

基于可变形卷积神经网络的遥感影像密集区域车辆检测方法

高鑫^① 李慧^{*①②} 张义^① 闫梦龙^① 张宗朔^③
孙显^① 孙皓^① 于泓峰^①

^①(中国科学院电子学研究所空间信息处理与应用系统技术重点实验室 北京 100190)

^②(中国科学院大学 北京 100049)

^③(中央兰开夏大学 英国 PR1 2HE)

摘要: 车辆检测是遥感图像分析领域的热点研究内容之一, 车辆目标的智能提取和识别, 对于交通管理、城市建设有重要意义。在遥感领域中, 现有基于卷积神经网络的车辆检测方法存在实现过程复杂并且对于车辆密集区域检测效果不理想的缺陷。针对上述问题, 该文提出基于端到端的神经网络模型DF-RCNN以提高车辆密集区域的检测精度。首先, 在特征提取阶段, DF-RCNN模型将深浅层特征图的分辨率统一并融合; 其次, DF-RCNN模型结合可变形卷积和可变形感兴趣区域池化模块, 通过加入少量的参数和计算量以学习目标的几何形变。实验结果表明, 该文提出的模型针对密集区域的车辆目标具有较好的检测性能。

关键词: 遥感影像; 车辆检测; 密集区域; 端到端卷积神经网络

中图分类号: TP751.2

文献标识码: A

文章编号: 1009-5896(2018)12-2812-08

DOI: [10.11999/JEIT180209](https://doi.org/10.11999/JEIT180209)

Vehicle Detection in Remote Sensing Images of Dense Areas Based on Deformable Convolution Neural Network

GAO Xin^① LI Hui^{*①②} ZHANG Yi^① YAN Menglong^① ZHANG Zongshuo^③
SUN Xian^① SUN Hao^① YU Hongfeng^①

^①(Key Laboratory of Technology in Geo-spatial Information Processing and Application System,
Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China)

^②(University of Chinese Academy of Sciences, Beijing 100049, China)

^③(University of Central Lancashire (UCLan), Preston PR1 2HE, United Kingdom)

Abstract: Vehicle detection is one of the hotspots in the field of remote sensing image analysis. The intelligent extraction and identification of vehicles are of great significance to traffic management and urban construction. In remote sensing field, the existing methods of vehicle detection based on Convolution Neural Network (CNN) are complicated and most of these methods have poor performance for dense areas. To solve above problems, an end-to-end neural network model named DF-RCNN is presented to solve the detecting difficulty in dense areas. Firstly, the model unifies the resolution of the deep and shallow feature maps and combines them. After that, the deformable convolution and RoI pooling are used to study the geometrical deformation of the target by adding a small number of parameters and calculations. Experimental results show that the proposed model has good detection performance for vehicle targets in dense areas.

Key words: Remote sensing images; Vehicle detection; Dense areas; End-to-end convolution neural network

1 引言

如何智能调控车辆数量、构建智能交通系统成

为了近些年的热门研究方向, 通过遥感影像车辆检测技术来掌握大范围、全路域的车辆数量及其分布特征是其核心内容之一。遥感领域中的车辆检测方法分为传统方法和基于卷积神经网络的(Convolution Neural Network, CNN)的方法。传统检测方法常用不同尺寸的滑窗遍历图像上的每个位置, 然后在每个滑窗内提取手工设计的特征并利用分类器进行分类。例如, Shao等人^[1]使用Harr特征和局部

收稿日期: 2018-03-02; 改回日期: 2018-06-11; 网络出版: 2018-07-16

*通信作者: 李慧 lihui@cas.ac.cn

基金项目: 国家自然科学基金(41501485)

Foundation Item: The National Natural Science Foundation of China (41501485)

二值模式(Local Binary Patterns, LBP)训练支持向量机(Support Vector Machine, SVM)分类器; Moranduzzo等人^[2]使用尺度不变特征转换(Scale Invariant Feature Transform, SIFT)算法、Klueckner等人^[3]和Thuckner等人^[4]使用方向梯度直方图(Histogram of Oriented Gradient, HOG)和积分通道特征(Integral Channel Features, ICF)训练Ada-boost分类器。但是, 上述传统方法使用穷搜法提取候选框需要巨大的计算量, 并且手工特征非常依赖特征的选择和设计, 当数据量很大或者是场景很丰富时, 手工特征表现出了不足。近年来, 基于CNN的方法被引入到了遥感领域车辆检测中, 另外, 为了减少由穷搜法带来的巨大计算量, 基于目标边缘、颜色和形状等信息的目标定位算法被提出^[5-9]。例如, Chen等人^[10]对多个阈值分割图像利用重心法定位车辆目标, 并提出混合神经网络模型, 使得网络能够学习到多尺度特征; Li等人^[11]利用车辆目标和周围背景颜色差异提取显著区域, 并利用灰度重心法定位车辆目标, 提出去噪神经网络模型(Denoizing-based Convolutional Neural Networks, DCNN), 模型对全连接层采用去噪自编码器的训练方法训练。以上基于CNN的方法都取得了较高的准确率, 但是存在以下问题: (1)定位目标算法中往往存在可变参数, 不同场景中的最优值可能不同, 并且难以处理密集区域目标, 例如上述采用重心法定位时, 得到的检测框可能包含多个车辆的部分, 造成一定程度的漏检现象。(2)检测框通常被设置为固定大小, 为了能包含各个角度的目标, 检测框尺寸通常较大, 但在密集区域一个检测框可能包含多个目标。(3)在对较大尺寸的影像进行搜索时, 需要着较大的计算量。(4)需要多个模型进行定位和检测, 难以进行整体优化。

