文章编号： 1673-3193（2024）05-0628-10

基于大核卷积和密集目标细化的遥感图像多尺度特征增强网络

王占魁，秦品乐，曾建潮

（中北大学 计算机科学与技术学院，山西 太原 030051）

摘　要： 针对遥感图像中目标尺度变化差异大、方向任意和分布密集，现有检测方法较少直接关注密集边缘信息且目标无法获得合适的感受野，遥感检测效果较差的问题，本文提出了一种基于大核卷积和密集目标细化的多尺度特征增强网络（LKCSFP-NET）来进行遥感图像的检测。该网络首先在SKNET基础上增加了空洞卷积形成大核卷积块（LKB），从而获得小目标的最佳感受野以及提升了网络对多尺度的适应性和准确度；其次在FPN基础上增加了集中空间特征金字塔CSFP模块，通过将全局语义信息与局部语义信息相结合，解决了遥感图像因目标分布密集以及背景复杂导致的检测效率较低的问题。实验结果表明，在DOTA和HRSC2016公开数据集上，所提算法在2个数据集上的平均检测精度分别为75.96%和96.60%，较基线网络提升了1.36百分点和0.63百分点，优于现有大多数模型。所提出的LKCSFP-NET在两个公开数据集中表现稳定，对小目标和密集排列的目标都有较好的检测效果，高于现有大多数模型的检测精度，可以很好地应用于遥感目标的检测。

关键词： 目标检测；遥感图像；多尺度；大核卷积；密集检测；特征融合
中图分类号： TP391.41；TP242　　**文献标识码：** A　　**doi:** 10.3969/j.issn.1673-3193.2024.05.009
引用格式： 王占魁，秦品乐，曾建潮．基于大核卷积和密集目标细化的遥感图像多尺度特征增强网络[J]．中北大学学报（自然科学版），2024，45(5)：628-637.
WANG　Zhankui，QIN　Pinle，ZENG　Jianchao. Multi-scale feature enhancement network based on large kernel convolution and dense object refinement for remote sensing images[J]. Journal of North University of China(Natural Science Edition)，2024，45(5)：628-637.

Multi-Scale Feature Enhancement Network Based on Large Kernel Convolution and Dense Object Refinement for Remote Sensing Images

WANG Zhankui，QIN Pinle，ZENG Jianchao

（School of Computer Science and Technology，North University of China，Taiyuan 030051，China）

Abstract： Due to the large difference in the scale of the object，the arbitrary direction and the dense distribution of the object in the remote sensing image，the existing detection methods rarely pay direct attention to the dense edge information and the object cannot obtain a suitable receptive field，so it is difficult to have good detection results in remote sensing detection. In order to solve the above problems，this paper proposed a multi-scale feature enhancement network based on large kernel convolution and dense object refinement（LKCSFP-NET）for remote sensing image detection. Firstly，the network based on SKNET

收稿日期： 2024-01-03
基金项目： 山西省科技重大专项计划“揭榜挂帅”项目（202101010101018）
作者简介： 王占魁（1995—），男，硕士生，主要从事目标检测的研究。
通信作者： 秦品乐（1978—），男，教授，博士，主要从事机器视觉、大数据的研究。E-mail：qpl@nuc.edu.cn。

added a cavity convolution to form a large kernel convolution block (LKB) to obtain the best sensitivity field for small targets and improve the adaptability and accuracy of the network to multiple scales. Secondly, on the basis of FPN, the centralized spatial feature pyramid CSFP module was added to solve the problem of low detection efficiency of remote sensing images due to dense object distribution and complex detection background by combining global semantic information with local semantic information. The experimental results show that on the DOTA and HRSC2016 public datasets, the average detection accuracy of the proposed algorithm on the two datasets is 74.90% and 96.60%, respectively, which is 1.36 and 0.63 percentage points higher than that of the baseline network, which is better than most existing models. The proposed LKCSFP-NET has stable performance in the two public datasets, and has good detection results for small objects and densely arranged objects, which is higher than the detection accuracy of most existing models, and can be well applied to the detection of remote sensing objects.

Key words: object detection; remote sensing image; multi-scale; large kernel convolution; intensive detection; feature fusion

0 引 言

遥感图像（Remote Sensing Imagery，RSI）的获取和应用越来越多样化[1-3]，其中遥感目标检测算法是遥感分析领域的研究热点之一，它不仅针对RSI中感兴趣的目标区域进行定位，还针对多目标进行分类，已被广泛用于灾害响应[4]、城市监测[5]、交通控制[6]等领域。尽管已经有很多遥感检测的算法，特别是有很多针对大规模RSI的检测算法，但是对密集复杂的场景和多尺度目标的检测仍存在很多问题。

与自然场景图像不同，RSI通常是从高空卫星或无人机上获取的，由于图像采集高度的变化，不同RSI中的物体具有不同的尺度[7]。某些类别的物体通常密集分布在RSI中，如船舶和车辆[8]。上述问题是RSI目标检测的主要障碍，这使得大多数自然图像算法不能很好地适应RSI。

