弱监督深层神经网络遥感图像目标检测模型

周明非¹, 汪西莉^{1,2*}

1. 陕西师范大学计算机科学学院, 西安 710119

2. 陕西师范大学现代教学技术教育部重点实验室, 西安 710119

* 通信作者. E-mail: wangxili@snnu.edu.cn

收稿日期: 2018-01-02; 接受日期: 2018-01-19; 网络出版日期: 2018-08-08

国家自然科学基金 (批准号: 41471280, 61701290, 61701289) 资助项目

摘要 本文提出了一种遥感图像目标检测框架, 克服了遥感图像中由于目标较小且背景复杂造成的目标检测任务中的困难. 所提框架包含两种深层神经网络模型, 分别是全卷积网络模型和卷积神经网络模型. 首先, 使用全卷积网络提取遥感图像中可能存在待检测目标的候选区域, 避免了对图像的穷举搜索. 其次, 使用深层卷积神经网络对候选区域分类, 通过提取高层特征提高分类正确率. 然后, 提出了新的遥感图像目标检测数据集, 模型的训练全部使用图像级的标签, 提出简化弱监督训练方法解决遥感图像目标检测领域目标级标签缺乏的问题. 最后, 提出一种候选框融合算法, 合并重叠候选框的同时调整候选框的位置. 提出的模型在本文所提数据集 satellite aircrafts dataset 和公开数据集 aircrafts dataset 上进行了测试. 实验结果表明, 提出的目标检测框架和其他使用深层神经网络的框架相比提高了目标检测的正确率, 并具有更高的检测效率.

关键词 遥感图像, 目标检测, 卷积神经网络, 全卷积网络, 弱监督训练

1 引言

遥感图像的目标检测指的是在大幅高分辨遥感图像中检测出一类或者几类目标并标注其位置, 飞机是常见的检测目标之一. 早期的遥感图像目标检测主要是基于模板匹配^[1]和形状先验^[2]的, 随着机器学习研究的发展, 目标检测问题逐渐被转换为分类问题^[3]. 如图1蓝色区域所示, 在使用机器学习方法的目标检测框架中, 需要依次进行: 获取候选区域、筛选候选区域、对候选区域进行特征提取、分类候选区域和后处理 (处理重叠结果) 等操作. 其中最重要的是候选区域的获取以及特征提取和分类. 候选区域是指待检测图像中可能包含目标的区域, 对其进行特征提取, 并把特征送入分类器, 就可以使用分类器判定候选区域属于目标还是背景. 以往采用人工设计并提取的特征, 不足以应对遥感图像中的较小目标和复杂背景带来的挑战. 随着卷积神经网络 (convolutional neural networks, CNNs) 等深度

引用格式: 周明非, 汪西莉. 弱监督深层神经网络遥感图像目标检测模型. 中国科学: 信息科学, 2018, 48: 1022–1034, doi: 10.1360/N112017-00208
Zhou M F, Wang X L. Object detection models of remote sensing images using deep neural networks with weakly supervised training method (in Chinese). Sci Sin Inform, 2018, 48: 1022–1034, doi: 10.1360/N112017-00208

图1 (网络版彩图) 结合机器学习和深度学习的遥感图像目标检测框架

Figure 1 (Color online) The object detection framework of remote sensing images with machine learning and deep learning

学习方法的发展和在图像分类^[4]、目标检测^[5]等领域的成功运用,深度学习的方法也被用于遥感图像目标检测。但是由于遥感图像领域样本的缺乏,主流检测框架^[5]无法直接应用在遥感图像的目标检测上。

2015年,Wu等^[6]首次使用浅层CNN模型(5层)完成飞机检测任务。浅层CNN模型直接使用少量的带图像级标签的遥感图像样本进行训练,可以在自动提取特征的同时完成分类。和使用支持向量机对人工设计的特征进行分类相比,浅层CNN模型可以自动提取遥感图像的本质特征,并提高识别精度。同年,Wu等^[7]采用类似方法使用两个独立训练的浅层CNN模型同时实现了飞机检测和储油罐检测。2016年,Zhang等^[8]使用深层CNN模型(11层)实现了遥感图像机场检测。Zhang等^[8]使用相邻平行线段搜索提取候选区域,并使用CNN提取的特征减少了复杂背景对目标检测的影响。如图1中绿色部分所示,由于遥感图像样本的缺乏,训练深层CNN模型需要两个步骤。第1步是在包含大量常见物体样本的数据集上预训练深层CNN模型,使得其权值可以被训练充分。完成预训练的模型可以提取遥感图像的特征,并使用另外的分类器进行分类。第2步是使用遥感图像样本微调预训练模型,完成微调的模型可以同时完成对遥感图像的特征提取和分类。同年,Cao等^[9]借助区域卷积神经网络(regions with CNN features, RCNN)^[10]的方法进行遥感图像飞机检测,使用selective search^[11]提取候选区域,并使用深层CNN模型完成特征提取和分类。2016年,Cheng等^[12]使用CNN模型学习具有旋转不变性的特征并进行候选区域分类,进一步提升了遥感图像目标检测的准确率。同时,结合滑窗^[13]和显著性检测^[14]的深度信念网络也被用于飞机检测。

