基于卷积神经网络的遥感图像目标检测

欧攀, 张正*, 路奎, 刘泽阳
北京航空航天大学仪器科学与光电工程学院, 北京 100191

摘要 针对遥感图像中的目标检测问题,采用基于卷积神经网络的目标检测框架对目标进行提取,针对该网络制作了包含三类遥感图像中常见目标的目标检测数据集。为了解决遥感图像目标旋转角度较大的问题,将空间变换网络融入超快区域卷积神经网络,提出了一种具有旋转不变性自学习能力的目标检测模型。通过与传统的目标检测方法进行对比分析,探究了不同方法对遥感图像目标检测的实际效果。相对于传统的目标检测方法,融合了空间变换网络的卷积神经网络所提取的特征具有更好的旋转不变特性,从而能够达到更高的检测精度。

关键词 图像处理; 卷积神经网络; 空间变换网络; 目标检测; 深度学习
中图分类号 TP183 **文献标识码** A **doi**: 10.3788/LOP56.051002

Object Detectionin of Remote Sensing Images Based on Convolutional Neural Networks

Ou Pan, Zhang Zheng*, Lu Kui, Liu Zeyang
School of Instrumentation Science and Opto-Electronic Engineering, Beihang University, Beijing 100191, *China*

Abstract Aiming at the problem of object detection in remote sensing images, the Faster-Rcnn network based on the convolutional neural network models is used to extract the features of the object area. An object detection dataset containing three kinds of common targets in remote sensing images is made to train this network. In addition, in order to solve the problem of large rotation angle of remote sensing images, a target detection model with a rotation invariance self-learning ability is proposed, which integrates the spatial transformation network into the Faster R-CNN framework. By the analysis and comparison with the traditional object detection methods, the true effects of object detection in remote sensing images by different methods are explored. The features extracted by the convolutional neural networks based on the spatial transformation networks possess stronger orientation robustness than those by the traditional methods, which makes it possible to obtain a high detection precision.

Key words image processing; convolutional neural networks; spatial transformation networks; object detection; deep learning
OCIS codes 100.4996; 100.3008; 100.4999; 100.2000

1 引 言

随着传感器技术和航天遥感技术的进步,研究人员现在可以方便地获取具有高空间分辨率或光谱分辨率的遥感图像。这些遥感图像提高了研究者对图像内容的理解程度,特别是在分析图像内容时。因此,遥感图像目标检测技术具有很高的研究价值。目标检测技术可以对图像目标进行准确的识别与定位,其流程大致为图像预处理、图像特征提取、使用分类器对特征进行分类。传统的图像特征提取方法主要是手工特征提取,针对不同的图像识别与检测任务,往往需要设计不同的特征提取方法,而不同的手工特征可以获得不同的效果。在遥感图像中,目标以俯视图视角呈现,且方向角变化范围较大。这就要求提取的特征对于目标的旋转要有较好的稳健性。为了实现旋转不变特征的提取,研究者们进行了许多尝试,如尺度不变量特征变换(SIFT)[1]、方向梯度直方图(HOG)[2]和显著性[3]等。然而,这些手工特征在遥感目标检测任务上并没有达到令人满意的表现,原因在于

收稿日期: 2018-08-06; **修回日期**: 2018-09-12; **录用日期**: 2018-09-21
* **E-mail**: seirios@buaa.edu.cn

这些特征提取方法并非是为提取旋转不变特征而设计的，因此，在某些领域（如行人检测）表现较好的特征描述符应用于遥感图像目标检测时仍需进行加强或者改进。在此基础上，Xiao 等[4]提出了一种基于极坐标傅里叶变换的旋转不变性图像特征提取方法，使用椭圆傅里叶变换（EFT）来增强 HOG 描述符的旋转不变稳健性。

