场景耦合的空对地多任务遥感影像智能检测算法

刘星, 陈坚, 杨东方*, 贺浩
火箭军工程大学导弹工程学院, 陕西 西安 710025

摘要 在空对地遥感检测中,目标所占视场比例小、视角单一、易受背景干扰且视场高度变化大,这给传统深度学习检测算法带来了挑战。针对该问题,提出一种场景耦合的多任务目标检测算法。首先,设计了一种新的场景耦合目标检测网络结构,将场景分类特征图和目标检测特征图在同一尺度上进行镜像融合,丰富了网络特征描述的细粒度;其次,设计了差异化激活模块,实现特征通道的重要性筛选;然后,推导了多任务耦合的网络优化函数,实现了目标检测损失和场景分类损失的同步优化;最后,建立了空对地目标检测多任务数据集,对所提方法的有效性进行验证。实验证明,本文算法有效提升了空对地小目标检测的精度和稳健性,同时能够自适应不同高度的识别检测多任务需求,为空基无人平台对地智能检测提供了新的思路和方法。

关键词 机器视觉; 多任务耦合; 深度学习; 目标检测; 场景感知; 空基无人平台
中图分类号 TP751.1 **文献标识码** A **doi**: 10.3788/AOS201838.1215008

Scene-Coupled Intelligent Multi-Task Detection Algorithm for Air-to-Ground Remote Sensing Image

Liu Xing, Chen Jian, Yang Dongfang*, He Hao
Missile Engineering College, Rocket Force University of Engineering, Xi'an, Shaanxi 710025, *China*

Abstract In air-to-ground remote sensing detection, the object has the characteristics of small field of view and single viewing angle, which is susceptible to background interference. At the same time, the height of the field of view varies greatly, which brings challenges to the traditional deep learning detection algorithm. To solve the problem, a scene-coupled multi-task object detection algorithm is proposed. First, a new scene-coupled object detection network structure is designed, which mirrors and fuses the scene classification feature map and the object detection feature map on the same scale to enrich the fine-grain of the feature description. Second, a differentiated activation module is designed to realize the importance screening of feature channels. Then, the optimization function of multi-task coupling is derived, which can simultaneously optimize the scene classification loss and object detection loss. Finally, an air-to-ground detection multi-task dataset is established to verify the effectiveness of proposed method. The experimental results show that the proposed algorithm effectively improves the accuracy and robustness of air-to-ground small object detection, and can adapt to different heights to identify multi-task requirements, which provides a new idea and method for space-based unmanned platform intelligent detection.

Key words machine vision; multi-task coupling; deep learning; object detection; scene understanding; unmanned aerial vehicle
OCIS codes 150.1135; 110.4155; 150.1135; 150.015

1 引 言

近年来,空基平台对地感知技术因在典型军事和民用领域具有重要价值而备受关注,其中对地目标检测是空基无人平台遂行侦察监视等典型任务的一项共性关键技术。

在前深度学习时代,目标检测和识别任务通常建立在对目标的人工描述基础上,通过提取目标的关键特征,如灰度、轮廓、纹理等具体特征,或尺度不变特征(SIFT)[1]、方向梯度直方图(HOG)[2]、快速稳健特征(SURF)[3]等抽象特征,完成目标的完备描述。然后采用支持向量机(SVM)等机器学习方

收稿日期: 2018-06-28; **修回日期**: 2018-07-19; **录用日期**: 2018-08-07
基金项目: 国家自然科学基金(61673017)、陕西省自然基金(2017JM6077)、陕西省重点研发计划项目(2018ZDXM-GY-039)
* **E-mail**: yangdf301@163.com

法，完成不同类型目标的分类检测或者不同目标个体的识别。这类方法简单易行，在处理常规的简单目标检测任务中，可以取得不错的效果。然而，所有人工设计特征因其维度低、目标描述细粒度不足，在复杂背景或细微差异的目标检测识别应用过程中难以获得满意的效果。近年来，深度学习的兴起为图像目标的描述提供了一种抽象有效的方式。深度神经网络由于在目标特征高维度表示方面具有独特的优势，因而可以很方便地提取更高细粒度特征。短短几年的发展，AlexNet[4]、VGGNet[5]、GoogLeNet[6]、ResNet[7]和MobileNets[8]等特征提取网络相继被提出，也为利用深度学习完成目标检测任务奠定了基础。在特征提取网络的基础上，基于深度学习的目标检测技术的发展经历了两个阶段：基于区域建议的两步检测算法和基于预选边框回归的一步检测算法。前一阶段以区域卷积神经网络(RCNN)[9]为代表，随后出现的Fast R-CNN[10]也属于这一类方法，而Faster RCNN[11]在RCNN基础上对Fast RCNN的特征图提取机制进行改进，同时将多任务损失优化与区域建议网络(RPN)结合，实现了端到端的检测网络，能够获得优于RCNN的目标检测实时性。为了进一步提高目标检测的实时性，在RCNN框架外提出了YOLO(You Only Look Once)[12]算法，该方法摒弃了RCNN的区域建议过程，通过对随机生成的目标框进行回归计算，实现了端到端的快速目标检测，进一步提高了目标检测算法的实时性。然而，受到其网络结构的制约，典型的YOLO算法在进行小目标检测时遇到了困难。SSD(Single Shot MultiBox Detector)[13]算法的出现，克服了YOLO算法的先天缺陷，有效提升了小目标检测性能。典型的SSD检测网络可以划分为两部分：前端的特征提取网络和后端的目标检测网络。在目标检测网络端，SSD的目标检测网络将Faster RCNN网络的锚点思想和YOLO的回归思想结合，分别在6个不同尺度特征图上产生先验框来进行预测。由于不同特征图经过卷积操作后的感受野不同，将不同尺度的特征图进行融合可以实现更多尺度的目标检测。