上述研究探索了深度学习在遥感图像目标检测中的应用,但依然存在一些不足。首先,上述研究大都使用滑动窗口、显著性检测、selective search^[11]等方法进行候选区域选取,深度学习方法仅用于候选区域的特征提取和分类。提取候选区域时需要对图像进行大规模的穷举搜索,并需要对搜索出的区域进行筛选,增加了时间消耗并提高了误检率。其次,由于样本数量的限制,使用传统的监督训练方法无法完成对结构更加复杂的深度模型的训练,这使得更加深层的深度模型难以迁移到遥感图像的目标检测上。本文提出的弱监督深度神经网络(weakly supervised deep neural networks, WS-DNN)遥感图像目标检测模型的创新主要体现在4个方面。第一,使用全卷积网络(fully convolutional networks, FCN)^[15]进行候选区域选取,可以处理任意规模的图像输入。全卷积网络为区域选取提供显著性先验,避免了对图像的穷举搜索,省略了对候选区域的筛选,提升了检测速度。第二,使用深层CNN模型进

行候选区域的特征提取和分类, 模型深度远远大于现有方法使用的 CNN 模型. 更加深层的 CNN 模型提取到的特征更具鲁棒性, 有利于遥感图像的目标检测. 第三, 提出了新的用于遥感图像目标检测的数据集, 并添加了图像级的标注. 提出简化弱监督训练方法, 使用图像级的标签, 而不是像素级的标签就可以完成对 FCN 模型的训练. 另外, 由于样本数量的扩充, 深度模型的训练可以更加充分. 第四, 提出了一种新的检测框微调算法对检测结果进行后处理. 新算法可以在融合重叠区域的同时微调新的检测区域的位置, 使检测结果更加精确.

2 FCN 模型和弱监督训练

CNN 模型的输入是图像, 卷积层和池化层负责特征提取和特征映射, 全连接层负责把特征转化为类别信息, 模型的输出是代表类别置信度的向量. FCN 模型^[15] 通过把 CNN 模型中的全连接层转化为卷积层, 并进行上采样, 可以得到和原图规模相同的逐像素分类结果. FCN 模型具有两种良好的性质. 首先, 可以处理任意规模的输入, 并产生相应大小的输出. 由于在进行处理之前不需要调整图像大小, 使用 FCN 模型可以最大限度保留待处理图像的真实信息. 其次, FCN 模型可以被端到端地训练, 输入的是原图像和像素级的标签图像, 输出的是有类别信息的分割图像, 原图像中的每一个像素都会被分类为一个类别. FCN 模型被广泛应用于语义分割领域, 其良好性质使其可被迁移到目标检测上. 训练常规的目标检测模型需要大量目标级标签, 即使用矩形框在整幅遥感图像中标注出所有目标及其位置. FCN 模型的训练需要使用像素级的标签, 即每一个像素都要被标记为一个类别. 目前遥感图像处理领域十分缺乏可用的目标级标签和像素级标签, 且遥感图像本身的规模大、目标小, 目标和背景杂糅等特性使得人工标注十分困难. 但是用于图像分类的图像级标签 (只标注样本中是否含有目标) 却相对容易得到, 因此弱监督训练的思想被引入到遥感图像目标检测中.

弱监督训练方法的主要思想是基于样本不足的前提, 使用容易得到的图像级标签代替目标级标签和像素级标签进行检测模型的训练. 2015 年, Zhang 等^[16] 使用基于显著性的自适应分割以及负样本挖掘算法得到训练样本进行弱监督训练, 使用经过迭代训练的支持向量机分类候选区域. 同年, Han 等^[17] 使用 Bayes 框架实现弱监督训练, 使用深度 Boltzmann 机提取高层特征改善了检测效果. 2016 年, Zhang 等^[18] 使用弱监督训练方法训练连结卷积神经网络并实现目标检测. 连结卷积神经网络包括区域生成网络和定位网络, 两个 CNN 模型交替训练并自动挖掘和标注新的训练样本. 由于训练样本的不断增多和提炼, 可以使 CNN 模型训练得更加充分, 同时提高识别正确率. 2016 年, Cheng 等^[19] 在遥感图像目标检测综述中指出, 深度学习特征提取和弱监督训练方法将成为未来遥感图像目标检测领域的两大研究热点. 虽然上述研究中使用弱监督训练方法解决了样本缺乏的问题, 但是迭代的训练方法和复杂的样本挖掘、评价机制使得深度模型的训练变得困难. WS-DNN 模型使用简化弱监督训练方法进行训练, 大幅度提升了深度模型的训练效率.

