基于密集连接网络的遥感图像检测方法

杜泽星*，殷进勇，杨建
江苏自动化研究所计算机事业部，江苏 连云港 222002

摘要 针对传统遥感图像检测算法中人为干预多、速度慢、检测精度低等问题，提出一种基于深度学习的遥感图像检测方法。采用密集连接的网络结构，充分利用每层网络提取的特征，减少网络推理时间；采用具有更大感受野的扩张块结构；使用扩张块结构和反卷积网络结构将浅层特征图和深层特征图进行信息融合，从而增强遥感图像中多尺度目标的检测能力。实验结果表明，该检测方法具有更高的准确率和更短的检测时间，尤其在小目标物体的检测上表现出更好的性能。

关键词 图像处理；遥感图像；小目标检测；密集连接网络；特征融合
中图分类号 TP183 **文献标识码** A **doi**：10.3788/LOP56.222803

Remote Sensing Image Detection Based on Dense Connected Networks

Du Zexing*, Yin Jinyong, Yang Jian
Computer Division of Jiangsu Automation Research Institution, Lianyungang, Jiangsu 222002, *China*

Abstract This study proposes a remote sensing image detection method based on deep learning to solve the issues of human intervention, slow speed, and low accuracy associated with the traditional remote sensing image detection algorithm. A dense connected network is considered to completely use the features extracted from each layer and reduce the network inference time. Further, an expanding block structure with a large perceptive field is adopted, and the low- and high-level feature informations of the network are combined based on the expanding block structure and deconvolution network. Thus, the performance of multiscale object detection for remote sensing images is improved. The experimental results denote that the proposed method exhibits high accuracy and short detection time, especially during the detection of small objects.

Key words image processing; remote sensing image; small object detection; dense connected structure; feature fusion
OCIS codes 100.3008; 100.4996; 100.1830

1 引 言

近年来，遥感图像具有受限条件少、效益高、更新周期短、使用范围广等优势，在军事和民用领域取得了广泛的应用。目前，遥感图像在天文探索、农作物和森林检测、军事侦察等领域发挥着巨大的作用。而如何在遥感图像中提取所需要的信息也成为了目前广泛研究的对象，其中包括了恶劣环境下对沼泽、沙漠、山脉等目标的检测，及在军事方面对飞机、船舶、机场等目标的检测。

传统的目标检测算法由于检测精度不足、适用范围小、需要人为大量干预等因素，难以满足遥感图像检测的智能化、快速化的需求。而随着深度学习的快速发展，各种基于深度卷积神经网络(CNN)的目标检测算法在各种检测任务中都取得良好的效果。目前主流的深度学习检测算法有 2 种，包括两阶段检测网络和端到端检测网络。其中两阶段网络结构的典型代表有 Fast-RCNN[1] 和 Faster-RCNN[2] 等，这种结构分为两个部分，第一部分是进行特征提取，利用传统的特征提取算法，在原始图像中提取所需要的特征，第二部分是将提取到的特征输入到卷积神经网络中进行检测和分类。然而，这种检测方法受限于网络结构，无法适用于多尺度多场景的遥感图像目标检测，难以在检测速度上达到

收稿日期：2019-05-06；**修回日期**：2019-05-08；**录用日期**：2019-05-17
* **E-mail**：duzexing@outlook.com

最优性能。因此人们提出将特征提取和目标检测融合为一阶段的端到端网络结构，其中，典型的代表为YOLO[3-5]系列和单步多框检测(SSD)算法[6]的检测网络。基于端到端的识别网络，在目标检测的精度和速度方面都得到了极大的提升。然而，由于遥感图像具有分辨率高、目标尺寸小等特点，上述的网络结构对小目标物体的检测精度较差，因而难以在遥感图像的检测中广泛应用。

近年来，针对多尺度多场景的遥感图像目标检测方法的研究成为热点。其中朱明明等[7]通过特征融合和软判决的非极大值抑制的方式加强了Faster-RCNN对小目标的检测性能，从而较好地实现了对遥感图像中飞机的检测。辛鹏等[8]通过融合多层特征来缓解深层网络对小目标检测不足的问题，然而，所采用的两阶段网络结构导致其检测时间较长。单倩文等[9]通过多尺度融合和轻量级的压缩型双线性融合方法对SSD模型进行优化，然而其选用VOC数据集，所含小目标物体的数量较少，不能很好地评估改进后的模型在小目标物体上的检测效果。陈立里等[10]使用ResNet-34替换SSD的基础网络，提升了网络的检测速度，但这种网络难以满足以小目标为主的遥感图像的检测需求。