随着深度学习的发展，卷积神经网络（CNN）在图像分类与识别领域取得了突破性进展，图像分类任务上 AlexNet[5]的出色表现使得卷积神经网络受到研究者们的广泛关注。在此基础上，目标检测领域也取得了重大进步。Girshick 等[6]提出的区域卷积神经网络（R-CNN），将 VOC2012 目标检测任务的平均精度均值（mAP）提升了 30%。超快区域卷积神经网络（Faster R-CNN）[7]改进提取候选区域（ROIs）的方式，实现了目标检测任务端到端的训练，并且提升了区域提取效率。使用卷积神经网络可以提取图像的高层特征，相比于手工提取的特征更加抽象，且具有更好的稳健性，对不同的目标检测任务都能达到不错效果。得益于池化层，卷积神经网络对平移和缩放具有一定的学习能力，但仍然缺乏对输入数据旋转不变性的学习能力。空间变换网络（STN）[8]是一个变换模块，通过显式学习，可以使传统卷积层获得平移、缩放、旋转等特性，在字符识别等应用上取得良好效果。本文通过将目标检测模型 Faster R-CNN 与 STN 模块相结合，提出了一种对目标旋转更具稳健性的检测框架，从而提高遥感图像目标检测的准确性。

2 超快区域卷积神经网络原理

2.1 区域推荐网络

Faster R-CNN 使用区域推荐网络（RPN）进行候选区域的推荐。此前，R-CNN、快速区域卷积神经网络[9]等框架均使用选择搜索[10]进行区域检测。RPN 基于滑动窗口的思想，使用卷积神经网络直接产生推荐区域，可以和分类网络共享特征提取网络，这就有效地减少了计算量，从而提高了整个目标检测过程的效率。

传统滑窗在原图上进行滑动，将每个窗口作为一个候选区域，计算量较大，而 RPN 通过卷积神经网络对原图进行一系列的卷积和池化过程，提取了原图的深层信息，并且进一步缩小了原图的尺寸。在此基础上使用 3×3 的卷积核在特征图上进行滑动，并且以该窗口的中心点映射回原图的对应点。在原图的对应点设置 9 个不同大小、中心重合的 anchor，由于特征图由原图进行多次池化得来，因此 9 个不同大小的 anchor 可以覆盖特征图上每一点在原图上的对应区域。使用 RPN 获得的候选区域在保留了原图有效信息的同时，减小了候选区域的生成数量，从而减少了计算量。

2.2 特征提取及分类网络

特征提取及分类网络使用 RPN 提取的候选区域作为输入，使用卷积神经网络进行特征提取后送入分类器得到检测结果。卷积神经网络可以使用视觉几何群网络（VGG-16）[11]、残差网络（ResNet）[12]等目前领先的分类网络，分类器使用归一化指数函数。由于生成候选区域的生成与分类网络都使用了卷积神经网络进行特征提取，因此 Faster R-CNN 通过共享特征提取网络，避免了重复计算，同时实现了端到端的训练。

3 空间变换网络

3.1 旋转与仿射变换

旋转是仿射变换的一种具体形式，如图 1 所示，点 $A(x_A, y_A)$ 旋转 θ 角得到点 $B(x_B, y_B)$，且 B 与水平方向夹角为 α。

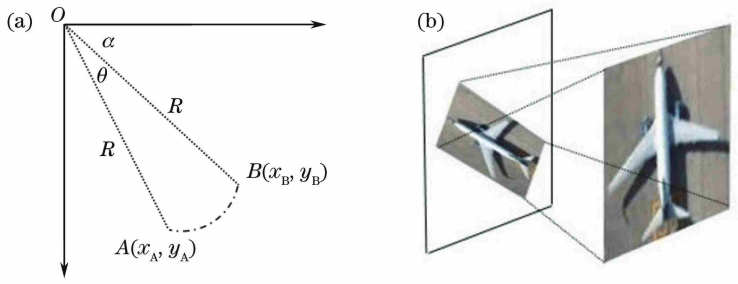

图 1 旋转变换示意图。(a)旋转原理；(b)目标旋转

Fig. 1 Schematic of rotation transformation. (a) Rotation principle; (b) target rotation

由对应关系可得：

$$x_B = x_A\cos\theta + y_B\sin\theta, \tag{1}$$

$$y_A = -x_A\sin\theta + y_B\cos\theta, \tag{2}$$

转换为矩阵形式，即

$$\begin{pmatrix} x_B \\ y_B \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_A \\ y_A \end{pmatrix}, \tag{3}$$

由此可以得到两点之间的旋转变换关系。更一般地，对于平面坐标系上两点(x, y)和(x', y')，通常使用 2×3 的矩阵来表示仿射变换关系，即

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \tag{4}$$

式中 a、b、c、d、e、f 为矩阵参数，通过特定的参数取值，可以实现平移、缩放、旋转等不同变换关系。当 $a=\cos\theta$，$b=\sin\theta$，$d=-\sin\theta$，$e=\cos\theta$，$c=f=0$ 时，两点之间即为旋转变换关系。