现有研究中给出了很多由自然场景图像方法演化而来的新方法，其中基于RCNN系列框架的方法应用最广泛，这些方法首先生成大量的水平边界框作为感兴趣区域（RoI），然后根据区域特征预测分类结果和回归结果。但是，水平RoI通常会受到边界框和定向目标之间严重错位的影响[9-10]。例如，航空图像中的目标通常具有任意方向且密集排列，也存在将多个目标当作单个目标作为水平RoI[11]，从而导致特征不对齐问题。因此，提取准确的视觉特征变得很困难。也有一些方法[10-12]利用定向边界框作为锚来处理旋转目标，但这些方法计算复杂，因为其中包含许多精心设计的具有不同角度、尺度和纵横比的锚。RoI-Trans[11]通过旋转RoI学习器并使用旋转的位置敏感RoI对齐模块提取旋转不变区域的特征，将水平RoI转换为定向RoI。然而，这种方法仍然需要精心设计锚，并且不够灵活。

由于航空图像中的大多数物体尺寸较小且尺度变化差异大，仅凭其外观难以识别。如果在识别这些物体时能够加入背景等先验，可能会取得更好的效果，因为周围的环境信息可以提供关于小物体的形状、方向及其它重要特征，从而有利于更好地获取上下文信息。因此，准确检测任意方向的目标仍然面临以下挑战：1）不同图片中同一种目标的距离不同，而且不同类型目标所需的上下文信息是不同的，在单一尺度下无法很好地检测。2）在单张图片中有大量目标且排列密集，存在将多个密集目标检测为单个目标的问题，使目标间的边界无法很好地分离。

为了解决上述问题，本文提出一种基于大核卷积和密集目标细化的多尺度特征增强网络（LKCSFP-NET）进行遥感图像检测。LKCSFP-NET在SK-NET的基础上引入空洞卷积构建大感受野块（LKB），并在聚集多尺度信息的同时添加空间选择机制来构造多尺度增强模块。目前基于任意方向的网络主流方法是加入角度分支并进行特征对齐，但由于遥感图像的目标排列密集，无法很好地将目标完全包围，导致最终得到的旋转框不精确。因此，为了进一步提升旋转框的准确性，在CFP[13]结构的基础上进行了优化，引入了CSFP来专注于密集检测，可以在增加少量参数的基础上达到比较好的效果。该分支首先使用

特征金字塔来提取图像的不同尺度特征，然后通过全局级联模块和局部级联模块进行优化。全局级联模块通过轻量级MLP获得特征金字塔顶层的全局长距离依赖关系，即全局特征；局部级联模块则通过可学习视觉中心机制来获得局部信息，即局部特征。其次，为了能更好地提取遥感图像的密集特征，使用带有空间信息的MLP结构来进行细化。最后，对不同层之间的特征进行融合来获取多尺度信息。

为了验证LKCSFP-NET的性能，本文分别在DOTA数据集和HRSC2016数据集上进行实验，并进行了一系列的消融实验。在此之后将一些检测结果可视化，以评估LKB与CSFP的有效性。通过定量指标以及可视化结果分析，证明LKCSFP-NET可以准确地检测出目标。

1 相关工作

1.1 基于特征金字塔的遥感目标检测

Lin等[14]提出了一种特征金字塔网络，该结构可以有效地获得多尺度特征。Liu等[15]添加了自下而上的路径，以更好地集成多尺度特征图。Guo等[16]提出了3个模块来解决FPN的3个缺点，并在检测性能上获得了明显的提高。这些FPN引入了复杂的模块来提高检测精度，但无法保持运行速度。Li等[17]提出了一种动态特征选择模块，用于根据新锚点的位置和大小来选择像素。这些方法的目的是在目标级别上选择合适的特征。为了提取细粒度特征，SKN[18]使用不同的核在每个位置选择具有不同感受野的特征。在遥感检测领域，有关FPN的方法如GL Net[19]、CANet[20]、SB-MSN[21]、FSoD-Net[22]、CF2PN[23]和ASSD[24]，给出了多种FPN方法来检测不同的遥感图像中的目标。但是，由于特征混淆或结构复杂，它们不能同时解决检测聚类目标和快速检测的问题。与之不同的是，本文提出了一种改进的FPN，增加了少量的参数以保持可观的检测速度，它可以有效地检测多尺度和密集的物体。

1.2 旋转目标密集检测

得益于深度学习的优势，遥感图像的研究在过去几年中取得了巨大进展，特别是在定向目标检测方面[25-27]。目前，高性能的遥感目标检测器大多依赖于RCNN，该框架由区域建议网络和区域CNN检测头组成。两阶段RoI transformer[11]在第一阶段使用全连接层来完成水平框到旋转框的转换，然后提取框内的特征进行进一步的回归和分类。SCRDet[28]使用注意力机制来减少背景噪声，并采用采样融合网络解决拥挤目标和小目标的检测精度低的问题。S2ANET[29]网络通过对齐卷积来解决特征不对齐的问题。DRN[30]利用动态精细化网络来调整神经元的感受野，以实现更准确的预测。与定向RCNN相比，RSDet[31]通过引入调制损失来解决回归损失的不连续性。AOPG[32]和R3Det[25]采用渐进回归方法，从粗粒到细粒度细化边界框。尽管上述方法在一定程度上提高了定向目标检测的性能，但它们仍然存在边界框不对齐的问题。

