基于特征融合与软判决的遥感图像飞机检测

朱明明^{1*},许悦雷²,马时平¹,李帅¹,马红强¹ ¹空军工程大学研究生院,陕西西安 710038; ²西北工业大学无人系统技术研究院,陕西西安 710072

摘要提出了一种特征融合结合软判决的飞机检测方法。以区域卷积神经网络为基本框架,依次采用 L2 范数归一 化、特征连接、尺度缩放和特征降维来融合多层特征。为了降低网络在目标高度重叠时的漏检率,引入软判决来改进 传统的非极大值抑制方法。实验结果表明,所提方法能够准确快速地检测到飞机,得到检测率为 94.25%、虚警率为 5.5%、平均运行时间为 0.16 s 的实验结果。与现有的其他检测方法相比,所提方法的各项指标均得到显著提升。 关键词 图像处理;飞机检测;特征融合;软判决;区域卷积神经网络 中图分类号 TP183; TP751.1 文献标识码 A doi: 10.3788/AOS201939.0210001

Airplane Detection Based on Feature Fusion and Soft Decision in Remote Sensing Images

Zhu Mingming^{1*}, Xu Yuelei², Ma Shiping¹, Li Shuai¹, Ma Hongqiang¹

¹ Graduate School, Air Force Engineering University, Xi'an, Shaanxi 710038, China; ² Unmanned System Research Institute, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China

Abstract An airplane detection method is proposed based on feature fusion and soft decision, in which the regionbased convolutional neural network is used as the basic framework and the L2 normalization, feature connection, scaling, and dimensionality reduction are in turn used to fuse the multi-layer features. The soft decision, which can improve the traditional non-maximum suppression method, is introduced in order to reduce the detection-omissionrate of grids in the case of significant overlap of targets. The experimental results show that the proposed method can be used to detect airplanes accurately and quickly with a detection rate of 94.25%, a false alarm rate of 5.5%, and the average running time of 0.16 s. Compared with those of the other existing detection methods, each index of the proposed method is significantly improved.

Key words image processing; airplane detection; feature fusion; soft decision; region-based convolutional neural network

OCIS codes 100.2000; 100.4996

1 引 言

随着高分辨率卫星遥感技术的快速发展,遥感 图像的分辨率变得越来越高,获取方式也更加方便、 多样,这推动了遥感图像中目标检测的发展^[1]。飞 机作为典型的军事和民用目标,在战场监视、航空管 制和交通运输等领域发挥着重要作用。然而,由于 飞机类型、姿态、尺寸和复杂背景的不确定性,在遥 感图像中检测飞机目标一直是一个具有挑战性的研 究课题。 为了解决这个难题,前人不断提出了各种方法。 Li等^[2]提出了将视觉显著性和对称性相结合的方法,利用形状匹配进行飞机分类。Bo等^[3]提出了一种基于区域分割的飞机检测方法。Liu等^[4]基于稀疏编码提出了一种飞机特征表达的方法。上述方法均采用人工设计特征结合分类器训练的传统固定模式,存在着滑动窗口冗余、特征表征不足等局限性,导致了飞机检测的准确率不高。近年来,深度学习被广泛用于自动学习目标特征,这是因为卷积神经网络(CNN)在图像目标分类和检测领域中的性能

收稿日期: 2018-07-04; 修回日期: 2018-08-02; 录用日期: 2018-09-11

基金项目: 航空科学基金(20175896022)

^{*} E-mail: ming_paper@163.com

遥遥领先于传统方法。Chen 等^[5]将深度置信网络 (DBN)用于飞机检测,但也只是将 DBN 作为目标 分类器,其检测性能不高且时间成本过高。Wu 等^[6]提出了一种二值化梯度(BING)结合 CNN 的 检测方法,但 BING 产生 2000~3000 个候选区域, 导致计算成本很高,严重影响了算法的速度。辛鹏 等^[7]融合了 CNN 中的多层特征来进行飞机检测, 但存在易漏检相邻目标的缺陷。Li 等^[8]提出了一 种基于强化学习和 CNN 的飞机检测架构,但是与 前沿的算法相比,这种方法的检测时间过长。

