

改进的 SSD 算法及其对遥感影像 小目标检测性能的分析

王俊强^{1,2}, 李建胜^{1*}, 周学文², 张旭¹

¹信息工程大学地理空间信息学院, 河南 郑州 450000;

²78123 部队, 四川 成都 610000

摘要 针对以 Faster R-CNN 为代表的基于候选框方式的遥感影像目标检测方法检测速度慢, 而现有 SSD 算法在小目标检测中性能低的问题, 提出一种改进的 SSD 算法, 综合利用现有基于候选框方式和一体化检测方式的优势, 提升检测性能。该算法利用密集连接网络替换原有的 VGGNet 作为骨干网络, 并且在密集连接模块之间构建特征金字塔, 代替原有多尺度特征图。为验证所提算法的精度及性能, 设计样本数据在线采集系统, 并采集飞机及运动场目标样本集作为实验样本, 通过对改进 SSD 算法的训练, 验证了其网络结构的稳定性, 在无迁移学习支持下依然能够达到良好效果, 且训练过程不易发散。通过对比以 101 层的残差网络 (ResNet101) 作为基础网络的 Faster R-CNN 算法和 R-FCN 算法可知, 改进 SSD 算法较 Faster R-CNN 算法和 R-FCN 算法的 MAP 在测试集上分别提升了 9.13% 和 8.48%, 小目标检测的 MAP 分别提升了 14.46% 和 13.92%, 检测单张影像耗时 71.8 ms, 较 Faster R-CNN 和 R-FCN 算法分别减少 45.7 ms 和 7.5 ms。

关键词 遥感; 小目标检测; 深度学习; 多尺度预测; 特征金字塔; 平均准确率均值

中图分类号 P209

文献标识码 A

doi: 10.3788/AOS201939.0628005

Improved SSD Algorithm and Its Performance Analysis of Small Target Detection in Remote Sensing Images

Wang Junqiang^{1,2}, Li Jiansheng^{1*}, Zhou Xuewen², Zhang Xu¹

¹*Institute of Geospatial Information, Information Engineering University,
Zhengzhou, Henan 450000, China;*

²*78123 Troops, Chengdu, Sichuan 610000, China*

Abstract An improved single shot multibox detector (SSD) algorithm is proposed aiming at the problems of slow detection speed of the target proposal based remote sensing image target detection method represented by faster regions with convolutional neural network (R-CNN) and the low performance in small target detection by the SSD algorithm. The algorithm can combine the advantages of the existing detection methods based on target proposal and one-stage target detection to improve the target detection performance. Furthermore, the algorithm replaces the original visual geometry group net with a densely connected network as the backbone network and constructs a feature pyramid between the densely connected modules instead of the original multi-scale feature map. A sample data online acquisition system is designed to verify the accuracy and performance of the proposed algorithm. A sample set of aircraft and playground target is collected as the experimental sample. The network structure stability is verified by training the improved SSD algorithm. Consequently, good results can be achieved without the support of transfer learning. Moreover, the training process is not easy to diverge. By comparing the Faster R-CNN algorithm using ResNet101 as the backbone network and the R-FCN (region-based fully convolutional networks) algorithm, we find that the mean average precision (MAP) of the improved SSD algorithm is 9.13% and 8.48% higher than that of the faster R-CNN and R-FCN algorithms in the test set, respectively. The proposed SSD algorithm improves the MAP in the small target detection by 14.46% and 13.92% compared to the faster R-CNN and R-FCN algorithms, respectively. Detecting a single image takes 71.8 ms, which is 45.7 ms and 7.5 ms less than that of the faster R-CNN and R-FCN algorithms, respectively.

收稿日期: 2019-01-16; 修回日期: 2019-02-25; 录用日期: 2019-03-12

基金项目: 国家自然科学基金(41876105)、国家重点研发计划资助(2017YFF0206000)

* E-mail: xindawangjunqiang@163.com

Key words remote sensing; small target detection; deep learning; multi-scale prediction; feature pyramid; mean average precision

OCIS codes 100.3008; 100.2000; 110.2970

1 引 言

遥感影像目标检测是遥感影像解译的重要分支之一,对于挖掘感兴趣信息至关重要。传统的方法有基于统计的目标检测^[1]、基于知识的目标检测^[2]以及基于模型的目标检测^[3]等,这些方法需要人工设定先验条件,稳健性差,智能程度低,难以应用于大范围自动化作业的需求。

近年来,随着卷积神经网络在 2012 年的 ImageNet^[4] 图像分类比赛中取得巨大成功后,基于深度学习的目标检测方法开始应用于图像目标检测领域。2013 年起,基于深度学习的目标检测算法沿着两条主线发展:一条是以 Fast R-CNN^[5]、Faster R-CNN^[6]、R-FCN^[7] 算法为典型的基于候选框方式的检测主线;另外一条是以 YOLO(You Only Look Once)^[8]、SSD^[9]、Retina-Net^[10] 为代表的一体化卷积神经网络检测算法。一体化检测方式的速度要明显快于基于候选框的检测算法,但检测精度要略逊于后者。文献[11]对当前主流的 Faster R-CNN、R-FCN 及 SSD 三种检测算法在 Microsoft COCO(common objects in context)^[12] 公开数据集上的性能进行详细对比,结果表明 Faster R-CNN 的精度最高,而 SSD 的速度最快。随着深度学习技术的发展,不少学者开始研究利用深度学习方法进行遥感影像目标检测,文献[13]全面比较了 Fast R-CNN、Faster R-CNN 和 R-FCN 三种算法对飞机目标检测的识别性能,其结果表明 R-FCN 算法的准确率和检测速度较优。文献[14]将深度学习引入飞机检测中,构建了基于深度信念网络及基于卷积神经网络的两种飞机检测结构模型,实现了飞机目标高精度检测。文献[15]基于 Faster R-CNN 方法对遥感影像中飞机、油罐等目标进行验证实验,取得较好效果。文献[16]提出了适应空中目标检测任务的特点和需求的 Faster R-CNN 改进策略,弥补了 Faster R-CNN 算法对弱小目标和被遮挡目标不敏感的缺陷并提升了检测精度。以上研究表明,基于候选框的算法是当前应用于遥感影像目标检测的主流方法,尤其是 Faster R-CNN 算法的应用最广泛,但以上研究均未对遥感影像小目标检测性能进行全面分析,而遥感影像相对于常规自然场