随着目标检测技术的发展, 计算机视觉领域提出许多的端到端的网络模型, 将目标定位和检测统一到同一网络框架中, 能够进行端到端的训练和测试。其中, Faster-RCNN^[12]在目标检测方面表现出了良好的检测性能, 该模型首先利用多层神经网络提取特征, 并应用于目标定位和检测, 最终输出多尺寸的检测框。如果直接将Faster-RCNN应用于遥感影像的车辆检测中, 在密集区域的检测效果不理想, 其主要存在以下问题: (1)遥感影像中的车辆尺寸较小^[13,14], 而Faster-RCNN对小目标检测效果不理想^[13]。(2)卫星通过垂直或者某一倾斜视角捕获遥感影像^[14], 在倾斜视角下的密集区域标注框造成了一定程度的重叠, 影响了密集区域的检测结果。针对问题(1), Deng等人^[13]结合最后3层特征图

获得丰富的特征, 实验证明这种方法取得了较好的检测性能, 但是用于定位和检测的特征图分辨率太低, 在密集区域出现了大量的漏检现象。针对问题(2), 在计算机视觉领域中, Dai等人^[15]提出了可变形卷积模块和可变形RoI池化模块, 基本的思想就是用带偏移的采样代替原来的固定位置采样, 最终能够以自适应的方式学习密集区域目标的几何形变, 并在密集区域取得了较好的检测效果。

根据遥感影像中车辆目标的特点和上述已有模型方法, 本文基于Faster-RCNN提出DF-RCNN(Deformable Faster-RCNN)模型。DF-RCNN模型针对Faster RCNN在密集区域检测效果不理想的问题, 主要在以下两方面做出了改进: (1)针对小尺寸车辆目标, 为了提高输出特征图的分辨率并充分利用深浅层特征图的信息, DF-RCNN模型将深浅层特征图统一到较高的分辨率并融合。(2)DF-RCNN模型引入计算机视觉领域提出的可变形卷积和可变形RoI(Region of Interest)池化模块以学习密集区域目标的几何形变。实验结果表明, DF-RCNN模型针对密集区域的车辆目标表现出了较好的检测性能。

本文结构安排如下: 第1节介绍研究现状和本文研究内容; 第2节介绍相关技术; 第3节介绍本文改进方法和本文模型结构; 第4节描述试验内容及试验结果; 第5节总结全文及未来的工作计划。

2 相关技术

2.1 CNN

CNN是近年来发展起来并引起广泛重视的一种高效识别方法, 主要用于识别位移、缩放以及其他形式扭曲不变性的2维图形^[16-18], CNN的特征检测层通过隐式地从训练数据中学习从而避免了显式的特征提取, 因此CNN提取的特征更具有表达性。CNN的主要操作有卷积、非线性化、池化、全连接:

(1) 卷积层: 卷积层利用卷积核提取特征, 一个卷积核生成一个特征图;

(2) 非线性化: 非线性化增加了网络的表达能力, 通常使用的非线性函数有relu, tanh, sigmoid等;

(3) 池化层: 池化层降低了各个特征图的维度, 并保留了大部分重要信息;

(4) 全连接层: 全连接层将学到的特征把输入图像基于训练数据进行分类。

2.2 Faster-RCNN

Faster-RCNN是一个实时的检测模型, 其首次提出利用神经网络产生候选区域, 并将候选区域提取、特征提取和分类整合为一个整体的框架, 实现

了端到端的训练。Faster-RCNN模型结构如图1所示,其主要由共享卷积层、RPN网络、RoI池化和分类4部分组成。共享卷积层由一组基础的卷积、非线性化和池化构成,目的是提取输入图像的特征图,提取到的特征图被共享用于RPN网络 and 全连接层,常用VGG^[19], ResNet^[20]和ZFNet^[21]等网络。由共享卷积层提取的特征被送入RPN网络中,RPN采用一个 $n \times n$ 大小的卷积核在共享卷积层输出

特征图上滑动,在每个滑动位置映射为一个低维特征向量,这些低维向量被送入两个全连接层,两个全连接层输出多种尺寸和多种比例候选框的位置和得分。由于RPN网络输出不同尺寸的候选区域,RoI池化层将大小不同的候选区域映射为长度相同的特征向量。最后将固定尺寸的特征向量送入两个全连接层,并接入得分层和回归层进行分类以及微调检测框输出位置。