由于飞机的姿态具有多方向性,学习旋转不变 性对于分类器是至关重要的。针对飞机的这种特 点,Zhang 等^[9]使用扩展直方图的梯度来获得具有 旋转不变性的特征;Wang 等^[10]使用一个旋转不变 矩阵来获得旋转不变性;Liu 等^[4]提出了一种基于 稀疏编码的特征提取方法,虽然能在一定程度上提 取旋转不变性,但对其他目标的适用性不强。本文 利用数据增强的旋转、翻转技术来实现飞机姿态的 多方向性,正如 CNN 学习图像的其他特征一样,直接学习这个特性。

由于大数据的出现和软硬件平台性能的不断提升,在自然图像中,基于深度学习的目标检测技术的性能也不断提升,并持续处于领先水平。然而,遥感图像的数据采集相对困难且人工标注的数据集较少,因此遥感图像中的目标检测技术水平一直滞后于自然图像。而迁移学习技术^[11]的出现,为将深度学习中的目标检测模型应用到遥感图像上提供了可能。因此,本文以更快的区域卷积神经网络^[12](Faster R-CNN)为基本架构,针对飞机目标的特点对网络模型进行改进和优化,提出了一种特征融合结合软判决的飞机检测方法。

2 更快的区域卷积神经网络

Faster R-CNN 结构主要由两部分组成,分别为 区域建议网络^[12](RPN)和检测网络,这两个网络共 享卷积层,如图1所示。





2.1 区域建议网络

RPN 利用 CNN 来生成候选框,并与后续的检测网络共享卷积层,这大大缩短了候选区域的生成时间。共享卷积层输出的最后一层卷积特征图作为 RPN 的输入,RPN 的具体结构如图 2 所示。



RPN 使用一个 3×3 的滑动窗口在输入的特征 图上进行卷积运算,将每个位置的计算结果映射到 一个低维向量。候选框的参数化表示称为 anchor, 即每个 anchor 以滑动窗口的中心为中心对应一组 尺寸和长宽比。对于卷积特征图上的每个位置,考 虑 3 种尺寸和 3 种长宽比的 anchors,以适应多尺度 的目标。将低维向量同时输入到分类层和回归层, 以进行分类和位置回归。

2.2 检测网络

检测网络主要由感兴趣区域^[13](ROI)池化层、 全连接层以及分类与回归层组成。对于来自同一幅 图像的候选区域,如果进行全部特征的提取,将导致 网络的计算成本很高。由于候选区域对应的特征图 与完整特征图存在固定的映射关系,可以先对整幅 图像进行特征提取,然后通过 ROI 池化层从完整特 征图上直接得到不同候选区域对应的特征矢量,实 现候选区域的特征共享。由于全连接层需要固定尺 寸的输入,因此 ROI 池化层采用最大池化,将特征 矢量固定为同一尺寸(512×7×7)。

3 多层特征融合与软判决

3.1 多层特征融合

在选取的实验图像中,飞机目标的尺寸很小,大 致像素范围为 32 pixel × 32 pixel ~ 64 pixel × 64 pixel。经过一系列的卷积层处理后,32 pixel × 32 pixel大小的飞机在最后一层卷积特征图上大概 只有 2 pixel×2 pixel。这样的特征图是很粗糙的, 同时相邻区域可能会相互重叠,不利于飞机的检测。 高层特征包含丰富的语义信息,更利于目标分类;低 层特征空间分辨率较大,更利于目标定位。目标检测可以理解为分类加定位,因此一个好的目标检测 系统应该同时利用低层和高层特征,尤其针对小目标物体时。

Kong 等^[14]从不同的卷积层提取了同一个位置的特征矢量,其中较低层特征的激活响应明显高于较高层,表明同一特征在不同层通常具有不同的响应尺度。因此融合多层特征,不能简单直接地将多层特征进行相加或合并。融合多层特征首先是利用 ROI 池 化层从不同卷积层("conv3"、"conv4"和"conv5")提取 候选区域对应的固定特征向量,然后依次进行 L2 范 数归一化以及各层特征连接、尺度变换和特征降维, 最后将融合后的特征输入到全连接层,如图 3 所示。



