

基于改进 YOLOv3 的目标识别方法^①



陈正斌¹, 叶东毅¹, 朱彩霞¹, 廖建坤²

¹(福州大学 数学与计算机科学学院, 福州 350108)

²(中国移动通信集团 福建有限公司漳州分公司, 漳州 363000)

通讯作者: 叶东毅, E-mail: yiedy@fzu.edu.cn

摘要: 在复杂的自然场景中, 目标识别存在背景干扰、周围物体遮挡和光照变化等问题, 同时识别的目标大多拥有多种不同的尺寸和类型. 针对上述目标识别存在的问题, 本文提出了一种基于改进 YOLOv3 的非限制自然场景中中等或较大尺寸的目标识别方法 (简称 CDSP-YOLO). 该方法采用 CLAHE 图像增强预处理方法来消除自然场景中光照变化对目标识别效果的影响, 并使用随机空间采样池化 (S3Pool) 作为特征提取网络的下采样方法来保留特征图的空间信息解决复杂环境中的背景干扰问题, 而且对多尺度识别进行改进来解决 YOLOv3 对于中等或较大尺寸目标识别效果不佳的问题. 实验结果表明: 本文提出的方法在移动通信铁塔测试集上的准确率达 97%, 召回率达 80%. 与 YOLOv3 相比, 该方法在非限制自然场景中的目标识别应用上具有更好的性能和推广应用前景.

关键词: 神经网络; 深度学习; 目标识别; YOLOv3; 多尺度

引用格式: 陈正斌, 叶东毅, 朱彩霞, 廖建坤. 基于改进 YOLOv3 的目标识别方法. 计算机系统应用, 2020, 29(1): 49-58. <http://www.c-s-a.org.cn/1003-3254/7230.html>

Object Recognition Method Based on Improved YOLOv3

CHEN Zheng-Bin¹, YE Dong-Yi¹, ZHU Cai-Xia¹, LIAO Jian-Kun²

¹(College of Mathematics and Computer Science, Fuzhou University, Fuzhou 350108, China)

²(Zhangzhou Branch, China Mobile Communications Group Fujian Co. Ltd., Zhangzhou 363000, China)

Abstract: In complex natural scenes, object recognition encounters the problems such as background interference, occlusion of surrounding objects, and illumination changes. At the same time, most of the identified objects have different sizes and types. In view of the above-mentioned problem of object recognition, this study proposes a medium or large size object recognition method based on improved YOLOv3 in unrestricted natural scenes (CDSP-YOLO). This method uses CLAHE image enhancement preprocessing method to eliminate the influence of illumination changes on object recognition in natural scenes, and uses stochastic spatial sampling pooling (S3Pool) as the downsampling method of feature extraction network to preserve the spatial information of feature map to solve the background interference problem in complex environment, and improves multi-scale recognition to solve the problem that YOLOv3 is not effective for medium or large size object recognition. The experimental results show that the proposed method has an accuracy rate of 97% and a recall rate of 80% on the mobile communication tower test set. Compared with YOLOv3, the algorithm has better performance and application prospects in object recognition applications in unrestricted natural scenes.

Key words: neural network; deep learning; object recognition; YOLOv3; multi-scale

① 基金项目: 国家自然科学基金 (61672158); 福建省自然科学基金 (2018J1798); 福建省高校产学研合作项目 (2018H6010)

Foundation item: National Natural Science Foundation of China (61672158); Natural Science Foundation of Fujian Province (2018J1798); Higher Education's Industry-University-Research Cooperation Project of Fujian Province (2018H6010)

收稿时间: 2019-06-11; 修改时间: 2019-07-05; 采用时间: 2019-07-15; csa 在线出版时间: 2019-12-27

随着在限制环境下目标识别算法在识别精度和实时性方面已经接近实用性, 研究人员越来越关注对非限制环境下的目标进行准确、实时的识别. 特别是针对自然场景中不同尺度和类型的目标进行准确并实时的识别是最具挑战性的任务^[1].

移动通信铁塔作为移动通信网络建设中的基础设施, 架设在环境极为复杂的自然场景中, 为了满足不同场景下的信号需求, 移动通信铁塔还具有多种不同的设备类型. 因此, 为了实现准确、实时的对移动通信铁塔目标识别, 其难点主要体现在以下几个方面: (1) 移动通信铁塔包含 6 大类, 并且有的类与类之间的差别极小; (2) 不同设备类型之间的尺度差别很大; (3) 目标铁塔的周围还存在强遮挡、背景干扰等问题; (4) 采集的图像还存在光照变化、分辨率不统一等问题. 由于背景复杂、周围物体遮挡、光照变化等因素引起的干扰导致在复杂环境下的目标识别问题更具有很大的挑战性^[2].

近年来, 深度学习在计算机视觉中得到了显著的发展, 特别是在目标识别方面^[3]. 其主要的方法可以分为两类. 一类是基于候选区域的目标识别算法. 首先使用区域推荐产生候选目标区域, 随后使用卷积神经网络进行分类、回归处理; 虽然其精度较高但实时性较差, 难以满足实时性需求的应用. 其中具有代表性的算法有 R-CNN^[4], Fast R-CNN^[5], Faster R-CNN^[6], Mask R-CNN^[7]等. 另一类是基于回归方法的目标识别算法, 将识别问题作为一个回归问题处理, 直接预测目标位置和类别. 这类方法具有较快的识别速度, 但相对而言精度偏低, 具有代表性的算法有 YOLO^[8]、SSD^[9]、YOLOv2^[10]、YOLOv3^[11]等.

