面向长尾分布的遥感图像细粒度目标检测

周波[1, 2]，杨明珠[1, 2]，周奥临[1, 2]，张若楠[3]，刘立波[1, 2*]
（1. 宁夏大学信息工程学院， 宁夏银川 750021；2. 宁夏“东数西算”人工智能与信息安全重点实验室， 宁夏银川 750021；
3. 宁夏大学信息与网络空间安全学院， 宁夏中卫 755000）

摘　要：遥感图像细粒度目标检测在军事侦察、机场监测及资源勘查等领域具有重要应用价值。然而，现有方法仍存在尾部类别样本学习不充分、多尺度融合过程中细节信息易丢失以及类间判别表征不足等问题。针对上述问题，本文提出一种面向长尾分布的遥感图像细粒度目标检测方法（LT-FGRSOD， Fine-Grained Remote Sensing Object Detection with Long-Tailed Distribution）。首先，设计类别感知的动态重采样策略（CADRS， Class-Aware Dynamic Resampling Strategy），结合类别重复因子与图像内实例数量设置采样优先级，增加尾部类别的采样次数，同时控制头部类别的重复采样，并通过随机颜色扰动增强样本多样性；其次，构建邻层自适应特征融合模块（ALAF， Adjacent-Level Adaptive Feature Fusion Module），通过相邻层级特征对齐与门控融合，联合浅层细节信息与深层语义信息，增强模型对细粒度旋转目标的表征能力；最后，引入类间对比判别模块（CCDM， Inter-class Contrastive Discriminative Module），在 RoI 嵌入空间构建高效对比组，保留判别约束同时降低头部类别冗余对比、增强类内紧凑性和类间可分性，从而提升长尾分布下的细粒度目标检测性能。为验证所提方法的有效性，在 MAR20 和 ShipRSImageNet 数据集上进行实验，结果显示，所提方法的 mAP_{50} 分别达到 89.5% 和 78.2%，较最优方法　PETDet 分别提升 4.1 个百分点和 2.4 个百分点。表明该方法能够有效提升长尾分布下遥感图像细粒度目标检测性能。

关键词：遥感图像；目标检测；长尾分布；多尺度特征融合
中图分类号：TP391.41
文献标志码：A
doi：10.11959/j.issn.2096−3750.WLW26154

Fine-grained remote sensing object detection with long-tailed distribution

ZHOU Bo[1,2], YANG Mingzhu[1,2], ZHOU Aolin[1,2], ZHANG Ruonan[3], LIU Libo[1,2*]
1. School of Information Engineering, Ningxia University, Yinchuan 750021, China
2. Ningxia Key Laboratory of Artificial Intelligence and Information Security for East-Data-West-Computing, Yinchuan 750021, China
3. School of Information and Cyberspace Security, Ningxia University, Zhongwei 755000, China

Abstract: Fine-grained object detection in remote sensing images is important for military reconnaissance, airport monitoring, and resource exploration. However, existing methods suffer from insufficient learning of tail classes, detail loss during multi-scale feature fusion, and weak inter-class discrimination. To address these issues, this paper proposes a fine-grained remote sensing object detection method for long-tailed distributions, termed LT-FGRSOD. First, a class-aware dynamic resampling strategy (CADRS) is designed. It sets sampling priority by class repetition factor and intra-image instance count, increases tail-class sampling frequency while curbing repeated head-class sampling, and enhances sample di-

收稿日期：XXXX−XX−XX；修回日期：XXXX−XX−XX
通信作者：刘立波，liulib@nxu.edu.cn
基金项目：宁夏自然科学基金（No. 2024AAC02010）；宁夏科技创新领军人才项目（No. 2022GKLRLX03）；银川市科技项目（No. 2025RC09）
Foundation Items: Natural Science Foundation of Ningxia, China（No. 2024AAC02010）, Science and Technology Leading Talent Training Projects of Ningxia (No. 2022GKLRLX03), Yinchuan City Science and Technology Projects（No. 2025RC09）

versity via random color perturbation. Second, an adjacent-level adaptive feature fusion module (ALAF) is developed. It dynamically adjusts adjacent-level fusion weights and combines shallow detail with deep semantics through feature alignment and gated fusion, enhancing the representation of fine-grained rotated objects. Finally, an inter-class contrastive discriminative module (CCDM) is introduced. It constructs efficient contrastive groups in the RoI embedding space, preserving discriminative constraints while reducing redundant contrast for head classes. This enhances intra-class compactness and inter-class separability. Experiments on MAR20 and ShipRSImageNet show that the proposed method achieves mAP_{50} scores of 89.5% and 78.2%, respectively, improving by 4.1 and 2.4 percentage points over PETDet. These results demonstrate its effectiveness for long-tailed fine-grained object detection in remote sensing images.

Key words: Remote sensing images; Object detection; Long-tailed distribution; Multi-scale feature fusion

0　引言

遥感图像细粒度目标检测不仅需要完成目标定位，还需要进一步识别物体的子类别。然而，遥感数据集中不同类别的样本数量差异显著，头部类别拥有大量标注样本而尾部类别样本有限，呈现明显的长尾分布。均匀采样导致训练样本以头部类别为主，尾部类别监督不足，其特征表示不稳定，分类边界也容易模糊。因此，如何缓解长尾分布带来的训练偏置、提升尾部类别的特征表征能力与类别判别能力，是遥感图像细粒度目标检测中亟待解决的重要问题。

长尾分布是遥感数据集的典型特征，现有研究主要从重采样和损失重加权两个方面展开。重采样方法通过改变训练样本的类别构成，增加尾部类别参与训练的机会。Gupta 等在大词汇实例分割（LVIS，Large Vocabulary Instance Segmentation）基准数据集上的相关实验表明，基于类别频次的重复因子采样策略，通过适当提高低频类别样本的采样频率，有效缓解了长尾分布带来的训练偏置[1]；Ouyang 等在 PCLDet[2] 中设计类别均衡采样器（CBS，Class-Balanced Sampler），在候选区域采样阶段平衡不同类别的采样数量，增加尾部类别参与训练的机会。损失重加权方法则通过调整各类别损失的相对权重，抑制头部类别对模型训练的过度支配，从而缓解梯度失衡。Lin 等提出焦点损失（Focal Loss）[3]，通过抑制易分类样本的损失权重，强化模型对困难样本的学习能力；Tan 等提出均衡损失（Equalization Loss）[4]，将长尾检测中的梯度失衡问题显式引入损失设计，进一步改善了模型对尾部类别的学习效果。Wang 等提出 Seesaw Loss[5]，通过缓解因子和补偿因子动态调整各类别的正负梯度贡献，改善长尾分布下的类别学习效果。然而，上述方法虽有效缓解了长尾分布带来的训练偏置，但主要针对粗粒度类别的样本不均衡问题，难以刻画同一大类下相似子类之间的细微差异。随着遥感成像分辨率的提高，图像中目标的纹理、形状和部件结构等细节信息更加丰富，仅区分目标大类已难以支撑精细化分析，细粒度识别成为该任务的关键环节。

细粒度目标检测在完成目标定位与粗粒度分类的基础上，进一步判别目标所属的子类别，如飞机型号、舰船类型。该任务更依赖局部形态、纹理与部件结构等细节信息，为此，已有研究主要从多尺度特征表示与判别约束设计两方面提升细粒度判别能力。Lin 等提出特征金字塔网络（FPN，Feature Pyramid Network）[6]，通过自顶向下的特征传递路径融合深层语义信息与浅层细节信息；Liu 等提出路径聚合网络（PANet，Path Aggregation Network）[7]，进一步增强自底向上的特征传递能力；Woo 等提出卷积块注意力模块（CBAM，Convolutional Block Attention Module）[8]，通过通道注意力和空间注意力突出关键特征并抑制背景干扰。除多尺度特征表示外，判别约束同样影响细粒度目标的识别效果。Schroff 等提出三元组损失（Triplet Loss）[9]，通过构造三元组样本关系，在压缩类内距离的同时增大类间间隔，从而增强特征判别性。Wen 等提出中心损失（Center Loss）[10]，通过引入类别中心约束，提高了同类样本的特征紧凑性。van den Oord 等提出对比预测编码（CPC，Contrastive Predictive Coding）[11]框架，并采用 InfoNCE 损失进行对比表示学习，为后续对比学习方法提供了重要基础。Li 等提出候选区域增强两阶段细粒度检测器（PETDet，Proposal Enhancement for Two-stage Fine-grained Object Detection）[12]，分别从原型对比建模与候选区域增强等方面提升了模型对细

