果



(b) 测试样本2的检测结果



(c) 测试样本3的检测结果



图 9 测试样本的检测结果: 左边为模型 A 的检测结果, 右边为模型 B 的检测结果

目标检测为汽车,把自行车等其他目标也检测为汽车。这应该是训练样本较小,同时样本中汽车数量比较多、人的数量较少造成的。

表 2 是三种方法在测试集上的测试结果。 $mAP(IoU=0.50)$ 表示 IoU 阈值为 0.5 时的平均准确率, $mAP(IoU=0.75)$ 表示 IoU 阈值为 0.75 时的平均准确率。可以看到,相比原始的 Faster RCNN 预训练模型,模型 A 的 $mAP(IoU=0.50)$ 提高了 7 倍, $mAP(IoU=0.75)$ 达到了 14 倍以上。模型 B 的 $mAP(IoU=0.50)$ 提高了 10 倍, $mAP(IoU=0.75)$ 更是达到了 24 倍以上。

从实验结果来看,采用迁移学习方法将可见光图像中的两级式目标检测框架 Faster RCNN 迁移到红外图像的目标检测领域是可行的。而且通过有针对性的训练以后,模型的小目标检测能力大幅提升;同时,训练数据的数量也会影响模型的精度。一般来说,训练数据的种类越多、数量越多,模型的检测效果就越好。

表 2 测试样本的测试结果

检测方法	mAP ($IoU=0.50$)	mAP ($IoU=0.75$)
Faster RCNN	0.0807	0.0178
模型 A	0.615	0.256
模型 B	0.858	0.434

4 结束语

本文针对传统方法难以解决城市背景下红外图像多目标检测的问题,提出了一种基于深度学习的红外图像小目标检测方法。在分析红外图像目标特点的基础上,采用迁移学习方法将可见光图像中的两级式目标检测框架 Faster RCNN 迁移到红外图像领域,同时比较了不同训练模型的检测结果。

本文的结果表明,可以通过迁移学习方法将深度学习中的目标检测框架移植到不同领域。同时,在训练过程中,数据的选择也非常重要。数据的多样性、数据的数量、数据的质量等都会影响模型的好坏。

参考文献

- [1] 刘刚,张丹. 红外成像制导图像处理技术 [M]. 北京: 科学出版社, 2016.
- [2] 王晨. 基于深度学习的红外图像语义分割技术研究 [D]. 上海: 中国科学院上海技术物理研究所, 2017.
- [3] 侯旺,孙晓亮,尚洋,等. 红外弱小目标检测技术研究现状与发展趋势 [J]. 红外技术, 2015, 37(1): 1-10.
- [4] Schmidhuber J. Deep Learning in Neural Networks: An Overview [J]. *Neural Networks*, 2015, 61: 85-117.
- [5] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2015, 39(6): 1137-1149.
- [6] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection [C]. Las Vegas: 29th IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- [7] 王文秀,傅雨田,董峰,等. 基于深度卷积神经网络的红外船只目标检测方法 [J]. 光学学报, 2018, 38(7): 0712006.
- [8] Liu L, Ouyang W L, Wang X G, et al. Deep Learning for Generic Object Detection: A Survey [J/OL]. *Arxiv*: 1809.02165, 2018.
- [9] He K, Gkioxari G, Dollar P, et al. Mask R-CNN [C]. Honolulu: 30th IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [10] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector [C]. Las Vegas: 29th IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- [11] Pan S J, Yang Q. A Survey on Transfer Learning [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22(10): 1345-1359.
- [12] Yosinski J, Clune J, Bengio Y, et al. How Transferable Are Features in Deep Neural Networks? [C]. Montreal: 28th Annual Conference on Neural Information Processing Systems, 2014.
- [13] Krizhevsky A, Sutskever I, Hinton G. Imagenet

- Classification with Deep Convolutional Neural Networks [C]. Lake Tahoe: 26th Annual Conference on Neural Information Processing Systems, 2012.
- [14] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J/OL]. *arxiv*: 1409.1556, 2014.
- [15] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition [C]. Las Vegas: 29th IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- [16] Szegedy C, Liu N W, Jia N Y, et al. Going Deeper with Convolutions [C]. Boston: 28th IEEE Conference on Computer Vision and Pattern Recognition, 2015.
- [17] Zeiler M D, Fergus R. Visualizing and Understanding Convolutional Networks [C]. Columbus: 27th IEEE Conference on Computer Vision and Pattern Recognition, 2014.
- [18] Zhou B, Khosla A, Lapedriza A, et al. Learning Deep Features for Discriminative Localization [C]. Las Vegas: 29th IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- [19] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization [C]. Honolulu: 30th IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [20] Zhang W, Cong M, Wang L. Algorithms for Optical Weak Small Targets Detection and Tracking: Review [C]. Nanjing: 2003 International Conference on Neural Networks and Signal Processing, 2003.