YOLOv3 在各个领域应用时相对 R-CNN 具有更好的泛化性, 基于 YOLOv3 的目标检测方法被迅速推广应用到各个领域, 如视频监控、零件检测、交通监控等. Qu 等人^[12]在 YOLOv3 的基础上, 结合图像增强技术, 提出了一种用于行人检测的方法. Song 等人^[13]提出了将 YOLOv3 算法应用于列车轨道表面的小目标缺陷检测. Bilel 等人^[14]利用无人飞行器采集的车辆图像对 Faster R-CNN 和 YOLOv3 算法进行比较, 得出 YOLOv3 在灵敏度和处理时间方面优于 Faster R-CNN 的结论.

YOLOv3 利用深度残差网络提取图像特征, 并实现多尺度预测, 获得了目前最好的识别精度与速度的平衡, 但是 YOLOv3 提取特征获得的最小特征图尺寸为 13×13 , 相对于 SSD 中的 1×1 仍然偏大, 造成 YOLOv3

对于中等或较大尺度的物体识别效果不好, 会产生误检、漏检或重复检测的问题. 因此, 本文对 YOLOv3 进行改进, 提出了针对自然场景中中等或较大尺寸的目标识别方法 CDSP-YOLO. 该方法对 YOLOv3 进行以下两大改进: (1) 设计了使用随机空间采样池化 (S3Pool^[15]) 的特征提取网络 DSP (DarkNet based on S3Pool), 以减少因下采样而导致特征图空间信息的丢失. (2) 对 YOLOv3 的多尺度识别策略进行改进, 增强方法对中等或较大尺寸的目标识别效果, 同时增加锚点框的数量, 以提高目标识别的召回率. 此外, 针对非限制自然场景中目标识别任务存在光照变化等原因而造成的误检或漏检的问题, 采用限制对比度直方图均衡化 (CLAHE) 进行图像增强预处理. 最后使用自行设计并标注的移动通信铁塔数据集对 CDSP-YOLO 方法进行性能评估, 实验结果表明本文提出的方法在非限制自然场景中对中等或较大尺寸的目标识别上能够取得较好的效果: 准确率达 97%, 召回率达 80%.

1 相关工作

在感兴趣目标识别的方面, 近年来得到了飞速的发展, 在精确性和实时性上都有了大幅度的提升, 主要原因除了 GPU 计算能力的提升之外, 深度卷积神经网络的引入也使得该领域取得了重大突破^[16]. 深度学习目标识别算法已经广泛地应用于各个目标识别领域, 同时针对特定问题的改进算法也不断提出, 并取得显著的效果. 以下将简要介绍利用深度卷积神经网络进行自然场景中中等或较大尺寸目标检测的最新研究.

基于候选区域的目标识别算法首先在图像中产生候选框, 然后对候选框中的目标进行分类. Cai 等人^[17]针对街道环境中的目标识别任务, 对 Faster R-CNN 进行适应性的结构微调, 使其对街道环境中的目标识别任务可以获得较好的效果. 但是该算法对于具有图像背景干扰及周围物体遮挡的目标识别效果不佳. Weidong Min 等人^[18]提出一种改进的 ViBe 算法, 用于多个目标的鲁棒性与准确性识别. 该算法采用良好的后处理方法抑制动态噪声, 同时使用两个分类器进一步解决无法识别具有遮挡和干扰目标的问题, 但是实验中对比的方法不是目前最好的方法, 不能完全说明该方法的优越性. Cao 等人^[19]提出具有知识引导训练和预测感兴趣区域的快速深度神经网络, 在显著降低整体计算复杂度的同时提高目标识别性能, 与传统 SSD 算法相

比, 该算法的检测速度显著提高, 但是检测精度却没有明显提升.

基于回归问题的目标识别算法使用端到端的神经网络, 直接预测图像中目标的位置和类别, 获得了目前最好的识别精度与速度的平衡. 针对不同场景下的任务需求, 基于 YOLOv3 算法提出了许多改进方法. 文献[12-14]已经表明基于改进 YOLOv3 的目标识别算法, 对于特定场景下的小目标识别能够取得较好的效果, 但是对于非限制自然场景中中等或较大尺寸的目标识别任务存在的背景干扰和周围物体遮挡的问题, 现有改进算法的识别性能仍然有待提升.

2 YOLOv3

YOLO 目标识别算法的核心思想是将目标识别问题转化为一个回归问题, 实现了仅用一个深度卷积神经网络, 同时在保证准确率较高的前提下实现快速的目标识别^[20].

YOLOv3 在 YOLOv2 的基础上进行了一些适应性的改进, 包括多尺度识别、多标签分类等, 并使用基于 ResNet 网络改进的 DarkNet-53 网络作为特征提取器, 使 YOLO 系列方法不擅长识别小物体的缺陷得到了改善. YOLOv3 也因此成为目前为止最优秀的目标识别算法之一. 如图 1 为 DarkNet-53 特征提取器的网络结构.