图 3 特征融合的结构 Fig. 3 Structure of feature fusion

首先使用 L2 范数对每层卷积层输出的第 d 个 通道的特征 $\mathbf{x} = (x_1 \cdots x_d)$ 进行归一化:

$$x'_{i} = \frac{x_{i}}{\parallel \mathbf{x} \parallel_{2}}, \qquad (1)$$

$$\| \mathbf{x} \|_{2} = \left(\sum_{i=1}^{d} |x_{i}|^{2} \right)^{1/2},$$
 (2)

式中*i* 表示特征的第*i* 个通道。(1)式为归一化特征,(2)式为 L2 范数的表达式。

将不同层的归一化特征 $\mathbf{x}' = (x'_1 \cdots x'_{d_1})$ 和 $\mathbf{y}' = (y'_1 \cdots y'_{d_2})$ 按照

$$z = (x', y') = (x'_1 \cdots x'_{d_1}, y'_1 \cdots y'_{d_2})$$
 (3)
进行融合连接,由于各层特征融合后的特征尺度很
小,不利于网络的训练^[15],因此有必要对特征进行
缩放处理:

$$z_i' = \gamma z_i, \qquad (4)$$

式中 γ 为缩放参数,具体的数值为最后一层卷积特 征的 L2 范数值。

利用预训练模型 VGG16 网络在遥感数据上进

行迁移学习,融合后的特征大小需与原网络保持一致。使用1×1卷积核进行降维,使输入到全连接层的特征大小为512×7×7。1×1卷积核已在GoogLeNet、ResNet等网络中得到广泛应用,它不仅能在保持特征尺度不变的前提下实现降维,而且能利用后续非线性激活函数来增加网络的非线性特性,最终实现跨通道的特征信息融合。

3.2 软判决的非极大值抑制

由于没有强制要求检测网络只产生唯一检测 框,因此会出现多个检测框包围同一目标的情况。 为了避免产生误检,利用非极大值抑制^[16](NMS)来 剔除多余的检测框,从而留下最优的检测框,如图 4 所示。

NMS 实现流程为:1)将集合 A 中的检测框依据分类得分从高到低排序。2)将得分最高的检测框 M 移到集合 B,同时计算检测框 M 与其他检测框的 IoU(Intersection-over-Union)值。若 IoU 值大于阈值 T,则剔除得分较低的检测框,否则认为存



图 4 非极大值抑制 Fig. 4 Non-maximum suppression

在多个目标。3)从集合 A 剩余的检测框中再选取 得分最高的,重复步骤 2)直到集合 A 中无检测框, 最后得到集合 B 中的检测框即为输出。如果两个 目标的检测框重叠面积过大,可能会导致其中一个 目标的检测框被剔除,从而造成目标漏检,如图 5 所 示。虽然这在一定程度上能够缓解阈值 T 的增大, 但大大增加了误检率。传统的 NMS 将邻近检测框 的得分强制归零,这是一种硬判决做法。信道译码 算法可以分为硬判决译码算法和软判决译码算法, 其中软判决的纠错能力远优于硬判决。因此,借鉴 软判决的思想对传统 NMS 算法进行改进,提出了 一种软判决的 NMS 算法。



图 5 漏检结果 Fig. 5 Results of omission detection

传统的 NMS 算法对邻近检测框的判决函数可 表示为

$$s_{i} = \begin{cases} s_{i}, & I_{oU}(M, a_{i}) < T \\ 0, & I_{oU}(M, a_{i}) \geqslant T \end{cases}$$
(5)

(5)式设置一个硬性阈值 T 来判断邻近检测框 a_i 的得分 s_i 是保持原值或归零, $I_{ou}(M, a_i)$ 表示邻 近检测框 a_i 与检测框M 的 IoU 值。

引入软判决对判决函数式进行改进,改进后的 判决函数可表示为

 $s_{i} = \begin{cases} s_{i}, & I_{oU}(M, a_{i}) < T\\ s_{i} [1 - I_{oU}(M, a_{i})], & I_{oU}(M, a_{i}) \ge T \end{cases}$ (6)

