

# 一种用于遥感图像检测的特征增强金字塔网络

陶建<sup>1</sup> 严华<sup>1\*</sup>  
TAO Jian YAN Hua

## 摘要

目前，特征金字塔结构在遥感旋转目标检测任务中发挥着关键作用，但仍存在一定缺陷：传统方法多采用粗糙的近邻插值方式，而遥感图像中存在大量高纵横比、小尺寸目标，易导致此类目标的特征信息无法有效传递与融合。基于上述问题，文章提出一种高效的特征增强金字塔网络 FEFPN。FEFPN 与传统特征金字塔网络存在本质差异：一是引入面向遥感高纵横比目标的特征增强模块，强化对不同宽高比目标的特征捕捉能力；二是增加小目标空间上下文感知处理，有效补强遥感小目标的特征表达。通过在 DOTAV1.0 数据集上开展大量对比实验，结果表明，所提 FEFPN 可显著提升遥感图像检测性能，兼顾检测精度与推理速度，实现二者的有效平衡。

## 关键词

特征增强；遥感；图像检测；金字塔网络

doi: 10.3969/j.issn.1672-9528.2026.05.015

## 0 引言

近年来，遥感目标检测凭借其在无人机、卫星航拍图像分析中的重要应用价值，受到广泛关注<sup>[1-2]</sup>。现阶段主流检测网络大多基于旋转边界框设计，以此精准覆盖待检测目标。针对边界不连续、方框表征局限以及长距离依赖信息捕捉困难等问题<sup>[3-6]</sup>，现有研究已取得诸多突破，陆续涌现出性能更优的旋转目标检测算法，涵盖 IoU 模拟损失函数、骨干网络优化、目标表征方式改进<sup>[7]</sup>等多个研究方向。

尽管现有研究已取得长足进展，但高纵横比小尺寸目标的精准检测依旧是遥感目标检测领域的一大难题。遥感场景中广泛存在大量细长、微型目标，而现有遥感图像检测方法<sup>[4-5]</sup>对细长目标的适配能力普遍不足，检测性能会随目标纵横比的增大出现明显下降。造成该类目标检测难度较大的原因主要分为两方面。一是遥感图像视野范围广阔，适用于大范围对地监测，但场景内目标分辨率低、成像质量有限，普遍存在尺寸微小（小于 32 px×32 px）、特征微弱、对比度低、有效语义信息匮乏等问题，大幅增加了检测难度。二是遥感观测条件可控性弱，成像过程易受平台运动、大气环境等复杂因素干扰，易造成目标与背景特征混淆、边界重叠，进一步加大了小目标的识别与区分难度。

基于卷积神经网络的传统检测器大多在方形窗口内提取输入特征图。这种设计极大地限制了其有效捕获遥感图像中常见的不同方向数据特征的能力，导致包含来自周围地区的无关信息在回归任务中带来了较大挑战。

本文旨在设计一种高精度、推理速度快的遥感小目标检测网络，缓解特征表示不足和背景混淆问题的关键在于特征增强和融合，检测效果对比如图 1 所示。

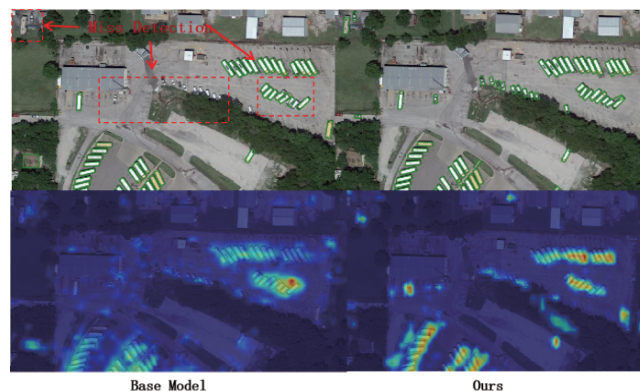


图 1 基准模型使用 FPN 与 FE-FPN 的检测效果比较

本文的主要贡献如下：

（1）为遥感应用设计了一种高效的小目标检测器（FE-FPN）。与几种基准模型和最新（SOTA）方法相比，FE-FPN 可在推理能力有限的小目标检测任务中具有较高的性能。