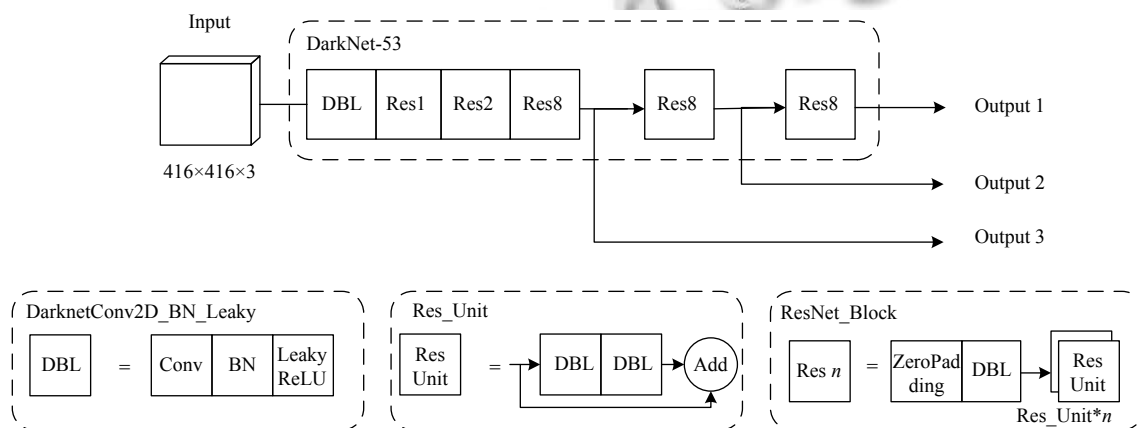


图 1 YOLOv3 算法的 DarkNet-53 网络结构

YOLOv3 首先将原始图片缩放到 416×416 的大小, 使用了类似 FPN 网络^[21]的尺度金字塔结构. 其中的 DarkNet-53 特征提取网络将原始图像按照特征图的尺度大小划分为 $S \times S$ 个等大的单元格, 对应的 3 个尺度的特征图大小为 13×13 、 26×26 、 52×52 . 接着采用浅层次特征和深层次特征的融合获得更具有鉴别性的深层特征. 最后在回归预测部分每一个单元格借助 3 个锚点框 (anchor box) 预测 3 个边框. 如图 2 所示为其中一个网格预测的边框示意图.

YOLOv3 在每一个单元格上会为每一个边框预测 4 个值, 即边框左上角坐标 (x, y) 与目标的宽和高 (w, h) , 记为 (t_x, t_y, t_w, t_h) . 若目标中心在单元格中相对于图像左上角有偏移 (C_x, C_y) , 并且锚点框具有宽度和高度 (P_w, P_h) , 则修正后的边框为:

$$b_x = \sigma(t_x) + C_x$$

$$b_y = \sigma(t_y) + C_y$$

$$\begin{aligned} b_w &= P_w \cdot e^{t_w} \\ b_h &= P_h \cdot e^{t_h} \end{aligned} \quad (1)$$

其中, 锚点框的选择采用维度聚类的方法, 传统的聚类算法包括层次聚类和 k 均值聚类, 以及基于模型的方法^[22]. 而在 YOLOv3 中利用 K-means 聚类算法对训练集中目标边框的尺寸进行聚类, 获取最佳锚点框的尺寸, 从而能够预测出更加准确的目标边框. 其中的 K-means 聚类算法的距离度量为:

$$d(\text{box}, \text{centroid}) = 1 - \text{IOU}(\text{box}, \text{centroid}) \quad (2)$$

其中, box 指数数据集中边框尺寸样本, centroid 指类簇中心尺寸.

YOLOv3 使用逻辑回归预测锚点框中包含物体的概率. 如果锚点框与真实目标边框的重叠率大于任何其他锚点框, 则这个锚点框的概率为 1; 如果锚点框与真实目标的边框的重叠率大于 0.5, 但又不是最大的, 则忽略这个预测. YOLOv3 在训练时仅对一个物体分配一个锚点框. 如果锚点框中没有包含被认为的物体,

则使其在损失函数中不起作用. YOLOv3 在训练过程中使用二元交叉熵损失和逻辑回归来进行类别预测, 这种方式使得 YOLOv3 可以对一个目标进行多标签分类.

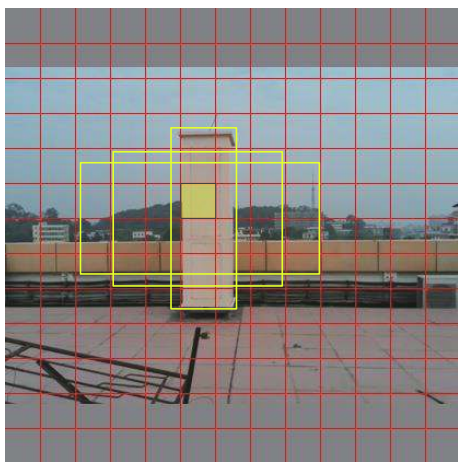


图2 YOLOv3 在 13×13 的单元格中的预测边框示意图

3 随机空间采样池化 (S3Pool)

卷积神经网络 (CNN) 中的特征池化层主要用于减少参数计算数量, 保持尺度的不变性, 并增加特征图的感受野大小^[23]. 在中规模的网络, 例如: VGG^[24]和 GoogLeNet^[25]中使用最大或者平均池化的方式; 而在大规模的网络, 例如: ResNet^[26]和 DenseNet^[27], 使用带步长的卷积方式进行特征图的压缩. 由于以上的方式, 仅从某个邻域内通过线性的方式下采样成一个点, 而没有更多的考虑特征图的空间关系, 使得网络在前向传播时丢失更多的信息. 在文献[15]中 Zhai 等人提出了随机空间采样池化 (S3Pool). 该算法先使用步长为 1 的最大池化方式, 得到一张与原特征图尺寸相同的特征图; 然后将特征图划分成边长为 g 的网格, 再使用随机选择行列的方式确定池化窗口选择的点进行特征图的下采样. 如图 3 为使用池化窗口为 2、步长为 2、网格大小为 2 的 S3Pool 过程.

S3Pool 通过将池化过程分解为两个步骤来扩展标准最大池: 使用步幅 1 进行最大池化和使用非确定性的空间下采样, 即从特征映射中随机采样行和列. 实际上, S3Pool 在每个池化阶段隐式地增加了训练数据, 这使得学习模型具有更强的泛化能力. 并且与带步长的卷积相比, S3Pool 不会引入任何附加参数, 并且可以插入任何现有的池层. 如表 1 为文献[15]中利用两种流行

的图像分类基准评估 S3Pool: CIFAR-10 和 CIFAR-100. 文献[15]选用两个目前通用的深度神经网络架构, 即 NIN^[28]和 ResNet. 对比两个架构中使用步长为 2 的卷积核 (Pool-2) 和 S3Pool 的下采样方法对于算法性能的影响. 实验结果表明 S3Pool 的随机性使得模型具有更好的泛化性. 同时相比于使用带步长的卷积方式, S3Pool 不会增加任何参数, 能够使得模型更易于训练.