当邻近检测框 a_i 与检测框 M 的 IoU 小于阈值 T,检测框得分保持不变;当邻近检测框 a_i 与检测 框 M 的 $I_{ou} \ge T$ 时,检测框得分呈线性衰减。(6)式 表明 $I_{oU}(M,a_i)$ 越大(即重叠面积越大),检测框得 分衰减越严重; $I_{oU}(M,a_i)$ 很小(即重叠面积很小), 检测框得分不受影响。

(6)式减小了 I_{ou}(M,a_i)大于阈值的检测框得 分而非完全归零剔除,尽管该检测框的分数降低,但 有效避免了漏检,同时设置检测框得分阈值为0.45, 以剔除得分较小的检测框,避免造成多余的误检。 这里仅通过修改判决函数来实现对传统 NMS 的改 进,所提算法实现简单且不会增加额外的计算成本。

4 实验与结果分析

所提算法的实验环境为:处理器为 i7-7700,主 频为 3.6 GHz,内存为 16 G,显卡为 NVIDIA GTX1060,操作系统为 Ubuntul4.04,网络框架为 Caffe。实验中采用检测率(DR)、虚警率(FAR)和 平均运行时间作为评价指标。

4.1 实验数据

所有的实验数据均来自 Google Earth 中的机 场遥感影像,比如北京首都国际机场、美国波士顿国 际机场等共 300 个机场,空间分辨率为 1.19 m。选 取 150 个机场中的飞机图像作为训练和验证集,图 像大小范围为 375 pixel × 375 pixel ~ 400 pixel × 400 pixel,共 1500 张。将剩余 150 个机场中的飞机 图像作为测试集,图像大小范围为 900 pixel × 800 pixel~1000 pixel×900 pixel,共 150 张。为了 避免出现过拟合和学习飞机的旋转不变性特征,在 此采用数据增强来扩充训练验证集图片的数量。将 1500 张图片随机旋转 90°、180°和 270°,同时以 0.5 的概率进行水平或垂直翻转,最终使得训练验证集 的图片数量为 5000 张,其中训练集和验证集的图片 各 2500 张。所有的图片标签均为手工标注,图 6 为 部分实验数据。



图 6 数据示例。(a)训练数据;(b)测试数据

Fig. 6 Examples of data. (a) Training data; (b) test data

4.2 网络训练

虽然遥感图像相对于自然图像具有尺度多样、

多方向、目标小、背景复杂等特点,但其含有的边缘、 颜色等低级语义特征和抽象的高级语义特征等目标 特征是类似的。卷积神经网络同样可以从遥感图像 中学习和提取低级或高级语义特征。对于尺度多样 性、多方向性、背景复杂等特点,卷积神经网络可以 通过模型训练从实验数据中学习得到,就如同学习 边缘颜色等语义特征一样。深度学习模型的目标检 测需要大规模的数据训练,而目前的遥感数据样本 有限,且数据的标注也需要大量的人工成本,这就是 深度学习在遥感图像领域应用较少的主要原因,因 此从 Google Earth 搜集遥感图像并进行人工标注 数据。使用飞机遥感数据重新训练一个新的 CNN 是低效的,甚至可能会导致过拟合。低层卷积层学 到的都是边缘纹理以及颜色等低级语义特征,飞机 遥感图像和普通自然图像上的这些特征是一致的, 如图7所示。因此,可以在大型自然图像数据库上 预先训练 CNN,再利用迁移学习将网络学到的特征 迁移到新的飞机检测任务中,事实证明这个过程十 分高效。共享卷积层的权重用预先训练的网络 VGG16进行初始化,其余层的权重用均值为 0、标 准差为 0.01 的高斯分布随机进行初始化。网络的 基本学习率为 0.001,动量为 0.9,权重衰减为 0.0005。为了实现 RPN 和检测网络共享卷积层,采 用交替优化策略^[12]来训练整个网络。



Fig. 7 Transfer learning

4.3 结果分析

为了分析融合不同层的特征对实验结果的影响,在其他条件不变的情况下,采用控制变量法进行 对比实验,结果如表1所示。

由表1可知,融合多层特征的DR均优于使用 单层特征(表1第1行),验证了融合多层特征的有 效性。正如第3.1节所述,经过多层的卷积和池化

表1 特征融合的结果

Tab	le 1	Results	of	feature	fusion
-----	------	---------	----	---------	--------