3.2 STN 的结构

STN 是一个显式的特征学习模块，它明确地针对特征的空间信息进行学习，使原网络在不进行额外监督训练的情况下获得空间变换的能力。作为一个独立且可导的模块，它可以灵活地插入到现有的卷积结构中，具有很强的可移植性。如图 2 所示，STN 由定位网络、网格生成器和采样网络三部分构成。

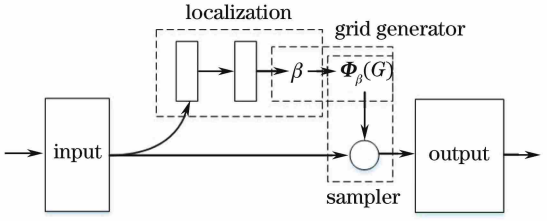

图 2 空间变换网络结构图

Fig. 2 Structural diagram of spatial transformation network

定位网络由卷积层和全连接层后接一个回归层构成，对输入的特征图进行变换，训练并且输出仿射变换的参数。网格生成器利用了仿射变换原理，使用输出的参数 β 对特征图进行变换。输入特征图与输出特征图的映射关系为

$$\begin{pmatrix} x_i^{(s)} \\ y_i^{(s)} \end{pmatrix} = \boldsymbol{\Phi}_\beta(G_i) = \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \end{pmatrix} \begin{pmatrix} x_i^{(t)} \\ y_i^{(t)} \\ 1 \end{pmatrix}, \tag{5}$$

式中 $x_i^{(s)}$、$y_i^{(s)}$ 为输入坐标，$\boldsymbol{\Phi}_\beta$ 表示以参数 β 为变换矩阵的仿射变换映射，$G_i=(x_i^{(t)}, y_i^{(t)})$ 代表对应的输出坐标，$G=\{G_i\}$ 代表输出坐标集合。进一步可以转换为仿射变换矩阵与输出坐标相乘的形式，通过对仿射变换参数 β_{11}、β_{12}、β_{13}、β_{21}、β_{22} 和 β_{23} 的学习，卷积神经网络可以获得包括旋转不变性在内的多个变换能力，稳健性获得增强。

经过仿射变换的输出特征图中可能会出现非整数像素坐标点，采样网络使用双线性插值的方式将非整数坐标点映射到整数位置，即

$$V_i^{(c)} = \sum_n^H \sum_m^W U_{nm}^{(c)} \times \max(0, 1-|x_i^{(s)}-m|)\max(0, 1-|y_i^{(s)}-n|), \tag{6}$$

式中 $V_i^{(c)}$ 对应输出特征图 c 通道上某点的灰度值，$U_{nm}^{(c)}$ 对应输入特征图 c 通道上点(n, m)的特征值。双线性插值使用输入坐标点$(x_i^{(s)}, y_i^{(s)})$周围的 4 个点的灰度来确定输出点的灰度值。$|x_i^{(s)}-m|$ 和 $|y_i^{(s)}-n|$ 表明，插值点距离 $(x_i^{(s)}, y_i^{(s)})$ 越近，权重越大。由于采样网络公式对 $U_{nm}^{(c)}$ 和 $(x_i^{(s)}, y_i^{(s)})$ 是可导的，因此 STN 的参数可以通过卷积神经网络的训练进行不断修正。

4 改进的检测框架

RPN 基于滑动窗口思想，使得 Faster R-CNN 具有平移不变性的学习能力。为了使检测框架获得旋转不变性的学习能力，提高 Faster R-CNN 对旋转目标的检测稳健性，从而使目标检测模型更适用于遥感图像目标检测任务。基于 STN 灵活的可移植特点，将 Faster R-CNN 与 STN 相结合，提出一种能够自主学习旋转不变特征的遥感目标检测模型。

4.1 结构调整

STN 作为一个独立模块，可以插入卷积神经网络的各个位置。在分类任务里，由于每张图片里只包含单个目标图像，不需要进行目标的定位，因此 STN 可以作为输入对原图进行变换，对输出图像进行后续的特征提取和分类，也可以将 STN 放置在卷积和池化等操作的后面对特征图进行变换。