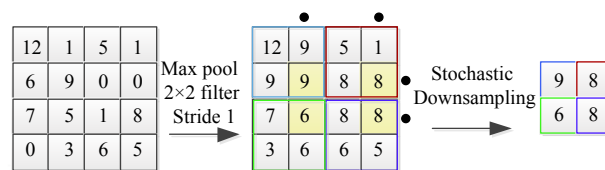


图3 使用窗口为 2、步长为 2、网格大小为 2 的 S3Pool 过程

表 1 S3Pool 与 Pool-2 在 CIFAR-10 和 CIFAR-100 的对比

Model	CIFAR-10		CIFAR-100	
	train err	test err	train err	test err
NIN(Pool-2)	0.63	10.68	6.15	35.24
NIN+S3Pool	1.85	9.30	9.25	33.85
ResNet(Pool)	0.00	14.07	0.02	42.32
ResNet+S3Pool	0.82	8.86	3.97	32.78

4 基于 YOLOv3 的改进

针对现有的目标识别算法对非限制自然场景中中等或较大尺寸目标识别中无法取得较好性能的问题, 本文对 YOLOv3 算法在网络结构和多尺度识别策略上进行改进, 提出了使用 CLAHE 图像增强预处理的 CDSP-YOLO 目标识别方法.

4.1 YOLOv3 网络改进

YOLOv3 中 Redmon 等人借鉴了 ResNet 提出 DarkNet-53 特征提取网络. 虽然 DarkNet-53 利用残差网络结构使得网络的训练难度得到降低, 并使用了大量的 1×1 卷积核与使用 3×3 大小步长为 2 的卷积核替代最大池化减少网络总体的参数数量, 但是由于通过带步长卷积的下采样方式仍是局部邻域的线性计算, 没有考虑邻域之间的空间关系, 因此该方法同样不能有效的解决因最大池化而导致特征图空间信息丢失的问题, 同时带步长的卷积会使得网络的参数增加, 导致训练复杂、并且减慢识别速度.

本文的任务在于完成非限制自然场景中中等或较大尺寸的目标识别. 使用在非限制自然场景下采集的移动通信铁塔图像作为算法性能的评估. 由于自然场

景下的移动通信铁塔图像存在背景复杂、光照变化和周围物体遮挡等问题,直接采用 YOLOv3 无法有效的解决上述问题,因此,本文提出了 DSP (DarkNet based on S3Pool) 特征提取网络. DSP 采用随机空间采样池化 (S3Pool) 的方法对特征图进行下采样. 该方法首先使用步长为 1 的最大池化方法,保留特征图之间更多

的空间关系,再使用随机采样的方式,在每个 2×2 的区域内进行采样,最后输出下采样之后的特征图. 本文的 DSP 特征提取网络将 DarkNet-53 中卷积核大小为 3×3 、步长为 2 的卷积层替换成 S3Pool 下采样层,起到最大概率保留特征图空间信息的作用. 如图 4 为改进后的 DSP 特征提取网络.

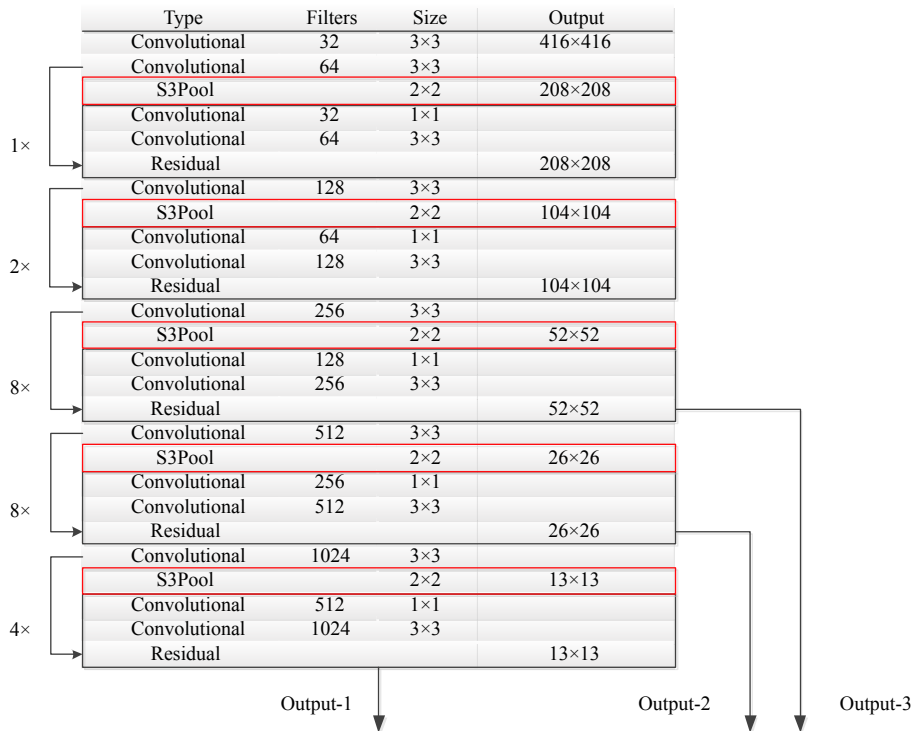


图 4 改进后的 DSP 特征提取网络

4.2 多尺度识别改进

YOLOv3 中,引入 FPN 网络,同时利用低层特征的高分辨率和高层特征的高语义信息,并通过上采样融合不同层的特征,在 3 个不同尺度的特征层上识别物体.如图 5 为移动通信铁塔图像在 13×13 、 26×26 、 52×52 的 3 个尺度层下的感受野图示,在移动通信铁塔目标较大的情况下,在 52×52 的尺度层输出特征图的感受野对应目标的部分极小.此时使用锚点框进行预测铁塔目标边框将很难保证锚点框对于大目标的覆盖率.同时,如图 6 所示移动通信铁塔图像中的铁塔目标大多数为大目标的情况,本文对 YOLOv3 中的尺度识别模块进行改进,将原有的 3 个尺度识别修改为 2 个尺度识别,在较小特征图给大目标分配更为准确的锚点框.