随着基于深度学习的目标检测技术研究日益升温，传统视场下的目标检测方法日趋成熟。然而，在空基无人平台对地检测识别领域，由于空对地平台视距远、视角单一，从空基平台上对目标成像通常会出现目标尺寸小、目标信息片面且容易受到遮挡等问题，这对于利用深度学习进行目标检测而言是严峻挑战，也是研究热点问题[14-16]。

事实上，人类在通过空对地视角观测地面时，不仅利用目标区域的信息进行目标描述，还充分利用了目标周围场景的信息、飞行高度的信息，乃至地标建筑物、道路网络等视场中可以看到的所有信息，而深度学习算法本身具有极高的场景感知能力[17]。受此启发，本文模拟了人类对地场景的认知过程，将场景的理解和目标的检测两个任务耦合到一起，建立了多任务耦合的深度学习目标检测网络框架，同时完成场景分类理解和目标检测的任务。因此，将该方法称为场景耦合的多任务目标检测算法，需要说明的是，此处的场景耦合不是简单场景分类结果的外部辅助，而是神经网络场景图像中所包含的所有信息和目标检测网络训练过程的同步耦合。

2　场景耦合的多任务目标检测算法

本文选择在目标检测领域取得重要成功的端到端目标检测框架SSD系列作为空对地耦合检测任务的基础模型，通过对SSD系列算法引入场景耦合的思想，提出了一种新的空对地目标检测算法。

2.1　空对地检测问题的网络化描述

对于空对地检测任务，受到观测高度影响，通常不能直接获得目标本身信息，而是需要对目标所在场景进行动态感知，例如军事目标侦察过程中军用机场等高价值场景搜索任务，因此多任务模型应运而生，多任务模型也是目前各类图像应用领域研究的热点问题[18-20]。该模型不仅能实现多个任务功能，而且其网络可以从多个任务中学习到不同任务共享和广义的特征表示，通过在不同任务中学习共享信息，从而提供更好的预测。

空对地目标检测区别于常规视场的一个显著特点是视角尺度不确定，甚至在工作过程中会发生剧烈变化，例如高空侦察无人机或遥感卫星，文献[14]通过栅格化大尺度高分辨率遥感图，将大图变为多个小图像来进行车辆检测。但该方法只能静态地处理固定尺度的目标信息，难以适应飞行器视场高度剧烈变化引起的任务变化。事实上，空对地识别任务通常存在高空大尺度场景与低空小尺度场景之间的过渡转换(图1)，因此检测视角需要具有明显的空间变分辨率特性。在执行空基对地侦察监视等感知任务时，对场景的感知同样具有重要的应用价值，一方面在较高空域可以利用序贯图像进行目标所属环境位置感知，逐步引导进行目标检测；另一方面通过对目标所在

环境进行感知，可以分析目标-场景的相关性，为执行后续任务提供判别依据。这也是本文提出场景耦合目标检测算法的初衷。

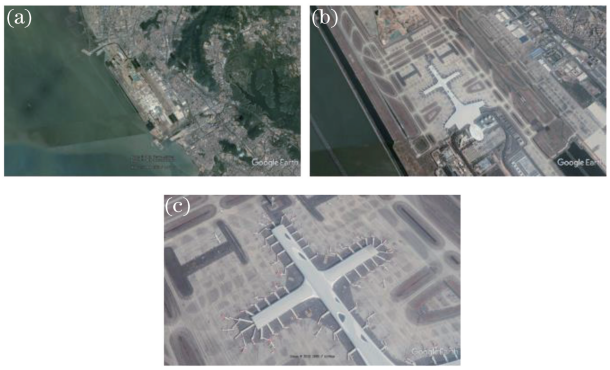

图 1　空对地变分辨率场景。(a)高空视场；(b)中空视场；(c)低空视场

Fig. 1　Air-to-ground variable resolution scene. (a) High-altitude vision; (b) middle-altitude vision; (c) low-altitude vision

当前 SSD 系列在小目标检测方面表现较好的方法是 FSSD(Feature Fusion Single Shot MultiBox Detector)模型[21]，该模型引入特征金字塔网络(FPN)[22]思想，将原 SSD 的 conv4_3、FC7、conv7_2(以 VGG16 为例)三个不同特征图进行串联作为目标检测网络的输入，以此丰富由于多次卷积操作而缺失的特征图细粒度，如图 2 所示。