粒度类别的区分能力。上述方法在细粒度识别方面已取得明显进展，但普遍基于类别分布均衡的数据。尾部子类样本稀少时，其判别能力难以保证，现有方法将细粒度识别与长尾目标检测相结合的研究仍不充分。

为此，本文提出 LT-FGRSOD 检测框架，从样本重平衡、特征融合与判别表征三个层面解决上述问题。该框架在缓解长尾分布带来的训练偏置的同时，增强模型对相似子类别的区分能力。首先，采用类别感知的动态重采样策略调整训练样本的类别构成，缓解长尾分布带来的样本不均衡；其次，构建邻层自适应特征融合模块，联合浅层细节与深层语义信息，提取更具判别性的多尺度特征；最后，引入类间对比判别模块，在嵌入空间拉开类间距离、增强类内紧凑性。上述模块在同一框架内联合处理样本重平衡与细粒度特征判别，以同时缓解类别不均衡并提升目标子类别的判别能力。主要贡献如下：

1. 提出类别感知的动态重采样策略（CADRS, Class-Aware Dynamic Resampling Strategy），通过类别重采样与颜色扰动，增强尾部样本的采样频次和多样性，缓解长尾分布下尾部样本学习不足的问题。

2. 设计邻层自适应特征融合模块（ALAF, Adjacent-Level Adaptive Feature Fusion Module），通过特征对齐和门控融合，结合浅层细节与深层语义信息，增强相似子类别的特征表征能力。

3. 构建类间对比判别模块（CCDM, Inter-class Contrastive Discriminative Module），并在 RoI 嵌入空间构建高效对比组，压缩类内方差、扩大类间边界，增强模型对易混淆类别的判别能。

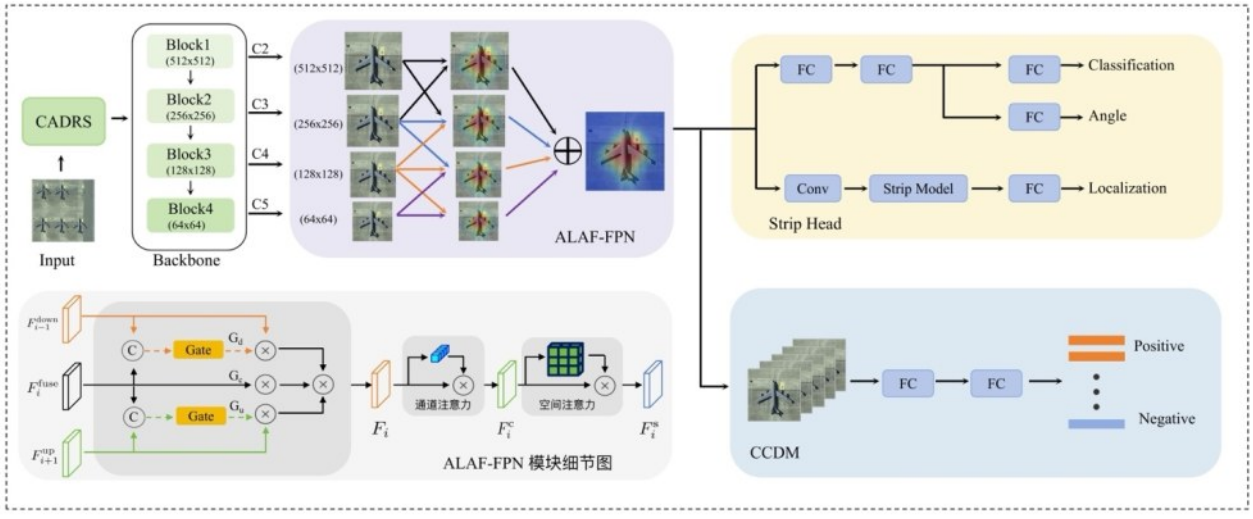

图 1　LT-FGRSOD 模型结构

1　模型与方法

1.1　网络模型概述

Strip R-CNN[13] 是一种两阶段目标检测模型，该方法基于面向旋转目标的区域卷积神经网络（Oriented R-CNN, Oriented Region-based Convolutional Neural Network）[14]构建而成，整体结构包括主干网络[15]、颈部网络和检测头三部分。不同于常见的大核方形卷积，Strip R-CNN 采用“小方核卷积与正交大条带卷积”串行设计。该设计在扩大单方向感受野的同时，减少特征冗余和背景干扰，更好地表征目标的轮廓信息。此外，检测头采用分类-定位-角度解耦的 Strip Head，并在定位分支中引入条带卷积，增强了对高宽比目标边界形状和方向变化的建模能力。在此基础上，本文提出了面向长尾分布的遥感图像细粒度目标检测方法（LT-FGRSOD, Fine-Grained Remote Sensing Object Detection with Long-Tailed Distribution），整体结构如图 1 所示。

1.2　类别感知的动态重采样策略

在遥感图像细粒度目标检测任务中，数据集通常呈现明显的长尾分布，不同类别样本数量差异较大。若采用图像级均匀采样，训练过程易被头部类别主导，尾部类别难以获得充分监督，导致其召回

率与分类精度偏低。直观上，增加尾部类别图像的采样次数可提高尾部类别的出现频次；然而，遥感图像中目标分布密集，单幅图像常包含多个类别的实例；重复采样含尾部类别的图像时，其中的头部类别实例会被采样，因此，重采样策略需要在提高尾部类别曝光次数的同时，抑制头部类别实例的附带重复。

针对上述问题，本文提出类别感知的动态重采样策略（CADRS）。该策略先在类别层面依据类别频率确定各类别的重复因子与目标采样次数，再在图像层面依据图像的实例组成确定采样优先级，优先选择包含尾部类别且头部类别实例较少的图像，以增加尾部类别的训练机会，并抑制头部类别实例的附带重复。本文定义与类别频率呈负相关的类别重复因子 R_c，并引入频率阈值 t 以区分尾部类别与头部类别：

$$R(c)=\begin{cases}\sqrt{\dfrac{t}{F(c)}}, & F(c)<t\\ 1, & F(c)\geqslant t\end{cases} \tag{1}$$

式中，t 为频率阈值，本文设置 t=0.15。当类别 c 的出现频率 $F(c)$ 低于阈值 t 时，将其判定为尾部类别并提高其重复因子；否则，将其视为头部类别并将重复因子截断为 1。

然而，仅提高尾部类别的采样频率容易导致模型反复学习相同的样本，从而引发过拟合。为此，该方法在动态重采样模块中集成了随机颜色扰动机制（Random Color Perturbation）。其核心思想是对被重采样的尾部图像分别施加独立的随机变换，使得同一张图像在不同 epoch 中在特征空间中的映射点均有所不同，从而扩充尾部类别的样本多样性。其变换过程如下：

$$\begin{aligned}f_b(\mathbf{x};\delta_{\mathrm{b}})&=\mathrm{clip}(\mathbf{x}+\delta_{\mathrm{b}},0,255)\\ f_c(\mathbf{x}';\delta_{\mathrm{c}})&=\mathrm{clip}(\delta_{\mathrm{c}}\cdot\mathbf{x}',0,255)\end{aligned} \tag{2}$$