3 WS-DNN 目标检测模型

WS-DNN 遥感图像目标检测模型的框架如图 2 所示. 模型使用简化弱监督训练方法进行训练, 主要包括 4 个部分, 分别是 FCN 模型、滑动窗口、CNN 模型和检测框微调 (bounding-box fine-tuning, BBFT) 算法. FCN 模型负责把待检测图像转化为粗分割图像, 可能含有飞机的区域在粗分割图像中会被高亮标记. 然后使用滑动窗口方法在标记区域提取出候选区域. CNN 模型用于对所有候选区域进行特征提取和分类. 分类的输出是二值的, 即是否是含有整个目标的区域. 对一幅待检测图像中的所

图 2 (网络版彩图) WS-DNN 目标检测模型框架

Figure 2 (Color online) The framework of the WS-DNN object detection model

有分类为目标的区域进行标注, 就得到了检测框和分类结果. 由于分类结果中一个目标可能被多个检测框覆盖, 并且检测框不一定能覆盖整个目标, 因此需要使用 BBFT 算法对检测框进行融合和位置调整. 经过 BBFT 算法处理的检测结果即是最终的检测结果.

3.1 实验数据集

WS-DNN 目标检测框架使用飞机检测数据集进行训练和测试. 为了使 CNN 和 FCN 模型的训练更加充分, 提出了飞机检测数据集 satellite aircrafts dataset. Satellite aircrafts dataset 数据集和公开数据集 aircrafts dataset^[6] 都包含 3 个部分, 分别是训练集、验证集和测试集, 所有图像均来源于 Google Earth. 训练集和验证集用来进行模型的训练和验证, 使用带有图像级标签的 RGB 三通道遥感图像, 所有图像归一化到 64×64 大小. 其中, satellite aircrafts dataset 数据集的训练集包括 2000 个飞机样本和 4000 个背景样本, 验证集包括 117 个飞机样本和 117 个背景样本. Aircrafts dataset^[6] 数据集的训练集包括 2000 个飞机样本 (经过 4 个 90 度旋转增加了样本数量) 和 5000 个背景样本, 验证集包括 453 个飞机样本和 453 个背景样本. 两个数据集中训练集和验证集的部分样本如图 3 所示.

两个数据集的测试集都来自整幅机场的截图, 包含带有目标级标签的 RGB 三通道遥感图像, 用来测试 WS-DNN 模型的性能. 所有的测试数据均来自 Google Earth 截图, 图像大小不一, 其范围一般覆盖整个机场区域. Satellite aircrafts dataset 数据集的测试集包括 5 幅图像, 共包含 117 个飞机目标, 图像大小范围是由 803×737 像素到 1421×754 像素. Aircrafts dataset^[6] 数据集的测试集包括 26 幅图像, 共包含 453 个飞机目标, 图像大小范围是由 565×369 像素到 1484×865 像素. 两个数据集的测试集的部分样本如图 4 所示.

图 3 (网络版彩图) (a) Satellite aircrafts dataset 和 (b) aircrafts dataset^[6] 数据集中训练集、验证集图像示例

Figure 3 (Color online) The image samples of training set and validating set in (a) satellite aircrafts dataset and (b) aircrafts dataset^[6]

图 4 (网络版彩图) (a) Satellite aircrafts dataset 和 (b) aircrafts dataset^[6] 数据集中测试集图像示例

Figure 4 (Color online) The image samples of testing set in (a) satellite aircrafts dataset and (b) aircrafts dataset^[6]

3.2 弱监督训练

WS-DNN 模型需要对 FCN 和 CNN 模型进行训练. 由于 CNN 模型执行的是图像级的分类任务, 因此提出的简化弱监督训练方法主要用于 FCN 模型的训练, 即使用图像级标签替换像素级标签完成 FCN 模型的训练. 在进行训练时, 把飞机样本图像中的所有像素点的标签置为 1, 把背景样本图像中的所有像素点的标签置为 0, 即获得 FCN 模型需要的像素级标签. 也就是在获取 FCN 训练样本时, 既不用把表示飞机的像素区别开来, 也不需要矩形框定位飞机的位置, 只需要按照图像的内容为所有像素打上统一标签即可. 另外, 和使用目标级标签进行训练的常规目标检测模型相比, 所提简化弱监督训练方法不需要标注目标的位置, 直接使用样本图像即可完成训练. 该方法有两个层面的优越性. 一方面, 在遥感图像处理领域, 该训练方法所需训练样本相对容易获得, 可以获得足够的训练样本并避免了消耗巨大且不十分可靠的人工标注. 另一方面, 由于采用了只包含单一目标且目标突出的训练样本, 并提出新的数据集扩充了样本数量, 使得新的弱监督训练方法可以舍弃文献 [16~19] 采用的复杂的样本判定和挖掘过程以及迭代训练的过程, 在简化训练过程的同时提升了深度模型的训练效率. CNN 和 FCN 模型的训练过程如图 5 所示.