景图片具有幅面大、分辨率低等特点,但需要在较小分辨率下进行检测才能实现在大范围内自动化高效作业,在此情况下,目标大小在影像中常表现为中小型特征,因此,需要更多地关注中小型目标的检测。Faster R-CNN 算法对于小目标的检测性能比 SSD 算法好,但检测时间更长,无法满足大范围自动化检测的需求。Feature Pyramid Networks (FPN)方法^[17]通过将网络中最顶层的特征图像逐层地反馈并与前层的特征图进行融合,提升低层次特征的语义强度,并且在不同尺度上进行目标检测,可提升小目标的检测优势,因此,为平衡速度和精度,可将 FPN 特征提取方法应用至 SSD 算法中,但存在问题是仅在 SSD 算法特征提取网络部分引入 FPN 特征提取方法后,检测速度将会进一步降低。

为解决以上问题,本文针对现有 SSD 算法进行改进,参考密集连接的 DenseNet (densely connected convolutional networks) 网络^[18] 设计特征提取网络,并将其替代原有的 16 层 VGGNet(VGG-16)^[19] 网络,并引入 FPN 特征提取方法,进行多尺度特征融合,以替代原有的多尺度预测方法。

2 改进的 SSD 算法

2.1 SSD 算法

SSD 是 Liu 等^[9] 提出的一种目标检测算法,是目前较流行的检测框架之一,相比于 Faster R-CNN 算法,SSD 算法在速度上具有明显优势,而相比于 YOLO 算法,SSD 算法在精度上又有明显的优势。SSD 算法框架如图 1 所示,该算法以 VGG-16 为基础网络,在此基础上添加辅助卷积、池化层等结构,得到的多尺度特征图均用来进行目标检测,较大特征图用于检测相对较小目标,而尺寸较小特征图负责大目标检测。

SSD 算法参照 YOLO 算法的一体化检测方式,但与 YOLO 算法最后采用全连接层预测不同的是,SSD 算法采用卷积对不同尺度特征图进行检测,同时借鉴了 Faster R-CNN 中 anchor 的思想,在特征图中的每个单元设置长宽比不同的先验框,预测时目标框以这些先验框为基准计算偏差,可在一定程度上降低训练的难度。

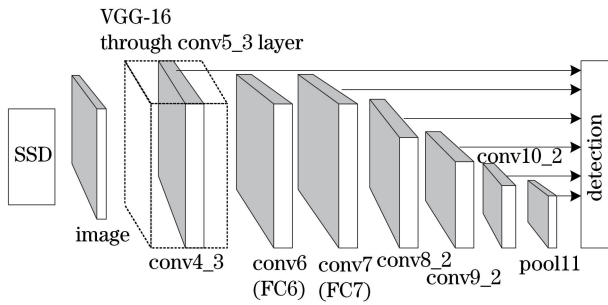


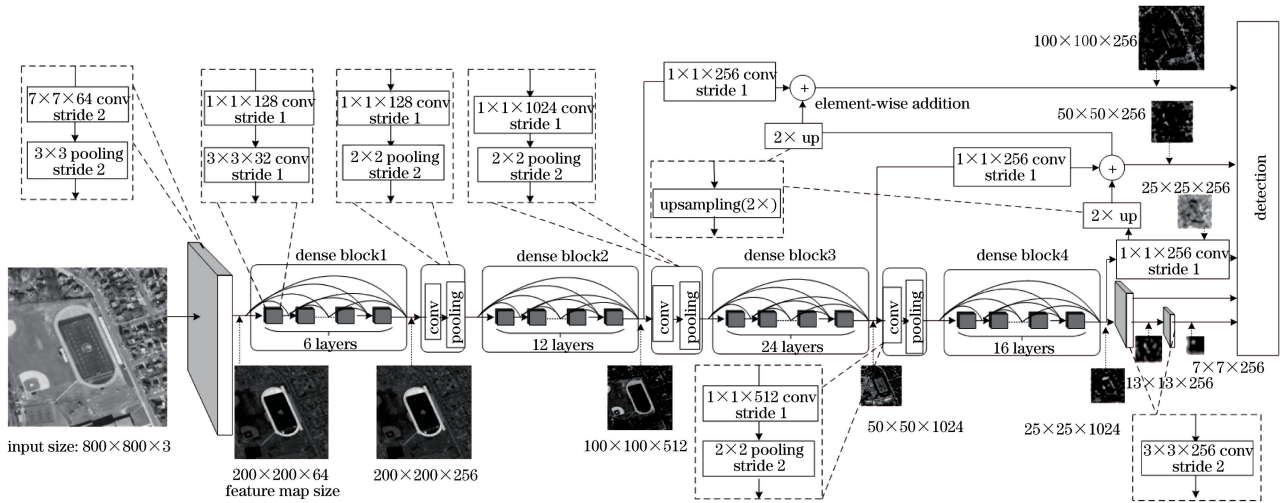
图 1 SSD 算法框架

Fig. 1 Framework of SSD algorithm

2.2 改进的 SSD 算法

SSD 算法虽采用多尺度预测,但不同尺度特征相互独立,其低层次特征位置信息较好,但分类精度差。为降低低层次特征带来的误差,SSD 算法从

VGG-16 网络中偏后的 conv4_3 卷积层开始构建多尺度特征图,其检测小目标性能较差。此外,SSD 算法的输入图像尺寸较小,文献[9]选取的输入图像尺寸为 300 pixel×300 pixel,小尺寸下能够提升目标检测速度,但对于遥感图像自动化检测任务,将图像固定至小尺寸进行检测,易造成目标特征信息的丢失。为实现良好检测性能,本研究固定输入图像为 800 pixel×800 pixel,导致计算量大幅增加,故采用原来网络来降低检测速度。为解决这些问题,对现有 SSD 算法做以下改进:1)参考密集连接的 DenseNet 网络结构,重新设计 SSD 算法中的基础网络;2)利用 FPN 方法对密集连接网络的全局池化模块构建特征金字塔,再进行多尺度目标检测。改进后的算法框架如图 2 所示。