其中 $\delta_c\sim\mathcal{U}(1-\Delta_c,1+\Delta_c)$，$\Delta_c$ 为对比度缩放比例。为调整饱和度和色相，先将图像从 RGB 空间映射到 HSV 空间：

$$[H,S,V]=\mathcal{H}(\mathbf{x}'') \tag{3}$$

随后在 HSV 空间中对 S 和 H 通道分别施加乘性和加性扰动：

$$\begin{aligned}S_{\mathrm{new}}&=\mathrm{clip}(S\cdot\delta_s,0,1)\\ H_{\mathrm{new}}&=(H+\delta_h)\bmod 360\end{aligned} \tag{4}$$

其中 $\delta_s\sim\mathcal{U}(1-\Delta_s,1+\Delta_s)$，$\delta_h\sim\mathcal{U}(-\Delta_h,\Delta_h)$。最终将增强后的 HSV 图像映射回 RGB 空间，得到增强后的图像：

$$\mathbf{x}_{\mathrm{aug}}=\mathcal{H}^{-1}([H_{\mathrm{new}},S_{\mathrm{new}},V]) \tag{5}$$

通过上述方式，网络在训练过程中能够减少对固定色彩模式的依赖，转而学习对颜色扰动更鲁棒的特征，从而缓解模型对尾部样本特征学习不足的问题。

1.3　邻层自适应特征融合模块

在细粒度识别中，相似子类别之间的差异通常体现在目标的局部结构、纹理及部件形态等细节，因此，多尺度特征融合需要兼顾浅层细节与深层语义信息。然而，传统特征金字塔网络在跨层信息传递与融合过程中，容易弱化目标的局部结构与纹理细节，难以充分保留并增强区分相似子类别所需的细粒度特征。为此，本文在 LT-FGRSOD 的颈部网络中引入邻层自适应特征融合模块（ALAF）。该模块以主干网络输出的多层级特征 C2、C3、C4、C5 为输入，通过相邻层特征对齐与门控融合，结合浅层细节与深层语义信息，并利用通道—空间协同注意力强化目标关键区域的结构与纹理响应，从而增强相似子类别的特征表征能力。

设当前层级特征图为 F_i，其空间尺寸为 $H_i\times W_i$，通道数为 C_i。为保证不同层级特征在融合前具有一致的空间分辨率和通道维度，需要分别对相邻浅层特征 F_{i-1} 和相邻深层特征 F_{i+1} 进行尺度与通道对齐，具体过程表示为：

$$\begin{aligned}F_{i-1}^{\mathrm{down}}&=\mathrm{Conv}_{1\times1}(\mathrm{Down}(F_{i-1}))\\ F_{i+1}^{\mathrm{up}}&=\mathrm{Conv}_{1\times1}(\mathrm{Up}(F_{i+1}))\end{aligned} \tag{6}$$

其中，上采样与下采样均采用最近邻插值，将相邻层特征调整至当前层的空间尺寸，再通过 1×1 卷积匹配通道数。完成匹配后，相邻层特征与当前层特征具有相同的空间尺寸和通道维度，进行拼接与逐元素运算。

与单向自顶向下的逐层传递不同，ALAF 仅在相邻层之间建立动态对齐路径，以减少冗余信息并保留必要的层间语义联系；最顶层和最底层特征仅与唯一相邻层融合。完成特征对齐后，ALAF 将当前层特征与对齐后的相邻层特征按通道拼接，再通过门控单元生成三个像素级权重图 G_d、G_u 和 G_c，分别对应浅层信息、深层信息以及当前层自身信息。具体公式如下：

$$\begin{aligned} G_{\mathrm{d}} &= \sigma\left(L_{\mathrm{gate}}\left(\mathrm{concat}\left(F_i, F_{i-1}^{\mathrm{down}}\right)\right)\right) \\ G_{\mathrm{u}} &= \sigma\left(L_{\mathrm{gate}}\left(\mathrm{concat}\left(F_i, F_{i+1}^{\mathrm{up}}\right)\right)\right) \\ G_{\mathrm{c}} &= 1 - \max\left(G_{\mathrm{d}}, G_{\mathrm{u}}\right) \end{aligned} \tag{7}$$

门控网络沿通道维拼接当前层与已对齐的邻层特征，使权重随样本、通道与空间位置自适应变化；同一层的上下邻层分支共享参数，不同层分别学习。由此得到多层级融合特征：

$$F_i^{\mathrm{fuse}} = G_{\mathrm{d}} \odot F_{i-1}^{\mathrm{down}} + G_{\mathrm{u}} \odot F_{i+1}^{\mathrm{up}} + G_{\mathrm{c}} \odot F_i \tag{8}$$

其中，$\odot$ 表示逐元素乘法。门控融合完成后，模型进一步引入通道-空间协同注意力机制。在通道注意力分支中，对 F_i^{fuse} 分别施加全局平均池化和全局最大池化，以聚合空间上下文信息，从而生成通道注意力权重 M_i^c，并得到通道增强特征：

$$F_i^{\mathrm{c}} = M_i^{\mathrm{c}} \odot F_i^{\mathrm{fuse}} \tag{9}$$

随后，在空间注意力分支中，对 F_i^c 沿通道维进行平均池化与最大池化，以生成空间注意力权重 M_i^s，进而得到最终的增强特征：

$$F_i^{\mathrm{s}} = M_i^{\mathrm{s}} \odot F_i^{\mathrm{c}} \tag{10}$$

式中，F_i^s 表示当前层特征经过通道注意力与空间注意力串联增强后的特征表示。通过上述过程，ALAF 模块在门控融合的基础上进一步突出目标关键区域并抑制背景噪声干扰，从而获得更具判别性的多尺度融合特征。

1.4　类间对比判别模块

传统两阶段检测框架主要依赖交叉熵损失优化分类分支，对易混淆类别的判别能力有限；长尾分布下，尾部类别样本不足，而交叉熵监督缺少对 RoI 特征分布结构的显式约束，容易导致类内特征分散、类间边界模糊。对比学习通过在嵌入空间中约束样本间的相似度，从而增强特征判别性。然而，现有方法通常在同一批次的所有前景 RoI 之间两两计算对比损失，计算开销较大；同时，长尾分布下尾部类别可用的正样本对稀少，损失容易受头部类别主导。为此，本文引入类间对比判别模块（CCDM, Inter-class Contrastive Discriminative Module），在独立的 RoI 嵌入空间中，按类别为每个 RoI 采样固定数量的正、负样本组成对比组，并在组内计算对比损失。具体而言，给定 RoI 特征 v_i，首先通过独立的非线性投影头（Projection Head）将其映射至嵌入空间。其映射形式为：

$$z_i = \mathcal{H}(v_i) = W^{(2)}\phi\left(W^{(1)}v_i + b^{(1)}\right) + b^{(2)} \tag{11}$$

其中，$W^{(1)} \in \mathbb{R}^{D \times D}$、$W^{(2)} \in \mathbb{R}^{d \times D}$ 分别表示投影头中两层线性变换矩阵，ϕ 为 ReLU[16]激活函数。随后，对投影头输出实施 L_2 归一化，将其映射至单位超球面：

$$\hat{z}_i = \frac{z_i}{\left\| z_i \right\|_2} \tag{12}$$

在归一化嵌入空间中，本文以每个前景 RoI 为锚点，从同类 RoI 中取 1 个作正样本、从异类 RoI 中取 3 个作负样本，与锚点共同构成对比组。若锚点没有同类正样本或没有异类负样本，该锚点不参与对比损失的计算，仅依靠分类与回归损失完成训练。对比损失定义如下：

$$\mathcal{L}_{\mathrm{con}}^{i} = -\log \frac{\exp\left(\hat{z}_i^{\mathrm{T}} \hat{z}_{s_i}^{+} / \tau\right)}{\sum_{a \in A_i} \exp\left(\hat{z}_i^{\mathrm{T}} \hat{z}_a / \tau\right)} \tag{13}$$

其中，$\hat{z}_{s_i}^{+}$ 表示从正样本集合 S_i^+ 中采样得到的正样本特征，A_i 表示由采样得到的正样本和负样本共同构成的对比集合，τ 为温度参数，本文设为 0.07。整体目标函数定义为：