为提升遥感图像的多尺度目标的检测精度和神经网络的检测速度，本文设计了基于密集连接网络[11]的端到端网络结构，该结构充分利用每层网络提取的特征，具有很少的参数量和优异的检测速度。通过设计具有更大感受野的扩张块结构来减少下采样所带来的特征损失，并且采用特征融合[12]的思想，将浅层特征通过所设计的扩张块进行下采样，将深层特征通过反卷积进行上采样，融合两个采样特征实现对多类别目标的检测。

2 网络结构

2.1 扩张块设计

图1(a)为一个高效的网络模型，该结构采用多分支不同大小的卷积核。首先使用1×1的卷积来对网络深度进行压缩，之后使用3×3的普通卷积和膨胀率为2的膨胀卷积来增大感受野，如图1(b)所示，这种结构的设计会使对中心区域的检测增强，并且没有过多增加网络参数，从而使后续设计的网络仍然满足实时性的要求。每个卷积都伴随一个Batch Normalization(BN)和ReLU激活函数。

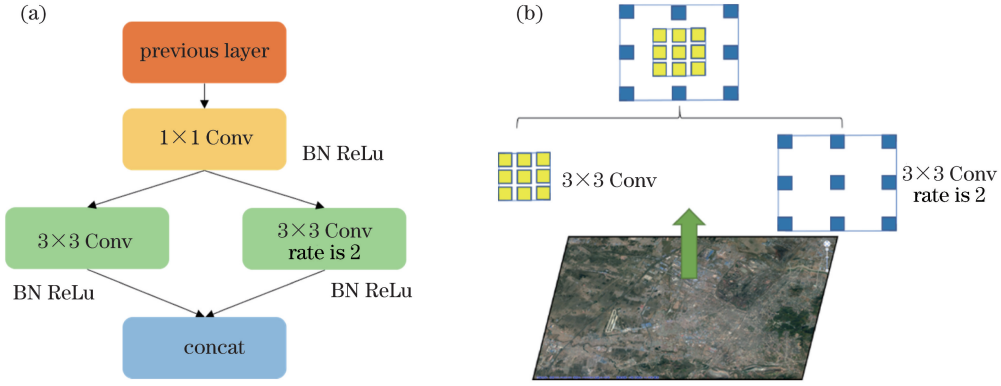

图1 扩张块设计。(a)扩张块结构图；(b)扩张块的感受野

Fig. 1 Design of expanding block. (a) Structure of expanding block; (b) receptive field of expanding block

2.2 密集块结构

密集块的结构参考文献[11]，如图2所示。在单个密集块中，每层网络都和在其之前的所有网络层进行连接。这种连接方法很好地解决了在训练过程中梯度消失的问题，这是因为这种密集连接方式使得每一个网络层都直接连接输入和损失。并且采用这种连接方法使得每层网络所提取到的特征得到了充分的复用，提升了网络效率，从而减少网络的参数量。其运算过程[11]为

$$x_l = H_l([x_0, x_1, \ldots, x_{l-1}]), \quad (1)$$

式中，$x_0, \ldots, x_{l-1}$代表前$l-1$层连接，H_l代表第l层的非线性转换函数，其中包括BN，ReLU，Conv操作。

所使用的密集网络结构中，层与层之间的连接除卷积等操作外，还加入1×1的卷积核来减少卷积的深度，这是由于在经过多次连接后，网络的深度还会变得很深，因此引入1×1的瓶颈层(Bottleneck)来达到降低维度的作用。在每个密集块结束后采用过渡层来进行块与块之间的连接，该结构除使用1×1卷积减少网络深度外，还采用池化操作进行下采样来减少参数数量。