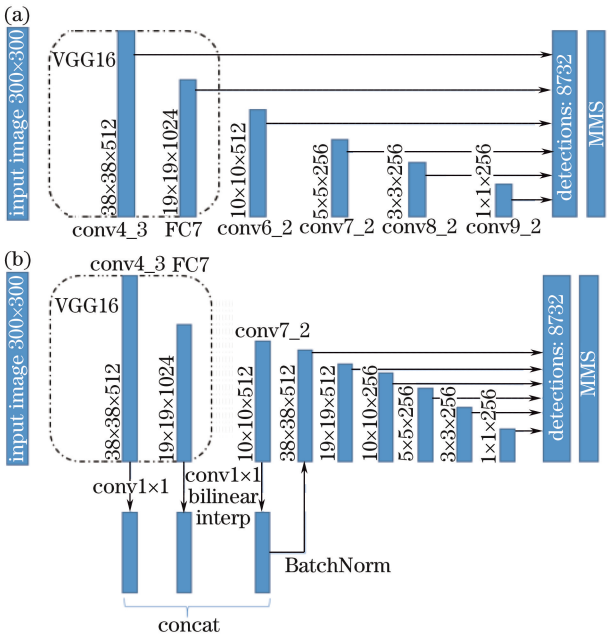

图 2　模型图。(a) SSD 模型；(b) FSSD 模型

Fig. 2　Models. (a) SSD model; (b) FSSD model

本文提出的场景耦合目标检测算法是基于 FSSD 架构的，将场景分类网络和目标检测网络进行一体化设计，并引入不同通道差异化激活机制，通过场景和目标特征的耦合，提高目标检测的精度，具体网络结构如图 3 所示。图中，左侧为输入图像，此处以 300×300 大小的图像为例，首先经过 FSSD 将 VGG16 特征提取网络的 conv4_3、FC7、conv7_2 层特征图通过通道串联后，进行归一化和一次 3×3 卷积运算，形成通道数为 512、大小为 38 pixel×38 pixel的融合特征图。上述模型中，检测网络部分仍然采用 SSD 基本结构，通过多次 3×3 卷积运算，生成大小不同、感受野不同的 6 个检测特征图。场景分类网络同样将 38 pixel×38 pixel 的融合特征图经过多次 3×3 卷积运算，与检测网络形成一种镜像对称结构，这样便于从相同尺度特征图中同时学习场景和目标的特征。在场景特征图和目标检测网络特征图融合方面，本文选择通道串联的特征图融合方式，因为场景和目标属于两种不同的语义层次，直接对其进行求和，容易破坏各类模态语义信息的内部拓扑关系。而通道串联方式能够同时保留场景和目标这两类不同模态语义特征信息，实现视场图像的分层理解。该网络中，检测网络模型和场景分类模型组成了镜像结构，可以使得目标检测和场景分类在进行特征融合时，两种任务对应位置的特征图具有相同的感受野和深度信息，方便进行两种特征描述的融合。

2.2　多模态特征图差异化激活模型

将场景和目标特征图融合后得到的双倍通道特征图中存在冗余的特征信息，不满足网络结构的正则化需求，也会在后续网络梯度优化计算过程中引入特征干扰。对此，本文借鉴文献[23]的思路，提出了通道信息激活(IA)模块，实现场景和目标特征图通道的筛选。该模块在网络结构中引入激活模块参数，利用激活参数的学习结果实现通道权重的调节、冗余信息的筛选和通道差异化选择。

在 IA 模式的选择方面，主要有同步激活和异步激活两种，两者的差异体现在如何对待场景和目标这两种模态的特征图通道方面。为了对两种激活模式进行对比，本文同时对其进行研究，这两种激活方式如图 4 所示，图 4(a)为场景特征通道和目标检测通道的同步激活，图 4(b)为异步激活。

网络中 IA 模块工作的过程是先进行特征图压缩操作，利用全局平均池化将特征图多个通道的三维张量信息压缩(F_{squeeze})成一维向量信息，即将每一个二维单通道特征图变成一个实数，这个实数可以近似表征该全局感受野的信息。其次是深度特征信息的训练过程，采用全连接层学习参数 w 来建模

特征通道间的关联性。最终网络通过学习到的 w 参数为每个特征通道生成权重，并进行通道的差异化激活标定（F_{scale}）。这种场景和目标模态特征图激活模块对应图 3 中目标检测网络的右上角部分。