所提 WS-DNN 框架中使用的 CNN 模型是 VGG-16 模型^[20]. VGG-16 模型是深层 CNN 模型, 有 13 个卷积层和 3 个全连接层, 和文献 [6~9] 使用的 CNN 模型相比具有更强的特征提取和分类能力. 同时, 和 VGG-19^[20] 或 GoogLeNet^[21] 等模型相比, VGG-16 模型层数不至于太多且结构不过于复杂. 在使用相对少量的遥感图像样本训练时, VGG-16 模型可以较好地完成任务且不容易发生过拟合. VGG-16 模型首先使用 ILSVRC (ImageNet large scale visual recognition competition)^[22]-2012 数

图 5 (网络版彩图) CNN 和 FCN 模型的训练过程

Figure 5 (Color online) The training process of CNN and FCN model

数据集完成预训练. ILSVRC-2012 数据集包含 1000 个类别的 100 万幅图像, 图像的内容均为日常常见物体. 预训练可以使深层 CNN 模型的权值训练得更加充分. 经过预训练后, VGG-16 模型在 satellite aircrafts dataset 和 aircrafts dataset^[6] 数据集组成的合并数据集上进行微调训练和验证. 经过微调训练, 模型的特征提取和分类能力会从日常常见物体转移到遥感目标上. 由于新提出的数据集增加了样本数量, 使得深层 CNN 模型的微调成为可能并可以进行得更加充分, 进而提高 CNN 模型的识别正确率. 同时, 经过微调训练, 该深层 CNN 模型可以同时完成候选区域的特征提取和分类, 不需要借助额外的分类器. 经过微调, VGG-16 模型在合并数据集上的分类正确率为 97.12%.

直接使用遥感图像样本训练 FCN 模型会造成网络模型的过拟合, 因此需要在 VGG-16 模型的基础上进一步进行微调训练, 就可以得到 FCN 模型. 首先, 在 VGG-16 模型的基础上, 去掉最后一个全连接层, 把剩余的全连接层转化为相应的卷积层. 然后, 把由全连接层转化得来的卷积层进行 32 倍上采样, 即可让网络模型产生和输入相同规模的输出. 最后, 保留未发生更改的模型权值, 使用简化弱监督训练方法, 在 satellite aircrafts dataset 和 aircrafts dataset^[6] 数据集组成的合并数据集上重新训练 VGG-16 模型, 就得到 FCN 模型 VGG_FCIN-32. 在 VGG_FCIN-32 模型的基础上, 融合倒数第 2 个池化层 Pool4 的信息, 再次使用简化弱监督训练方法进行微调训练, 就得到模型 VGG_FCIN-16. 同理, 进一步融合倒数第 3 个池化层 Pool3 的信息, 再次使用简化弱监督训练方法进行微调训练, 就得到模型 VGG_FCIN-8. 从 VGG_FCIN-32 到 VGG_FCIN-16 再到 VGG_FCIN-8 模型, 越来越多的底层细节信息被融合进来, 因此 FCN 模型的细节刻画能力越来越强.

3.3 候选区域的提取和分类

WS-DNN 模型使用 FCN 模型和滑动窗口方法来实现候选区域的提取. FCN 模型的输入可以是任意的规模, 这使得 FCN 模型十分适合完成候选区域提取任务. 待检测图像输入训练好的 FCN 模型后, 得到的是一幅粗分割图, 其中可能包含飞机的区域被置为 1, 可能是背景的区域被置为 0. 这种类

图 6 (网络版彩图) 不同 FCN 模型处理结果对比

Figure 6 (Color online) The comparison of processing results of different FCN models. (a) Original image; (b) Pascal-FCN-32 model; (c) Pascal-FCN-16 model; (d) Pascal-FCN-8 model

似于显著性检测的机制为目标区域的选取提供了先验信息, 可以避免在原图像中的穷尽搜索.