1.3 大核卷积网络

大核卷积在自然场景下的检测已经证明了其有效性，但是缺乏在遥感检测特定领域的研究。文献[33-36]已经证明，大的感受野是提升检测结果的一个关键因素。研究也表明，设计良好的具有大感受野的卷积网络也可以与基于Transformer的模型具有同等的效果。ConvNeXt[37]在主干中使用7×7深度卷积，从而显著提高了下游任务的性能。RepLKNet[38]通过使用31×31卷积核进行重参数化，达到了自然场景下检测的最好性能。SLaK[39]通过核分解和稀疏群技术将核大小扩展为51×51。VAN[40]引入了大核的有效分解作为卷积注意力。SegNeXt[41]和Conv2Former[42]证明了大核卷积在利用更丰富的上下文调制卷积特征方面发挥着重要作用。由于航空图像具有独特的特征，大内核特别适合遥感任务的检测。受以上工作的启发，本文设计了自适应调整神经元的感受野模块，可以为不同角度、形状和尺度的各种物体重新组合适当的特征。

2 本文方法

本文将Oriented R-CNN作为基线网络，分别介绍LKCSFP-NET网络结构、大核卷积模块、改进的集中空间特征金字塔和损失函数。

2.1 网络结构

LKCSFP-NET的具体结构由主干网络ResNet50、金字塔FPN、大核卷积模块(LKB)、

密集目标细化模块（CSFP）和定向检测头五部分组成，如图 1 所示。其中，主干网络 ResNet50 和金字塔 FPN 用于提取多尺度特征{P2、P3、P4、P5}，LKB 用于选取最佳感受野区域，CSFP 用于细化目标的边界，定向检测头用于得到分类与回归结果。

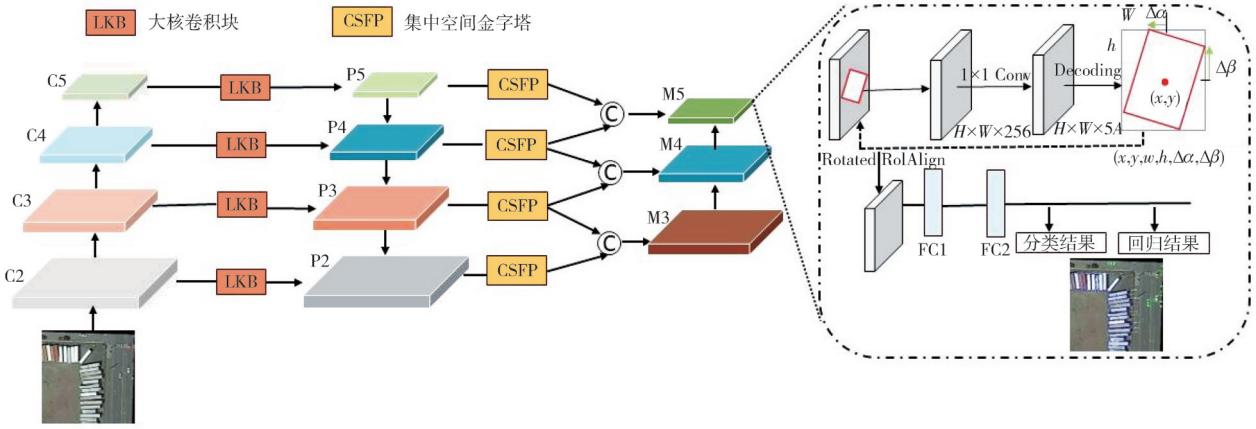


图 1　LKCSFP-Net 概述图

Fig. 1　Overview of LKCSFP-Net

2.2　大核卷积模块

遥感图像不同目标之间尺度的巨大差异导致了检测效果不理想。通道注意力 SE 块使用全局平均信息来重新加权特征通道，而 GENet 空间注意力模块通过空间掩码增强了网络对上下文信息建模的能力，CBAM 将通道注意力和空间注意力结合起来利用了两者的优势。

ASPP 并联使用多层空洞卷积或者池化层等稀疏采样的方法，但会丢失像素之间的联系，尤其空间位置关系。由于通道选择无法对图像空间中不同目标的空间方差进行建模，而空间关系的编码信息对于遥感任务来说更直观、更有效，所以，本文在 SKNET 的基础上进行了优化，在空间维度上聚合了大内核的信息，可以进一步增强不同感受野中的目标区域，最终实现通过大内核获得最佳感受野以及通过空间选择机制来空间选择特征图。