建特征金字塔的第 3 个特征图。相比于 SSD 算法,改进后的 SSD 算法融入了多个高层次特征的语义信息,可弥补 SSD 算法低层次特征语义信息差的问题,而大部分小目标的检测是利用低层次特征图实现,因此可提升小目标检测精度。另一方面,由于采用了密集连接的网络结构,大幅度减少网络中的参数量,可在一定程度上弥补输入图像尺寸增大带来的计算效率的降低,而相比于以 Faster R-CNN 为代表的基于候选框的检测方式,该方法无须进行区

域建议网络的训练过程,因此更易于训练。图 3 为改进后的 SSD 算法在 dense 模块融合前后特征图输出的对比,选取了三个通道绘制,图 3(e)是图 3(f)融合图 3(d)后的特征图,图 3(d)是图 3(b)融合图 3(c)后的特征图,显然图 3(d)和图 3(e)相比于图 3(b)和图 3(c)飞机目标高亮显示,具备更强的语义信息,理论上在图 3(d)和图 3(e)上预测结果会明显优于图 3(b)和图 3(c),尤其是绝大部分小目标在低层次特征图上预测,因此可提升小目标检测精度。

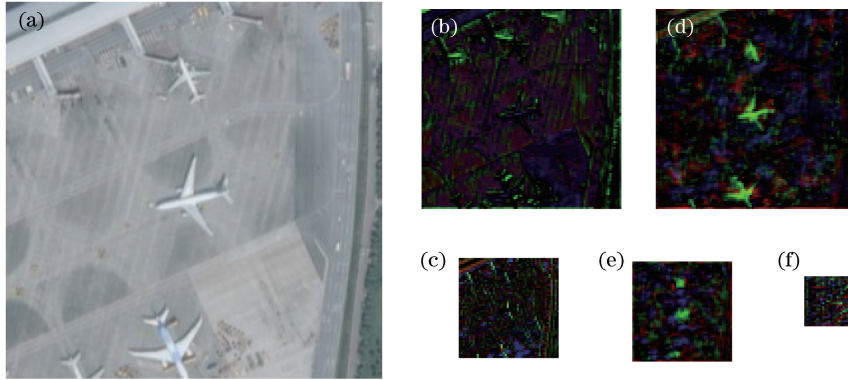


图 3 融合前后的特征图对比。(a) 输入图像;(b) dense 模块 2 输出;(c) dense 模块 3 输出;(d) dense 模块 2 融合特征后的输出;(e) dense 模块 3 融合特征后的输出;(f) dense 模块 4 输出

Fig. 3 Comparison of feature maps before and after integration. (a) Input image; (b) output of dense block2; (c) output of dense block3; (d) output of dense block2 with feature integration; (e) output of dense block3 with feature integration; (f) output of dense block4

2.3 损失函数及训练方法

将 SSD 算法的损失函数表示为位置损失与分类损失的加权和^[9]:

$$L(x, c, l, g) = \frac{1}{N} [L_{\text{class}}(x, c) + \alpha L_{\text{loc}}(x, l, g)], \quad (3)$$

式中: x 表示真实框; c 表示预测框; l 为预测的位置信息; g 为真实框的位置信息; N 为与真实框相匹配的先验框个数(正样本数量); $L_{\text{class}}(x, c)$ 为分类损失; $L_{\text{loc}}(x, l, g)$ 为位置损失; α 为权值系数,本文设定为 1。 $L_{\text{loc}}(x, l, g)$ 借鉴 Faster R-CNN 的位置回归函数 $\text{smooth}_{\text{L}_1}$, 表示为

$$L_{\text{loc}}(x, l, g) = \sum_{i \in N_{\text{pos}}} \sum_{m \in \{C_X, C_Y, w, h\}} x_{ij}^{(k)} \times \text{smooth}_{\text{L}_1}(l_i^{(m)} - \hat{g}_j^{(m)}), \quad (4)$$

式中: $x_{ij}^{(k)} \in \{0, 1\}$, 当 $x_{ij}^{(k)} = 1$ 时表示第 i 个先验框与第 j 个真实框相匹配, 并且类别为 k , 否则为 0; N_{pos} 表示正样例集合; (C_X, C_Y, w, h) 分别表示边界框中心像素坐标以及宽高; $\hat{g}_j^{(m)}$ 为编码后的真实框位置参数; $l_i^{(m)}$ 表示先验框的预测值。 $\text{smooth}_{\text{L}_1}$ 函数可表示为

$$\text{smooth}_{\text{L}_1}(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & \text{else} \end{cases}. \quad (5)$$

$L_{\text{class}}(x, c)$ 函数采用交叉熵损失函数^[23]来表示:

$$L_{\text{class}}(x, c) = - \sum_{i \in N_{\text{pos}}} x_{ij}^{(p)} \log(\hat{c}_i^{(p)}) - \sum_{i \in N_{\text{neg}}} \log(\hat{c}_i^{(0)}), \quad (6)$$