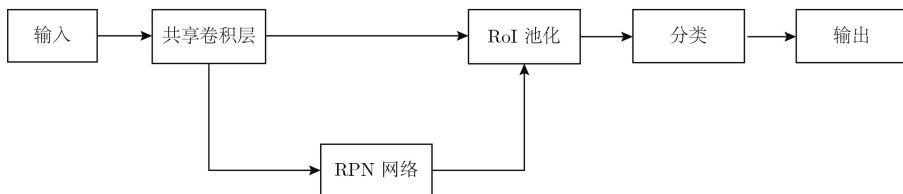


图1 Faster-CNN结构图

3 DF-RCNN模型

为了解决传统模型在密集区域检测时遇到的难点,本节在Faster-RCNN基础上提出DF-RCNN模型,该模型将特征提取模块中深浅层特征图统一到一个较高分辨率并融合产生融合特征图,解决原模型密集区域漏检现象严重的问题,并且采用了可变形卷积和可变形RoI池化模块以学习目标的形状信息,使得卷积核的形状适应目标形状。DF-RCNN模型对于遥感图像中密集区域车辆目标表现出较好的检测性能。

3.1 DF-RCNN总框架

DF-RCNN总框架如图2所示,模型输入图像大小为 500×500 ,网络调整为 1000×600 ,模型首先在VGG基础上加入可变形卷积层以学习偏移参数,并将深浅层特征结合产生融合特征图作为RPN模块和检测模块的输入。为了产生候选区域

框,本文在融合特征图上使用一个 3×3 的滑窗,在每一个滑动窗的位置,同时映射原图3种尺寸和3种长宽比的候选区域。本文数据采用0.1 m分辨率,遥感影像中车辆尺寸大约为 $20 \times 30 \sim 30 \times 40$ 像素,因此采用候选区域的大小为 20×20 , 30×30 , 40×40 和3种长宽比为2:1, 1:2和1:1,候选区域尺寸如表1所示。由于候选区域尺寸不同,DF-RCNN模型利用加入偏移参数的可变形RoI池化层从候选框和融合特征图提取定长(本文选用 7×7 大小)的特征向量,最后将这些特征和两个全连接层相连,由两个全连接层对候选区域调整 and 分类。

3.2 特征图融合

VGG的结构如图3所示,VGG结构包含5组卷积,前2组中每组包含2个卷积层,后3组中每组包含3个卷积层,每组分别命名为Conv1_x—Conv5_x。图像输入大小为 1000×600 ,输入到第1组中的第

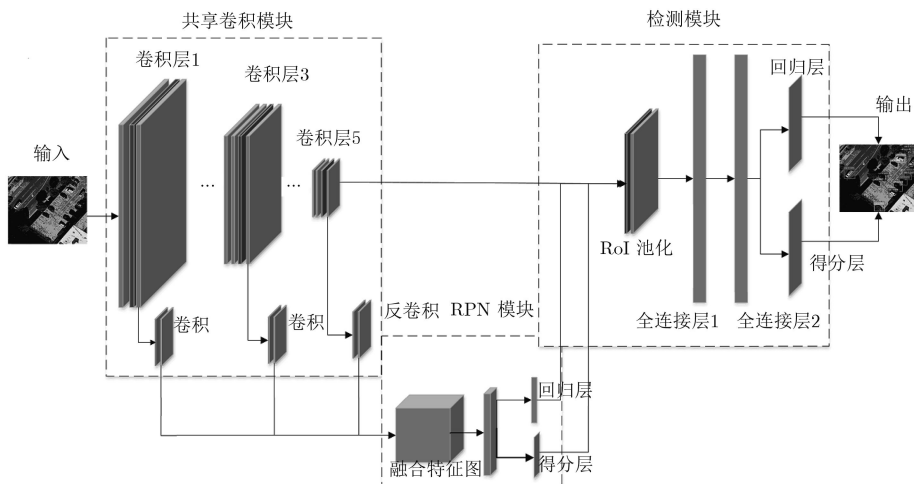


图2 DF-RCNN模型结构

表 1 3 种尺度和 3 种比例映射候选区域尺寸

区域尺寸和比例	20 ² , 1:1	20 ² , 1:2	20 ² , 2:1	30 ² , 1:1	30 ² , 1:2	30 ² , 2:1	40 ² , 1:1	40 ² , 1:2	40 ² , 2:1
映射区域大小	20×20	14×28	28×14	30×30	21×42	42×21	40×40	28×56	57×28