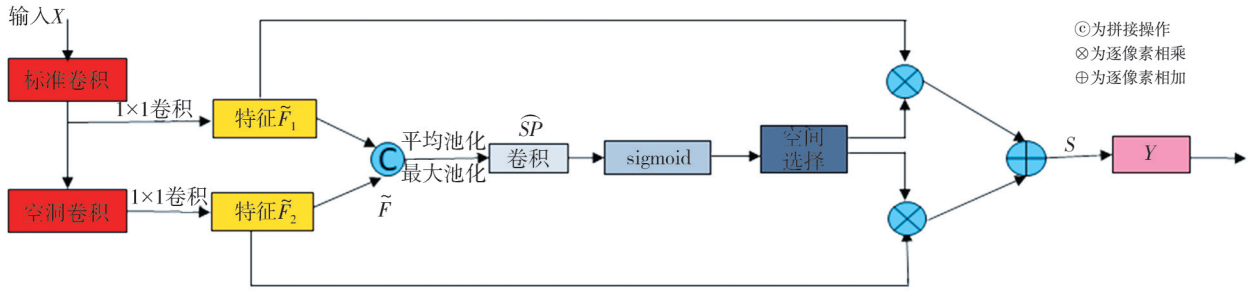


图 2　大核卷积模块

Fig. 2　Large kernel convolution module

如图 2 所示，首先将输入特征 X 通过标准卷积与空洞卷积分别获得两个不同的特征图，并通过卷积核大小为 1*1 的卷积将二者的通道数转换为相同的大小，即 $\widetilde{F_1}$、$\widetilde{F_2}$，然后将来自不同感受野卷积核的特征拼接为 $\tilde{F}$，并通过平均池化 P_{avg} 和最大池化 $P_{\max}$ 操作提取空间信息，接着对不同空间的描述信息通过卷积操作将空间池化特征拼接，并将 2 个通道的池化特征转换为 N 个空间注意力特征图 $\widehat{SP}$，之后通过 Sigmoid 函数应用到每一个空间注意力特征图，可获得每个解耦的大卷积核所对应的独立的空间选择掩膜 $\widehat{SP}_i$，最后将解耦后的大卷积核序列的特征与对应的空间选择掩膜进行加权处理，获得注意力特征 Y。具体操作如式（1）～式（6）。

$$F_0 = X, F_{i+1} = T_i^{dk}(F_i), \tag{1}$$

$$\widetilde{F_i} = T_i^{1*1}((F_i)),\ i \in [1, N], \tag{2}$$

$$\tilde{F} = \left[\widetilde{F_1}; \cdots; \widetilde{F_i}\right], \tag{3}$$

$$\widehat{SP} = S^{2\to N}\left(\left[P_{\text{avg}}(\tilde{F}); P_{\max}(\tilde{F})\right]\right), \tag{4}$$

$$\widehat{SP}_i=\sigma\left(\widehat{SP}_i\right), \tag{5}$$

$$Y=S\left(\sum_{i=1}^{N}\widehat{SP}_i\cdot\widetilde{F}_i\right), \tag{6}$$

式中：d为空洞率；k为卷积核大小；σ为sigmoid激活函数。

2.3 集中空间特征金字塔

针对遥感图像目标排列密集的问题，现有基于水平框或者旋转框的方法都无法很好地将目标分离。本文对原结构中的通道MLP结构进行了改进，提出了全新的CS-MLP结构，该结构在第一个残差块中构建通道C-MLP，在第二个残差块中构建空间S-MLP。其中C-MLP使用可变形卷积，该卷积结构相对标准卷积加了一个偏移量，添加偏移量后可以应对如目标移动、尺寸缩放、旋转等各种情况，从而能够更好地检测遥感场景目标。然后，将C-MLP的输出作为S-MLP的输入，S-MLP使不同空间位置之间的信息交互，从而更好地获得遥感图像的空间信息。因此，本文通过在FPN中加入CSFP模块，不仅能够学习层间的信息，更主要的是能够学习层内以及边缘的特征，这对于遥感图像密集任务的检测非常重要。

在得到FPN输出的特征图时，将特征图经过一个stem模块(由7×7卷积、批量归一化、ReLU激活函数组成)平滑处理，对特征图上的噪声进行抑制，从而保留特征图的具体细节，之后输入到CSFP模块中，CSFP是由轻量级MLP与视觉中心LVC并行连接的模块组成的，如图3所示。对于CSFP模块中的MLP而言，其作用是捕获全局的长距离依赖关系，使用该模块可以更全面、更准确地获得遥感图像中尺度差异大的目标的特征表示，从而提高目标分类和定位的准确性。该过程可以表示为