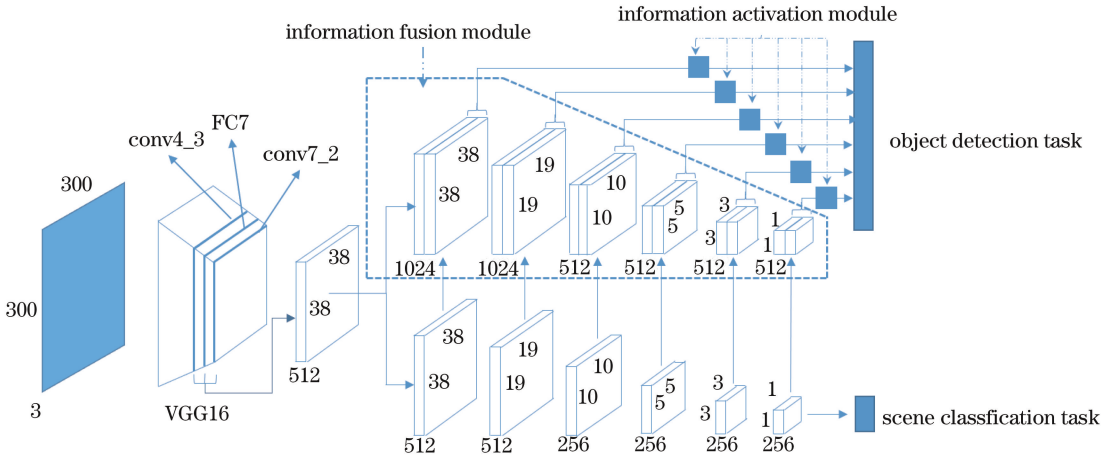

图 3　场景耦合的多任务目标检测模型

Fig. 3　Scene-coupled multi-task object detection model

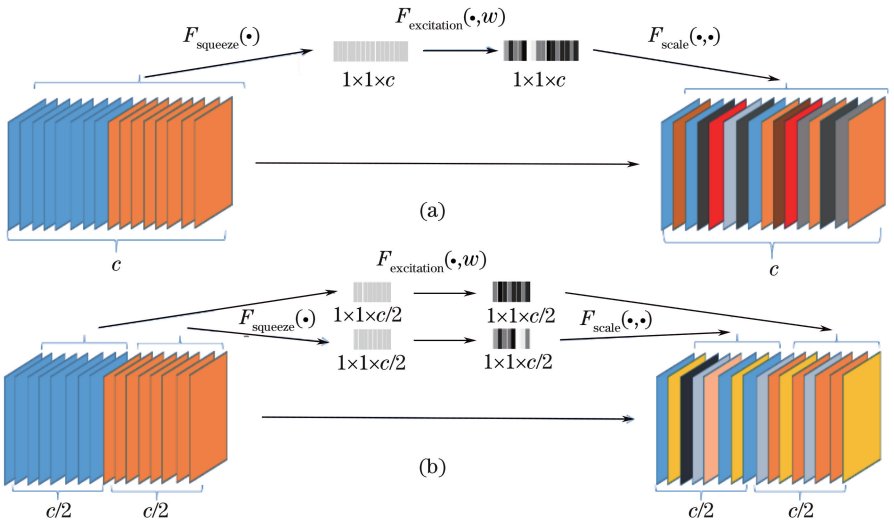

图 4　IA 模块。(a)同步激活；(b)异步激活

Fig. 4　Information activation module. (a) Synchronous activation; (b) asynchronous activation

2.3　多任务损失函数

场景耦合目标检测模型中涉及到场景分类和目标检测两个任务，因此，对应的损失函数也和传统的目标检测网络损失函数有所不同。在目标检测网络中，典型的损失函数，例如在 SSD 框架下，将建议框的位置和目标类别的置信度作为损失函数，记 $x_{ij}^{p}\in\{1,0\}$ 表示类 p 第 i 个默认框匹配到第 j 个标记框的指示函数，满足 $\sum_{i}x_{ij}^{p}\geqslant 1$。目标函数包含坐标框的定位损失和目标分类的置信度损失两部分，此时目标函数可以描述为

$$L_{\text{dec}}(x,c,l,g)=\frac{1}{N}[L_{\text{conf}}(x,c)+\alpha L_{\text{loc}}(x,l,g)], \tag{1}$$

式中：N 是匹配到默认框的数量，如果 $N=0$，设置损失为 0；c、g 是标注的类标签和边框坐标参数；l 是预测的类标签和边框参数；α 是权衡系数。坐标框的定位损失函数定义为

$$L_{\text{loc}}(x,l,g)=\sum_{i\in P_{\text{pos}}}^{N}\sum_{m\in\{cx,cy,w,h\}}x_{ij}^{k}\times \text{smooth}_{L1}(l_i^m-\hat{g}_j^m), \tag{2}$$

$$\hat{g}_j^{cx}=\frac{g_j^{cx}-d_i^{cx}}{d_i^{w}},\hat{g}_j^{cy}=\frac{g_j^{cy}-d_i^{cy}}{d_i^{h}},$$

$$\hat{g}_j^{w}=\ln\left(\frac{g_i^{w}}{d_i^{w}}\right),\hat{g}_j^{h}=\ln\left(\frac{g_i^{h}}{d_i^{h}}\right), \tag{3}$$

其中，以文献[10]提出的 smooth_{L1} 范数作为边框的回归学习方式，其定义为

$$\text{smooth}_{L1}(x)=\begin{cases}0.5x^2, & \text{if}\ |x|<1\\ |x|-0.5, & \text{otherwise}\end{cases}。\quad(4)$$

目标分类损失采用 softmax 损失函数描述，即

$$L_{\text{conf}}(x,c)=-\sum_{i\in P_{\text{pos}}}^{N}x_{ij}^{p}\ln(\hat{c}_i^{p})-\sum_{i\in N_{\text{neg}}}\ln(\hat{c}_i^{0}),\quad(5)$$

式中：$\hat{c}_i^{p}=\dfrac{\exp(c_i^{p})}{\sum_{p}\exp(c_i^{p})}$；$\hat{c}_i^{0}$ 表示背景 softmax 概率。场景分类目标函数同样采用 softmax 损失函数，即

$$L_{\text{sce}}=-y_i\ln\left[\frac{\exp(s_i)}{\sum_i\exp(s_i)}\right],\quad(6)$$

式中 y_i 是标记类别，s_i 是网络场景预测类别输出。最终得出场景辅助多任务耦合检测模型的损失定义，ε 为检测任务和场景分类任务的权衡系数，满足如下关系：

$$L=L_{\text{dec}}(x,c,l,g)+\varepsilon L_{\text{sce}}。\quad(7)$$

本文将上述损失函数作为整个耦合网络的损失函数，可以看出，该函数同时考虑了场景分类任务和目标检测任务，两种任务对应的网络参数是同时进行优化计算的，因此，两者可以在利用样本进行学习和训练的过程中，完成场景和目标之间耦合关系的描述。