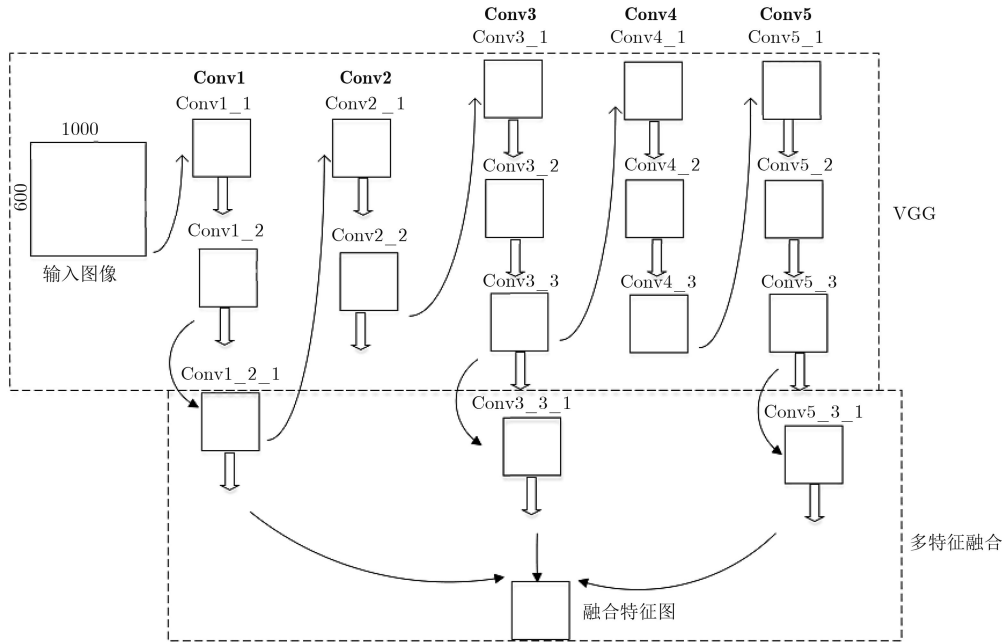


图 3 融合特征图结构

1个卷积层Conv1_1, Conv1_1后面接relu层, 第2个卷积层Conv1_2将前一层输出作为输入, 并和Conv1_1有相同的配置, Conv1_2后面接relu层和池化层, 剩下4组的卷积层与第1组类似。每组之间使用核大小为2×2的池化层进行分割, 因此VGG输出层Conv5_3分辨率为输入的1/16, 针对密集区域目标, 输出特征图分辨率太低。

DF-RCNN模型将深浅层的特征图统一到较高的分辨率并融合, 不同深浅层层融合后将会产生不同的输出特征。以融合1, 3, 5组中最后一层特征图为例, 由图3可知, 卷积层Conv1_2, Conv3_3和Conv5_3输出特征图尺寸不一致, 为了将各层特征图统一到相同尺寸, 本文在卷积层Conv1_2和Conv3_3后加入卷积操作, 在卷积层Conv5_3后加入反卷积操作, 并加入局部响应归一化(Local Response Normalization, LRN)操作, 其它深浅层特征图融合操作过程与图3类似。由图3可以看出共享卷积的融合特征图大小为输入的1/4。这对于遥感图像中的目标更为适合, 另外, 此方法没有产生更多的计算量。

3.3 可变形卷积和可变形RoI池化

为了学习密集区域车辆目标的形变信息, 本文引入可变形卷积和可变形RoI池化模块^[15]来增强网络对于几何变换的建模能力。两个模块的基本思想

都是用带偏移的采样代替原来的固定位置采样, 并且增加的偏移量是网络结构的一部分, 可以通过另外一个平行的标准卷积单元计算得到, 进而也可以通过梯度反向传播进行端到端的学习。加上该偏移量的学习之后, 可变形卷积核的大小和位置可以根据当前需要识别的图像内容进行动态调整, 不同位置的卷积核采样点位置会根据图像内容发生自适应的变化, 从而适应车辆目标的几何形变。两个模块的定义如下:

(1) 可变形卷积模块: 对于一个3×3大小、膨胀系数为1的卷积核, 本文将其定义为 $\mathbf{R}$:

$$\mathbf{R} = \{(-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\} \quad (1)$$

其中, 卷积核权重为 $\mathbf{w}$, 输入特征图 $\mathbf{x}$, 对于输出特征图 $\mathbf{y}$ 上的 p_0 点, 传统卷积表示为

$$\mathbf{y}(p_0) = \sum_{p_n \in \mathbf{R}} \mathbf{w}(p_n) \cdot \mathbf{x}(p_0 + p_n) \quad (2)$$

可变形卷积通过对卷积核中每个采样点的位置都增加一个2维偏移量使卷积核形状适应物体形状, 此偏移量利用一个卷积层从输入特征图学到, 偏移量和输入特征图共同输入到后一个卷积层, 结构如图4(a)所示。可变形卷积为每个特征点增加偏移量 Δp_n , 输出可表示为

$$\mathbf{y}(p_0) = \sum_{p_n \in \mathbf{R}} \mathbf{w}(p_n) \cdot \mathbf{x}(p_0 + p_n + \Delta p_n) \quad (3)$$