（2）提出了两种创新的轻量级即插即用模块：MSFEM 和 GCAM 两个模块分别提高了网络的局域特征融合和全局关联跨信道和空间的能力。可以用作插入任何检测网络的通用模块，以增强小物体的弱特征表示并抑制易混淆的背景。

## 1 相关工作

遥感目标探测。通用对象检测通常依赖于水平边界框来检测对象<sup>[8]</sup>。但在遥感场景中，物体的方向是任意的，水平

1. 四川大学电子信息学院 四川成都 610065

框通常无法精确定位物体,往往包含背景信息或其他物体<sup>[9]</sup>。因此,旋转边界框通常用于对象表示。旋转边界框的早期表示使用五个参数  $(x,y,w,h,\theta)$ <sup>[10]</sup>。

由于角度的周期性,基于这一表示的训练模型在回归中经常面临边界不连续性问题<sup>[3]</sup>。为了解决这个问题,几种方法提出了旋转边界框的改进表示<sup>[7,11]</sup>。例如,OR-CNN<sup>[12]</sup>面向对象检测通过采用旋转边界框来推进传统的水平对象检测范式,使其成为一项更精确的任务。

尽管现有研究已取得显著进展,但高纵横比目标与小尺寸目标的精准检测仍是遥感目标检测领域的共性难题。针对该类目标检测存在的固有难点,本文方法进行了针对性设计:通过结合多分支结构与条带卷积,提升模型对细长难样本目标的泛化能力;同时引入轻量化自注意力模块,强化网络对小目标特征的提取与识别能力,有效改善小目标检测效果。

## 2 FE-FPN 模型

在本节中,详细描述所提出的 FE-FN 网络的架构。目标是应用多分支条带卷及空间自注意力机制,提出一种增强遥感图像中小目标物体特征的颈部网络,以便生成的模型可以在不同宽高比、尺寸较小的目标上表现良好。

### 2.1 整体结构

基于遥感图像检测中常用的 FPN 架构,在 FPN 的每一层应用 MSFEM (multi-branch strip feature enhancement Module) 和 GCAM (global context-aware module) 模块,网络结构如图 2 所示,具体来说:MSFEM 模块通过多分支结构,应用条带卷积 StripConv,通过一种轻量级的有限元方法来提高网络的局域感知能力,以捕捉不同宽高比对象的特征;GCAM 模块通过空间自注意力机制及前馈神经网络,提高全局关联跨信道和空间的能力以捕获小目标周围的语义信息。

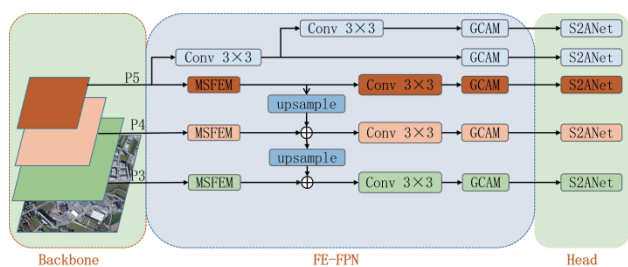


图 2 FE-FPN 网络结构图

### 2.2 MSFEM 模块

因遥感图像的复杂性,通常物体较小,且存在较多细长的目标,比如机场、货车、跑道等,常用的颈部网络 FPN 的提取能力是有限的,所提取的特征包含较少的语义信息和狭窄的感受野,使难将小物体与背景区分开来,容易出现具有相似特征的误报,且对于不同宽高比对象的特征较难提取。

因此,所提出的方法从两个角度考虑增强小物体的特征。从增加特征丰富度的角度来看,采用多分支卷积结构来提取多个判别语义信息。从扩大感受野的角度来看,水平和垂直大条带卷积的结合使网络能够收集两个空间轴上的方向特征,增强空间维度中细长或狭窄结构的表示。空洞卷积被应用于获得更丰富的局部语境信息。网络的整体结构如图 3 所示。每个分支对输入特征图执行平方卷积运算,以初步调整其输入特征通道,然后中间的分支执行条带卷积及空洞卷积运算,以提取不同方向上的特征。