式中: $\hat{c}_i^{(0)}$ 表示正确且类别为背景预测框的概率; N_{neg} 表示负样例集合; $\hat{c}_i^{(p)}$ 为利用 softmax 函数计算的概率值, 可表示为

$$\hat{c}_i^{(p)} = \frac{\exp(c_i^{(p)})}{\sum_p \exp(c_i^{(p)})}. \quad (7)$$

采用数据增强方式对图像进行增强, 丰富样本数量, 迁移学习带来的特征也优于直接从随机初始化学习的特征^[24], 因此, 基于 ImageNet 数据集上预训练的模型进行迁移学习, 利用随机梯度下降算法 (SGD) 对损失函数(3)式进行优化, 寻求最优解, 在训练过程中学习率逐渐递减, 学习动量为 0.9。

3 精度评估模型

以平均准确率均值 (MAP)^[25] 作为衡量模型训

练结果精度的指标。每一个类别都可以根据召回率和准确率绘制一条曲线,那么准确率均值(AP)就是该曲线下的面积,而 MAP 是多个类别 AP 的平均值, AP 计算可表示为

$$R_{AP} = \int_0^1 p(r) dr. \quad (8)$$

利用预测框与真实框的交并比(R_{IoU})作为判

定预测真假的前提,Pascal VOC2010 数据集^[25]以 0.5 为交并比阈值,再根据预测框内类别置信度进行判定,从而确定预测为真或假样例。本研究根据更加全面客观的 COCO 数据集衡量标准^[12]进行衡量。COCO 数据集依然采用 MAP 作为衡量标准,但衡量更加多样化,其主要衡量指标如表 1 所示。

表 1 主要衡量指标
Table 1 Main metrics

Metric	Remarks
MAP	MAP at R_{IoU} in $\{0.5+0.05 \times m, m=0,1,\dots,9\}$ (primary challenge metric)
$MAP^{R_{IoU}=0.50}$	MAP at $R_{IoU}=0.50$ (pascal VOC metric)
$MAP^{R_{IoU}=0.75}$	MAP at $R_{IoU}=0.75$ (strict metric)
MAP^{small}	MAP for small targets: $S_{area} < (32 \text{ pixel})^2$
MAP^{medium}	MAP for medium targets: $(32 \text{ pixel})^2 \leq S_{area} \leq (96 \text{ pixel})^2$
MAP^{large}	MAP for large targets: $S_{area} > (96 \text{ pixel})^2$

4 实验分析

4.1 实验平台与数据

实验硬件采用联想 P920 工作站,配置 32 G 内存及 NVIDIA TITAN Xp 显卡,操作系统为 Ubuntu 16.04。在 Python 中基于 Tensorflow 深度学习框架构建算法模型。

为验证小目标的检测性能,参考 COCO 数据集的大、中、小目标所占比例,采集制作飞机和运动场 2 类典型目标数据集,样本采集工作利用基于 ArcGIS API for JavaScript 组件在 B/S 架构下设计开发的遥感影像目标检测训练样本数据在线采集系统开展,系统界面如图 4 所示,通过叠加 DigitalGlobe 公司的公开多源影像数据、公开民用机场点位数据(网址:<http://ourairports.com>),快速定位机场位置及主要城市,采集 4 个影像瓦片层级影像数据(第 15~18 层)并自动生成对应的 XML 标签数据,相对于主流的 LabelImg 目标检测样本标记软件,该方法更加高效、便捷。

采集不同场景下的样本集图片共计 2574 张,其中验证集 412 张,测试集 326 张,统计样本集目标分布情况如表 2 所示,依据 COCO 数据集定义的大、中、小目标划分,各类数据集均以中、小目标为主,尤其是测试集的中、小目标占比最大,可用于对小目标的检测验证,图 5 所示为目标实际大小及其对应的目标等级。

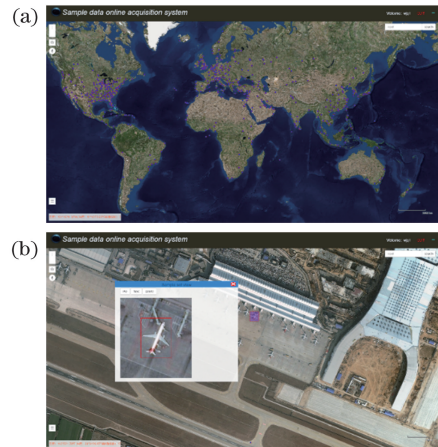


图 4 训练样本在线采集系统界面。(a) 叠加主要机场点位数据;(b) 飞机样本采集

Fig. 4 Interface of training sample online acquisition system. (a) Superimposed main airport point data; (b) aircraft sample collection

4.2 改进 SSD 算法训练分析

数据增强采用随机裁切、随机旋转及缩放操作,初始学习率设定为 0.01,采用余弦函数衰减法^[26]对学习率进行衰减,学习率衰减曲线如图 6 所示。

采用随机梯度下降法对总损失函数进行优化训练,设定批处理尺寸为 4,训练曲线及 $MAP^{R_{IoU}=0.50}$ 如图 7 所示,同时也利用 ImageNet 数据集上预训练的 DenseNet 对改进 SSD 算法基础网络的部分权值参数进行迁移学习的训练曲线进行绘制,作为对比实验。由图 7(a)可知,两种训练方式的总损

表 2 样本集统计
Table 2 Sample set statistics

Data set	Class	Target amount				Percentage /%		
		Small	Medium	Large	Total	Small	Medium	Large
Training set	airplane	1204	1542	431	4401	27.36	35.04	9.79
	playground	178	516	530		4.04	11.72	12.04
Validation set	airplane	390	427	74	1169	33.36	36.53	6.33
	playground	27	120	131		2.31	10.27	14.21
Test set	airplane	940	366	111	1892	49.68	19.34	5.87
	playground	133	294	48		7.03	15.54	2.54