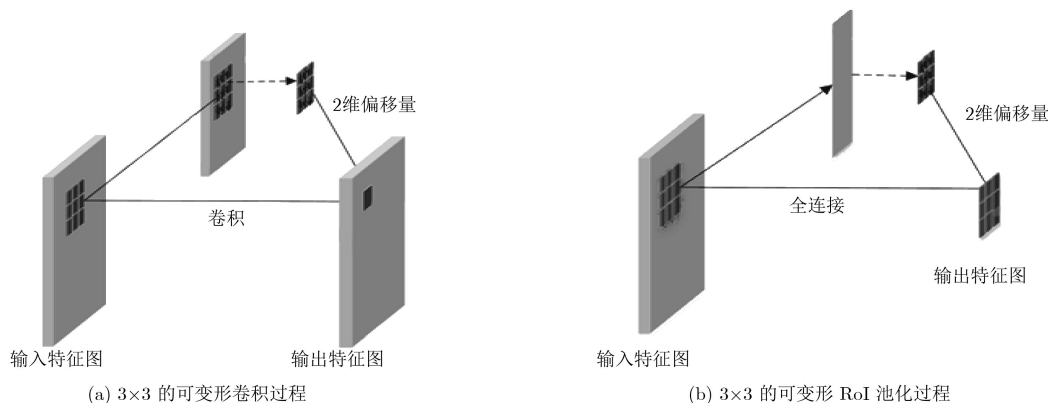


图4 可变性卷积和可变性RoI池化模块学习过程

(2) RoI池化的作用是将尺寸不同的候选区域映射为长度相同的特征向量, 具体实现是将候选区域映射为输出特征图分辨率大小(由前文可知, 映射分辨率大小为输入的1/16)的特征图, 将特征图平均分为 $k \times k$ 块, 并对每块进行最大化池化操作。对于第 (i, j) ($0 \leq i, j < k$) 特征图, 传统RoI池化输出为

$$y(i, j) = \sum_{p_n \in \text{bin}(i, j)} x(p_0 + p) / n_{i, j} \quad (4)$$

其中, x 为输入特征图, $n_{i, j}$ 为每块区域大小, p_0 为RoI左上角坐标。可变形RoI池化在RoI池化基础上对每块增加2维偏移, 与可变形卷积相同, 偏移量通过一个全连接层从输入特征图和RoI学到, 结构如图4(b)所示。其计算方式与可变形卷积类似, 输出为

$$y(i, j) = \sum_{p_n \in \text{bin}(i, j)} x(p_0 + p + \Delta p_{i, j}) / n_{i, j} \quad (5)$$

其中, $\Delta p_{i, j}$ 为每块的偏移量。

4 实验结果与分析

本文数据集是从Google Earth上收集了10000幅大小为 500×500 、分辨率为0.1 m的图像, 包含了所有场景如停车场、高速公路、居民区等, 并且场景不重叠。数据集包括了密集区域和非密集区域, 其中密集区域图像5000幅, 非密集图像5000幅。本文随机选取4000幅密集区域图像和4000幅非密集区域图像作为训练, 1000幅密集区域图像和1000幅非密集区域图像作为测试。为了评价车辆的检测性能, 本文选用3种常用的评价指标: 准确率、召回率和 F_1 指标, 3种指标的定义如式(6)~式(8):

$$\text{准确率} = \frac{\text{正确检测的个体总数}}{\text{检测出的个体目标}} \quad (6)$$

$$\text{召回率} = \frac{\text{正确检测的个体总数}}{\text{测试集中存在的个体总数}} \quad (7)$$

$$F_1 \text{ 指标} = \frac{2 \times \text{准确率} \times \text{召回率}}{\text{准确率} + \text{召回率}} \quad (8)$$

4.1 融合不同层特征图

考虑到相邻特征图之间相似性较高, 本文针对融合了3, 5层, 2, 4层和1, 3, 5层的模型进行试验, 并采用前文提出的方法将特征图统一到输入的1/4大小。表2展示了不同层检测性能结果, 从实验结果可以看出: 尽管浅层特征图的分辨率较高, 然而深层特征提取到更有效的特征, 显示出更好的检测性能; 多层融合比单层检测性能更好, 其中融合1, 3, 5层模型 F_1 指标最高。

4.2 结合可变形模块

4.1节证明了融合1, 3, 5层综合指标最高, 在此基础上, 本文分别对可变形卷积、可变形RoI池化、可变形卷积和可变形RoI池化结合的模型进行验证。首先, 我们在1, 3, 5层之前加入可变形卷积模块; 其次, 在传统的RoI池化之前加入了可变形RoI池化模块; 最后将可变形卷积和可变形RoI池化模块同时加入到网络中。为了评估可变形模块的普适性, 本文随机选取4.1节中的模型并结合可变形模块。表3展示了不同模型的检测结果, 从表3中可以看出单独使用可变形卷积和可变形RoI池化检

表2 不同层检测性能比较

模型	检测区域	召回率	准确率	F_1 指标
层1	密集区域	0.461	0.611	0.525
层3	密集区域	0.623	0.754	0.682
层5	密集区域	0.655	0.923	0.766
层3+5	密集区域	0.714	0.930	0.808
层2+4	密集区域	0.709	0.925	0.803
层1+3+5	密集区域	0.725	0.927	0.814

测总体指标均有提升，但是结合两者得到了最好的检测结果；可变形模块具有普适性，结合不同模型的检测结果均有提升。图5展示了融合1, 3, 5层特征图和加入可变形模块的检测结果，其中白色框表

示检测框，黑色框表示漏检目标。由图5可以看出，在车辆目标倾斜并且密集的区域，采用融合特征图的模型存在着少量的漏检现象，当模型再结合可变形模块，有效地改善了这一现象。