Layer 2	Layer 3	Layer 4	Layer 5	DR /%	
			\checkmark	86.28	
		\checkmark	\checkmark	90.30	
	\checkmark	\checkmark	\checkmark	91.90	
\checkmark	\checkmark	\checkmark	\checkmark	91.90	

后,在最后一层即第5层卷积特征图上飞机的特征 尺度很小,这样不利于飞机的定位。粗糙、高语义的 高层特征有利于目标的分类,而精细、高分辨率的低 层特征有利于目标的定位,因此融合多层特征能够 结合不同层的优势,从而提高针对飞机这样的小目 标的网络检测能力。逐渐增加融合的层数能不断提 高 DR, 但融合层数增加到第 2 层时, DR 并没有继 续提高,这表明融合 3,4,5 层特征的 DR 最优,继续 融合其他层的特征无法提供额外的有用信息,即网 络性能的提升已经饱和。以下实验中若无特殊说 明,默认融合3,4,5 层特征。为了更直观地说明特 征融合的重要性,给出不同方法的检测结果对比图, 如图 8 所示。图 8(a)是使用第 5 层特征的检测结 果,图 8(b)是融合 3,4,5 层特征的检测结果。可以 看出,融合多层特征可以使网络能够检测到更多的 较小目标,从而提高了检测率。



图 8 不同方法的网络检测结果。(a)第 5 层的结果; (b)融合 3,4,5 层的结果

Fig. 8 Network detection results by different methods.(a) Results of 5th layer; (b) results after fusion of 3, 4, 5 layers

为了说明改进 NMS 算法的有效性,在飞机测 试集上进行对比实验。采用软判决 NMS 算法得到 的 DR 为 91.90%,所提算法得到的 DR 为 94.25%。 采用软判决 NMS 算法测试集时 DR 提高了约 2%。 这是因为传统的 NMS 算法将与检测框 M 重叠面 积大于阈值的所有检测框全部剔除,可能会导致一 些相邻目标的漏检。而改进的 NMS 算法是使用一 个关于重叠面积的函数来减小相邻检测框的分类得 分,并不同于传统 NMS 算法将邻近检测框的分类 得分归零,这样能有效地降低漏检率。图9呈现的 是 NMS 改进前后对不同飞机的检测结果,可以看 出 NMS 算法错误地将相邻检测框抑制掉,而所提 方法能够准确地检测到目标。



图 9 NMS算法改进前后的飞机检测结果。(a)传统 NMS算法;(b)改进 NMS算法

Fig. 9 Airplane detection results before and after improvement of NMS algorithm. (a) Traditional NMS; (b) improved NMS

4.4 方法对比

为了说明所提方法的优越性,选取已有的飞机 检测方法进行对比实验。在本文实验环境下使用网 上公开的代码进行仿真,记录每种方法在测试集上 的检测率、虚警率和平均运行时间,并进行比较,结 果如表2所示。文献[5]是基于深度置信网络的方 法,文献[6]是基于卷积神经网络的方法,Faster R-CNN 是既不融合多层特征也不改进 NMS 算法的 原始方法。

表 2 不同飞机检测方法的结果对比

 Table 2
 Comparison among airplane detection results

 by different methods

Mothod	DR / %	FAD /0/	Average running
Methou		$\Gamma A K / / 0$	time /s
Ref. [5]	79.54	22.52	171.25
Ref. [6]	84.25	17.66	6.41
Faster R-CNN	86.28	8.15	0.15
Proposed	94.25	5.50	0.16

由表2结果可知,所提方法的检测率、虚警率和 平均运行时间均明显优于以往的飞机检测方法。文 献[5]的方法依旧采用手工设计特征加分类器的方 式,只是用深度置信网络代替了以往的简单分类器 时,依旧存在手工特征表征能力不足和滑动窗口冗 余的问题。文献[6]的方法先利用 BING 技术产生 候选区域,再使用卷积神经网络对飞机进行检测。 然而,与 RPN 相比,BING 产生的候选区域质量差 且时间成本较高。Faster R-CNN 将候选区域生成、 特征提取、目标分类和边界回归整合到一个深度卷 积神经网络框架中,充分利用 CNN 强大的特征表 达能力实现端到端的飞机检测,因此 Faster R-CNN 和所提方法的准确性和速度均优于前两种方法。正 如第 4.3 节所述,通过融合多层特征和软判决的 NMS 算法进一步提升了 Faster R-CNN 的飞机检 测性能,且改进方法的同时并没有造成过多的时间 成本。因此,所提方法的速度与 Faster R-CNN 相 当,且检测率和虚警率要优于 Faster R-CNN。综上 所述,与以往的飞机检测方法相比,所提方法的性能 最优。