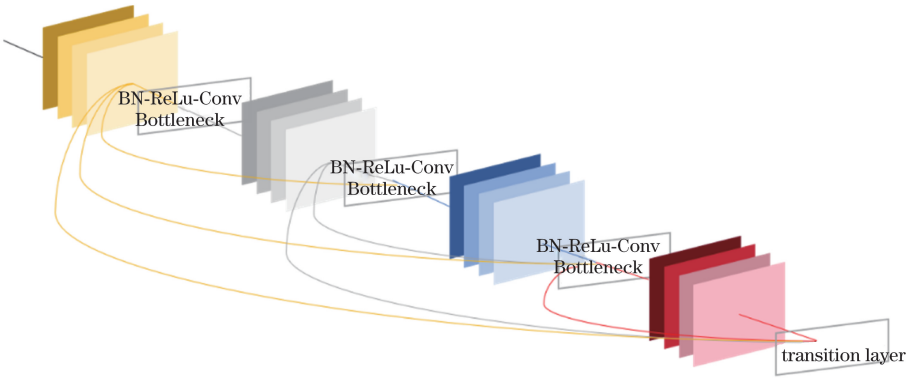

图 2 密集连接网络结构

Fig. 2 Structure of dense connected network

2.3 网络结构

图 3 为本文所采用的网络结构，将图像经过预处理后输入到多个密集块所连接的网络中。由于浅层网络拥有更加精准的位置信息，有利于目标的定位，深层网络拥有更加丰富的语义特征，有利于目标的分类[12]。遥感图像中所含目标的尺寸一般较小，用传统的方法难以进行检测。因此，通过设计的扩展块对浅层网络进行下采样，这种采样方法相比于传统的池化采样具有更大的感受野，对小目标更加友好。对深层网络采用反卷积的方法进行上采样，这种采样方法具有可训练的参数，可以更好地保留所提取的语义信息。

在进行网络预测时，采取文献[2]中所提到的 anchor boxes 方法。Anchor boxes 的设置对网络的精度和速度有着很深的影响，为了能选择合适大小和比例的先验框，采用 K-means 算法，对训练数据集的边框进行聚类，并且选取 5 个聚类中心作为先验框。在经过聚类后，得到的先验框大小为 30×30，20×40，40×20，53×67，67×53。

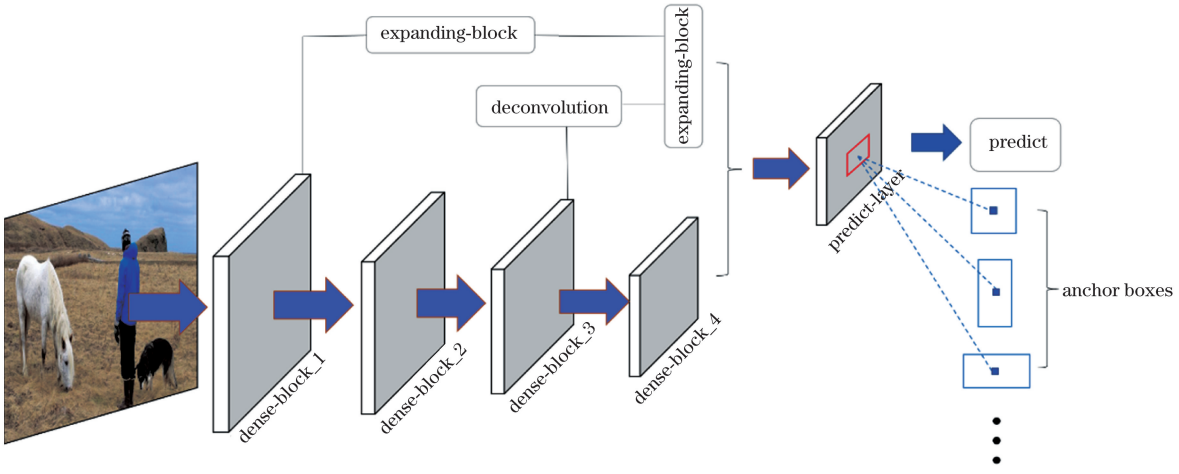

图 3 网络结构

Fig. 3 Structure of network

2.4 损失函数

在网络的预测过程中，直接预测边框相对于先验框的偏移量，会导致模型不稳定，尤其是在刚开始训练时。因此采用和文献[4]中类似的处理方法，取预测边框中心对每个 cell 左上角位置的相对偏移值，并使用 Sigmoid 函数将坐标预测归一化处理，坐标表达式分别为

$$b_x=\sigma(t_x)+c_x, \tag{2}$$

$$b_y=\sigma(t_y)+c_y, \tag{3}$$

$$b_w=p_w e^{t_w}, \tag{4}$$

$$b_h=p_h e^{t_h}, \tag{5}$$

式中：b_x，b_y，b_w，b_h 代表边框坐标和长宽的实际预测结果；t_x，t_y，t_w，t_h 代表每个边框的预测值；c_x，c_y 代表 cell 左上角相对于图片左上角的距离；p_w，p_h 代表 anchor boxes 的宽和高；$\sigma(x)=\frac{1}{1+e^{-x}}$ 代表 Sigmoid 函数。