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{cls}} + \mathcal{L}_{\mathrm{reg}} + \lambda \mathcal{L}_{\mathrm{con}} \tag{14}$$

其中，L_{cls} 和 L_{reg} 分别表示分类交叉熵损失和旋转框回归 Smooth L1[17] 损失，L_{con} 表示类间对比损失，λ 为其权重系数，设为 0.1。该联合损失同时约束目标分类、边界框回归和 RoI 嵌入特征分布，使同类目标特征更加紧凑、异类目标特征间隔更大，从而提升细粒度类别的判别能力。

2　实验结果与分析

2.1　数据集

实验采用 MAR20[18] 和 ShipRSImageNet[19] 两个遥感图像细粒度目标检测数据集，覆盖军事飞机和舰船两类典型遥感目标。如图 2 所示，MAR20 和 ShipRSImageNet 均呈现典型的长尾分布特征，模型训练过程容易受头部类别主导，导致尾部类别特征学习不足，从而影响模型对少样本细粒度类别的识别能力与泛化性能。

MAR20 数据集由 Yu 等（2023）发布，是面向遥感图像军事飞机识别与检测的公开基准数据集。训练集包含 1331 张图像和 7870 个目标实例，测试集包含 2511 张图像和 14471 个目标实例。该数据集同时提供水平框与旋转框标注。

ShipRSImageNet 数据集由 Zhang 等（2021）

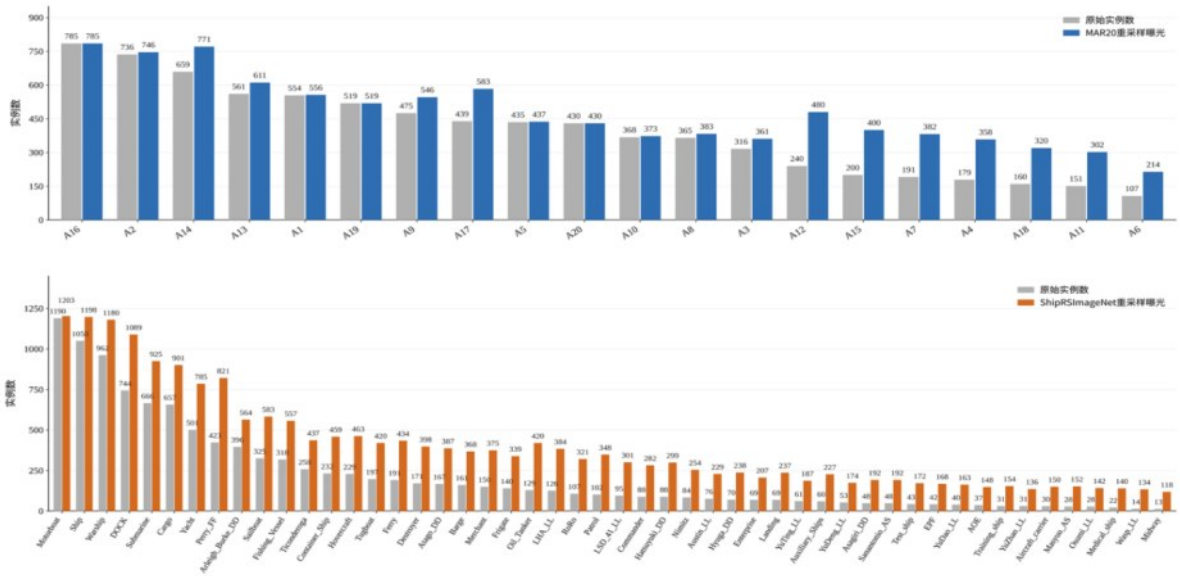

图 2　MAR20 与 ShipRSImageNet 数据集训练样本分布图

发布，是遥感领域具有代表性的高分辨率舰船细粒度目标检测数据集之一。该数据集涵盖近岸港口、远海航行及复杂海况等多样化场景，其中训练集包含 2198 张图像和 10860 个舰船实例，涵盖 50 个细粒度舰船类别。

2.2　实验设置

所有实验均基于 MMRotate[20] 框架，在 4 张 NVIDIA GeForce RTX 3090 GPU（24 GB 显存）上进行训练与测试。训练阶段单卡 GPU 的批次大小为 2，经多卡并行后总批次大小为 8；优化器采用 AdamW[21]，初始学习率设为 1×10^{-4}，动量参数取 $\beta_1 = 0.9$、$\beta_2 = 0.999$，权重衰减系数为 0.05。骨干网络采用 ImageNet[22] 预训练权重初始化，训练采用 3x 调度策略，共计 36 个 epoch，并在第 24 轮和第 33 轮将学习率按 0.1 系数衰减。

2.3　评价指标

为全面评估 LT-FGRSOD 模型在遥感图像细粒度目标检测任务中的性能，本文从检测精度和模型复杂度两个方面进行评价。采用平均精度（Average Precision，AP）及其均值（mean Average Precision，mAP）作为主要评价指标，并以 IoU 阈值为 0.5 时的 mAP 作为主要性能指标，以分析模型在不同类别上的检测性能。在模型复杂度方面，采用参数量（Params）和计算量（GFLOPs）衡量模型的参数规模与计算开销，并结合检测精度分析模型的综合性能。相关评价指标的计算公式如下：

$$
\begin{aligned}
P &= \frac{TP}{TP + FP} \\
R &= \frac{TP}{TP + FN} \\
\mathrm{mAP} &= \frac{1}{K}\sum_{i=1}^{K} AP_i
\end{aligned}
\tag{15}
$$

式中，TP、FP 和 FN 分别表示真正例、假正例和假反例；P 和 R 分别表示精确率与召回率。通过改变置信度阈值可得到精确率-召回率曲线，并进一步计算各类别 AP 及整体 mAP。

2.4　对比实验

本文在 MAR20 数据集上与 Faster R-CNN[23]、R3Det[24]、RoI Transformer[25]、Gliding vertex[26]、PCLDet[2] 和 PETDet[12] 等多种典型方法[27]进行对比实验，结果如表 1 所示。LT-FGRSOD 在所有对比方法中取得最优性能。具体而言，LT-FGRSOD 的 mAP_{50} 达到 89.5%，相比当前性能最强的 PETDet，提升 4.1 个百分点；相比 Strip R-CNN，进一步提升了 4.8 个百分点，表明所提出方法能够有效提升模型在遥感场景下的整体目标检测性能。

从各类别的检测结果来看，LT-FGRSOD 在 20 个飞机类别上均表现出较强的检测能力。对于 B-52、P-3C 和 FA-18 等外观特征较为明显的类别，检测精度均超过 89%。对于类别间外观相似、识别难度较高的目标，LT-FGRSOD 同样表现出较强的细粒度类别判别能力。其中，SU-24 和 SU-34 的 AP_{50} 分别达到 89.1% 和 88.5%，较 PETDet 分别提升 8.4% 和 3.7%；对于 F-15、F-22 和 KC-10，其