3　实验验证

3.1　场景耦合多任务数据集

在目标检测任务中，当前基于深度学习的目标检测数据集如 Pascal VOC、COCO 等通常以常规视场为主，主要利用手机、相机等工具进行拍摄，特点是数量多、视角全，为卷积神经网络特征学习提供了巨大训练资源。而空对地数据集的搜集需要特殊采集工具（例如无人机、遥感卫星等），这导致数据样本少，样本质量差异化大，也给利用深度学习进行空对地目标检测增加了难度。为此，利用深度学习进行目标检测首先需要建立空对地目标数据集，本文通过公开数据集[24]、谷歌地球（Google Earth）、无人机航拍等不同手段进行了场景辅助多任务检测数据的搜集工作。空对地目标数据集划分为场景-目标检测数据集和遥感场景分类数据集两类，其中场景-目标数据集主要来自公开数据集和谷歌地球，在该数据集上进行了目标和场景的多任务标注，即同一个样本，同时标注其场景类型和目标的位置类别信息，将场景和目标进行关联；低、中、高空变尺度遥感场景数据集来自谷歌地球，其特点是具有多尺度变换特性。结合空对地目标特点和应用需求，对数据集进行了预处理，预处理主要分为数据集尺度大小统一和数据集增强。数据尺度统一到 300 pixel×300 pixel～500 pixel×500 pixel，数据增强主要通过对比度、亮度的调节来模拟不同光照，同时空对地目标具有先天的旋转不变性，因此将图片进行多角度旋转。本文所涉及的场景-目标数据集总数为 3000，包括城镇、机场、水域三类场景，而目标涵盖小型汽车、卡车、飞机、船舶等。遥感场景分类数据集搜集了谷歌地球上不同视场高度的城市遥感图像，类别划分与场景-目标数据集场景类别一致，包括城镇、机场、水域三类场景，每类各 1000 张，数据集范例如图 5 所示。最终将制作的 3000 张场景-目标数据集和 3000 张遥感场景数据集分别按照 7∶3 的比例随机划分为训练集和验证集。

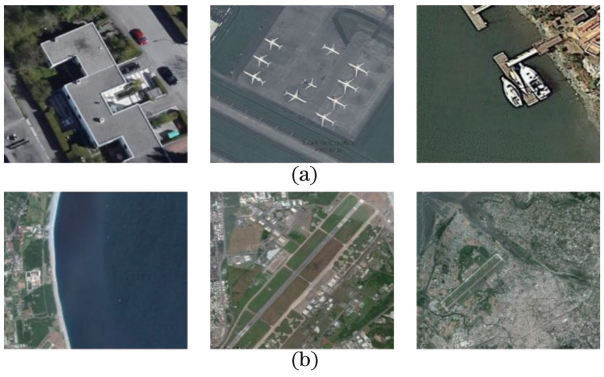

图 5　场景辅助多任务数据集。(a)场景-目标数据集；(b)遥感场景数据集

Fig. 5　Scene-assisted multi-task datasets. (a) Scene-object datasets; (b) remote sensing scene datasets

3.2　场景耦合目标检测实验

由于本文研究场景分类数据集类别相对较少，因此将场景辅助多任务耦合检测模型的损失权衡系数 ε 设置为 0.3，最终的损失函数为

$$L=L_{\text{dec}}(x,c,l,g)+0.3L_{\text{sce}}。\quad(8)$$

针对多目标类别检测识别，实验验证采用 Pascal VOC2010[25]评价体系，该评价体系在深度学习领域具有广泛的通用性和参考性[9-13]，主要有精确率、召回率、平均正确率（AP）、精度（mAP）4 个评价指标。AP 衡量模型在单个检测类别上的性能优劣，该值[精确率-召回率（PR）曲线与横轴的面积]可以解决精确率、召回率等指标单点值的局限性，通过提高对低召回率性能的可视性，提高了对模型度量的灵敏度，AP 指标能够在更大程度上突出方法之间的差异[25]。其中精确率和召回率的表达式为

$$P=\frac{N_{TP}}{N_{TP}+N_{FP}},\tag{9}$$

$$R=\frac{N_{TP}}{N_{TP}+N_{FN}},\tag{10}$$

式中：N_{TP}表示目标正确识别的数量；N_{FP}表示非目标当成目标的数量；N_{FN}表示没有识别出来的目标数量；$N_{TP}+N_{FP}$表示识别出来的目标数量；$N_{TP}+N_{FN}$表示目标的总数量。AP定义为

$$P_{AP}=\int_0^1 P(R)\mathrm{d}R。\tag{11}$$

mAP衡量模型在多个类别上的综合性能，即在类别数为N的检测中，取所有类别AP的均值，即

$$P_{mAP}=\frac{1}{N}\sum_1^N P_{APi},i\in N。\tag{12}$$

实验基于HP-Z840图形工作站，采用Nvidia 1080Ti(11G)图形显卡，并基于64位Ubuntu16.04操作系统，采用Nvidia CUDA并行计算工具包，网络的搭建采用pytorch深度学习框架，实验包括模型结构研究和模型对比两部分。