采用的损失函数参考 YOLO v2[4] 的定义，其中

包含边界框坐标损失、边界框尺寸损失、检测目标的类别损失、置信度损失 4 部分，可表示为

$$x_{\text{loss}}=\lambda_{\text{coord}}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{\text{obj}}\left[(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2\right]+\lambda_{\text{coord}}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{\text{obj}}\left[\left(\sqrt{w_i}-\sqrt{\hat{w}_i}\right)^2+\left(\sqrt{h_i}-\sqrt{\hat{h}_i}\right)^2\right]+$$
$$\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{\text{obj}}(C_i-\hat{C}_i)^2+\lambda_{\text{noobj}}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{\text{obj}}(C_i-\hat{C}_i)^2+\sum_{i=0}^{S^2}1_{ij}^{\text{obj}}\sum_{c\in\text{classes}}(p_i(c)-\hat{p}_i(c))^2,\tag{6}$$

式中：λ_{coord} 代表边框损失所占的权重，取 5；λ_{noobj} 代表不含目标的置信度损失权重，取 0.5；S^2 代表网格数量；B 代表先验框的数量；x_i，y_i，w_i，h_i 代表边框的预测结果；$\hat{x}_i$，$\hat{y}_i$，$\hat{w}_i$，$\hat{h}_i$ 代表边框标记的真实坐标信息；C_i 代表检测目标的置信度；$\hat{C}_i$ 代表标记样本的置信度；p_i 代表预测类别；$\hat{p}_i$ 代表标记样本的类别。

3 网络训练

3.1 预训练

采用自标记的 3391 张遥感数据集，若使用设计好的网络直接进行训练，容易由于数据规模小而造成过拟合，因此采用预训练-微调的方法对网络进行训练。由于所使用的遥感数据集的目标尺寸较小，为使预训练模型具有较好的小目标提取能力，因此在预训练时采用同样含有较多小尺寸的 Cifar10 数据集。

Cifar10 是一个含有 6 万张分辨率为 32×32 彩色图像的分类数据集。因此在预训练网络时，用全连接网络和 Softmax 回归代替图 3 中的边框预测部分，所使用的网络如图 4 所示。在预训练时由于是分类任务，所以并没有使用扩张块和上采样。在层之间使用瓶颈层，块之间使用过渡层来减少网络参数。另外在瓶颈层和过渡层中都添加 dropout 层，防止数据集训练过程中出现过拟合问题。在设计好网络结构后，使用 Cifar10 数据集进行 300 轮的训练得到预训练模型。

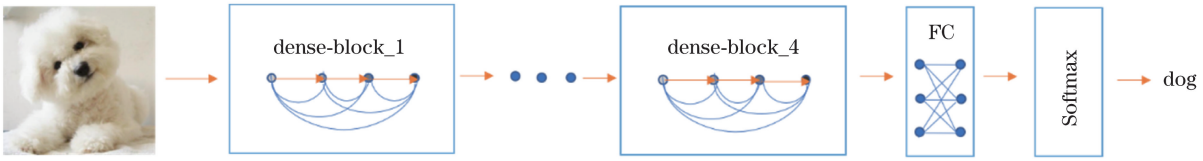

图 4 预训练网络模型

Fig. 4 Structure of pre-trained network

3.2 微调

在预训练之后，将得到的密集网络参数加载到图 3 的模型中进行训练，没有经过预训练的权重则采用高斯分布进行随机初始化。实验采用的电脑配置 CPU 为 E5-2650v2 2.6 GHz，GPU 为 Tesla P4，深度学习框架为 Tensorflow。数据集中各类别的目标数 Airplane 为 6681、Airport 为 479、Boat 为 3825、Bridge 为 1102、Harbor 为 961、Oilcan 为 2210。图像均拍摄于不同的地点，拍照时的光照强度和天气等因素各不相同，这也增强了训练模型的稳健性，可以适应各种环境情况。数据集中，Airport 和 Harbor 类别的每张图片仅含有一个目标，Airplane、Boat、Oilcan 类别的每张图片含有的平均目标数为 8，Bridge 类别的每张图片平均含有的目标数为 2。