表 1　MAR20 数据集各类别 AP50 对比结果

Method	SU-35	C-130	C-17	C-5	F-16	TU-160	E-3	B-52	P-3C	B-1B	E-8	TU-22	F-15	KC-135	F-22	FA-18	TU-95	KC-10	SU-34	SU-24	mAP_{50}
R3Det	70.5	80.8	49.5	46.6	57.4	79.3	53.8	83.4	22.1	78.9	50.2	82.3	59.0	77.6	33.9	86.5	89.2	9.5	72.2	73.6	62.9
Faster R-CNN	85.2	89.9	88.9	80.1	75.4	89.1	90.2	89.6	90.1	90.9	86.5	87.6	67.1	88.7	47.2	88.4	90.4	66.0	77.6	76.6	82.3
RoI Transformer	87.0	81.7	88.7	79.3	78.2	90.7	90.1	89.7	89.2	90.9	85.5	88.6	67.8	88.8	45.2	89.0	90.6	71.4	84.4	78.5	82.8
Gliding-V	82.9	80.4	86.3	68.5	72.7	89.4	89.7	86.7	90.4	88.4	82.5	85.4	64.9	79.2	51.9	87.2	89.9	54.6	77.6	71.2	79.0
ORCNN	86.4	88.2	88.6	75.5	79.8	90.6	90.4	89.5	89.4	90.9	88.0	87.5	65.9	87.9	49.6	89.2	90.4	72.5	83.2	79.0	83.1
PCLDet	85.5	81.7	89.4	78.2	79.8	90.9	90.5	87.4	89.9	90.8	87.2	89.3	66.1	88.0	52.4	88.7	90.4	73.4	82.0	79.4	83.1
Strip R-CNN	87.4	88.4	91.0	80.5	79.3	90.3	90.2	89.9	88.9	90.9	85.8	90.5	70.5	88.4	64.6	88.8	90.4	73.5	84.2	79.4	84.7
PETDet	88.2	89.2	89.9	82.9	82.7	91.0	90.7	89.0	90.1	90.9	87.8	90.3	73.2	89.3	61.1	88.8	90.6	78.0	84.8	80.7	85.4
Ours	88.6	90.8	93.0	81.7	87.5	90.9	90.9	91.4	90.9	90.9	88.7	90.6	87.3	90.3	85.5	89.6	91.8	92.0	88.5	89.1	89.5

AP_{50} 分别达到 87.3%、85.5% 和 92.0%，提升幅度分别为 14.1%、24.4% 和 14.0%。结果表明，LT-FGRSOD 在外观相似、样本数量较少的类别上具有较强的特征辨别能力，能够有效缓解长尾分布下细粒度类间混淆对检测性能造成的影响。

此外，为全面评估模型的泛化能力，本文进一步在舰船细粒度目标检测数据集 ShipRSImageNet 上开展实验，并与多种代表性方法[28]进行比较。结果如表 2 所示，在 ShipRSImageNet 数据集上，LT-FGRSOD 的 mAP_{50} 与 $mAP_{50:90}$ 分别达到 78.2% 和 58.3%，较次优方法 PETDet 分别提升 2.4 和 2.6 个百分点。在模型复杂度方面，LT-FGRSOD 体现出较优的综合性能。，该模型参数量为 45.1 M，较 PETDet（47.7 M）减少 5.5%；计算量为 187.6 GFLOPs，较 PETDet（204.1 GFLOPs）降低 8.1%。与基线 Strip R-CNN 相比，LT-FGRSOD 的参数量增加 2.6 M，计算量降低 22.8 GFLOPs，并取得了更高的检测精度。上述结果表明，所提方法不仅在飞机细粒度目标检测场景中有效，在舰船细粒度目标检测任务中同样具有较强的泛化能力。

表 2　ShipRSImageNet 数据集上复杂度与精度对比

方法	Params (M)	GFLOPs	mAP_{50}	$mAP_{50:90}$
Gliding-V	**41.1**	211.3	58.7	28.7
RoI - Transformer	55.0	225.2	64.1	33.6
Oriented - RCNN	43.7	198.5	71.8	51.9
Strip R-CNN	42.5	210.4	73.2	52.1
PETDet	47.7	204.1	75.8	55.7
Ours	45.1	**187.6**	**78.2**	**58.3**

2.5　消融实验

本节在 MAR20 数据集上进行消融实验，以验证 CADRS、ALAF 和 CCDM 三个模块的有效性。所有对比实验均采用相同的实验设置，以保证实验结果的公平性。结果如表 3 所示。基线模型的 mAP_{50} 为 84.71%，单独引入 CADRS、ALAF 和 CCDM 后分别提升至 86.85%、85.78% 和 85.23%，表明三个模块均具有独立增益，其中 CADRS 的提升最为明显。在此基础上，进一步验证不同模块组合的有效性。其中，CADRS+CCDM 的 mAP_{50} 达到 87.59%，较基线提升 2.88 个百分点；CADRS+

ALAF 的 mAP_{50} 达到 88.73%，提升 4.02 个百分点；ALAF+CCDM 的 mAP_{50} 达到 87.21%，提升 2.50 个百分点。三模块联合时模型 mAP_{50} 取得最优，达到 89.50%，较基线提升 4.79 个百分点。表明样本再平衡、多尺度特征增强和对比判别模块能够形成有效互补，共同提升长尾分布下模型的检测性能。

表 3　MAR20 数据集消融实验结果

CADRS	ALAF	CCDM	mAP_{50}
-	-	-	84.71
□	-	-	86.85
-	□	-	85.78
-	-	□	85.23
□	-	□	87.59
□	□	-	88.73
-	□	□	87.21
□	□	□	89.50

2.6　负样本采样策略敏感性分析

为验证负样本构建策略对 CCDM 判别性能的影响，本文在 MAR20 数据集上比较了不同负样本数量与采样方式，结果如表 4 所示。随机抽取 1、3、7 个负样本时，mAP_{50} 分别为 87.1%、88.5% 和 87.2%，$mAP_{50:90}$ 分别为 64.2%、64.6% 和 63.7%。负样本过少时，类间对比信息不足，难以形成有效的判别约束；负样本过多则会引入较多易区分样本，稀释对比损失的有效梯度并造成冗余对比，因此每个正样本抽取 3 个负样本时性能最优。在采样方式方面，本文进一步实现了基于余弦相似度的困难负样本挖掘，即对每个正样本选取嵌入相似度最高的 3 个异类样本作为负样本，其 mAP_{50} 和 $mAP_{50:90}$ 分别为 88.1% 和 64.5%，与随机抽取的 88.5% 和 64.6% 基本持平，未表现出额外优势。综合检测性能与实现复杂度，CCDM 最终采用随机抽取 3 个负样本的构建方式。

表 4　CCDM 负样本策略消融结果

方法	负样本	mAP_{50}	$mAP_{50:90}$
随机	1	87.1	64.2
随机	3	88.5	64.6
随机	7	87.2	63.7
困难	3	88.1	64.5

2.7　类别频率阈值敏感性分析

在式（1）中，t 为类别频率阈值，用于划分低频类别。为确定合理的阈值，本文统计 MAR20 训练集中各类别的频率分布，并将 t 的候选范围设置为 0.05～0.25，以 0.05 为间隔进行敏感性实验，结果如表 5 所示。随着 t 从 0.05 提升至 0.15，模型 mAP_{50} 由 87.5% 上升至 89.5%，在 t = 0.15 处取得 89.5% 的最优性能；阈值超过 0.15 后，模型精度开始下降，阈值过高或过低都会降低尾部类别优化效果。因此，本文在后续消融实验与对比实验中均选定 t = 0.15 作为后续实验的固定参数，并在后续实验中保持不变。

表 5　阈值 t 取值分析

t	0.05	0.10	0.15	0.20	0.25
mAP_{50}	87.5	88.3	**89.5**	87.9	87.1

2.8　不同长尾学习策略对比

为验证 CADRS 在长尾分布下的有效性，本文以 Strip R-CNN 为基线，并与 Seesaw Loss[5]、Focal Loss[3] 及 Equalization Loss[4] 进行实验对比，结果如表 6 所示。具体而言，参照 LVIS 数据集的长尾类别划分方式，按各类别在训练集中的实例数量将其划分为 Tail（尾部类）、Medium（中部类）、Head（头部类），分别统计各组的 mAP_{50}。基线 Strip R-CNN 的整体 mAP_{50} 及 mAP_{Tail}、mAP_{Medium}、mAP_{Head} 分别为 84.7%、81.9%、84.7% 和 85.5%。引入不同长尾学习策略[29]后，各方法性能差异明显。Seesaw Loss 和 Focal Loss 的整体 mAP_{50} 仅小幅提升至 84.9% 和 85.1%，mAP_{Tail} 分别提高 1.7 和 1.5 个百分点，但增益有限；Equalization Loss 的 mAP_{50} 和 mAP_{Head} 低于基线，分别为 83.8% 和 85.1%。相比之下，CADRS 将整体 mAP_{50} 提升至 86.9%，mAP_{Tail} 和 mAP_{Medium} 分别达到 87.6% 和 88.2%，同时保持头部类别检测性能。在此基础上融合 ALAF 和 CCDM 模块得到 LT-FGRSOD，其 mAP_{50} 达到 89.5%，mAP_{Tail}、mAP_{Medium}、mAP_{Head} 分别为 88.9%、89.1%、90.1%。表明 CADRS 在显著增强尾部类与中部类特征学习的同时，不会损害头部类的检测性能，从而有效缓解了长尾类别不均衡问题；在此基础上引入特征融合与判别表征模块后，模型的整体检测能力得到进一步提升。