$$MLP_{out}=SMLP\left(GN(Y)\right)+Y, \tag{7}$$

其中，$$Y=DCN\left(GN(X_{in})\right)+X_{in},$$

$$X_{in}=\left(BN\left(Conv_{7*7}(X_{top})\right)\right),$$

式中：$SMLP$表示空间MLP；GN表示组归一化；DCN表示可变形卷积；BN表示批归一化；top表示金字塔顶层P_5。

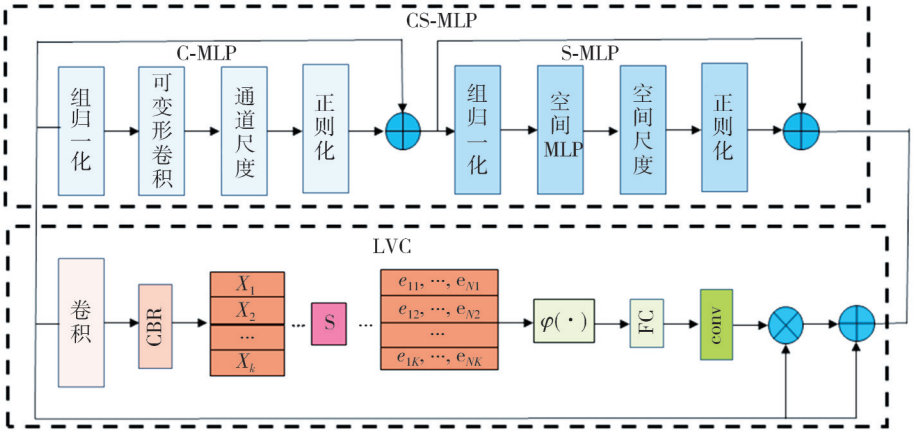


图3 密集目标细化CSFP模块

Fig. 3 Intensive object refinement of CSFP module

对于LVC而言，该模块主要由卷积层、全连接层以及字典编码器组成。通过卷积层对输入特征进行编码，并使用具有归一化的卷积和Relu激活函数组成的CBR模块对编码进行处理并输送到字典编码器中，使用编码器能够获得关于编码的整个图像的完整信息，之后通过将编码器的输出馈送到全连接层和卷积层以预测突出关键类的特征。在遥感图像中使用LVC模块能够捕获图像的局部角落区域，通过该模块可以更好地辨别出遥感图像中需要检测的目标以及与目标相似的背景和建筑，而且对于遥感图像而言，需要检测的目标不一定在图像的中心，该模块还可以调整网络的关注区域，避免出现漏检的现象，从而提高目标检测的准确性。该过程可以表示为

$$X_{all}=X_{in}+Z。 \tag{8}$$

其中，$$Z=X_{in}\otimes\left(\delta\left(\mathrm{Conv}_{1*1}\left(\sum_{k=1}^{K}\phi(e_k)\right)\right)\right),$$

$$e_k=\sum_{i=1}^{N}\frac{e^{-s_k\|X_i-b_k\|^2}}{\sum_{j=1}^{K}e^{-s_k\|X_i-b_k\|^2}}(X_i-b_k)。$$

最后通过并行连接MLP与LVC，得到最终的CMLP输出结果，CSFP不仅从全局的角度出发

获取了顶层特征图中目标的特征，而且还考虑了特征图的局部信息，从而使网络能够充分提取检测目标的特征信息，提高对密集目标的检测精度。

2.4　损失函数

为了训练定向RPN，对正样本和负样本进行定义。首先，为每个锚点分配一个二进制标签$p^*\in\{0,1\}$，其中，0表示正样本，1表示负样本。满足以下两个条件其一将视为正样本：1）与GT（地面真实）框的交并集（IoU）高于0.7，2）与GT框的IoU最高且IoU高于0.3。当IoU低于0.3时，锚被标记为负样本。既不是正样本也不是负样本的锚被视为无效样本，在训练过程中被忽略。上述GT框指定向边界框的外部矩形，定义损失函数L_1为

$$L_1=\frac{1}{N}\sum_{i=1}^{N}F_{\text{cls}}\left(p_i,p_i^*\right)+\frac{1}{N}p_i^*\sum_{i=1}^{N}F_{\text{reg}}(\delta_i,t_i^*),\tag{9}$$

式中：i为锚的索引；N为一个批量中的样本总数，默认为256。分类分支中，F_{cls}为交叉熵损失；p_i^*为GT值标签；p_i表示锚点为前景的概率，是分类分支的输出。回归分支中，F_{reg}为平滑$L1$损失；δ_i为输出建议相对于锚的偏移量；t_i^*为锚点的真实值偏移量，分别用式（10）表示。

$$\begin{aligned}\delta_i&=\left(\delta_x,\delta_y,\delta_w,\delta_h,\delta_\alpha,\delta_\beta\right),\\ t_i^*&=\left(t_x^*,t_y^*,t_w^*,t_h^*,t_\alpha^*,t_\beta^*\right),\end{aligned}\tag{10}$$

$$\begin{cases}\delta_\alpha=\Delta\alpha/w,\delta_\beta=\Delta\beta/h,\\ \delta_w=\log(w/w_a),\ \delta_h=\log(h/h_a),\\ \delta_x=(x-x_a)/w_a,\ \delta_y=\left(y-y_a\right)/h_a,\\ t_\alpha^*=\Delta\alpha_g/w_g,\ t_\beta^*=\Delta\beta_g/h_g,\\ t_w^*=\log\left(w_g/w_a\right),\ t_h^*=\log\left(h_g/h_a\right),\\ t_x^*=\left(x_g-x_a\right)/w_a,\ \ t_y^*=\left(x_g-x_a\right)/h_a,\end{cases}\tag{11}$$

式中：(x_g,y_g)，w_g，h_g分别为外接矩形的中心点坐标、宽、高；$\Delta\alpha_g$和$\Delta\beta_g$分别为顶部和右侧顶点相对于顶部中点和左侧中点的偏移量。

3　实验结果

在DOTA数据集和HRSC2016数据集中进行实验，以验证本文所提方法的有效性。

3.1　数据集

1）DOTA数据集是一个大规模旋转目标检测数据集，包含2 806张照片和188 282个带有定向边界框注释的实例。该数据集包含15个对象类：飞机（PL）、棒球场（BD）、桥梁（BR）、地面田径场（GTF）、小型车辆（SV）、大型车辆（LV）、船舶（SH）、网球场（TC）、篮球场（BC）、储槽（ST）、足球场（SBF）、环岛（RA）、海港（HA）、游泳池（SP）和直升机（HC）。DOTA数据集图像像素的变化范围为800×800到4 000×4 000。实验将训练集和验证集结合起来进行训练，使用测试集进行测试，最终提交到DOTA官网用于获得检测结果。