然而在目标检测任务里，一张图片往往包含多个目标区域，且背景也占据了原图的部分区域，因此不能使用 STN 直接处理原图，而应该将提取到的候选框作为输入，使得 STN 的输入只包含单个目标区域。

在传统 Faster R-CNN 的 RPN 与分类网络之间插入一个 STN，便得到了改进后的目标检测框架，如图 3 所示。左边的虚线框里为候选区域提取网络，与原Faster R-CNN网络保持一致，即原卷积

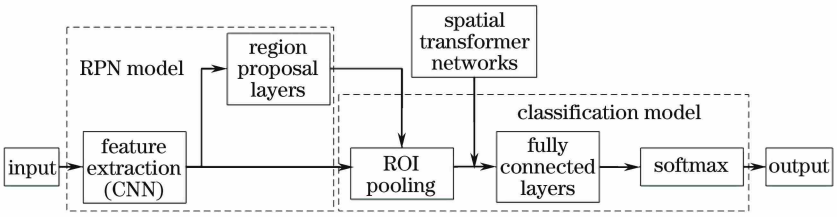

图 3 改进方法示意图

Fig. 3 Schematic of improved method

神经网络提取输入原始图片的特征，输出的特征图分别被 RPN 和分类网络共享，其中 RPN 网络使用特征图提取候选区域。右边的虚线框里为分类网络，特征池化层(ROI pooling)则利用候选区域从特征图中提取特征区域送入后续的全连接和归一化指数函数网络进行分类。将 STN 设置在特征池化层之后，包含目标的特征区域作为 STN 的输入，经过 STN 增强之后再送入全连接层进行分类操作。通过调节 STN 输入和输出参数来适应原网络，可以避免原有其他层参数的修改，由于 STN 只对输入进行仿射变换，因此不会改变输入特征图的其他特征。

4.2 训练策略调整

常规目标检测任务不需要对数据集进行过多分类处理，而解决目标转角变化较大的目标检测任务，通常使用的策略是数据增强[13]。具体做法是在数据集的制作过程中，需要确定一个基准方向，并且根据目标与基准方向的夹角对样本进行均匀的分类，从而保证目标检测网络得到充分且均匀的学习，对各个朝向的目标均有较好的识别能力。图 4 所示为本研究所使用数据集三种样本的角度分布直方图。设目标朝向与竖直方向夹角为 0°时为基准方向，按照样本朝向与基准方向的夹角均匀分为 6 组，且每组之内的样本数量分布较均匀。制作训练集时，在每个分组内均匀选取样本，保证训练集里各个朝向的目标分布均匀。

5 对比实验

5.1 实验数据

实验数据集部分来自网络公开数据集，部分截取自谷歌地球。空间分辨率为 0.5～2 m，样本总数为 17000，分为飞机、汽车、船三类[14]。其中飞机样本和汽车样本分别为 7000，船样本为 3000，每一类样本按照角度分布均匀分为 6 组。数据集以 VOC2007 的格式制作，且训练集和测试集的比例为 7∶3。由于设置了手工特征 HOG 算子的对比实验，因此同时截取不含目标的遥感图像 1000 张，用于负样本区域的提取。部分训练样本及负样本如图 5 和图 6 所示。

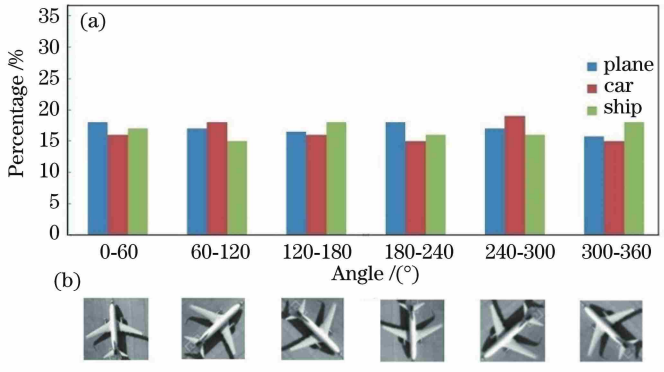

图 4 角度分布统计图。(a)角度分布直方图；(b)不同朝向的目标

Fig. 4 Statistical figure of angle distribution. (a) Histogram of angle distribution; (b) targets with different orientations