2.9　颜色扰动参数敏感性分析

本节分析颜色扰动及其幅度对检测性能的影响，在其他实验设置保持一致的条件下，对不同色相与饱和度进行消融实验，结果如表 7 所示。表中

表 6　MAR20 数据集上不同长尾学习策略的性能比较

方法	mAP_{50}	mAP_{Tail}	mAP_{Medium}	mAP_{Head}
Strip R-CNN	84.7	81.9	84.7	85.5
Strip R-CNN + Seesaw Loss	84.9	83.6	85.5	86.2
Strip R-CNN + Focal Loss	85.1	83.4	85.3	86.8
Strip R-CNN + Equalization Loss	83.8	82.5	86.9	85.1
Strip R-CNN + CADRS	86.9	87.6	88.2	88.2
LT-FGRSOD	89.5	88.9	89.1	90.1

“—”表示未启用对应扰动，即仅施加另一项扰动，或两项均不施加。单独采用色相扰动时，±12°设置的 mAP_{50} 达到88.2%，较±6°和±18°分别提高0.5和0.2个百分点；单独调整饱和度时，0.65–1.35范围取得最高的 mAP_{50}，为88.9%，较0.8–1.2和0.5–1.5分别提高1.1和0.5个百分点。将±12°色相扰动与0.65–1.35饱和度调整结合后，mAP_{50}进一步达到89.5%，较不使用颜色扰动提高2.3个百分点，mAP_{Tail}、mAP_{Medium} 和 mAP_{Head} 分别提高1.1、3.5和2.3个百分点。上述扰动均在较小幅度内进行，色相扰动不超过 ±18°，饱和度缩放限制在 0.5–1.5 之间，本文最终选定色相 ±12°、饱和度 0.65–1.35 组合，以模拟光照和大气变化引起的颜色偏移。该扰动仅在颜色空间进行，不改变目标的几何形状与纹理，也不对目标材质或光谱特性作物理建模。结果表明，二者联合扰动取得最优检测性能，验证了该策略的有效性。

表 7　颜色扰动参数敏感性分析结果

H	S	mAP_{50}	mAP_{Tail}	mAP_{Medium}	mAP_{Head}
—	—	87.2	86.8	87.4	86.9
±6°	—	87.7	87.6	87.5	88.3
±12°	—	88.2	87.9	88.3	88.7
±18°	—	88.0	87.5	88.4	87.9
—	0.8–1.2	87.8	87.9	88.1	87.7
—	0.65–1.35	88.9	88.3	89.2	89.0
—	0.5–1.5	88.4	88.0	88.6	88.9
±12°	0.65–1.35	89.5	87.9	90.9	89.2

2.10　多尺度融合结构消融实验

多尺度融合模块通过逐层融合浅层细节与深层语义信息，不同融合策略对这两类信息的利用方式存在差异，进而影响模型的细粒度判别能力。为验证ALAF对旋转目标细粒度判别性能的有效性，本文在相同实验设置下比较了FPN、BiFPN[30]、PA-Net、ALAF-ALL全层级融合和ALAF邻层融合，结果如表8所示。采用邻层融合的ALAF取得最优性能，mAP50达到89.5%，较FPN、BiFPN和PA-Net分别提升3.1、1.9和1.6个百分点。从分组精度看，ALAF的 mAP_{Tail} 和 mAP_{Head} 均为最高，分别为88.9%和90.1%；mAP_{Medium} 为89.1%，与PANet持平。尾部类别的表现尤为突出，ALAF的 mAP_{Tail} 达到88.9%，较FPN提升4.7个百分点，较BiFPN和PANet也分别高出1.5和2.4个百分点。此外，与ALAF-ALL全层级融合相比，ALAF的 mAP_{50}、mAP_{Tail}、mAP_{Medium} 和 mAP_{Head} 分别提升1.7、2.7、0.4和1.8个百分点。实验结果表明，ALAF在保留层级间互补信息的同时，抑制了冗余特征与背景干扰，增强了模型对遥感图像目标的表征能力。

表 8　MAR20 数据集上特征融合结构对比

方法	mAP_{50}	mAP_{Tail}	mAP_{Medium}	mAP_{Head}
FPN	86.4	84.2	88.2	86.3
BIFPN	87.6	87.4	88.8	86.5
PANet	87.9	86.5	89.1	87.8
ALAF-ALL	87.8	86.2	88.7	88.3
ALAF	89.5	88.9	89.1	90.1

2.11　对比损失参数敏感性分析

为分析对比损失权重 λ 和温度参数 τ 对特征判别能力的影响，本文对不同参数配置下的RoI嵌入特征进行t-SNE可视化，结果如图3所示。图3(a)为不使用对比损失的基线，图3(b)–(d)为不同λ、τ 配置下的可视化结果。λ 从0增至0.10，分别对应无对比损失、低权重和中权重三种情况。参考对比学习中的常用设置，本文选取 τ = 0.07 作为常用温度。固定 τ = 0.07，将 λ 由0.05增大至0.10后，A5与A15已明显分开，边界更清晰。固定 λ = 0.10，将 τ 由0.07增大至0.20后，A5与A15的特征簇间距明显缩小，并出现局部重叠。综合来看，λ = 0.10、τ = 0.07时部分易混淆类别的类间边界最为清晰，表明CCDM在常用参数范围内具有较好的参数鲁棒性。

2.12　CADRS 的类别均衡效果分析

表9给出了Strip R-CNN与引入CADRS后在两个随机种子下的分组精度及类别均衡结果。整体上，CADRS使 mAP_{50} 由83.71%、84.22%提升至86.79%、86.85%，分别提高3.08与2.63个百分点。分组来看，尾部（Tail）、中部（Medium）和

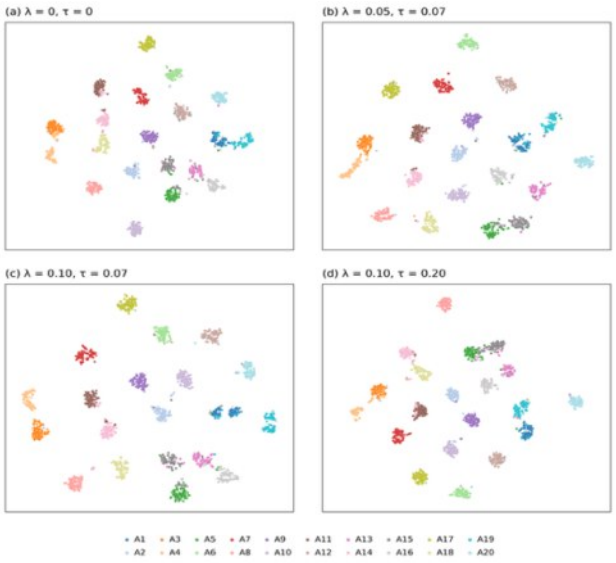

图 3　不同参数配置下 RoI 特征的 t-SNE 可视化

头部（Head）三组均获得正向增益，其中 Tail 组由 81.95%、82.17% 提升至 85.13%、85.30%，分别提高 3.18 与 3.13 个百分点，提升幅度最大。进一步看尾部类别指标，最差类别 AP 由 58.49%、60.87% 提升至 79.89%、80.25%，分别提高 21.40 与 19.38 个百分点；最差三类平均 AP 由 65.50%、67.63% 提升至 82.61%、82.77%，分别提高 17.11 与 15.14 个百分点，说明增益覆盖多个低精度类别，而非仅由单一类别驱动；类间方差由 45.54、43.41 降至 18.76、17.64，表明类别间 AP 的离散程度明显下降。综合来看，CADRS 在提升整体检测精度的同时，明显改善了尾部类别及其中低精度类别的检测效果，缩小了类别之间的性能差距，且未削弱头部类别的性能。该策略能够有针对性地缓解长尾分布带来的类别不均衡问题，为后续特征融合与判别模块提供了更均衡的样本基础。