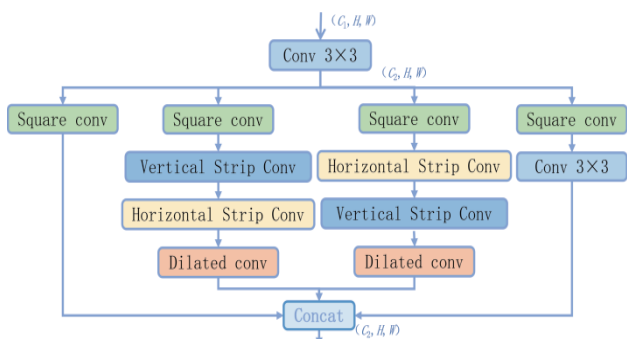


图 3 MSFEM 模块网络结构图

MSFEM 方法可以用公式表示为:

$$X1 = f_{\text{conv}}^{3 \times 3}(X) \quad (1)$$

$$W1 = f_{\text{conv}}^{3 \times 3} \{ f_{\text{conv}}^{1 \times 1}(X1) \} \quad (2)$$

$$W2 = f_{\text{dconv}}^{3 \times 3} \{ f_{\text{conv}}^{9 \times 1} \{ f_{\text{conv}}^{1 \times 9} [ f_{\text{conv}}^{1 \times 1}(X1) ] \} \} \quad (3)$$

$$W3 = f_{\text{dconv}}^{3 \times 3} \{ f_{\text{conv}}^{1 \times 9} \{ f_{\text{conv}}^{9 \times 1} [ f_{\text{conv}}^{1 \times 1}(X1) ] \} \} \quad (4)$$

$$W4 = f_{\text{conv}}^{1 \times 1}(X1) \quad (5)$$

$$Y = \text{Concat}(W1, W2, W3, W4) \quad (6)$$

式中:  $f_{\text{conv}}^{1 \times 1}(x)$ 、 $f_{\text{conv}}^{1 \times 9}(x)$ 、 $f_{\text{conv}}^{9 \times 1}(x)$  和  $f_{\text{conv}}^{3 \times 3}(x)$  分别表示核大小为  $1 \times 1$ 、 $1 \times 9$ 、 $9 \times 1$  和  $3 \times 3$  的标准卷积运算;  $f_{\text{dconv}}^{3 \times 3}(x)$  表示扩张率为 5 的空洞卷积操作;  $\text{Concat}(\cdot)$  表示特征图拼接操作;  $X$  表示输入特征图;  $W2$  和  $W3$  表示通过条带卷积和空洞卷积后前三个分支的输出特征图;  $Y$  表示 MSFEM 的输出特征图。

与 Transformer 相比,MSFEM 具有更轻的结构,以及更少的参数量,并使模型能够通过多分支空洞卷积学习更丰富的局部上下文特征,从而提高了小对象的特征表示能力。

### 2.3 GCAM 模块

在 MSFEM 模块之后,特征图已经考虑了局部上下文信息,并表示出不同宽高比物体的特征。在这一阶段对小目标和背景之间的全局关系进行建模比在主干中更有效。全局上下文信息可用于表示像素跨空间之间的关系,抑制了无用的背景,增强了对对象和背景之间的区分。因此引入了空间上下文感知模块,通过在每个级别内学习全局空间上下文来进一步增强特征。模块在通道上对输入特征图进行划分,分别通

过模块内部的两个分支，第一分支使用自注意力机制整合全局信息；第二个分支使用前馈神经网络提取根据代表性的高层特征，随后将第一和第二分支进行拼接。网络对应结构如图4所示。