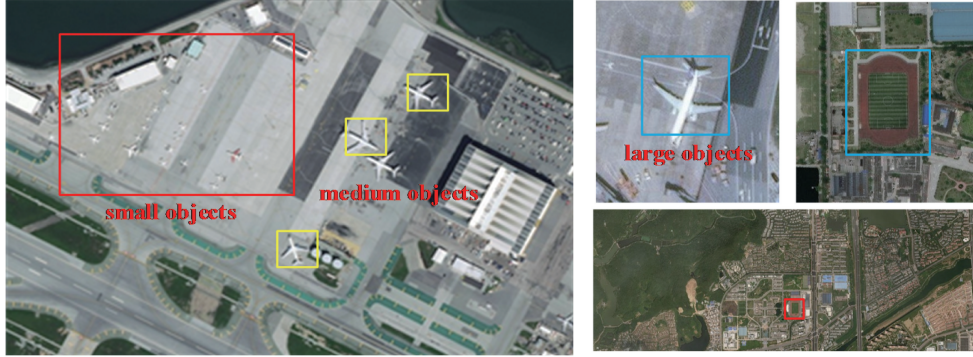


图 5 样本集各类大小目标示意图
Fig. 5 Size of each target in sample set

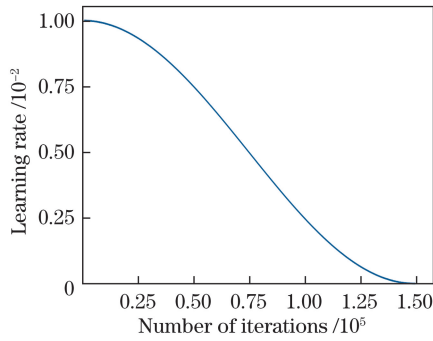


图 6 学习率衰减曲线
Fig. 6 Decay curve of learning rate

失最终均能收敛至同一水准(0.1 以内),并且收敛效果均较好,由图 7(b)可知,部分权值迁移训练方式在前期训练过程中精度增幅较快,但后期基本保持一致,最终 $MAP^{R_{IoU}=0.50}$ 为 90.55%,较随机初始化仅增加 0.53%,增幅较小。综上所述,改进后的 SSD 模型易于训练,在反向传播过程中梯度不易发散,并且在无迁移训练支撑下能够获得较理想的结果,验证了改进 SSD 算法网络结构的有效性。

4.3 改进 SSD 算法精度分析

为验证改进 SSD 算法的性能,同时利用训练集训练了基于候选框方式的 Faster R-CNN 和 R-FCN 两种算法作为对比,并在验证集上进行精度验证,

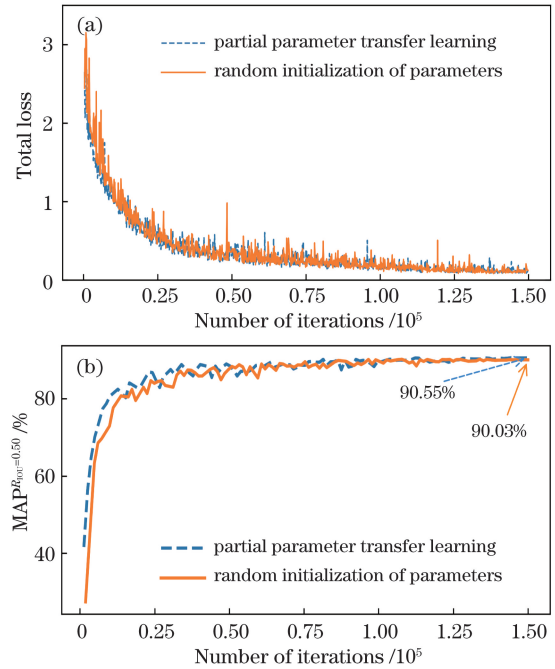


图 7 迁移训练和随机初始化两种方式总损失和精度对比。(a) 总损失随迭代次数变化;(b) $MAP^{R_{IoU}=0.50}$ 随迭代次数变化

Fig. 7 Comparison of total loss and precision between transfer training and random initialization. (a) Total loss varies with number of iterations; (b) $MAP^{R_{IoU}=0.50}$ varies with number of iterations

Faster R-CNN 分别以 ResNet50 和 ResNet101 两种网络作为基础网络, R-FCN 以 ResNet101 作为基础网络, 利用精度更高的 Inceptionv2^[27] 网络替换 SSD 算法的 VGG-16, 也参与对比实验。以上算法图像的输入尺寸均为 800 pixel × 800 pixel, 除改进 SSD 算法外, 其余算法均利用 COCO 数据集上的预训练模型

进行迁移学习, 改进 SSD 算法利用 4.2 节所提的迁移方式进行部分参数迁移, 以上算法均训练迭代 1.5×10^5 次, 并分别在验证集和测试集上进行精度分析统计。单张影像预测耗时及验证集精度指标的对比结果如表 3 所示, 验证集精度随迭代次数变化情况如图 8 所示, 测试集上的精度对比如表 4 所示。

表 3 计算耗时及验证集上精度对比

Table 3 Comparison of calculation time and precision on validation set

Method	Parameter /MB	Time overhead /ms	Metric / %					
			MAP	MAP ^{large}	MAP ^{medium}	MAP ^{small}	MAP ^{R_{IoU}=0.50}	MAP ^{R_{IoU}=0.75}
SSD+Inceptionv2	53.4	24.8	47.15	69.48	50.16	11.56	85.08	48.95
Faster R-CNN+ResNet50	173.3	108.6	47.12	72.81	48.91	10.16	82.72	50.59
Faster R-CNN+ResNet101	249.5	117.5	50.50	73.06	53.51	13.76	85.84	55.03
R-FCN+ResNet101	258.2	79.3	51.17	74.69	51.43	16.01	87.20	55.86
Improved SSD algorithm	59.8	71.8	54.14	73.31	54.32	21.16	90.55	57.38