在粗分割结果图上使用滑动窗口进行候选区域的选取. 滑窗的大小固定为 60, 步长为 15, 区域选取的阈值为 0.65, 即只有一个窗口中 65% 以上的像素点的值都为 1 时该窗口才能被确定为是候选区域, 这使得区域选取只在图 6 中的红色区域进行. 上述 3 个参数均是根据大量实验选取出的最优参数, 其中滑窗大小的选取是基于数据集中飞机的平均大小. 在粗分割结果上选取候选区域, 使滑动窗口具有先验知识和导向性, 避免了对整幅图像的穷举搜索并省略了候选区域的筛选过程. 同时, 使用 FCN 模型后显著减少了提取的候选区域的数量. 以 satellite aircrafts dataset 数据集为例, 使用滑动窗口方法平均每幅图像得到 5375 个候选区域, 而结合 FCN 模型之后平均候选区域数量为 2014 个. 较少的候选区域数量减少了后期区域分类的工作量, 大幅度提高了检测效率.

为了评估所提 FCN 模型的性能, 对几种使用不同训练方式的 FCN 模型进行了对比. 除了上述 VGG-FCN-32, VGG-FCN-16 和 VGG-FCN-8 模型, 这些模型还包括在日常常见物体分割数据集 PASCAL VOC 2011^[23] 上微调预训练 VGG-16 模型得到的 Pascal-FCN-32, Pascal-FCN-16 和 Pascal-FCN-8 模型. 6 个模型对同一数据的处理结果如图 6 所示.

从图 6 中可以看出, 随着融合的细节信息的增多, FCN 模型刻画细节的能力也逐步增强. 由于在训练时更多地学习了遥感数据的特征, 使用遥感数据进行微调的 VGG-FCN 模型系列的结果要优于使用 PASCAL VOC 2011^[23] 进行微调的 Pascal-FCN 模型系列. 所有模型中细节刻画能力最好的是 VGG-FCN-8 模型. 模型的细节刻画能力越强, 可能包含飞机的位置被刻画得越精确, 后期滑窗产生的候选区域包含飞机的可能性越大, 分类的工作量也越小.

在 WS-DNN 模型中, CNN 模型的作用是实现候选区域的特征提取和分类, 即判定候选区域中是否含有飞机 (含飞机为 1, 背景为 0). 另外, 在分类候选区域时对 CNN 模型设定了独立阈值, 只有含有飞机的置信度 (CNN 输出为 1 的概率) 大于 0.6 时该候选区域才能被保留. 独立阈值的设定可以对重叠检测结果和没有完全覆盖目标的检测结果起到过滤作用.

表1 区域提取算法对比

Table 1 The comparison of different proposal extracting algorithms

Algorithm	FPR-SAD (%)	MR-SAD (%)	Runtime-SAD (s)	FPR-AD (%)	MR-AD (%)	Runtime-AD (s)
Sliding window	91.21	0	0.59	95.29	0	1.01
Selective search ^[11]	90.27	0.03	27.96	87.85	0	24.06
FCN	78.16	0	7.73	84.08	0	11.15

3.4 检测框微调算法

候选区域提取和分类完成之后,会出现多个检测框覆盖了同一个目标的情况,并且多数检测框的位置并不是刚好完全覆盖住了飞机.传统的做法是对所有的检测框进行非极大值抑制处理.但是使用非极大值抑制时只能消除部分重复覆盖的检测框,无法修正检测框的位置,因此提出了BBFT算法来解决这一问题.

BBFT算法的步骤如下:(1)把一副图像中的所有检测框加入队列 $L1$;(2)复制队列 $L1$ 中的第1个检测框 A 的位置坐标 p ;(3)将 p 和 $L1$ 队列中的其他检测框计算IoU(intersection over union);(4)如果 p 和检测框 B 的IoU大于0.5,计算重叠区域的中心点 q ,以 q 为中心,检测框边长为边长构造新的正方形检测框并加入队列 $L1$ 末尾,从队列 $L1$ 中删除 p 对应的检测框和检测框 B ,跳转到步骤(2);(5)如果没有检测框和 p 的IoU大于0.5,则跳过 p 对应的检测框,复制队列 $L1$ 中下一个检测框的位置为新的 p ,跳转至步骤(3);(6)队列 $L1$ 中所有检测框都不和任意检测框有大于0.5的IoU时,算法终止,此时队列 $L1$ 中剩余的检测框就是最终检测结果.

BBFT算法在去除重复覆盖检测框的同时微调检测框的位置,使目标检测更加精确.精确的位置将提升检测结果和真值检测框的IoU,进而提升WS-DNN模型的检测精度.

4 实验结果分析

实验使用的是配备64位Ubuntu系统的工作站,硬件配置为Intel(R) Xeon(R) CPU E5-2690 v3 2.60 GHz处理器、256 GB内存和4 TB硬盘.CNN和FCN模型的训练使用Caffe^[24]深度学习平台,使用一块NVIDIA Tesla K40c 12 GB显存GPU进行加速.WS-DNN模型使用提出的satellite aircrafts dataset和aircrafts dataset^[6]数据集中的测试集进行测试.