5 结 论

以更快的区域卷积神经网络为基本框架,提出 了一种特征融合结合软判决的遥感图像飞机检测方 法。依次采用 L2 范数归一化、特征连接、尺度缩放 和特征降维来融合多层特征,引入软判决来改进传 统的非极大值抑制方法。实验结果表明,所提方法 能够提高小目标检测的性能,同时降低了网络在目标 高度重叠时的漏检率,最终得到检测率为94.25%、虚 警率为 5.5%、平均运行时间为 0.16 s 的实验结果。 所提方法优于现有的其他飞机检测方法,对于飞机的 实时、精确检测具有较好的理论指导和实际意义。

参考文献

- [1] Song M Z, Qu H S, Jin G. Weak ship target detection of noisy optical remote sensing image on sea surface [J]. Acta Optica Sinica, 2017, 37 (10): 1011004.
 宋明珠,曲宏松,金光.含噪光学遥感图像海面弱小舰船目标检测[J].光学学报,2017,37 (10): 1011004.
- [2] Li W, Xiang S M, Wang H B, et al. Robust airplane detection in satellite images [C]. IEEE International Conference on Image Processing, 2011: 2821-2824.
- [3] Bo S K, Jing Y J. Region-based airplane detection in

remotely sensed imagery [C]. International Congress on Image and Signal Processing, 2010: 1923-1926.

- Liu L, Shi Z W. Airplane detection based on rotation invariant and sparse coding in remote sensing images
 [J]. Optik-International Journal for Light and Electron Optics, 2014, 125(18): 5327-5333.
- [5] Chen X Y, Xiang S M, Liu C L, et al. Aircraft detection by deep belief nets [C]. Asian Conference on Pattern Recognition, 2013: 54-58.
- [6] Wu H, Zhang H, Zhang J F, et al. Fast aircraft detection in satellite images based on convolutional neural networks [C]. IEEE International Conference on Image Processing, 2015: 4210-4214.
- [7] Xin P, Xu Y L, Tan H, et al. Fast airplane detection based on multi-layer feature fusion of fully convolutional networks[J]. Acta Optica Sinica, 2018, 38(3): 0315003.
 辛鹏,许悦雷,唐红,等.全卷积网络多层特征融合 的飞机快速检测[J].光学学报, 2018, 38(3): 0315003.
- [8] Li Y, Fu K, Sun H, et al. An aircraft detection framework based on reinforcement learning and convolutional neural networks in remote sensing images[J]. Remote Sensing, 2018, 10(2): 243.
- [9] Zhang W C, Sun X, Fu K, *et al*. Object detection in high-resolution remote sensing images using rotation invariant parts based model[J]. IEEE Geoscience and

Remote Sensing Letters, 2014, 11(1): 74-78.

- [10] Wang G L, Wang X C, Fan B, et al. Feature extraction by rotation-invariant matrix representation for object detection in aerial image[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14 (6): 851-855.
- [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] 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.
- [13] Girshick R B. Fast R-CNN[C]. International Conference on Computer Vision, 2015: 1440-1448.
- [14] Kong T, Yao A, Chen Y, et al. HyperNet: towards accurate region proposal generation and joint object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 845-853.
- [15] Liu W, Rabinovich A, Berg A C, et al. ParseNet: looking wider to see better [EB/OL]. (2015-11-23) [2018-01-15]. https://arxiv.org/abs/1506.04579
- [16] Rothe R, Guillaumin M, Van Gool L, et al. Nonmaximum suppression for object detection by passing messages between windows[C]. Asian Conference on Computer Vision, 2014: 290-306.