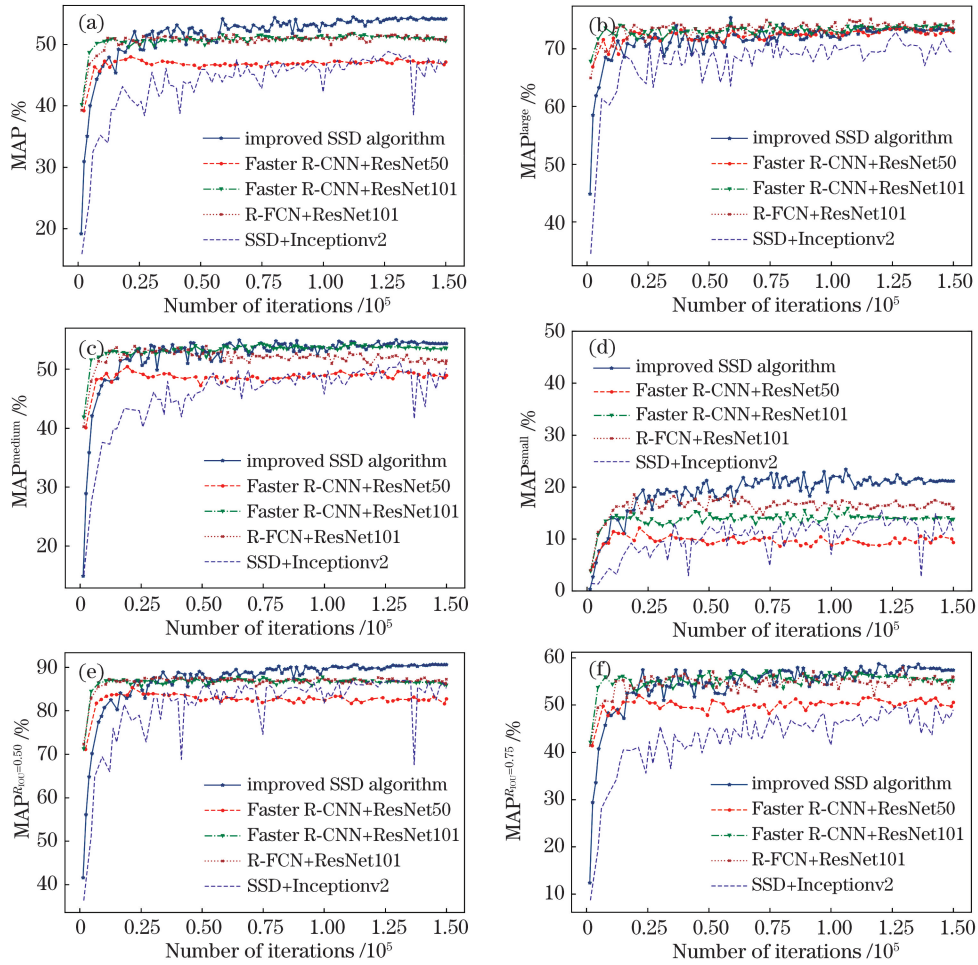


图 8 改进 SSD 算法与其他算法随迭代次数精度变化对比。(a) MAP; (b) MAP^{large}; (c) MAP^{medium}; (d) MAP^{small}; (e) MAP^{R_{IoU}=0.50}; (f) MAP^{R_{IoU}=0.75}

Fig. 8 Comparison of precisions of improved SSD algorithm and other algorithms varying with number of iterations. (a) MAP; (b) MAP^{large}; (c) MAP^{medium}; (d) MAP^{small}; (e) MAP^{R_{IoU}=0.50}; (f) MAP^{R_{IoU}=0.75}

表 4 测试集上精度对比
Table 4 Comparison of precision on test set

Methods	Metric / %					
	MAP	MAP ^{large}	MAP ^{medium}	MAP ^{small}	MAP ^{R_{IoU}=0.50}	MAP ^{R_{IoU}=0.75}
SSD+Inceptionv2	35.85	62.73	43.55	19.52	77.72	24.07
Faster R-CNN+ResNet50	28.72	61.53	34.67	13.43	68.87	17.60
Faster R-CNN+ResNet101	36.05	61.55	42.83	21.74	76.83	27.69
R-FCN+ResNet101	36.70	59.18	43.96	22.91	77.30	28.23
Improved SSD algorithm	45.18	65.31	50.68	31.65	83.95	42.15

由表 3 可知,改进 SSD 算法的模型参数量较 SSD+Inceptionv2 仅增加了 6.4 MB,但相对于 Faster R-CNN 和 R-FCN 算法降低较多,尤其是仅为 R-FCN+ResNet101 的 23%。从预测单张影像时间开销来看,改进 SSD 算法较 SSD+Inceptionv2 增加 47 ms,但相对于 Faster R-CNN 和 R-FCN 算法有所降低,尤其是相对于 Faster R-CNN+ResNet101 降低 45.7 ms。从精度指标来看,改进 SSD 算法除验证集上的大目标检测精度 MAP^{large} 相对于 R-FCN+ResNet101 降低了 1.38% 外,其余指标均优于其他方法,尤其是在检测小目标上优势更明显,在测试集上,MAP^{small} 相比于次高的 R-FCN+ResNet101 增加了 13.92%,说明改进 SSD 算法在检测中、小型目标上优势明显,如表 4 所示。由图 8 可知,Faster R-CNN 和 R-FCN 算法相对于改进 SSD 算法收敛较快,说明改进后的 SSD 算法收敛时间相对较长,这与其中部分参数随机初始化具有一定关系,对比改进 SSD 算法与 SSD+Inceptionv2 算法可知,SSD+Inceptionv2 精度随着迭代步数的变化起伏较大,尤其是图 8(d)中小目标的精度突变最为明显,说明 SSD 算法对小目标数据集训练效果差,进一步验证了 SSD 低层次特征对于小目标识别较差,但改进后的 SSD 算法能够较好适应小目标数据,改进 SSD 算法的各精度指标变化总体较为平稳。