4.1 FCN 区域提取算法评价

所提FCN候选区域提取算法将和经典的滑动窗口方法以及selective search^[11]算法进行比较.其中,滑动窗口方法使用和FCN候选区域提取算法相同的窗口大小和步长,selective search^[11]方法使用Matlab版本的selective search实现代码.为了方便进行对3种算法的比较,定义了两个评价参数,分别是FPR(false positive rate)^[18]和MR(missing ratio)^[18].其中,FPR计算的是不包含飞机的候选区域占候选区域总数的比例,MR计算的是未被候选区域包含的飞机数量占飞机总数的比例.实验结果如表1所示.表1中AD代表aircrafts dataset^[6]数据集,SAD代表satellite aircrafts dataset数据集.

从表1可见提出的FCN候选区域提取算法在两个数据集上都取得了最好的结果.滑动窗口方法在待检测图像上进行穷举搜索,虽然覆盖的范围广并且不需要进行判定和筛选使得用时较少,但是由于缺

图 7 (网络版彩图) BBFT 和 NMS 算法对比

Figure 7 (Color online) The comparison of BBFT and NMS algorithm. (a) Image region 1; (b) NMS and (c) BBFT results of image region 1; (d) image region 2; (e) NMS and (f) BBFT results of image region 2

乏先验知识的引导, 大量不包含飞机的区域被选取出来. selective search^[11] 算法虽然避免了对图像的穷举搜索, 但是其思想是找出所有目标而不针对飞机, 提取的候选区域中会包含建筑物、跑道、停车场等其他目标. 因此遥感图像中目标较多、较小且目标和背景杂糅在一起等特点会对 selective search^[11] 算法的效率产生较大影响. 另外, 由于没有尺度的要求, 在实际应用中 selective search^[11] 算法常会选取出包含多个飞机的候选区域, 提取结果不够精确. 提出的 FCN 模型经过了充分的训练, 学习到区分飞机和背景的高层特征, 进而产生能刻画大致包含飞机的区域的粗分割图. 基于粗分割图进行搜索时, 可以在先验知识的指导下集中搜索可能包含飞机的区域, 在降低搜索的复杂度的同时提升搜索的效率和精度. 此外滑窗大小是数据集中飞机的平均大小, 有了窗口大小的限制, 可以保证多个飞机不会出现在同一个检测框中.

4.2 BBFT 算法评价

为了评估所提 BBFT 算法的性能, 添加了 BBFT 算法和经典的非极大值抑制 (non-maximum suppression, NMS) 算法的对比. 对比结果如图 7 所示.

从图 7 可以直观地看出, 提取的候选区域经过 CNN 模型分类得到的结果会有较多重叠覆盖的情况. 这主要是由于在提取候选区域时选择了较小的步长, 使得在同一个目标上提取了多个候选区域. 但是如果使用较大的步长, 在提取候选区域时又会漏选目标. 因此一个合理的方式是使用小的步长提取候选区域, 在后期使用融合或抑制算法对 CNN 分类结果进行后处理. NMS 算法可以较好地抑制重叠结果, 但它只考虑筛选而不进行融合, 导致最终检测结果不精确. BBFT 算法在对区域筛选的同时实现融合, 其特有的融合机制使检测框的位置得到调整, 新的位置和原位置相比更靠近目标的中心. 另外, 新生成的融合检测框继续加入队列并和其他检测框计算 IoU, 由于新的更靠近目标中心的位置的引入, 覆盖同一目标但相距较远的检测框可以被进一步融合, 而这些相距较远的检测框使用 NMS 算法处理时可能被同时保留. 因此 BBFT 算法在精修位置的同时提升了算法去除重叠结果的能力. 从图 7 可以看出, 和 NMS 算法相比, BBFT 算法去除了更多的重叠结果并使得检测框的位置更靠近目

表 2 WS-DNN 模型在 satellite aircrafts dataset 数据集上的实验结果
Table 2 The experimental results of the WS-DNN model on satellite aircrafts dataset

Evaluating method	Results (%)
FAR	7.69
MR	3.42
PR	93.22

表 3 不同检测算法在 aircrafts dataset^[6] 数据集上的对比
Table 3 The comparison of different object detection methods on aircrafts dataset^[6]

Given MR (%)	FAR of BING-CNN ^[6] (%)	PR of BING-CNN ^[6] (%)	FAR of WS-DNN (%)	PR of WS-DNN (%)
25	7.28	90.50	4.65	94.12
20	9.27	90.00	8.74	90.53
15	17.66	84.00	10.26	89.33

标中心.