为了研究本文所提模型网络结构，分别对由原始SSD和FSSD两种检测框架以及多种特征提取网络组合构成的场景辅助多任务耦合模型进行了训练和对比实验。训练数据集包括2100张场景-目标和2100张遥感场景数据集，并利用场景-目标数据集的900张验证集进行验证。首先，以VGG16作为特征提取网络，利用同步激活方式分别将SSD和FSSD框架下的场景辅助多任务耦合检测模型在通道相加和通道串联两种融合方式下进行了对比，从表1结果看出，FSSD框架下的精度高于SSD框架的精度，通道串联方式下精度高于通道相加精度。相加方式在将两种不同模态信息进行融合时破坏了不同任务特征图自身的深度拓扑关系，导致检测精度出现下降。

表1　两种特征图通道融合方式(同步激活、VGG16)

Table 1　Two feature map channel fusion methods (synchronous activation, VGG16)

Base model	Channel addition mAP /%	Channel concatenation mAP /%
SSD	82.13	86.63
FSSD	86.78	90.45

其次，进行了IA模块的研究，IA模块学习的目的是从深度特征图的通道与通道之间学习到相关耦合特性，以VGG16作为特征提取网络，对比在不同检测框架下多任务模型有无IA模块的精度(表2)，可以发现激活模块对检测任务精度具有较好的提升作用。

表2　IA模块在不同框架模型上的效果对比

Table 2　Comparison of IA module on different framework models

Base model	Object detection mAP /%	Scene classification mAP /%
SSD-IA	86.63	98.21
SSD-none	83.12	98.31
FSSD-IA	90.45	98.69
FSSD-none	88.44	98.56

再次，本文将多任务情形下的IA方式划分为不同模态(任务)的异步激活方式和同步激活方式两种。通过训练后，两种方式均能学习到不同通道之间的耦合关系，但从表3结果看出，异步激活将两类任务独立开来进行，失去了多任务的耦合性，没能真正发掘出场景与目标之间的深度拓扑关系、有效利用两者共享深度特征信息，因此其精度低于同步激活方式的精度。

表3　同步激活与异步激活对目标检测精度的影响

Table 3　Effect of synchronous and asynchronous activations on accuracy of object detection

Base model	Synchronous activation mAP /%	Asynchronous activation mAP /%
SSD	86.63	84.31
FSSD	90.45	88.36

将IA模块与特征图的不同组合进行对比验证，最终选择模型的总体结构组成如图3所示。本文以此模型为基础在不同主流特征提取网络上进行了验证实验。从表4结果看出，对于目标检测任务，大型特征提取网络模型如VGG16、ResNet50总体精度高于轻量级提取网络如MobileNetsv2、Darknetv2，但实时性能却下降。通过研究不同特征提取网络在本文算法上的检测分类性能，权衡精度和速度，最终采用了VGG16网络作为本文算法前端特征提取网络。

最后，将最终模型与传统SSD和FSSD算法进行了对比，从表5结果看出，本文模型将场景信息耦合辅助目标检测，针对汽车、卡车这类容易错误分类的小目标也能进行较好的预测。

对本文模型验证集结果可视化如图6所示。从图6可以发现，本文模型对小目标具有较高的检测精度，对于树木和阴影的遮挡、背景干扰等情况能进行正确辨识，同时还能对目标所在场景进行正确的感知识别，提升了检测网络的应用能力和范围。

表 4　不同特征提取网络下场景耦合多任务模型检测结果

Table 4　Scene-coupled multi-task model detection results based on different feature extractions

Feature extraction	Object detection task AP /%				Scene classification task precision /%			Frame rate
	Car	Truck	Airplane	Boat	Town	Airport	Waters	
VGG16	91.97	77.51	98.42	93.91	98.32	98.75	99.01	30
ResNet50	93.12	84.23	99.17	94.67	99.31	99.12	99.52	14
MobileNetsv2	84.76	79.34	88.45	86.56	98.22	97.43	98.32	46
Darknetv2	83.13	77.21	85.31	82.14	97.51	98.21	98.77	40

表 5　本文算法与传统模型对比

Table 5　Comparison of proposed algorithm with traditional object detection models　%

Algorithm	AP				mAP
	Car	Truck	Airplane	Boat	
SSD-VGG16	84.24	67.22	98.31	89.77	84.89
FSSD-VGG16	88.87	69.45	97.62	92.28	87.05
Proposed-VGG16	91.97	77.51	98.42	93.91	90.45