在实验中，对所有训练数据集采用开源的标注方法，将目标的位置坐标生成为 XML 格式的数据标签，用数据集中的 70% 作为训练数据集，其余 30% 作为测试数据集。部分数据集如图 5 所示，将目标尺寸按照表 1 的标准划分为大、中、小 3 种，通过评价网络在 3 种尺寸上的表现来综合评价网络的性能。由表 1 可以看出，本次实验所用的数据集中，小目标的占比较高，因而可以很好地评估模型在小目标检测上的表现。

表 1 目标尺寸的划分标准

Table 1 Object size division standard

Area	$(0,32^2)$	$[32^2,96^2]$	$(96^2,\infty)$
Classes	small	medium	large
Percentage /%	28	32	40

在训练时使用数据增强对训练样本进行扩充，采用的方法有随机翻转、旋转、调整图像色彩、饱和度等来增强网络对各种角度和色彩目标的识别效果。训练过程中，采用动量梯度下降法对损失值进行优化，用指数式衰减的学习率设置方法，初始学习率设置为 0.001，训练迭代次数为 20000 次。

图 5 将目标划分为大,中,小三个尺寸

Fig. 5 Divide the target into large, medium, and small sizes

4 结果讨论

经过微调后得到训练好的网络模型,然后在测试数据集上对所得到的模型进行测试,部分检测结果如图 6 所示。采用 mAP 指标对网络模型进行评估,该指标综合考虑每个类别的准确率和召回率,是评价神经网络模型常用指标。

图 6 部分检测结果

Fig. 6 Partial detection results

设计一个实验来评估网络所含密集块的数量对网络检测性能的影响。如表 2 所示,在相同网络深度的前提下,改变网络密集块数量得到网络在不同大小的目标上的检测精度和检测时间。随着密集块数量的增加,网络的推理时间在不断减少,这是因为在相同网络深度的情况下,密集块数量的增加导致单个密集块中的网络层数减小,从而导致密集块中的连接复杂度降低,缩短了检测时间。此外密集块数量的增加也会导致网络下采样的次数增加,密集块数量由 4 增加到 5 时,下采样的倍数也由 16 增加到 32,而这种下采样倍数的增加使得网络对原本尺寸小于 32 的目标检测效果变差,从而也就造成在相同网络深度的条件下,网络对小目标的检测性能变差。因此在此后的测试中,综合

表 2 相同网络深度条件下密集块数对检测结果的影响

Table 2 Effect of number of dense blocks on detection results under the same network depth

Num-block	Large / %	Medium / %	Small / %	mAP / %	Time / s
3	84.58	80.28	78.43	81.48	0.014
4	83.67	79.34	75.90	80.11	0.010
5	84.25	57.41	50.65	66.25	0.008

考虑检测精度和检测时间,选取密集块数为 4 的网络模型。

表 3 通过改变每次卷积输出的特征层数(growth-rate)和网络深度来综合衡量网络的参数数量、推理速度及 mAP 的表现。由实验结果可以看出,随着特征层数和网络深度的增加,网络的检测性能也在不断提升,这是由于随着网络参数量的增加,网络提取特征的能力也在增强,但是网络的检测速度也在下降。而当这两个因素增加到一定值时,网络的检测能力增加缓慢,而网络的推理速度却在大幅降低。

表 3 密集块数均为 4 时,改变特征层数和网络深度得到的实验结果

Table 3 Experimental results obtained by changing the number of feature layers and network depth when number of dense block is 4

Growth-rate	Depth	Params /10^6	Time /s	mAP /%
12	40	0.1	0.010	80.11
12	100	0.6	0.021	85.99
24	40	0.6	0.020	85.24
24	100	2.4	0.070	86.59
40	190	22.6	0.234	87.72

为对比本实验所设计的网络与主流网络在小目标物体上的检测性能,选取主流目标检测神经网络 YOLO v3 和 SSD 进行对比。在相同的实验环境下,采用指数式衰减的学习率设置方法和动量梯度下降法对损失值进行优化,初始学习率设置为0.001,在经过 20000 轮的训练后,网络的损失值下降到 0.2 左右并保持,网络参数接近稳定,达到最优。