本文将 YOLOv3 的 3 个尺度修改为 2 个尺度后,需要对锚点框进行重新确定.通过权衡平均交并比与

锚点框的数量(如图 7)的关系,取 16 个锚点框,分别为: (12,46)、(16,111)、(22,76)、(39,72)、(22,166)、(30,113)、(75,79)、(33,198)、(49,148)、(40,282)、(57,327)、(291,53)、(64,205)、(94,294)、(135,154)、(194,310). 在每个尺度上的每一个单元格借助 8 个锚点框预测 8 个目标边框.

基于以上提出的改进思路,如图 8 中给出了本文提出的改进后的 DSP-YOLO (DarkNet based on S3Pool-YOLO) 网络结构.网络的输入为 416×416 的待识别图像,经过 DSP 特征提取,再利用多尺度识别,最后输出 2 个尺度的识别结果,分别为 $13 \times 13 \times 88$ 、 $26 \times 26 \times 88$, 其中的 88 维通道表示使用 8 个锚点框预测 8 个目标边框的信息,每个预测边框信息具体包括边框坐标 (x, y) 、大小 (w, h) 、置信度 (C) 及所属 6 个类别的概率 $(C_1, C_2, C_3, C_4, C_5, C_6)$.

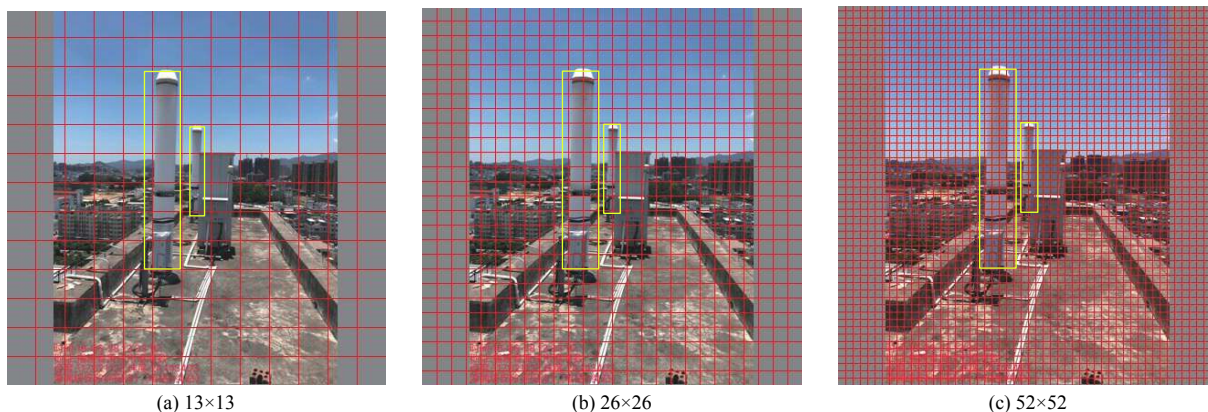


图5 移动通信铁塔图像在3个尺度层下的感受野图示

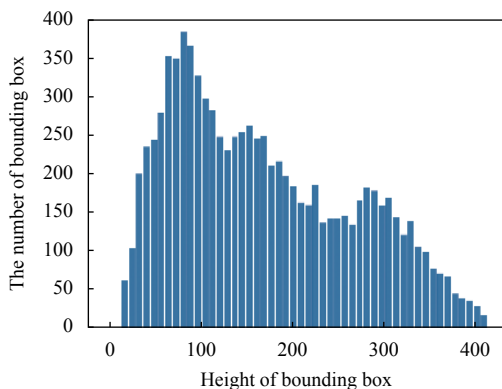


图6 数据集中边框高度的分布情况

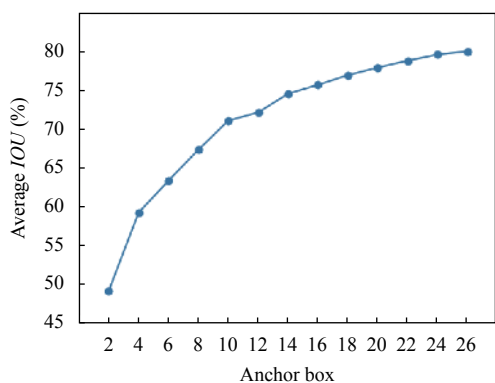


图7 锚点框与数据集边框平均交并比的关系图

4.3 图像增强预处理

由于非限制自然场景中的图像在采集过程中存在光线不足或者光照不均匀等的现象,该现象可能导致图像质量下降.因此,采用传统的YOLOv3算法无法很好适用于该场景下的移动通信铁塔目标识别问题.

为了克服以上提及的问题,本文提出了CDSP-

YOLO目标识别方法.该方法采用了对自适应直方图均衡化算法(AHE)^[29]改进的限制对比度直方图均衡化的图像增强预处理算法(CLAHE).其中CLAHE图像增强预处理算法的主要步骤为:

(1) 首先将图像的颜色空间由RGB转为HSV.

(2) 然后将图像分块,并对每个分块在HSV颜色空间中的V(亮度)分量进行直方图均衡化的操作,该操作将直方图中超过阈值的灰度级像素剪切均分给每个灰度级.

(3) 最后将V分量和原先的H、S分量拼接再转到RGB颜色空间得到增强后的图像.