4.3 WS-DNN 模型评估

WS-DNN 模型在 satellite aircrafts dataset 和 aircrafts dataset^[6] 数据集上进行了测试. 在 satellite aircrafts dataset 数据集上, 本小节将给出 WS-DNN 模型的各项实验结果. 在 aircrafts dataset^[6] 数据集上, WS-DNN 模型将和此数据集上目前获得最优结果的算法 BING-CNN^[6] 进行对比. BING-CNN^[6] 模型使用 BING 方法^[25], 利用二值化梯度特征在图像中搜索候选区域并使用浅层 CNN 模型分类. 所有实验中, 如果检测结果和图像目标级标注的 IoU 大于 0.5, 则视为该目标被成功检出. 使用 3 个指标评价检测结果, 分别是误检率 (false alarm rate, FAR)^[6], 漏检率 (missing ratio, MR)^[18] 和精度 (precision rate, PR)^[6]. 其中, 误检率计算的是误检为飞机的结果的数量占飞机总数的比例, 漏检率计算的是没有被检测出的飞机占飞机总数的比例, 精度计算的是含有飞机的检测框占检测框总数的比例. WS-DNN 模型在 satellite aircrafts dataset 数据集上的实验结果如表 2 所示.

实验结果表明, WS-DNN 模型在 satellite aircrafts dataset 数据集上达到了较好的检测效果, 对于大幅面背景复杂的遥感图像得到了较低的误检率、漏检率和较高的精度. 为了和 BING-CNN^[6] 方法在 aircrafts dataset^[6] 数据集上进行对比, 表 3 中给出了两种方法在指定漏检率下的结果.

从表 3 可以看出, 指定漏检率时, WS-DNN 模型从误检率和精度上均优于 BING-CNN^[6] 方法, 这主要是由 4 个原因造成的. 首先, 本文自行采集的数据集加大了训练样本的数量, 结合简化弱监督训练方法使得深度模型的训练成为可能. 其次, 提出的 FCN 候选区域提取模型使用遥感图像训练, 和 BING 方法^[25] 相比可学习到遥感图像具有区分性的高层特征. 该模型为滑动窗口提供带有显著性的粗分割图, 使候选区域的提取更加高效和精确, 并显著降低了飞机检测框架的漏检率. 再次, 所提模型采用的深层 CNN 分类模型和 BING-CNN^[6] 方法中的浅层 CNN 模型相比更能学习到飞机目标和其他目标间的本质区别, 进而帮助 WS-DNN 模型减少错分样本, 大幅度降低误检率. 最后, BBFT 算法可以有效去除重叠结果并修正检测框的位置, 进一步提升了飞机检测的精度. 另外, WS-DNN 模型具有较强的鲁棒性, 尽管两个数据集中的目标在大小、尺度、纹理、色彩等方面均有差别, WS-DNN 在两个数据集上均得到了较好的检测结果. WS-DNN 模型的部分检测结果如图 8 所示.

图 8 (网络版彩图) WS-DNN 模型的部分检测结果

Figure 8 (Color online) Some detected results of the WS-DNN model. Detections of image (a) 1, (b) 2, (c) 3

5 结论

本文提出了一种使用简化弱监督训练方法训练的遥感图像目标检测模型. 在减少对训练样本的标注要求的同时, WS-DNN 模型使用深层 FCN 模型提取候选区域, 使用深层 CNN 模型进行分类. 由于深度模型的引入, 可以提取到遥感目标的本质特征, 使 WS-DNN 模型在目标检测任务中取得了较好的检测结果. 提出的 BBFT 算法对检测结果进行精修, 进一步提升了检测精度.

未来的工作主要有两个方面, 首先是尝试解决 WS-DNN 模型无法处理尺度变化较大的遥感目标的检测的问题. 其次, WS-DNN 模型也适用于其他目标的检测, 未来将建立基于 WS-DNN 模型的多遥感目标检测模型.