表 9　引入 CADRS 前后的类别均衡效果

方法	种子	mAP_{50}	mAP_{Tail}	mAP_{Medium}	mAP_{Head}	最差类别	最差3类平均	类间方差
Strip R-CNN	3407	83.71	81.95	84.75	86.12	58.49	65.50	45.54
Strip R-CNN	1337	84.22	82.17	84.82	85.92	60.87	67.63	43.41
Strip R-CNN + CADRS	3407	86.79	85.13	87.73	88.07	79.89	82.61	18.76
Strip R-CNN + CADRS	1337	86.85	85.30	87.47	88.69	80.25	82.77	17.64

2.13　混淆矩阵分析

为深入分析各模型在细粒度舰船类别上的判别特性，图 4 展示了对比模型 PETDet 与本文 LT-FGRSOD 模型在 ShipRSImageNet 数据集上的混淆矩阵对比结果。LT-FGRSOD 模型在多个困难舰船类别上均取得较为明显的分类精度提升，具体来看，辅助舰船（Auxiliary Ships）、滚装船（RoRo）和货船（Cargo）类别的分类准确率分别达到 88%、93% 和 73%，较基线模型分别提升 21%、12% 和 9%。与此同时，LT-FGRSOD 对大尺度目标仍保持较高的识别精度，如航空母舰（Aircraft Carrier）和船坞（Dock）类别的准确率继续稳定在 94% 和 98%。从类别间混淆情况来看，PETDet 中 RoRo 被误判为 Cargo 的比例为 15%，而 LT-FGRSOD 中该类误分类比例降至 7%，降低了 8 个百分点；对于 Fishing_Vessel 类别，其被误判为 Tugboat 的比例由 12% 降至 5%，降低 7 个百分点。以上实验结果表明，所引入的动态重采样策略、邻层自适应特征融合模块与对比判别机制能够提升模型对相似外形、相近尺度及复杂背景条件下舰船类别的区分能力，从而提升检测精度。

2.14　ALAF 模块消融实验

为验证 ALAF 各组件的有效性，消融实验固定 CADRS 与 CCDM，仅改变颈部融合的组件构成，分别加入门控融合（G）、通道注意力（CA）、空间注意力（SA），并进行组合消融实验，结果如表 10 所示。加入门控融合后，mAP_M 与 mAP_S 分别为 85.07% 与 24.24%；加入通道注意力后，mAP_M 与 mAP_L 分别达到 86.51% 与 91.74%，在中大型目标上表现较好，对小目标的改善有限；仅加入空间注意力时 mAP_S 为 23.70%，说明缺少通道语义引导时小目标区域响应不足，而通道与空间注意力同时加入后 mAP_S 回升至 25.00%，表明二者需协同使用。同时引入三个组件后，ALAF 模块的 mAP_{50} 与 mAP_S 分别为 89.50% 与 25.00%，表明该模块通过融合浅层细节与深层语义，有效提升了相似子类别的特征区分能力。

2.15　可视化结果分析

图 5 给出了 Oriented R-CNN[14] 与 LT-FGRSOD 在 MAR20 数据集上的可视化对比结果。由图可

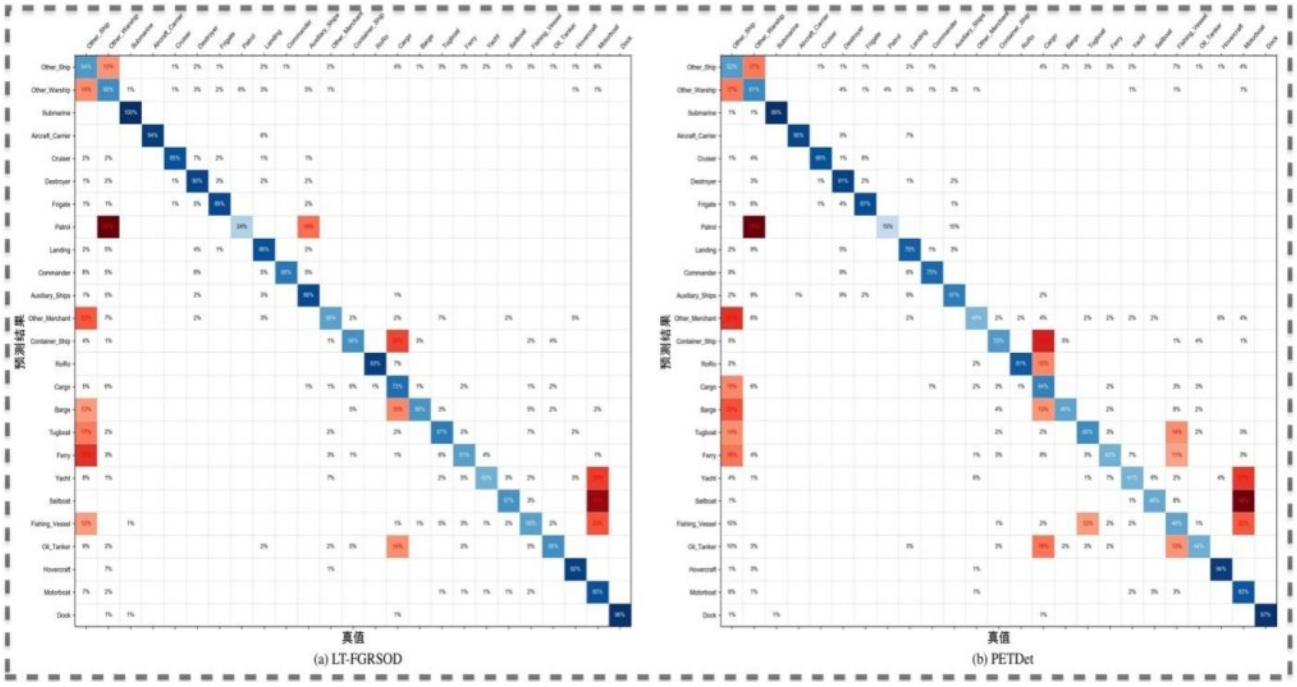

图 4　ShipRSImageNet 数据集的混淆矩阵

表 10　ALAF 各组件消融结果

方法	mAP_{50}	mAP_{Tail}	mAP_{Medium}	mAP_{Head}	mAP_{S}	mAP_{M}	mAP_{L}
FPN + G	89.28	88.39	88.92	89.31	24.24	85.07	90.65
FPN + CA	89.34	87.02	89.60	89.96	25.00	86.51	91.74
FPN + SA	89.12	87.88	88.79	89.48	23.70	86.18	90.73
FPN+ CA+SA	89.29	88.27	88.69	89.77	25.00	86.31	89.72
FPN+ G + CA+SA (ALAF)	89.50	88.91	89.13	90.15	25.00	85.84	90.80