通过CLAHE预处理后可以减小光照对图像质量的影响,同时由于提升了对比度还会增加了样本图像光照的多样性.在训练和测试的过程中对图像进行增强预处理,提高移动通信铁塔图像质量,从而提高模型识别的准确率和召回率.如图9为CDSP-YOLO目标识别方法流程图.

5 实验结果与分析

为了验证DSP-YOLO算法的在识别非限制自然场景中中等或较大尺寸目标的可行性以及采用基于CLAHE图像增强预处理的CDSP-YOLO方法进行目标识别的有效性.本文在中国移动集团福建有限公司漳州分公司提供图像构造的数据集上,通过对比实验的方法,将YOLOv3、DSP-YOLO和CDSP-YOLO方法应用于移动通信铁塔目标的识别.实验服务器的配置参数如下: Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10 GHz 处理器、128 GB 内存、16 GB NVIDIA Tesla P100 GPU、Ubuntu 16.04.3 操作系统.

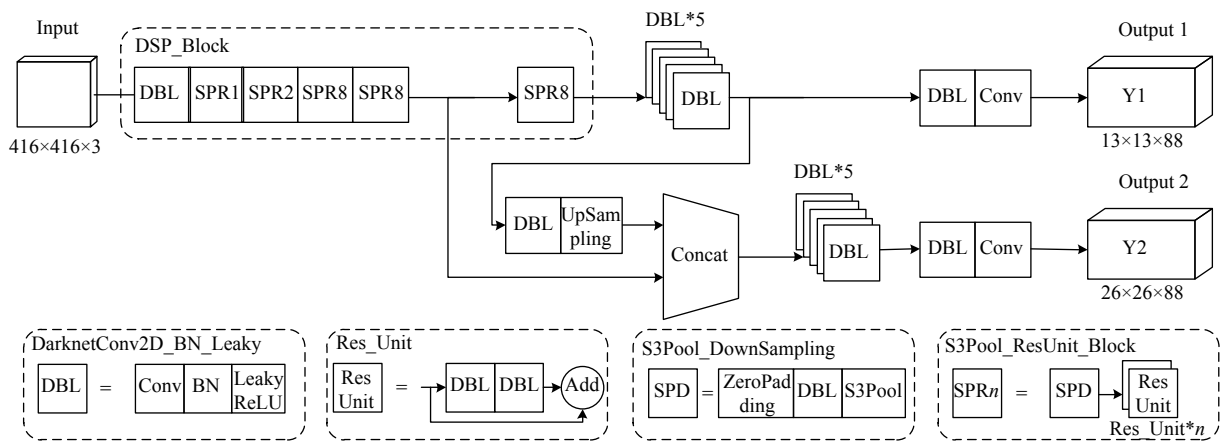


图8 DSP-YOLO网络结构

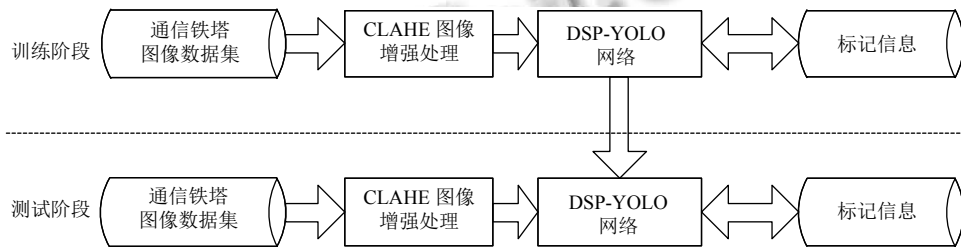


图9 CDSP-YOLO目标识别方法流程图

实验主要分为两个部分: 第一, 通过实验结果定量评估 YOLOv3、DSP-YOLO 和 CDSP-YOLO 方法对于中等或较大尺寸的移动通信铁塔目标识别的效果, 验证 CDSP-YOLO 方法的有效性. 第二, 通过实验结果定性的评估 YOLOv3、DSP-YOLO 和 CDSP-YOLO 方法在自然场景中对不同类别的移动通信铁塔的识别能力, 验证 CDSP-YOLO 方法具有良好的鲁棒性.

本次实验规定当模型预测的目标边框与手工标注的边框 IOU ≥ 0.5 时, 则认为移动通信铁塔目标识别正确, 否则错误. 实验选取准确率 P (Precision)、召回率 R (Recall) 和平均交并比 $mIOU$ 作为评价指标.

准确率:

$$P = \frac{TP}{FP + TP} \quad (3)$$

召回率:

$$R = \frac{TP}{FN + TP} \quad (4)$$

其中, TP 为真正例, FP 为假正例, FN 为假负例.

5.1 数据集

本文中提到的移动通信铁塔目标识别的原始实验数据由中国移动通信集团福建有限公司漳州分公司提供, 实验数据包含 6 个类别共 240 张图像.

由于原始移动通信铁塔图像数据数量较少, 本文采取了 3 种策略对数据进行扩增^[30], 分别为: (1) 随机图像平移: 将图像在水平或者垂直方向上进行平移 -0.2 倍~0.2 倍. (2) 随机图像缩放: 对图像进行随机的缩小或者放大 0.2 倍. (3) 随机图像镜像翻转: 将图像进行随机的翻转, 得到镜像图像. 通过以上的数据扩增方法, 本文得到 6691 张图像用于后续的工作. 如图 10 为中国移动通信集团福建有限公司漳州分公司提供的移动通信铁塔目标数据集扩增后的部分数据展示.

5.2 识别结果定量评估

将本文提出的基于 CLAHE 图像增强预处理的 CDSP-YOLO 与未采用图像增强处理的 DSP-YOLO 及 YOLOv3 进行比较, 对比实验结果如表 2 所示. 表中的数据是在 500 个样本的测试集上获得, 其中对网络结构和多尺度改进的 DSP-YOLO 算法得到了 91% 的准确率和 77% 的召回率, 相比于 YOLOv3 分别提升了 11% 和 20%. 更进一步的是, 在使用了基于 CLAHE 图像增强预处理的 CDSP-YOLO 相比于未使用图像增强预处理的 DSP-YOLO 在准确率和召回率又提高了 6% 和 3%, 最终取得 97% 的准确率和 80% 的召回率.