表 3 结合可变形模块检测性能比较

模型	检测区域	召回率	准确率	F_1 指标
层1+3+5, 可变形卷积	密集区域	0.731	0.932	0.819
层1+3+5, 可变形RoI池化	密集区域	0.726	0.925	0.814
层1+3+5, 可变形卷积, 可变形RoI池化	密集区域	0.744	0.940	0.831
层5, 可变形卷积, 可变形RoI池化	密集区域	0.725	0.924	0.812
层2+4, 可变形卷积, 可变形RoI池化	密集区域	0.715	0.922	0.805
层3+5, 可变形卷积, 可变形RoI池化	密集区域	0.729	0.928	0.817

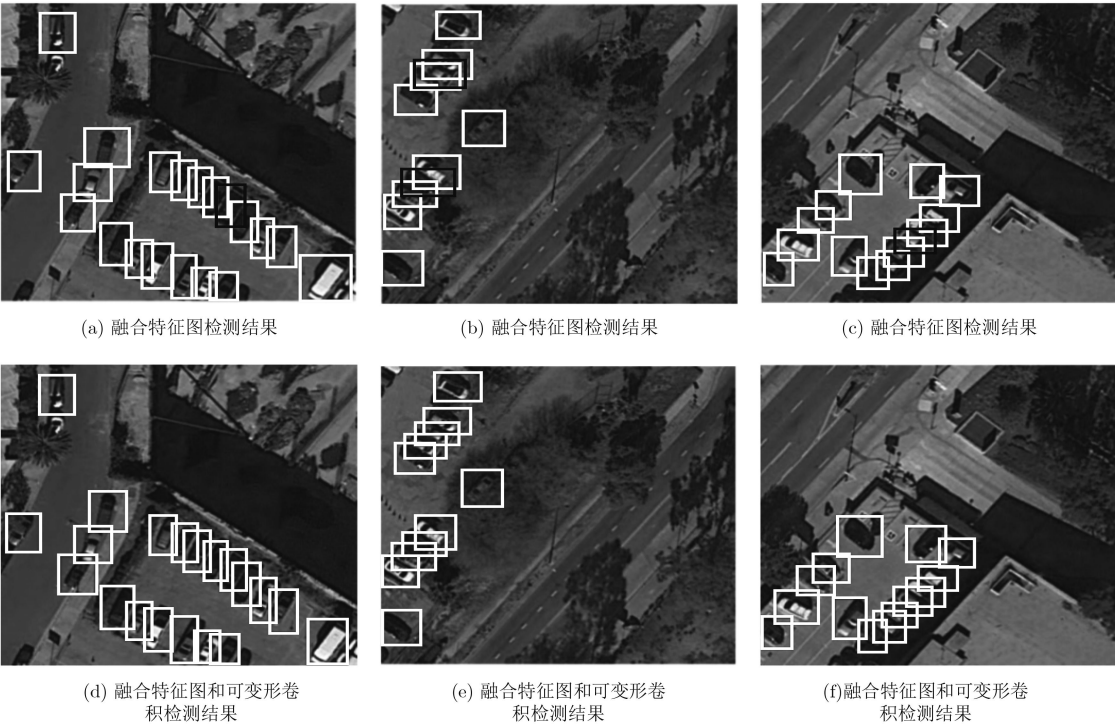


图 5 融合1, 3, 5层特征图和结合可变形模块的检测结果

4.3 模型对比

由4.1节和4.2节可知，当融合1, 3, 5层并结合可变形模块产生了最好的检测结果，并将此模型作为最终评估模型。本文将DF-RCNN, Faster-RCNN和DCNN分别在密集区域和非密集区域的检测性能进行比较，表4为3种模型在密集区域和非密集区域的检测结果。由表4可以看出：在密集区域，3个模型都有较高的准确率，Faster-RCNN在提取特征时，输出特征图分辨率较低而存在着大量的漏检，DCNN产生相同大小的检测框，不能适应目标的多尺寸，召回率较低，而DF-RCNN输出目标框更贴合目标形状且漏检率较低；在非密集区域，DF-RCNN

和Faster-RCNN利用网络提取候选区域，准确率都较高，而DCNN利用算法定位目标区域，受环境因

表 4 不同模型检测性能比较

模型	检测区域	召回率	准确率	F_1 指标
Faster-RCNN	密集区域	0.655	0.932	0.766
DCNN	密集区域	0.402	0.925	0.560
DF-RCNN	密集区域	0.744	0.940	0.831
Faster-RCNN	非密集区域	0.901	0.942	0.921
DCNN	非密集区域	0.962	0.785	0.865
DF-RCNN	非密集区域	0.910	0.952	0.931