2）HRSC2016数据集是一个船舶检测数据集，该数据集由1061张图像和2976个实例组成。图像像素范围为300×300到1 500×900。训练集、验证集和测试集中分别有436，541和444个图像。最终使用PASCAL VOC07和VOC12指标给出检测结果。

3.2　实施细节

本文方法通过在单个NVIDIA RTX 3090上使用PyTorch实现。选择带有FPN的ResNet50作为骨干网络，使用与Faster R-CNN-O中相同的超参数，批量大小设置为2。初始学习率为0.000 2，动量为0.90，权重衰减率为0.000 1，选择SGD来优化模型。对于DOTA数据集，所有模型都经过了12轮训练（多尺度下为36轮）。在第8轮和第11轮，学习率除以10。对于HRSC2016数据集，训练36轮，在第24轮和第33轮，学习率除以10。将DOTA数据集的图片裁剪成像素尺寸为1 024×1 024的新图片，步长为824像素。在训练过程中，两个数据集都使用了随机水平和垂直翻转。

3.3　各方法性能比较

在DOTA数据集和HRSC2016数据集上对本文方法与现有各方法的性能进行了评估。

DOTA数据集上的定性和定量结果分别如表1和图4所示。由表1可以看出，使用ResNet50 FPN作为骨干网络，本文的方法在单尺度下实现了74.90%的mAP，多尺度下实现了75.96%的mAP（表中MS代表多尺度），本文方法的性能超过了先进的RoI Transformer、R3Det和S2Anet等方法。由图4的定性结果可以看出，当检测高宽比密集物体时，本文检测方法产生的错误更少。

表1 DOTA数据集上各方法的比较

Tab. 1 Comparison of different methods in DOTA datasets %

方法	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	mAP
RSDet	89.80	82.90	48.60	65.20	69.50	70.10	70.20	90.50	85.60	83.40	62.50	63.90	65.60	67.20	68.00	72.20
SCRDet	89.98	80.65	52.09	68.36	68.36	60.32	72.41	90.85	87.94	86.86	65.02	66.68	66.25	68.24	65.21	72.61
DRN	89.71	82.34	47.22	64.10	76.22	74.43	85.84	90.57	86.18	84.89	57.65	61.93	69.30	69.63	58.48	73.23
R3Det	88.76	83.09	50.91	67.27	76.23	80.39	86.72	90.78	84.68	83.24	61.98	61.35	66.91	70.63	53.94	73.79
S2Anet	89.11	82.84	48.37	71.11	78.11	78.39	87.25	90.83	84.90	85.64	60.36	62.60	65.26	69.13	57.94	74.12
Rol Transformer	88.65	82.60	52.53	70.87	77.93	76.67	86.87	90.71	83.83	82.51	53.95	67.61	74.67	68.75	61.03	74.61
Mask OBB	89.61	85.09	51.85	72.90	75.28	73.23	85.57	90.37	82.08	85.05	55.73	68.39	71.61	69.87	66.33	74.86
SASM	86.42	78.97	52.47	69.84	77.30	75.99	86.72	90.89	82.63	85.66	60.13	68.25	73.98	72.22	62.37	74.92
DODet	89.34	84.31	51.39	71.04	79.04	82.36	88.15	90.90	86.88	84.91	62.69	67.63	75.47	72.22	45.54	75.49
Oriented RepPoints	87.02	83.17	54.13	71.16	80.81	78.40	87.28	90.90	85.97	86.25	59.90	70.49	73.53	72.27	58.97	75.97
Baseline(ORCNN)	95.10	75.70	48.70	68.30	73.40	87.90	90.00	94.20	73.40	80.80	59.60	62.50	81.50	59.90	69.00	74.60
LKCSFP-NET (ours)	96.30	78.30	48.70	68.40	72.90	86.60	**90.20**	94.10	73.50	81.40	62.00	66.90	82.30	54.10	67.50	74.90
LKCSFP-NET (MS)	**96.46**	80.21	48.53	**72.93**	72.92	86.68	89.20	**94.89**	75.67	83.80	62.04	65.55	**83.16**	60.29	67.14	75.96

图4 DOTA数据集检测结果示例

Fig. 4 Example of detection results of the DOTA dataset

HRSC2016数据集上的定性和定量结果如表2和图5所示，该数据集包含大量在不同角度且具有大纵横比的船只。由表2可以看出，本文的方法使用ResNet50 FPN分别实现了90.58%和96.60%的mAP，该结果优于大多数的检测器。由图5可以看出，本文的方法对于纵横比大的目标能产生准确的边界框。

表2 HRSC2016数据集上各方法的比较

Tab. 2 Comparison of different methods in HRSC2016 datasets

方法	mAP(07)/%	mAP(12)/%
PIoU	89.20	—
DRN	—	92.70
R3Det	89.26	96.01
DAL	89.77	—
S^2ANet	90.17	95.01
Rotated RPN	79.08	85.64
R2CNN	73.07	79.73
RoI Transformer	86.20	—
Gliding Vertex	88.20	—
CenterMap-Net	—	92.80
Baseline(ORCNN)	90.22	95.97
LKCSFP-NET(ours)	90.58	96.60