图 10 移动通信铁塔目标识别数据集展示

表 2 实验结果对比

Network	P (%)	R (%)	$mIOU$
YOLOv3	80	57	0.71
DSP-YOLO	91	77	0.74
CDSP-YOLO	97	80	0.75

5.3 定性评估与误差分析

为了更直观的展示实验的对比效果,如图 11 展示了 YOLOv3、DSP-YOLO 和 CDSP-YOLO 对中等或较大尺寸的移动通信铁塔测试集中 6 个类别的不同场景图像的识别结果. 实验主要从目标强遮挡、光照变化、图像对比度低、图像模糊和多目标识别等方面对 3 个方法进行全面的对比. 由图 11 的对比实验结果可知,直接使用 YOLOv3 算法进行移动通信铁塔目标识别存在漏检和识别错误的问题. 而使用适应性的 DSP-YOLO 算法在一定程度上能克服 YOLOv3 算法出现的漏检问题,但对于图像中出现光照变化、对比度低和模糊的现象,直接使用 DSP-YOLO 算法进行目标识别也很难有好效果. 由图 11 中第 3 列的 CDSP-YOLO 方法识别结果可知,使用 CLAHE 图像增强预处理方法可以很好的避免图像中出现光照变化、对比度低和模糊的现象对于目标识别的影响. 实验对比结果表明本文提出的 CDSP-YOLO 目标识别方法对于自然场景下的移动通信铁塔目标识别具有良好的鲁棒性.

图 11(a)、图 11(b)、图 11(c) 3 个对比图存在强遮挡、对比度低和光照变化的问题,仅有 CDSP-YOLO 方法可以全部识别移动通信铁塔目标. YOLOv3 算法相比于 DSP-YOLO 算法在特征图的下采样过程中丢失的空间信息更多,而导致在存在上述问题的图像中无法有效地将移动通信铁塔目标识别. 而 DSP-YOLO

算法虽然能够识别出一些铁塔目标,但是与 CDSP-YOLO 方法相比存在置信度较低的问题,如图 11(c) 中 DSP-YOLO 算法识别的置信度为 0.75,而 CDSP-YOLO 方法识别的置信度高达 0.94. 综上,CDSP-YOLO 目标识别方法在场景复杂的移动通信铁塔目标识别上具有较好的识别效果.

图 11(d)、图 11(e)、图 11(f) 3 个对比图为多目标识别问题,同时场景中仍存在遮挡、对比度低和图像模糊的问题,同样仅有 CDSP-YOLO 算法可以识别出所有的移动通信铁塔目标. YOLOv3 算法不仅存在漏检的问题,而且在移动通信铁塔目标类别较为相似的情况下还存在识别错误的问题,主要原因是 YOLOv3 的特征提取网络的对特征图空间信息的保留能力不足,同时由于没有对多尺度做适应性改进,每个尺度采用 3 个锚点框无法取得较好的覆盖率从而导致识别的召回率低,因此 YOLOv3 算法无法直接应用于自然场景中中等或较大尺寸的目标识别任务中. 而直接使用 DSP-YOLO 算法对于场景存在背景干扰、遮挡和光照变化的多目标识别仍存在少数的漏检情况,主要还是因为上述自然场景过于复杂引起. 所以采用基于 CLAHE 图像增强预处理的方法能够有效的提高 DSP-YOLO 的识别召回率和准确率.

6 总结

本文基于 YOLOv3 对网络结构和多尺度识别进行改进,提出了一种针对非限制自然场景中中等或较大尺寸的目标识别方法,并自行设计标注了一个移动通信铁塔目标的数据集,用于方法性能的评估. 实验结果

表明, 本文提出的 CDSP-YOLO 目标识别方法对自然场景中中等或较大尺寸的目标具有良好的鲁棒性, 识别的准确率达 97%, 召回率达 80%.