5.2 实验环境与模型参数

实验配置如下：显卡为 NAVIDA GTX 1060，CPU 为 Intel Core i5-8600k，3.60 GHz，内存为16 G，操作系统为 Ubuntu 16.04，框架为 Tensor Flow。

实验所用卷积神经网络均采用 ResNet101。ResNet 使用残差网络更容易优化，解决了网络层数增加带来的退化问题，可以达到比 VGG-16 更高的精度。模型的参数设置如下：learning_rate 为 0.001，step_size 为 30000，momentum 为 0.9，weight_decay为0.0001，batch_size为128。每次训

图 5 训练集样本

Fig. 5 Samples from training dataset

图 6 负样本

Fig. 6 Negative samples

练的迭代次数均为 10000。模型的性能由精确率、召回率以及通过调节阈值获得的平均精度(AP)值进行评估[15]。

5.3 实验设置及结果分析

5.3.1 改变数据集大小

不同的训练集大小对模型的检测性能有较大影响。使用不同大小的训练集分别训练 Faster R-CNN 与融合了 STN 的改进模型,使用飞机数据集作为训练与测试样本,评估数据集大小对检测结果 AP 值的影响。图 7 为平均精度随着正样本数量的增加得到的变化曲线。可以看出,基于 Faster R-CNN 的检测框架在正样本数量达到 1000 左右时平均精度趋于稳定,融合了 STN 的改进模型在正样本数量为 700 时平均精度已经趋于稳定。与 Faster R-CNN 相比较,本研究方法在达到同样检测精度的情况下所需的训练集较少。在同样训练集的情况下达到的检测精度较高。

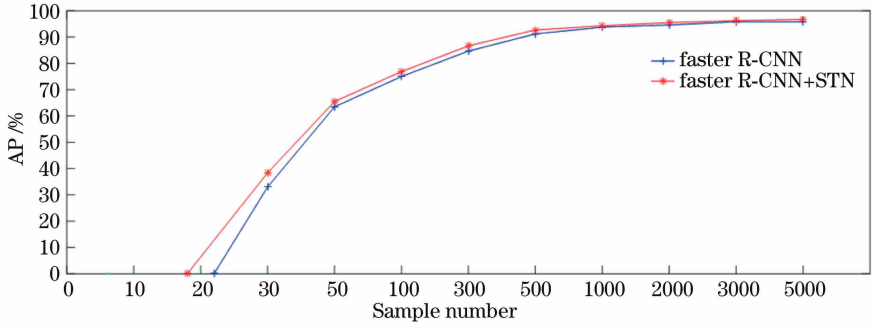

图 7 不同样本数量下的平均精度变化曲线

Fig. 7 Average precision versus sample number

5.3.2 不同检测方法的性能对比

采用传统的 HOG 算子加滑窗检测方法作为对比,评估三种方法的单张图像检测时间以及 AP 值。三种方法使用相同的数据集,种类为飞机,样本大小为 1500,样本图片大小均为 1280 pixel×659 pixel,检测结果如表 1 所示。传统的 HOG 特征配合支持向量机(SVM)分类,由于 HOG 特征不具备旋转不变性,因此检测精度较低;同时由于滑动窗口法提取特征区域的效率低于 RPN 网络,因此单张图片的检测耗时较长。与改进前的 Faster R-CNN 相比,融合了 STN 的改进模型的检测精度有所提高;同时单张图片的检测耗时基本不变,说明 STN 对检测网络的计算量影响较小。

表 1 不同检测方法性能对比

Table 1 Performance comparison among different detection methods

Detection method	Detection time of image /s	AP /%
HOG+SVM	1.03	64.61
Faster R-CNN	0.31	96.58
Faster R-CNN+STN	0.31	97.31

5.3.3 不同类别样本的检测结果

对于以上三种方法,取相同数据集大小的飞机、汽车、船三类样本分别进行实验,其中各类样本容量均为 1500,结果如表 2 所示。可以看出,同一检测方法的检测结果中,飞机的检测精度都高于汽车与船,这是由于飞机的目标较大,不易与建筑物混淆而

表 2 不同类别样本结果对比

Table 2 Comparison of detection results for different classes %

Detection method	AP of plane	AP of car	AP of ship
HOG+SVM	64.61	61.40	60.97
Faster R-CNN	96.58	88.35	84.87
Faster R-CNN+STN	97.31	89.71	84.76