图5 HRSC2016数据集检测结果示例

Fig. 5 Example of detection results for HRSC2016 dataset

3.4 消融实验

3.4.1 LKB模块的有效性验证

不同尺度的目标所需要的感受野大小是不同的，为了验证LKB模块的重要性，设计了验证实验。将大内核分解为一个普通卷积和空洞卷积内核，如表3中第一行(仅基线)和第二行(基线+大核卷积模块)所示，在基线网络的基础上增加LKB模块后mAP达到了74.83%。

表3 消融实验

Tab. 3 Ablation experiments

方法	mAP/%
基线	74.60
基线+LKB	74.83
基线+CSFP	74.69
LKCSFP-NET	74.90

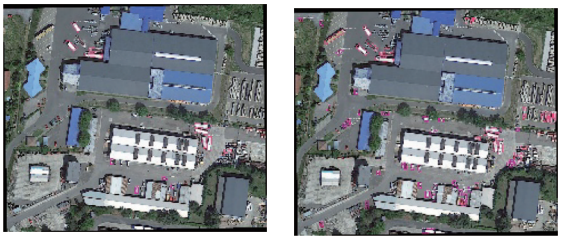
表4 不同感受野下的结果比较

Tab. 4 Comparison of results under different perceptions

(k_1, d_1)	(k_2, d_2)	感受野	mAP/%
(3, 1)	(5, 2)	11	74.70
(5, 1)	(7, 3)	23	74.83
(7, 1)	(9, 4)	39	74.74

表4显示，过小或过大的感受野会阻碍LKB模块的性能，感受野为23时是最有效的。图6为

LKB模块的可视化结果，图 7 为不同感受野下的可视化结果。



（a）基线结果　（b）本文结果

图 6　LKB模块的可视化结果

Fig. 6　Visualization of the LKB module

（a）感受野为11　（b）感受野为23

图 7　不同感受野下的可视化结果

Fig. 7　Visualization of different receptive field

3.4.2　CSFP模块的有效性验证

为了探究密集目标细化模块的作用，设计了多个实验，如表 4 中第一行（仅基线）和第三行（基线＋密集目标细化）所示，在基线网络的基础上增加 CSFP 模块后 mAP 达到了 74.69%。当 LKCSFP-NET 没有 LKB 与 CSFP 时，对密集目标的检测会出现明显的误判（多个目标检测为一个），图 8 为CSFP模块的可视化结果。定性和定量的实验表明，CSFP在密集排列下的情况下得到了较好的效果，并且将它与 LKB 结合后的 LKCSFP-NET 可以获得更好的性能。

（a）基线结果　（b）本文结果

图 8　CSFP模块的可视化结果

Fig. 8　Visualization of the CSFP module

4　结　论

本文提出了一种多尺度特征增强和密集目标细化网络以用于检测遥感图像中的物体。所提出的LKCSFP-NET有以下优势：1）LKCSFP-NET利用LKB获得最佳感受野并且使用空间选择机制以解决了小目标和多尺度检测的问题。2）所提出的CSFP可以有效地提取密集目标的特征信息，获得任意方向的目标特征。实验表明，该方法在具有挑战性的DOTA数据集和广泛使用的HRSC2016数据集中取得了优于其他方法的结果。后续将研究最佳边界框的表示方法，以对密集目标设计更加准确的包围框。

参考文献：

[1] WANG Q, GUO J Y, YUAN Y. Embedding Structured Contour and location prior in siamesed fully convolutional networks for road detection[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(1): 230-241 .

[2] XIE W Y, LEI J, FANG S, et al. Dual feature extraction network for hyperspectral image analysis [J]. Pattern Recognition, 2021, 118: 107992.

[3] XIE W Y, LEI J, CUI Y H, et al. Hyperspectral pansharpening with deep priors[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(5): 1529-1543.

[4] GANCI G, CAPPELLO A, BILOTTA G, et al. How the variety of satellite remote sensing data over volcanoes can assist hazard monitoring efforts: The 2011 eruption of Nabro Volcano[J]. Remote Sensing of Environment, 2020, 236: 111426.

[5] XIE W Y, ZHANG X, LI Y S, et al. Weakly supervised low-rank representation for hyperspectral anomaly detection[J]. IEEE Transactio- ns on Cybernetics, 2021, 51(8): 3889 - 3900.

[6] WANG Q, GAO J Y, YUAN Y. A joint convolutional neural networks and context transfer for street scenes labeling[J]. IEEE Transactions on Intelligent Transportation Systems, 2018, 19(5): 1457-1470.

[7] CEHNG G, HAN J W, ZHOU P C, et al. Multi-class geospatial object detection and geographic image classification based on collection of part detectors[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2014, 98: 119-132.

[8] LI K, WAN G, CHENG G, et al. Object detection in

optical remote sensing images: A survey and a new benchmark [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 159: 296-307.

[9] XIA G S, BAI X, DING J, et al. Dota: A large-scale dataset for object detection in aerial images[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2018: 3974-3983.