综上所述,改进 SSD 算法精度较 SSD、Faster R-CNN 和 R-FCN 算法有较大提升,尤其是在检测小目标上优势更明显;改进 SSD 算法对单张图像预测耗时较 Faster R-CNN 算法和 R-FCN 算法有所降低,尤其是相对 Faster R-CNN 算法降低较大,但由于特征金字塔的构建,相对于 SSD 算法有所增加;改进的 SSD 算法能够解决 SSD 算法对小目标数据集的适应效果差的问题。

4.4 检测效果对比

通过 4.3 节的分析可知,以 ResNet101 为基础

网络的 Faster R-CNN 算法和 R-FCN 算法的检测精度较高,本节以这两种算法作为对比,分析改进后的 SSD 算法检测小目标的效果,如图 9 所示,预测图像为 2 张分别带有飞机目标和运动场目标的大尺寸影像,缩小至 800 pixel×800 pixel 的输入尺寸后,目标将表现小目标特征,图 9 中列出了检测效果图的局部区域放大图。由图 9(a)和图 9(b)可知,Faster R-CNN 算法和 R-FCN 算法检测边框的置信度较高,但存在一些误检,误检框主要出现在对小目标的识别。由图 9(c)可知,所有小目标均被检测出,且没有误检框,尽管边界框置信度相对 Faster R-CNN 算法和 R-FCN 算法较低,但不影响最终结果的判断,主要原因是改进 SSD 算法属于一体化检测方式,没有进行候选框预训练这个过程,相对于基于候选框的检测方式较低。对比图 9 的所有放大图的检测边框位置信息可知,改进后的 SSD 算法预测边框位置回归准确度明显高于 Faster R-CNN 算法和 R-FCN 算法。

5 结 论

对现有 SSD 算法进行改进,大幅度提升了算法的计算精度,解决了原有 SSD 算法小目标检测精度低而现有基于候选框的方法检测速度慢的问题。通过对以飞机和运动场中小型样本为主的检测实验进行分析,可得到以下结论。1)改进后的 SSD 算法的网络结构设计更合理,且训练过程更容易收敛,在无迁移学习支撑下也能够获得较理想的结果。2)改进后的 SSD 算法相对于原 SSD 算法、Faster R-CNN 算法以及 R-FCN 算法总体精度有大幅度提升,尤其是在检测小目标上,效果更加明显,测试集上相比于精度次高的 R-FCN 算法可提升 13.92%,说明这种高层次特征与低层次特征融合具有较好的语义和位置信息。3)从目标检测时间开销来看,改进后的 SSD 算法预测单张影像相比于 Faster R-CNN 算法

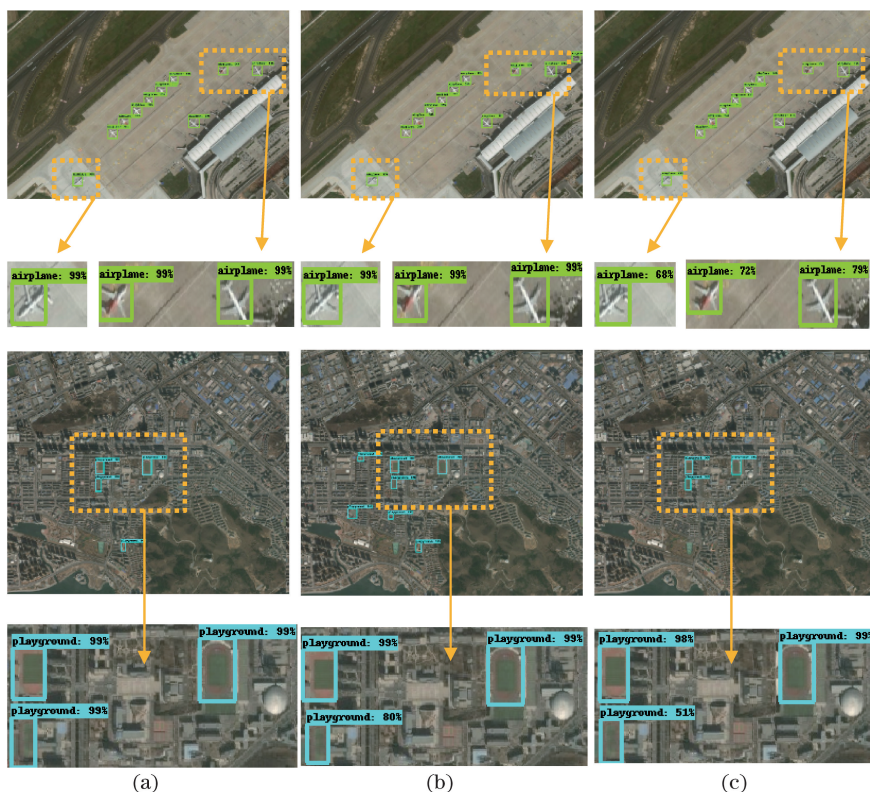


图 9 改进 SSD 与其他方法检测效果对比。(a) Faster R-CNN+ResNet101 方法;
(b) R-FCN+ResNet101 方法;(c) 改进 SSD 算法

Fig. 9 Comparison of improved SSD algorithm and other algorithms in detection effect. (a) Faster R-CNN+ResNet101;
(b) R-FCN+ResNet101; (c) improved SSD algorithm

降低 45.7 ms, 相比于速度较快的 R-FCN 算法降低 7.5 ms, 但相比于原 SSD 算法有所增加, 主要原因是特征金字塔构建耗时过多。4) 从预测效果来看, 改进后的 SSD 算法预测位置边框要优于原 SSD 算法、Faster R-CNN 算法以及 R-FCN 算法, 其置信度相比 Faster R-CNN 算法以及 R-FCN 算法更低, 但不影响最终的预测结果。