素的影响较大而准确率较低。综上,DF-RCNN模型在密集区域和非密集区域都表现出了良好的检测

性能。图6展示了不同模型在车辆密集区域的检测效果,其中白色框表示检测框。

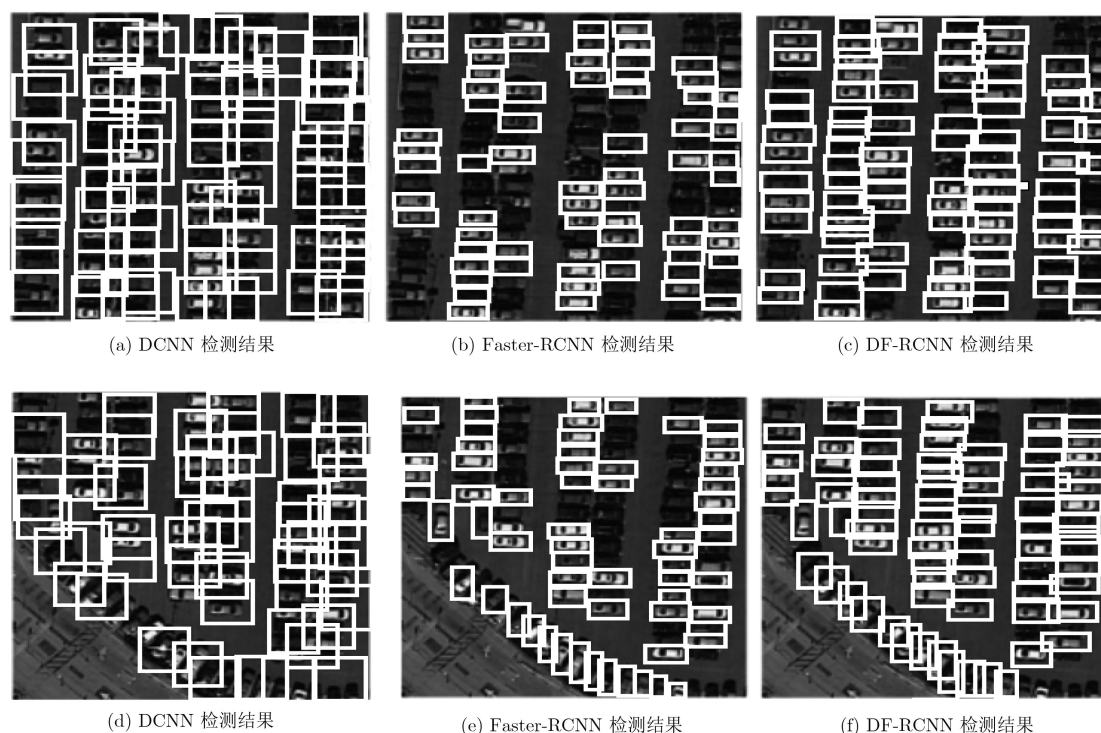


图6 不同模型的检测结果

5 结束语

本文中提出了DF-RCNN模型以提升遥感图像中密集区域的检测精度。本文模型采取两种策略:首先统一深浅层特征图的分辨率到合适尺度并进行融合提高模型在密集区域的召回率;其次结合了可变形卷积和可变形RoI池化模块学习密集区域的车辆形变信息,实验证明本文的模型对于密集区域车辆目标较为有效。在未来的学习中,将对CNN进行更深入的研究,争取更高的检测率和更短的训练时间。

参考文献

- [1] SHAO Wen, YANG Wen, LIU Guang, *et al.* Car detection from high-resolution aerial imagery using multiple features[C]. Geoscience and Remote Sensing Symposium, Munich, Germany, 2012: 4379–4382. doi: [10.1109/IGARSS.2012.6350403](https://doi.org/10.1109/IGARSS.2012.6350403).
- [2] MORANDUZZO T and MELGANI F. Automatic car counting method for unmanned aerial vehicle images[J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2013, 52(3): 1635–1647. doi: [10.1109/TGRS.2013.2253108](https://doi.org/10.1109/TGRS.2013.2253108).
- [3] KLUCKNER S, PACHER G, GRABNER H, *et al.* A 3D teacher for car detection in aerial images[C]. International Conference on Computer Vision, Rio de Janeiro, Brazil, 2007: 1–8. doi: [10.1109/ICCV.2007.440883](https://doi.org/10.1109/ICCV.2007.440883).
- [4] TUERMER S, KURZ F, Reinartz P, *et al.* Airborne vehicle detection in dense urban areas using HoG features and disparity maps[J]. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 2013, 6(6): 2327–2337. doi: [10.1109/JSTARS.2013.2242846](https://doi.org/10.1109/JSTARS.2013.2242846).
- [5] CHEN Ziyi, WANG Cheng, WEN Chenglu, *et al.* Vehicle detection in high-resolution aerial images via sparse representation and superpixels[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, 54(1): 103–116. doi: [10.1109/TGRS.2015.2451002](https://doi.org/10.1109/TGRS.2015.2451002).
- [6] 康妙, 计科峰, 冷祥光, 等. 基于栈式自编码器特征融合的SAR图像车辆目标识别[J]. 雷达学报, 2017, 6(2): 167–176. doi: [10.12000/JR16112](https://doi.org/10.12000/JR16112).
- [7] KANG Miao, JI Kefeng, LENG Xiangguang, *et al.* SAR target recognition with feature fusion based on stacked autoencoder[J]. *Journal of Radars*, 2017, 6(2): 167–176. doi: [10.12000/JR16112](https://doi.org/10.12000/JR16112).
- [8] 王思雨, 高鑫, 孙皓, 等. 基于卷积神经网络的高分辨率SAR图像飞机目标检测方法[J]. 雷达学报, 2017, 6(2): 195–203. doi: [10.12000/JR17009](https://doi.org/10.12000/JR17009).
- [9] WANG Siyu, GAO Xin, SUN Hao, *et al.* An aircraft detection method based on convolutional neural networks in high-resolution SAR images[J]. *Journal of Radars*, 2017, 6(2): 195–203. doi: [10.12000/JR17009](https://doi.org/10.12000/JR17009).