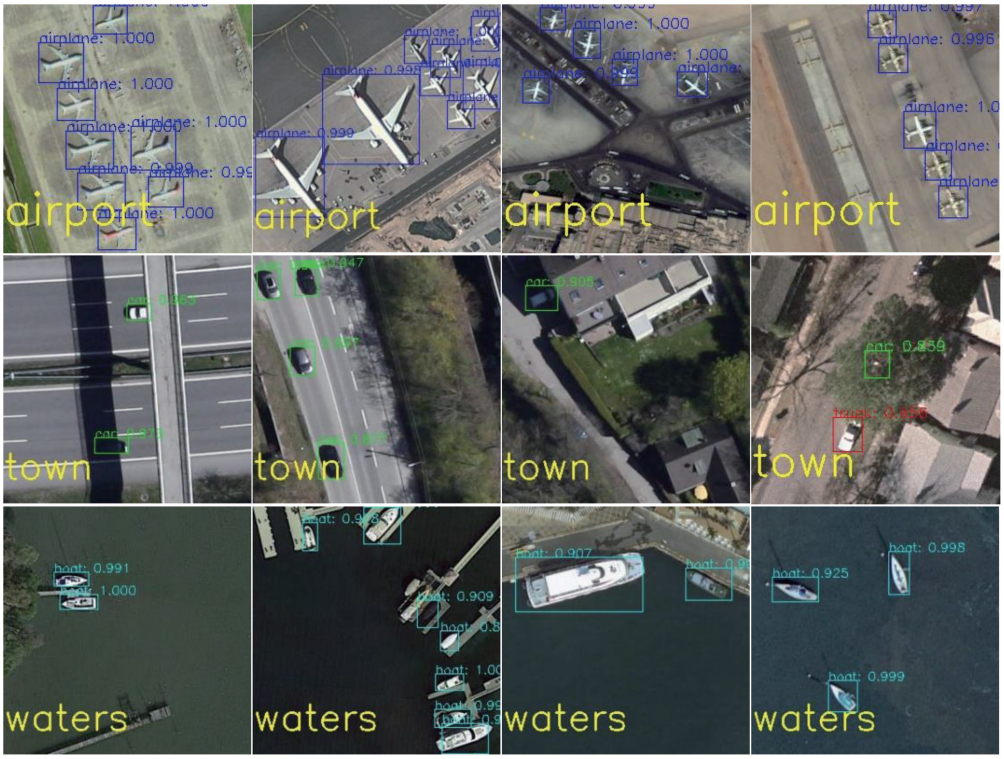

图 6　场景耦合多任务模型(VGG16)验证集可视化结果

Fig. 6　Visualization results of scene coupling multi-task model (VGG16) validation set

3.3　遥感场景栅格化感知实验

在不同尺度场景下，空对地检测适用的任务类型也在发生变化。当视野环境处在大尺度上(图 7)时，如高空观测日本某机场，单纯利用传统深度学习检测算法无法实现环境的定位感知搜索，目标检测倾向于失效，而对于场景类别(如机场、城市、港口等)的判断则为更为适用的任务；当视野环境聚焦于小尺度上时，目标变得逐渐清晰，这时目标检测才成为较为适用的任务。

为了验证网络的场景感知能力，本文利用验证集中的 900 张遥感场景对训练好的场景辅助目标检测模型进行了实验验证，验证结果如表 6 所示，从实验结果看出，本文算法在不同特征提取网络下均能取得较高的精度。

受到文献[14]的启发，利用本文算法的场景感知能力实现了空间变分辨率场景预测。首先将从谷歌地球获取的日本某机场不同遥感高度的视野图像进行栅格化，由于本文数据集尺寸大小为300 pixel×

表 6　不同特征提取网络下遥感场景分类的结果

Table 6　Classification results in remote sensing scenes under different feature extraction networks

Feature extraction	Precision /%		
	Town	Airport	Waters
VGG16	98.44	99.75	99.11
ResNet50	98.51	98.98	99.43
MobileNetsv2	97.32	97.93	98.92
Darknetv2	97.73	97.66	98.57

300 pixel ～500 pixel×500 pixel，采用 4×4 网格对视野图像进行切割，每个栅格可以限制在本文数据集尺寸范围之内。高空视角时，基于本文模型首先利用场景分类对目标所在场景进行动态感知，找出飞机所在特定场景即机场位置，逐步缩小搜索范围，如图 7(a)所示。当逐渐进入中低空范围时，网络在感知到目标后自动引入检测任务，一方面继续执行场景感知任务，另一方面同时进行目标检测，如图 7(b)所示。