改进的 SSD 算法对于将深度学习方法高效地用于大区域遥感影像目标检索具有一定参考价值, 尤其是在遥感影像小目标的检测上具有较大优势。尽管改进的 SSD 算法较基于候选框的目标检测方法时间开销少, 但相比于原 SSD 算法耗时依然过多, 下一步将对网络结构进一步优化, 提升其计算速度。

参 考 文 献

- [1] Wang G X, Huang X T, Zhou Z M. UWB SAR change detection of target in foliage based on local statistic distribution change analysis[J]. Journal of Electronics & Information Technology, 2011, 33(1): 49-54.
王广学, 黄晓涛, 周智敏. 基于邻域统计分布变化分

析的 UWB SAR 隐蔽目标变化检测[J]. 电子与信息学报, 2011, 33(1): 49-54.

- [2] Wu W. Research on knowledge-based target recognition and tracking techniques[D]. Harbin: Harbin Institute of Technology, 2007: 10-28.
吴畏. 基于知识的目标识别与跟踪技术研究[D]. 哈尔滨: 哈尔滨工业大学, 2007: 10-28.
- [3] Cao J Z, Song A G. Research on the texture image segmentation method based on Markov random field[J]. Chinese Journal of Scientific Instrument, 2015, 36(4): 776-786.
曹家梓, 宋爱国. 基于马尔科夫随机场的纹理图像分割方法研究[J]. 仪器仪表学报, 2015, 36(4): 776-786.
- [4] Russakovsky O, Deng J, Su H, *et al.* ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
- [5] Girshick R. Fast R-CNN[C] // 2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 1440-1448.
- [6] Ren S Q, He K M, Girshick R, *et al.* Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern

- Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [7] Dai J, Li Y, He K, *et al.* R-FCN: object detection via region-based fully convolutional networks[C]//NIPS'16 Proceedings of the 30th International Conference on Neural Information Processing Systems, December 5-10, 2016, Barcelona, Spain. USA: Curran Associates Inc., 2016: 379-387.
- [8] Redmon J, Divvala S, Girshick R, *et al.* You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 779-788.
- [9] Liu W, Anguelov D, Erhan D, *et al.* SSD: single shot multibox detector[M]//Leibe B, Matas J, Sebe N, *et al.* Computer Vision-ECCV 2016. Cham: Springer, 2016, 9905: 21-37.
- [10] Lin T Y, Goyal P, Girshick R, *et al.* Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018: 2858826.
- [11] Huang J, Rathod V, Sun C, *et al.* Speed/accuracy trade-offs for modern convolutional object detectors [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 3296-3297.
- [12] Lin T Y, Maire M, Belongie S, *et al.* Microsoft COCO: common objects in context[M]//Fleet D, Pajdla T, Schiele B, *et al.* Computer Vision-ECCV 2014. Cham: Springer, 2014, 8693: 740-755.
- [13] Xu Y Z, Yao X J, Li X, *et al.* Object detection in high resolution remote sensing images based on fully convolution networks[J]. Bulletin of Surveying and Mapping, 2018(1): 77-82.
徐逸之, 姚晓婧, 李祥, 等. 基于全卷积网络的高分辨率遥感影像目标检测[J]. 测绘通报, 2018(1): 77-82.
- [14] Zhang Z Y. Plane detection in optical remote sensing images based on deep learning[D]. Xiamen: Xiamen University, 2016: 20-30.
张志远. 基于深度学习的光学遥感图像飞机检测[D]. 厦门: 厦门大学, 2016: 20-30.
- [15] Wang J C, Tan X C, Wang Z H, *et al.* Faster R-CNN deep learning network based object recognition of remote sensing image[J]. Journal of Geo-Information Science, 2018, 20(10): 1500-1508.
王金传, 谭喜成, 王召海, 等. 基于 Faster R-CNN 深度网络的遥感影像目标识别方法研究[J]. 地球信息科学学报, 2018, 20(10): 1500-1508.
- [16] Feng X Y, Mei W, Hu D S. Aerial target detection based on improved faster R-CNN[J]. Acta Optica Sinica, 2018, 38(6): 0615004.
冯小雨, 梅卫, 胡大帅. 基于改进 Faster R-CNN 的空中目标检测[J]. 光学学报, 2018, 38(6): 0615004.
- [17] Lin T Y, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 936-944.
- [18] Huang G, Liu Z, Maaten L V D, *et al.* Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2261-2269.
- [19] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10)[2018-12-22]. <https://arxiv.org/abs/1409.1556>.
- [20] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]//ICML'15 Proceedings of the 32nd International Conference on International Conference on Machine Learning, July 6-11, 2015, Lille, France. Massachusetts: JMLR. org, 2015: 448-456.
- [21] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.
- [22] Parker J A, Kenyon R V, Troxel D E. Comparison of interpolating methods for image resampling [J]. IEEE Transactions on Medical Imaging, 1983, 2(1): 31-39.
- [23] de Boer P T, Kroese D P, Mannor S, *et al.* A tutorial on the cross-entropy method[J]. Annals of Operations Research, 2005, 134(1): 19-67.
- [24] Yosinski J, Clune J, Bengio Y, *et al.* How transferable are features in deep neural networks? [EB/OL]. (2014-11-06)[2018-12-22]. <https://arxiv.org/abs/1411.1792>.
- [25] Everingham M, van Gool L, Williams C K I, *et al.* The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.
- [26] Loshchilov I, Hutter F. SGDR: stochastic gradient descent with warm restarts[EB/OL]. (2017-03-03)[2018-12-25]. <https://arxiv.org/abs/1608.03983>.
- [27] Szegedy C, Vanhoucke V, Ioffe S, *et al.* Rethinking the inception architecture for computer vision[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 2818-2826.