- [8] 田壮壮, 占荣辉, 胡杰民, 等. 基于卷积神经网络的SAR图像目标识别研究[J]. 雷达学报, 2016, 5(3): 320–325. doi: [10.12000/JR16037](https://doi.org/10.12000/JR16037).
TIAN Zhuangzhuang, ZHAN Ronghui, HU Jiemin, *et al.* SAR ATR based on convolutional neural network[J]. *Journal of Radars*, 2016, 5(3): 320–325. doi: [10.12000/JR16037](https://doi.org/10.12000/JR16037).
- [9] YANG Zi and PUN-CHENG L S C. Vehicle detection in intelligent transportation systems and its applications under varying environments: A review[J]. *Image & Vision Computing*, 2017, 69: 143–154. doi: [10.1016/j.imavis.2017.09.008](https://doi.org/10.1016/j.imavis.2017.09.008).
- [10] CHEN Xueyun, XIANG Shiming, LIU Chenglin, *et al.* Vehicle detection in satellite images by hybrid deep convolutional neural networks[J]. *IEEE Geoscience & Remote Sensing Letters*, 2014, 11(10): 1797–1801. doi: [10.1109/LGRS.2014.2309695](https://doi.org/10.1109/LGRS.2014.2309695).
- [11] LI Hao, FU Kun, YAN Menglong, *et al.* Vehicle detection in remote sensing images using denoizing-based convolutional neural networks[J]. *Remote Sensing Letters*, 2017, 8(3): 262–270. doi: [10.1080/2150704X.2016.1258127](https://doi.org/10.1080/2150704X.2016.1258127).
- [12] REN Shaoqing, HE Kaiming, SUN Jian, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, 39(6): 1137–1149. doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031).
- [13] DENG Zhipeng, SUN Hao, ZHOU Shilin, *et al.* Toward fast and accurate vehicle detection in aerial images using coupled region-based convolutional neural networks[J]. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 2017, 10(8): 3652–3664. doi: [10.1109/JSTARS.2017.2694890](https://doi.org/10.1109/JSTARS.2017.2694890).
- [14] ELMIKATY M and STATHAKI T. Car detection in aerial images of dense urban areas[J]. *IEEE Transactions on Aerospace & Electronic Systems*, 2017, 54(1): 51–63. doi: [10.1109/TAES.2017.2732832](https://doi.org/10.1109/TAES.2017.2732832).
- [15] DAI Jifeng, QI Haozhi, XIONG Yunwen, *et al.* Deformable convolutional networks[OL]. <http://arxiv.org/abs/1703.06211>.2017.
- [16] LIU Derong, LI Hongliang, and WANG Ding. Feature selection and feature learning for high-dimensional batch reinforcement learning: A survey[J]. *International Journal of Automation and Computing*, 2015, 12(3): 229–242. doi: [10.1007/s11633-015-0893-y](https://doi.org/10.1007/s11633-015-0893-y).
- [17] YANG Bin, YAN Junjie, LEI Zhen, *et al.* Convolutional channel features[C]. IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 82–90. doi: [10.1109/ICCV.2015.18](https://doi.org/10.1109/ICCV.2015.18).
- [18] ZHONG Bin, ZHANG Jun, WANG Pengfei, *et al.* Jointly feature learning and selection for robust tracking via a gating mechanism[J]. *Plos One*, 2016, 11(8): e0161808. doi: [10.1371/journal.pone.0161808](https://doi.org/10.1371/journal.pone.0161808).
- [19] SIMONYAN K and ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[OL]. <http://arxiv.org/abs/1409.1556>.2014.
- [20] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Deep residual learning for image recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, United States, 2016: 770–778. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [21] ZEILER M D and FERGUS R. Visualizing and Understanding Convolutional Networks[C]. European Conference on Computer Vision, Zurich, Switzerland, 2014: 818–833. doi: [10.1007/978-3-319-10590-1_53](https://doi.org/10.1007/978-3-319-10590-1_53).
- 高 鑫: 男, 1966年生, 研究员, 研究方向为机载SAR信息处理应用、空间信息处理与应用系统技术研究。
- 李 慧: 女, 1992年生, 硕士生, 研究方向为图像处理与分析。
- 张 义: 男, 1987年, 助理研究员, 研究方向为阵列信号处理。
- 闫梦龙: 男, 1985年生, 副研究员, 研究方向为机器学习与遥感影像智能解译。
- 张宗朔: 男, 1995年生, 本科生。
- 孙 显: 男, 1981年生, 副研究员, 研究方向为机器学习与遥感影像智能解译。
- 孙 皓: 男, 1984年生, 副研究员, 研究方向为机器学习与遥感影像智能解译。
- 于泓峰: 男, 1991年生, 助理研究员, 研究方向为机器学习与遥感影像智能解译。