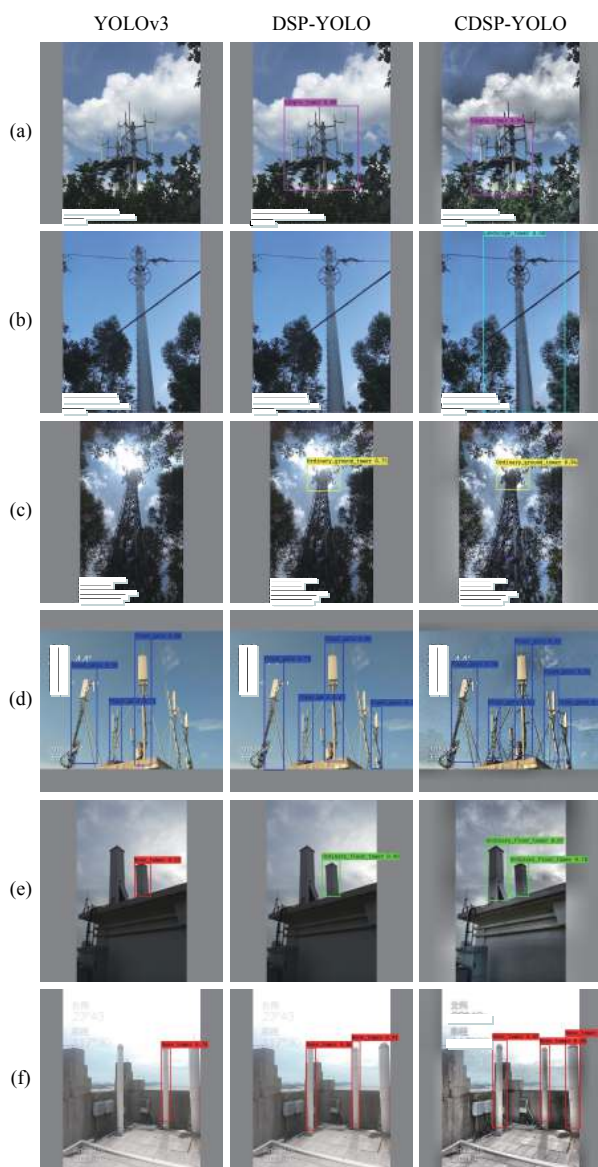


图 11 YOLOv3、DSP-YOLO、CDSP-YOLO 对比实验结果

参考文献

- 1 Li H, Fu K, Yan ML, *et al.* Vehicle detection in remote sensing images using denoizing-based convolutional neural networks. *Remote Sensing Letters*, 2016, 8(3): 262–270. [doi: 10.1080/2150704x.2016.1258127]
- 2 Li YD, Zhang XY, Li HG, *et al.* Object detection and tracking under Complex environment using deep learning-based LPM. *IET Computer Vision*, 2019, 13(2): 157–164.

[doi: 10.1049/iet-cvi.2018.5129]

- 3 Xu ML, Fang H, Lv P, *et al.* D-STC: Deep learning with spatio-temporal constraints for train drivers detection from videos. *Pattern Recognition Letters*, 2019, 119: 222–228. [doi: 10.1016/j.patrec.2017.09.040]
- 4 Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH, USA. 2014. 580–587.
- 5 Girshick R. Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*. Santiago, Chile. 2015. 1440–1448.
- 6 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. *Proceedings of the 28th International Conference on Neural Information Processing Systems*. Montreal, Canada. 2015. 91–99.
- 7 He KM, Gkioxari G, Dollár P, *et al.* Mask R-CNN. *Proceedings of IEEE International Conference on Computer Vision*. Venice, Italy. 2017. 2980–2988.
- 8 Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA. 2016. 779–788.
- 9 Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot MultiBox detector. *Proceedings of the 14th European Conference on Computer Vision*. Amsterdam, The Netherlands. 2016. 21–37. [doi: 10.1007/978-3-319-46448-0_2]
- 10 Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA. 2016. 6517–6525.
- 11 Redmon J, Farhadi A. Yolov3: An incremental improvement. *arXiv: 1804.02767*, 2018.
- 12 Qu HQ, Yuan TY, Sheng ZY, *et al.* A pedestrian detection method based on YOLOv3 model and image enhanced by retinex. *Proceedings of the 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics*. Beijing, China. 2018. 1–5.
- 13 Song YN, Zhang H, Liu L, *et al.* Rail surface defect detection method based on YOLOv3 deep learning networks. *Proceedings of 2018 Chinese Automation Congress*. Xi'an, China. 2018. 1563–1568.
- 14 Benjdira B, Khurshheed T, Koubaa A, *et al.* Car detection using unmanned aerial vehicles: Comparison between faster R-CNN and YOLOv3. *Proceedings of the 2019 1st*

- International Conference on Unmanned Vehicle Systems-Oman. Muscat, Oman. 2019. 1–6.
- 15 Zhai SF, Wu H, Kumar A, *et al.* S3Pool: Pooling with stochastic spatial sampling. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA, 2017: 4003–4011.
- 16 Napiorkowska M, Petit D, Marti P. Three applications of deep learning algorithms for object detection in satellite imagery. Proceedings of IGARSS 2018 —2018 IEEE International Geoscience and Remote Sensing Symposium. Valencia, Spain. 2018. 4839–4842.
- 17 Cai WD, Li JD, Xie ZZ, *et al.* Street object detection based on faster R-CNN. Proceedings of the 2018 37th Chinese Control Conference. Wuhan, China. 2018. 9500–9503.
- 18 Min WD, Fan MD, Guo XG, *et al.* A new approach to track multiple vehicles with the combination of robust detection and two classifiers. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(1): 174–186. [doi: [10.1109/TITS.2017.2756989](https://doi.org/10.1109/TITS.2017.2756989)]
- 19 Cao WM, Yuan JH, He ZH, *et al.* Fast deep neural networks with knowledge guided training and predicted regions of interests for real-time video object detection. IEEE Access, 2018, 6: 8990–8999. [doi: [10.1109/ACCESS.2018.2795798](https://doi.org/10.1109/ACCESS.2018.2795798)]
- 20 Lan WB, Dang JW, Wang YP, *et al.* Pedestrian detection based on YOLO network model. Proceedings of the 2018 IEEE International Conference on Mechatronics and Automation. Changchun, China. 2018. 1547–1551.
- 21 Lin TY, Dollar P, Girshick R, *et al.* Feature pyramid networks for object detection. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 936–944.
- 22 Ruan TJ. Multiple-objective clustering analysis[Ph. D. Thesis]. New York: State University of New York at Stony Brook, 2016.
- 23 Saeedan F, Weber N, Goesele M, *et al.* Detail-preserving pooling in deep networks. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 9108–9116.
- 24 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556, 2014.
- 25 Szegedy C, Liu W, Jia YQ, *et al.* Going deeper with convolutions. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 1–9.
- 26 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 770–778.
- 27 Huang G, Liu Z, van der Maaten L, *et al.* Densely connected convolutional networks. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 2261–2269.
- 28 Lin M, Chen Q, Yan S. Network in network. arXiv: 1312.4400, 2014.
- 29 Russakovsky O, Deng J, Su H, *et al.* Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 2015, 115(3): 211–252. [doi: [10.1007/s11263-015-0816-y](https://doi.org/10.1007/s11263-015-0816-y)]
- 30 Bargoti S, Underwood J. Deep fruit detection in orchards. Proceedings of International Conference on Robotics and Automation. Singapore. 2017. 3626–3633.