产生虚警，且常常位于机场等单一环境里，环境噪声小，而相对较小的汽车和舰船，由于目标大小与建筑物相近，且所处环境较为复杂，因此检测平均精度相对较低。对于同一种测试样本，传统的 HOG 算子加滑窗检测方法的检测精度都远低于另外两种方法，而本研究所用方法相比传统 Faster R-CNN 都有所提高，说明本研究方法对不同的样本检测均有较好的稳健性。

部分遥感图像的检测结果如图 8 和图 9 所示，可以看出，飞机检测的结果中，检测框的置信度均为 1，说明检测精度较高；而小目标[16]汽车的检测框的置信度普遍处于 0.9 左右，说明检测精度相对飞机而言略低一些。

图 8 飞机的检测结果

Fig. 8 Detection results of plane

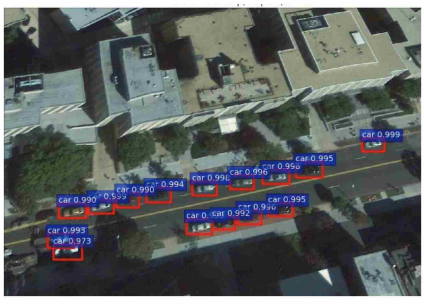

图 9 汽车的检测结果

Fig. 9 Detection results of car

6 结 论

以 Faster R-CNN 网络为基础，通过加入 STN 以提升对遥感图像目标的检测效果，通过与改进前网络在不同训练集大小下的平均精度进行对比，改进网络获得稳定精度所需的样本数量较改进前网络有所下降，飞机检测任务的精度提高到 97.31%，验证了 STN 对旋转目标有一定的自学习能力。同时检测小目标汽车与船的平均精度相对飞机较低，说明检测网络对小目标的检测还存在一些问题，对复杂环境的稳健性还有一定的提升空间。下一步将针对遥感图像中的小目标的检测进行研究，提升检测网络对多尺度目标的检测能力。

参 考 文 献

[1] 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[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(6): 3325-3337.

[2] Shao W, Yang W, Liu G, *et al*. Car detection from high-resolution aerial imagery using multiple features [C] // IEEE International Geoscience and Remote Sensing Symposium, 2012: 4379-4382.

[3] Zhang F, Du B, Zhang L P. Saliency-guided unsupervised feature learning for scene classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(4): 2175-2184.

[4] Xiao Z F, Liu Q, Tang G F, *et al*. Elliptic Fourier transformation-based histograms of oriented gradients for rotationally invariant object detection in remote-sensing images[J]. International Journal of Remote Sensing, 2015, 36(2): 618-644.

[5] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.

[6] Girshick R, Donahue J, Darrell T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.

[7] 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.

[8] Jaderberg M, Simonyan K, Zisserman A, *et al*. Spatial transformer networks [EB/OL]. (2016-02-04) [2018-07-25]. https: // arxiv. org/abs/1506.02025.

[9] Girshick R. Fast R-CNN[C]// IEEE International Conference on Computer Vision (ICCV), 2015: 1440-1448.

[10] Uijlings J R R, van de Sande K E A, Gevers T, *et al*. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104(2): 154-171.

[11] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10)[2018-07-25]. https://arxiv.org/abs/1409.1556.

[12] He K M, Zhang X Y, Ren S Q, *et al*. Deep residual learning for image recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770-778.

[13] Zhu H G, Chen X G, Dai W Q, *et al*. Orientation robust object detection in aerial images using deep convolutional neural network[C]// IEEE International Conference on Image Processing (ICIP), 2015: 3735-3739.

[14] Zhu M M, Xu Y L, Ma S P, *et al*. Airport detection method with improved region-based convolutional neural network[J]. Acta Optica Sinica, 2018, 38(7): 0728001.
朱明明, 许悦雷, 马时平, 等. 改进区域卷积神经网络的机场检测方法[J]. 光学学报, 2018, 38(7): 0728001.

[15] Wei Y M, Quan J C, Hou Y Q Y. Aerial image location of unmanned aerial vehicle based on YOLO v2[J]. Laser & Optoelectronics Progress, 2017, 54(11): 111002.
魏湧明, 全吉成, 侯宇青阳. 基于 YOLO v2 的无人机航拍图像定位研究[J]. 激光与光电子学进展, 2017, 54(11): 111002.

[16] 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.