在对比实验中,选取数据集 Airplane 类别进行测试。通过召回率和准确率这两个指标来对检测效果进行综合评估,分别表示为

$$R=\frac{X_{\mathrm{TP}}}{X_{\mathrm{TP}}+X_{\mathrm{FN}}},\tag{7}$$

$$P=\frac{X_{\mathrm{TP}}}{X_{\mathrm{TP}}+X_{\mathrm{FP}}},\tag{8}$$

式中,X_{TP}代表检测结果中被正确检测出来的目标,X_{FN}表示未被检测到的目标,而 X_{FP}表示被误检的

目标。

如表 4 所示，其中 Ours-40 代表设计的特征层数为 12，网络深度为 40 的网络结构；Ours-100 代表设计的特征层数为 12，网络深度为 100 的网络结构。通过对比可以看出 SSD 和 YOLO v3 作为主流的神经网络，由于其网络结构的限制，二者在小尺寸目标的数据集上表现较差，检测准确率和召回率均低于本文所设计的网络。

表 4 不同检测方法的检测效果对比

Table 4 Comparison of detection results of different algorithms

Algorithm	X_{TP}	X_{FP}	X_{FN}	P	R
SSD	5211	943	527	0.90	0.85
YOLO v3	5612	694	376	0.93	0.89
Ours-40	5652	639	390	0.94	0.90
Ours-100	6013	409	259	0.96	0.93

此外，还验证扩张块和反卷积结构对小目标检测的提升效果，如表 5 所示，Densenet-40 表示只有密集网络，不包含扩张块等部分的 40 层网络结构，Ours-40 代表经过优化后的 40 层网络结构。可以看出，所设计的优化方法使得密集网络对小目标的检测性能提升了 10.98%，并且依然保持着很短的检测时间。

由此可知，本文所设计的网络结构相比于主流的目标检测算法有着更加准确的检测率和更快的推理速度。这是因为网络主体框架采用密集连接的网络结构，这种结构充分利用每层网络所提取到的特征，网络参数也因此减少。通过扩张块和反卷积结构对深层网络结构和浅层网络结构进行融合，优化了网络对小目标的检测效果，并且由于采用高效的网络结构，推理时间并没有太多的增加。

表 5 网络改进效果

Table 5 Improvement effect of network

Algorithm	mAP_{large} /%	mAP_{medium} /%	mAP_{small} /%	mAP /%	Time /s
Densenet-40	82.37	78.12	64.92	76.12	0.008
Ours-40	83.67	79.34	75.90	80.11	0.010

5 结 论

以密集连接网络为主体框架，并通过扩张块结构和反卷积结构将网络的深层信息和浅层信息进行融合，优化网络结构。首先使用 K-means 聚类算法得到候选框，设计了边框的处理方法和损失函数，在 Cifar10 数据集上对网络进行训练。将目标分为大，中，小 3 种尺寸，在自标记的遥感数据集对网络进行微调和测试。实验表明，所设计的网络结构最高具有 87.72%的 mAP 值，检测精度高于其他网络，并且在网络深度值较小时，在没有大幅度损失精度的前提下具有最高的检测速度，性能远优于现有的其他检测算法，对遥感图像检测具有很好的理论和实际意义。

参 考 文 献

[1] 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.

[2] 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.

[3] 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.

[4] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 6517-6525.

[5] Redmon J, Farhadi A. YOLOv3: an incremental improvement[J/OL]. (2018-04-08)[2019-05-06]. https://arxiv.org/abs/1804.02767.

[6] 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. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.

[7] Zhu M M, Xu Y L, Ma S P, *et al*. Airplane detection based on feature fusion and soft decision in remote sensing images[J]. Acta Optica Sinica, 2019, 39(2): 0210001.
朱明明，许悦雷，马时平，等. 基于特征融合与软判决的遥感图像飞机检测[J]. 光学学报，2019，39(2)：0210001.

[8] Xin P, Xu Y L, Tang 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.

[9] Shan Q W, Zheng X B, He X H, *et al*. Fast object detection and recognition algorithm based on improved multi-scale feature maps [J]. Laser & Optoelectronics Progress, 2019, 56(2): 021002.
单倩文, 郑新波, 何小海, 等. 基于改进多尺度特征图的目标快速检测与识别算法[J]. 激光与光电子学进展, 2019, 56(2): 021002.

[10] Chen L L, Zhang Z D, Peng L. Real-time detection based on improved single shot multibox detector[J]. Laser & Optoelectronics Progress, 2019, 56(1): 011002.
陈立里, 张正道, 彭力. 基于改进SSD的实时检测方法[J]. 激光与光电子学进展, 2019, 56(1): 011002.

[11] Huang G, Liu Z, van der Maaten L, *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.

[12] 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.