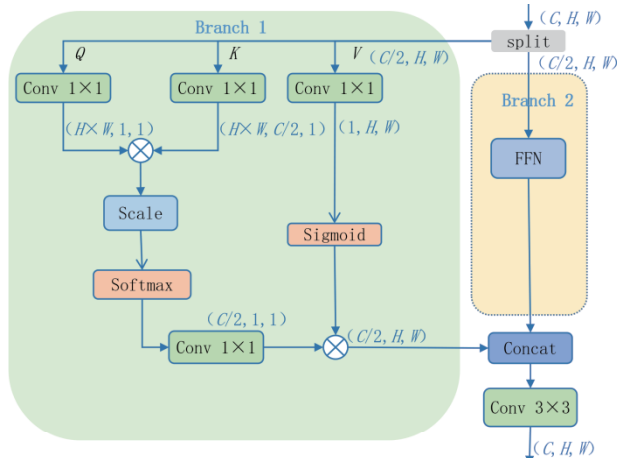


图4 GCAM 模块网络结构图

对于第一个分支，在实践中，首先使用  $1 \times 1$  卷积来执行映射，分别表示自注意力机制中的 Query、Key、Value。将映射得到的  $Q$  与  $K$  进行矩阵乘法，计算不同特征之间的相关性，得到的结果经过尺度缩放和归一化操作后，会生成一组权重，这些权重表示了每个位置相对于当前位置的注意力得分，即自注意力机制中 Value 所对应的权重，将权重与 Value 相乘得到最终输出，向量  $V$  包含每个位置的实际信息，加权求和的过程即根据位置之间的关联程度，聚合相关信息。这一操作简化了广泛使用的自注意力机制，大大降低了参数和计算成本。并且不需要从其他空间位置聚合特征。这种权重加权策略可以有效地融合全局特征，同时减少信息混淆。

$$Q_i^j = \text{FFN}(P_i^j) + a_i^j \sum_{j=1}^{N_i} \left[ \frac{\exp\left(\frac{\omega_q P_i^j \cdot \omega_k P_i^j}{\sqrt{d}}\right)}{\sum_{n=1}^{N_i} \exp\left(\frac{\omega_q P_i^n \cdot \omega_k P_i^n}{\sqrt{d}}\right)} \right] \quad (7)$$

$$a_i^j = \frac{1}{1 + \exp(-w_v P_i^j)} \quad (8)$$

式中： $P_i^j$  和  $Q_i^j$  分别表示  $i$  级特征图中第  $j$  个像素的输入和输出； $N_i$  表示像素总数； $\omega_q$ 、 $\omega_k$  和  $\omega_v$  表示用于投影特征图的线性变换矩阵，通过  $1 \times 1$  卷积简化； $d$  表示尺度缩放因子，通常选取数据的维度，这里选择  $d=256$ ，目的是更好地进行数值运算； $a_i$  表示一个与  $P_i$  和  $Q_i$  形状相同的重新加权矩阵，平衡每个像素的全局空间上下文的聚合程度。

FFN 前馈神经网络能够对输入的图像特征进行非线性变换。遥感图像的特征往往具有更小的尺寸、更复杂的背景。常规的线性模型难以充分捕捉小目标的特征信息。FFN 可以将特征映射到更高维的空间，从而学习到更复杂、更抽象的

特征表示，使模型能够更好地区分不同的目标和背景，以增强模型对于小目标的检测能力。

### 3 实验

#### 3.1 数据集

在常用遥感目标检测数据集中进行了广泛的实验。DOTA-v1.0<sup>[13]</sup> 是一个用于遥感探测的大规模数据集，包含 2 806 幅图像、188 282 个实例和 15 个类别，具有多种方向和尺度。该数据集由 1 411 张、458 张和 937 张分别用于训练、验证和测试的图像组成。

实验在 MMDetection 框架上进行。为了和其他方法对比，使用训练和验证集用于训练，在测试集进行测试。按照之前方法的设置，将原始图像裁剪成  $1\,024 \times 1\,024$  个补丁，DOTA-v1.0 数据集的重叠为 200。为防止过度拟合，图像在训练过程中会按照之前的方法进行随机调整大小和翻转。

#### 3.2 主要结果

DOTA-v1.0 的结果：首先将 FE-FPN 与目前主流的颈部网络在 DOTA v1.0 数据集上进行了比较，分别使用 ResNet-50、ResNet-101、GRA 作为网络的骨干，对不同模型进行了比较分析，并在 DOTA-v1.0 数据集上展示了平均精度 (mAP) 的详细结果。如表 1 所示，单尺度评估表明，与遥感图像检测常用的颈部网络 FPN 相比，平均精度提高了 1.84%~2.50%。

表1 与常用的颈部网络进行精度比较

| BACK       | NECK       | mAP/% |
|------------|------------|-------|
| ResNet-50  | FPN-Carafe | 72.66 |
|            | FPN        | 73.49 |
|            | FE-FPN     | 75.79 |
| ResNet-101 | FPN-Carafe | 73.75 |
|            | FPN        | 73.80 |
|            | FE-FPN     | 75.64 |
| GRA        | FPN-Carafe | 74.45 |
|            | FPN        | 74.71 |
|            | FE-FPN     | 77.55 |