参考文献

- 1 Xu C F, Duan H B. Artificial bee colony (ABC) optimized edge potential function (EPF) approach to target recognition for low-altitude aircraft. *Pattern Recogn Lett*, 2010, 31: 1759–1772
- 2 Liu G, Sun X, Fu K, et al. Aircraft recognition in high-resolution satellite images using coarse-to-fine shape prior. *IEEE Geosci Remote Sens Lett*, 2013, 10: 573–577
- 3 Sun H, Sun X, Wang H Q, et al. Automatic target detection in high-resolution remote sensing images using spatial sparse coding bag-of-words model. *IEEE Geosci Remote Sens Lett*, 2012, 9: 109–113
- 4 He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 2016. 770–778
- 5 Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intel*, 2017, 39: 1137–1149
- 6 Wu H, Zhang H, Zhang J F, et al. Fast aircraft detection in satellite images based on convolutional neural networks. In: *Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP)*, Quebec City, 2015. 4210–4214
- 7 Wu H, Zhang H, Zhang J F, et al. Typical target detection in satellite images based on convolutional neural networks. In: *Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Hong Kong, 2015. 2956–2961
- 8 Zhang P, Niu X, Dou Y, et al. Airport detection from remote sensing images using transferable convolutional neural networks. In: *Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN)*, Vancouver, 2016. 2590–2595
- 9 Cao Y S, Niu X, Dou Y. Region-based convolutional neural networks for object detection in very high resolution remote sensing images. In: *Proceedings of the 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, Changsha, 2016. 548–554
- 10 Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 2014. 580–587
- 11 Uijlings J R, Sande K E, Gevers T, et al. Selective search for object recognition. *Int J Comput Vision*, 2013, 104: 154–171
- 12 Cheng G, Zhou P C, Han J W. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. *IEEE Trans Geosci Remote Sens*, 2016, 54: 7405–7415
- 13 Chen X Y, Xiang S M, Liu C L, et al. Aircraft detection by deep belief nets. In: *Proceedings of the 2nd IAPR Asian Conference on Pattern Recognition (ACPR)*, Okinawa, 2013. 54–58
- 14 Diao W H, Sun X, Zheng X W, et al. Efficient saliency-based object detection in remote sensing images using deep belief networks. *IEEE Geosci Remote Sens Lett*, 2016, 13: 137–141
- 15 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 2015. 3431–3440
- 16 Zhang D W, Han J W, Cheng G, et al. Weakly supervised learning for target detection in remote sensing images. *IEEE Geosci Remote Sens Lett*, 2015, 12: 701–705
- 17 Han J W, Zhang D W, Cheng G, et al. Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning. *IEEE Trans Geosci Remote Sens*, 2015, 53: 3325–3337
- 18 Zhang F, Du B, Zhang L P, et al. Weakly supervised learning based on coupled convolutional neural networks for aircraft detection. *IEEE Trans Geosci Remote Sens*, 2016, 54: 5553–5563
- 19 Cheng G, Han J W. A survey on object detection in optical remote sensing images. *ISPRS J Photogramm Remote Sens*, 2016, 117: 11–28
- 20 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014. ArXiv:1409.1556
- 21 Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 2015
- 22 Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vision*, 2015, 115: 211–252
- 23 Everingham M, Gool L, Williams C K, et al. The pascal visual object classes (VOC) challenge. *Int J Comput Vision*, 2010, 88: 303–338

- 24 Jia Y Q, Shelhamer E, Donahue J, et al. Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, 2014. 675–678
- 25 Cheng M M, Zhang Z M, Lin W Y, et al. BING: binarized normed gradients for objectness estimation at 300 fps. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, 2014. 3286–3293

Object detection models of remote sensing images using deep neural networks with weakly supervised training method

Mingfei ZHOU¹ & Xili WANG^{1,2*}

1. School of Computer Science, Shaanxi Normal University, Xi'an 710119, China;

2. Key Laboratory of Modern Teaching Technology, Ministry of Education, Shaanxi Normal University, Xi'an 710119, China

* Corresponding author. E-mail: wangxili@snnu.edu.cn

Abstract An object detection framework of remote sensing images is proposed in this paper. It aims to bridge the difficulties in the object detection task of remote sensing images caused by small targets and complex backgrounds. The proposed framework consists of two deep neural network models: fully convolutional network model and convolutional neural network model. First, the fully convolutional network model is used to extract proposals that may contain targets to be detected in the remote sensing images, thus avoiding an exhaustive search in the whole image. Next, the convolutional neural network model is used to classify the proposals. High-level features are extracted to improve the rate of classification accuracy. Subsequently, a new dataset of remote sensing images used for object detection is provided. Image-level annotations are used to train all the models in the proposed framework. Simplified weakly supervised training method is used to solve the problem unachievable by object-level annotations in the field of object detection on remote sensing images. Finally, a novel proposal fusing algorithm is proposed, by which the positions of proposals are adjusted while overlapped proposals are fused. The proposed framework is tested on the proposed satellite aircraft dataset and the public aircraft dataset. The experimental results demonstrate that the proposed object detection framework improves the recognition rate as well as the detection efficiency of object detection when compared with other object detection frameworks using deep neural networks.

Keywords remote sensing image, object detection, convolutional neural networks, fully convolutional networks, weakly supervised training

Mingfei ZHOU was born in 1993. He is currently working toward his M.S. degree in the Machine Learning Laboratory, School of Computer Science, Shaanxi Normal University. His research interests include remote sensing image processing, deep learning, and deep neural networks.

Xili WANG received her M.S. degree in computer science and her Ph.D. degree from the National Lab of Radar Signal Processing in Xidian University, Xi'an, China, in 1994 and 2004, respectively. She is currently a professor at the School of Computer Science, Shaanxi Normal University. Her primary research interests include intelligent information processing, pattern recognition, remote sensing image processing, and image perception and comprehension.