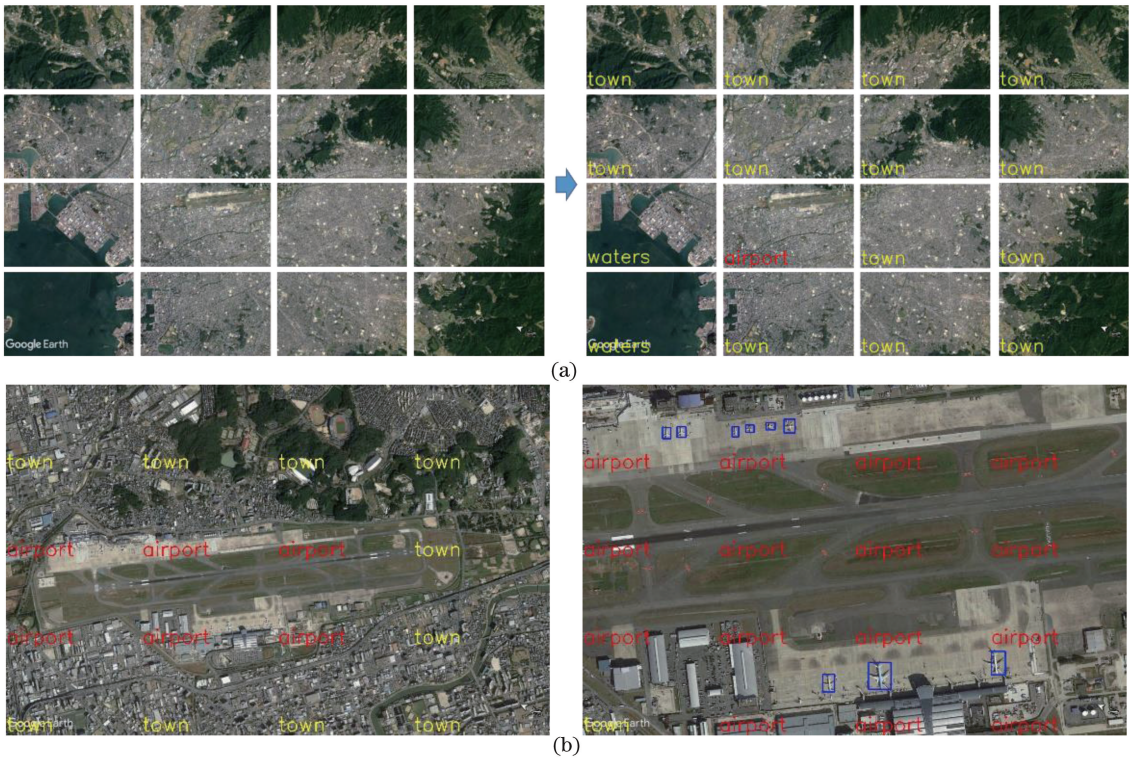

图 7　序贯场景变分辨率目标搜索。(a)远视距栅格化场景感知示意图；(b)高空场景感知引导的重要目标检测结果

Fig. 7　Sequential scene change resolution target search. (a) Far view rasterization scene perception schematic; (b) high-altitude scene-aware guided object detection

从实验结果看，本文模型能够有效地完成高空到低空的多尺度变分辨率场景感知任务，实现了高价值场景目标搜索和目标检测的多任务过程。

4　结　　论

当前主流深度学习检测算法通常执行单一的目标检测任务，不具有同时对场景进行变尺度动态感知的能力，将传统深度学习的方法直接应用于空对地目标检测任务时，受到视角单一、目标特征描述不全面等因素的影响，难以取得满意效果。针对该问题，提出了一种新的空基多任务耦合模型，将场景特征信息和目标特征检测信息耦合，实现了场景分类信息和目标检测任务的相互辅助，提高了空对地目标检测算法的性能。该算法在实现空对地目标检测的同时，能对场景进行动态感知，弥补了高空视场无法定位感知的缺陷，增加了检测识别的空间范围，根据不同遥感观测高度自适应执行需求的目标检测和场景感知任务，在遥感卫星的变分辨率任务搜索和空基无人平台执行对地任务上均有较高的应用价值。

参　考　文　献

[1] Lindeberg T. Scale invariant feature transform[J]. Scholarpedia, 2012, 7(5): 10491.

[2] Dalal N, Triggs B. Histograms of oriented gradients for Human detection[C]//2005 IEEE Computer

Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005: 886-893.

[3] Bay H, Ess A, Tuytelaars T, *et al*. Speeded-up robust features (SURF)[J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.

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

[5] Pfister T, Simonyan K, Charles J, *et al*. Deep convolutional neural networks for efficient pose estimation in gesture videos[C]// Asian Conference on Computer Vision, 2014: 538-552.

[6] Szegedy C, Liu W, Jia Y Q, *et al*. Going deeper with convolutions[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 1-9.

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

[8] Howard A G, Zhu M, Chen B, *et al*. MobileNets: efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv: 1704.04861, 2017.

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

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

[11] 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.

[12] 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), 2016: 779-788.

[13] Liu W, Anguelov D, Erhan D, *et al*. SSD: Single shot multibox detector[C]// European Conference on Computer Vision, 2016: 21-37.

[14] Deng Z P, Sun H, Zhou S L, *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 and Remote Sensing, 2017, 10(8): 3652-3664.

[15] 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.

[16] Hou Y Q Y, Quan J C, Wei Y M. Valid aircraft detection system for remote sensing images based on cognitive models[J]. Acta Optica Sinica, 2018, 38(1): 0111005.
侯宇青阳, 全吉成, 魏湧明. 基于认知模型的遥感图像有效飞机检测系统[J]. 光学学报, 2018, 38(1): 0111005.

[17] Liu D W, Han L, Han X Y. High spatial resolution remote sensing image classification based on deep learning[J]. Acta Optica Sinica, 2016, 36(4): 0428001.
刘大伟, 韩玲, 韩晓勇. 基于深度学习的高分辨率遥感影像分类研究[J]. 光学学报, 2016, 36(4): 0428001.

[18] Chu X, Ouyang W, Yang W, *et al*. Multi-task recurrent neural network for immediacy prediction[C]// 2015 IEEE International Conference on Computer Vision (ICCV), 2015: 3352-3360.

[19] Long M, Wang J. Learning multiple tasks with deep relationship networks[J]. Computer Science, 2017, arXiv: 1506.02117v1.

[20] Teichmann M, Weber M, Zoellner M, *et al*. MultiNet: real-time joint semantic reasoning for autonomous driving[J]. Computer Vision and Pattern Recognition, 2018, arXiv: 1612.07695.

[21] Li Z, Zhou F. FSSD: feature fusion single shot multibox detector[J]. Computer Vision and Pattern Recognition, 2017, arXiv: 1712.00960.

[22] 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), 2017: 936-944.

[23] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[J]. Computer Vision and Pattern Recognition, 2018, arXiv: 1709.01507.

[24] Liu K, Mattyus G. Fast multiclass vehicle detection on aerial images[J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12(9): 1938-1942.

[25] Dagan I, Glickman O, Magnini B. The PASCAL recognising textual entailment challenge[M]. Heidelberg: Springer, 2006: 177-190.