在常用的遥感图像数据集中，本文方法始终超越了以前最先进的方法，证明了其通用性和有效性。

### 4 结论

本文提出了一种简单有效的特征增强金字塔网络 (FE-FPN) 用于遥感图像的旋转目标检测，其中提出的多分支特征增强模块通过多分支结构，应用条带卷积 StripConv，通过一种轻量级的有限元方法来提高网络的局域感知能力，以捕捉不同宽高比对象的特征；提出的全局上下文感知模块通过空间自注意力机制及前馈神经网络，提高全局关联跨信道和

空间的能力以捕获小目标周围的语义信息。探索了在大尺寸图像上进行检测的方法,以获得更好的速度和准确性权衡。大量的实验表明,FE-FPN可以在常用遥感图像数据集DOTA中得到较大的精度提升。

#### 参考文献:

- [1] Li Weijie, Yang Wei, Hou Yuenan, et al. SARATR-X: towards building a foundation model for SAR target recognition[J]. IEEE Transactions on Image Processing, 2024, 34: 869-884.
- [2] Li Weijie, Yang Wei, Liu Tianpeng, et al. Predicting gradient is better: exploring self-supervised learning for SAR ATR with a joint-embedding predictive architecture[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2024, 218: 326-338.
- [3] Yang Xue, Yan Junchi. Arbitrary-oriented object detection with circular smooth label[C]//Computer Vision-ECCV 2020. Berlin: Springer, 2020: 677-694.
- [4] Yang Xue, Yan Junchi, Ming Qi, et al. Rethinking rotated object detection with gaussian wasserstein distance loss[C]//Proceedings of the 38<sup>th</sup> International Conference on Machine Learning. Berlin: Springer, 2021: 11830-11841.
- [5] Yang Xue, Yang Xiaojiao, Yang Jirui, et al. Learning high-precision bounding box for rotated object detection via kullback-leibler divergence[PP/OL]. (2022-04-18)[2026-04-27]. <https://doi.org/10.48550/arXiv.2106.01883>.
- [6] Cai Xinhao, Lai Qiuxia, Wang Yuwei, et al. Poly kernel inception network for remote sensing detection[PP/OL]. (2024-03-20)[2026-04-27]. <https://doi.org/10.48550/arXiv.2403.06258>.
- [7] Li Wentong, Chen Yijie, Hu Kaixuan, et al. Oriented reppoints for aerial object detection [PP/OL]. (2022-03-24)[2026-04-27]. <https://doi.org/10.48550/arXiv.2105.11111>.
- [8] Chen Yuming, Yuan Xinbin, Wu Ruiqi, et al. YOLO-MS: rethinking multi-scale representation learning for real-time object detection[PP/OL]. (2025-02-25)[2026-04-27]. <https://doi.org/10.48550/arXiv.2308.05480>.
- [9] Xia Guisong, Bai Xiang, Ding Jian, et al. DOTA: a large-scale dataset for object detection in aerial images[C/OL]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018[2026-04-27]. <https://ieeexplore.ieee.org/document/8578516>. DOI: 10.1109/CVPR.2018.00418.
- [10] Ding Jian, Xue Nan, Long Yang, et al. Learning RoI transformer for oriented object detection in aerial images[C/OL] 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019[2026-04-27]. <https://ieeexplore.ieee.org/document/8953881>. DOI: 10.1109/CVPR.2019.00296.
- [11] Xiao Zikai, Yang Guoye, Yang Xue, et al. Theoretically achieving continuous representation of oriented bounding boxes[C]// 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2024: 16912-16922.
- [12] Xie Xingxing, Cheng Gong, Wang Jiabao, et al. Oriented R-CNN for object detection[C/OL]// 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2021[2026-04-27]. <https://ieeexplore.ieee.org/document/9710901>. DOI: 10.1109/ICCV48922.2021.00350.
- [13] Trautner R, Vitulli R. Ongoing developments of future payload data processing platforms at ESA[C]//International Conference on On-Board Payload Data Compression (OBPDC). Noordwijk, The Netherlands: ESA/ESTEC, 2010: 1-12.

#### 【作者简介】

陶建(1999—),男,四川内江人,硕士,研究方向:计算机视觉, email: 2023222055154@stu.scu.edu.cn。

严华(1971—),通信作者(email: 4429175@qq.com),男,四川渠县人,博士,博士生导师、教授,研究方向:智能算法、存储系统、并行计算。

(收稿日期: 2025-09-03 修回日期: 2026-05-06)