见，Oriented R-CNN 在多种复杂场景下仍存在明显的误检、漏检与类别混淆问题。在（a）列中，相似机型混排场景中，Oriented R-CNN 受到相邻目标外观相似性的干扰，将真实类别为 E-8 的目标错误识别为 E-3；（b）列展示了目标邻近航站楼的场景，由于复杂背景和局部遮挡的共同影响，对比模型未能有效检测出 F-15 目标，出现漏检现象；从（c）列可以看出，在高密度紧密停放场景中，模型难以准确分离相邻飞机之间的特征响应，导致目标 KC-135 被误判为 E-8；（d）列中目标尺度差异较为明显，Oriented R-CNN 将真实类别为 C-17 的目标错误预测为 KC-135，反映出其在相似外观目标区分方面仍存在不足。相比之下， LT-FGRSOD 模型在上述场景中的检测结果与真实标注保持更高的一致性，在密集分布及多尺度变化条件下表现出更强的细粒度类别判别能力，验证了所提方法在复杂遥感场景中的鲁棒性与实际应用潜力。

3　结束语

本文针对遥感图像细粒度目标检测中的长尾分布、多尺度特征表达不足和相似类别混淆问题，提出了一种面向长尾分布的细粒度旋转目标检测方法 LT-FGRSOD（Fine-Grained Remote Sensing Object Detection with Long-Tailed Distribution）。该方法从样本重平衡、特征融合和判别表征三个层面提升模型检测性能。在 MAR20 和 ShipRSImageNet 数据集上的实验结果表明，所提方法的 mAP_{50} 分别达到 89.5% 和 78.2%，较 PETDet 提升 4.1 和 2.4 个百分点，且在尾部、中部和头部类别组上均取得稳定的性能增益；消融实验进一步表明，三个模块具有互补增益。后续工作将围绕模型轻量化、密集小目标检测以及复杂场景适应性进一步开展研究。

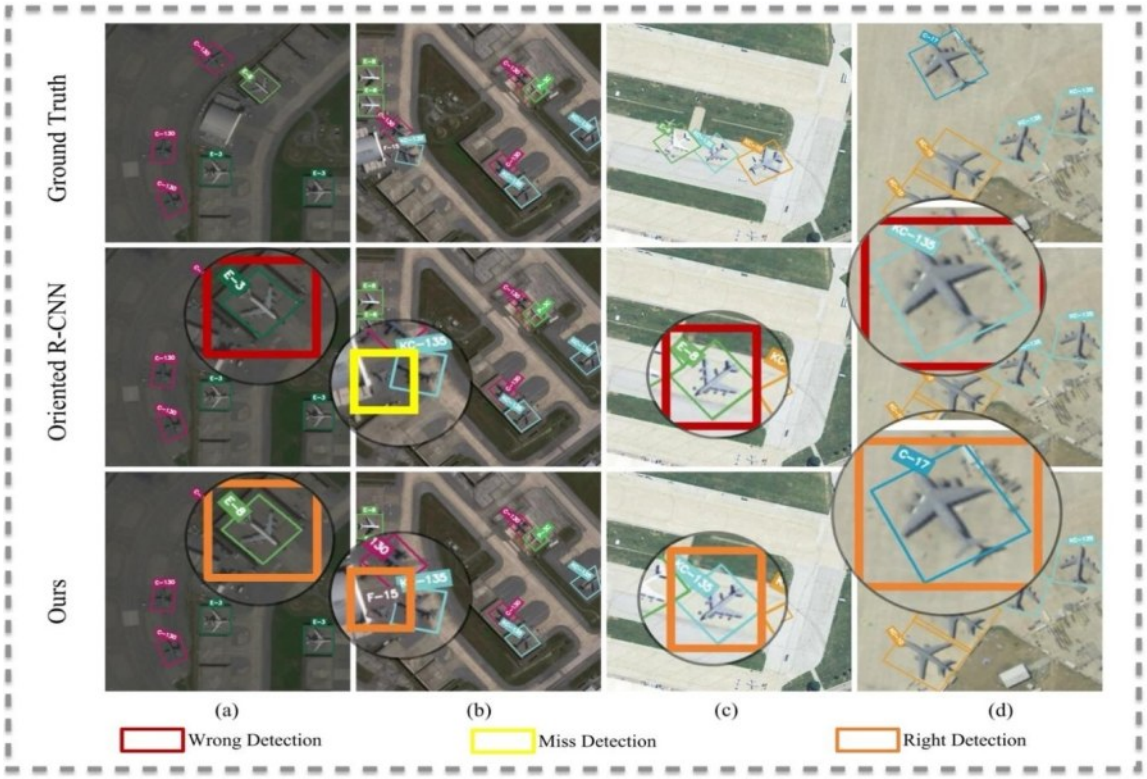

图 5　MAR20 数据集可视化结果

参考文献：

[1] GUPTA A, DOLLÁR P, GIRSHICK R. LVIS: a dataset for large vocabulary instance segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 5356-5364.

[2] OUYANG L H, GUO G M, FANG L Y, et al. PCLDet: prototypical contrastive learning for fine-grained object detection in remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-11.

[3] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2980-2988.

[4] TAN J R, WANG C B, LI B Y, et al. Equalization loss for long-tailed object recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 11662-11671.

[5] WANG J Q, ZHANG W W, ZANG Y H, et al. Seesaw Loss for Long-Tailed Instance Segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 9695-9704.

[6] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2117-2125.

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

[8] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision. Munich: Springer, 2018: 3-19.

[9] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a unified embedding for face recognition and clustering[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 815-823.

[10] WEN Y D, ZHANG K P, LI Z F, et al. A discriminative feature learning approach for deep face recognition[C]//Proceedings of the European Conference on Computer Vision. Amsterdam: Springer, 2016: 499-515.

[11] VAN DEN OORD A, LI Y Z, VINYALS O. Representation learning with contrastive predictive coding[J]. arXiv preprint arXiv: 1807.03748, 2018.

[12] LI W T, ZHAO D P, YUAN B, et al. PETDet: proposal enhancement for two-stage fine-grained object detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62: 1-14.

[13] YUAN X B, ZHENG Z H, LI Y X, et al. Strip R-CNN: large strip convolution for remote sensing object detection[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2026, 40(15): 12259-12267.

[14] XIE X X, CHENG G, WANG J B, et al. Oriented R-CNN for object detection[C]//Proceedings of the IEEE/CVF International

Conference on Computer Vision. Montreal: IEEE, 2021: 3520-3529.

[15] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770-778.

[16] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines[C]//Proceedings of the 27th International Conference on Machine Learning. Haifa: Omnipress, 2010: 807-814.

[17] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1440-1448.

[18] YU W Q, CHENG G, WANG M J, et al. MAR20: a benchmark for military aircraft recognition in remote sensing images[J]. National Remote Sensing Bulletin, 2023, 27(12): 2688-2696.

[19] ZHANG Z N, ZHANG L, WANG Y, et al. ShipRSImageNet: a large-scale fine-grained dataset for ship detection in high-resolution optical remote sensing images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 8458-8472.

[20] ZHOU Y, YANG X, ZHANG G F, et al. MMRotate: a rotated object detection benchmark using PyTorch[C]//Proceedings of the 30th ACM International Conference on Multimedia. Lisboa: ACM, 2022: 7331-7334.

[21] LOSHCHILOV I, HUTTER F. Decoupled weight decay regularization[C]//International Conference on Learning Representations. New Orleans: 2019.

[22] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Miami: IEEE, 2009: 248-255.

[23] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//Advances in Neural Information Processing Systems. Montreal: Curran Associates, 2015: 91-99.

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

[25] DING J, XUE N, LONG Y, et al. Learning RoI transformer for oriented object detection in aerial images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 2849-2858.

[26] XU Y C, FU M T, WANG Q M, et al. Gliding vertex on the horizontal bounding box for multi-oriented object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(4): 1452-1459.

[27] WEN L, CHENG Y, FANG Y, et al. A comprehensive survey of oriented object detection in remote sensing images[J]. Expert Systems with Applications, 2023, 224: 119960.

[28] 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.

[29] LI Y, WANG T, KANG B, et al. Overcoming classifier imbalance for long-tail object detection with balanced group softmax[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 10991-11000.

[30] TAN M, PANG R, LE Q V. EfficientDet: scalable and efficient object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 10781-10790.