[10] LIU Z K, WANG H Z, WENG L B, et al. Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(8): 1074-1078.

[11] DING J, XUE N, LONG Y, et al. Learning roi transformer for oriented object detection in aerial images [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2019: 2844-2853.

[12] LIU Z K, HU J G, WENG L B, et al. Rotated region based cnn for ship detection[C]//IEEE International Conference on Image Procecssing, 2017: 900-904.

[13] QUAN Y, ZHANG D, ZHANG L Y, et al. Centralized feature pyramid for object detection [J]. IEEE Transactions on Image Processing, 2023, 32: 4341-4354.

[14] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2017: 936-944.

[15] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.

[16] GUO C X, FAN B, ZHANG Q, et al. AugFPN: Improving multi-scale feature learning for object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2020: 12592-12601.

[17] LI S, YANG L X, HUANG J Q, et al. Dynamic anchor feature selection for single-shot object detection [C]//IEEE International Conference on Computer Vision, 2019: 6608-6617.

[18] LI X, WANG W H,M HU X L, et al. Selective kernel networks [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2019: 510-519.

[19] TENG Z, DUAN Y N, LIU Y, et al. Global to local: Clip-LSTM-based object detection from remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5603113.

[20] SHI L K, KUANG L Y, XU X, et al. CANet: Centerness-aware network for object detection in remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5603613.

[21] HAN W, FAN R Y, WANGL Z, et al. Improv- ing training instance quality in aerial image object detection with a sampling-balance-based multistage network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(12): 10575-10589.

[22] WANG G Q, ZHUANG Y, CHENG H, et al. FSoD-Net: Full-scale object detection from optical remote sensing imagery [J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5602918.

[23] HUANG W, LI G Y, CHEN Q Q, et al. CF2PN: A cross-scale feature fusion pyramid network based remote sensing target detection[J]. Remote Sensing, 2021, 13(5): 847.

[24] XU T, SUN X, DIAO W H, et al. ASSD: Feature aligned single-shot detection for multiscale objects in aerial imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5607117.

[25] YANG X, YAN J C, FENG Z M, et al. R3Det: Refined single-stage detector with feature refinement for rotating object[C]//AAAI Conference on Artificial Intelligence, 2021, 35(4): 3163-3171.

[26] YANG X, YAN J C. Arbitrary-oriented object detection with circu-lar smooth label[C]//European Conference on Computer Vision, 2020: 677-694.

[27] YANG X, HOU L P, ZHOU Y, et al. Dense label encoding for boundary discontinuity free rotation detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2021: 15814-15824.

[28] YANG X, YAN J C, LIAO W L, et al. Scrdet++: Detecting small, cluttered and rotated objects via instance-level feature denoising and rotation loss smoothing[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(2): 2384-2399.

[29] HAN J M, DING J, LI J, et al. Align deep features for oriented object detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5602511.

[30] PAN X, REN Y Q, SHENG K K, et al. Dynamic refinement network for oriented and densely packed object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2020: 11204 -11213.

[31] QIAN W, YANG X, PENG S L, et al. Learning modulated loss for rotated object detection[C]//AAAI Conference on Artificial Intelligence, 2021: 2458-2466.

[32] CHENG G, WANG J B, LI K, et al. Anchor-free

oriented proposal generator for object detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5625411.

[33] René Ranftl, Alexey Bochkovskiy, Vladlen Koltun. Vision transformers for dense prediction [C]//IEEE International Conference on Computer Vision, 2021: 12159-12168.

[34] YAN H T, LI Z, LI W J, et al. Contnet: Why not use convolution and transformer at the same time? [DB/OL]. (2021-05-10) [2024-01-03]. https://arxiv.org/abs/2211.11943v3.

[35] ZHENG S X, LU J C, ZHAO H S, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2021: 6877-6886.

[36] LUO W J, LI Y J, Raquel Urtasun, et al. Understanding the effective receptive field in deep convolutional neural networks [C]//International Conference on Neural Information Processing Systems, 2016: 4905-4913.

[37] LIU Z, MAO H Z, WU C Y, et al. A convnet for the 2020s [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2022: 11966-11976.

[38] DING X H, ZHANG X Y, HAN J G, et al. Scaling up your kernels to 31×31: Revisiting large kernel design in cnns [C]//IEEE Conference on Computer Vision and Pattern Recognition, 2022: 11953-11965.

[39] LIU S W, CHEN T L, CHEN X H, et al. More convnets in the 2020s: Scaling up kernels beyond 51x51 using sparsity [DB/OL]. (2023-03-03) [2024-01-03]. https://arxiv.org/abs/2207.03620v3.

[40] GUO M H, LU C R, LIU Z N, et al. Visual attention network [J]. Computational Visual Media, 2023, 9(4): 733-752.

[41] GUO M H, LU C Z, HOU Q B, et al. SegNeXt: Rethinking convolutional attention design for semantic segmentation [C]//Annual Conference on Neural Information Processing Systems, 2022: 1140-1156.

[42] HOU Q B, LU C Z, CHENG M M, et al. Conv2former: A simple transformer-style ConvNet for visual recognition [DB/OL]. (2022-11-22) [2024-01-03]. https://arxiv.org/abs/2211.11943.