难度与噪声的解耦：面向长尾遥感分类的不确定性驱动模型无关框架

Chi Ding[1*]、**Junxiao Xue**[1*]、**Xinyi Yin**[2],**Shi Chen**[3],**Yunyun Shi**[3]
Yiduo Wang[2],**Fengjian Xue**[3],**Xuecheng Wu**[3†]空间计算系统研究中心，之江实验室，中国杭州[2]网络空间科学与工程学院，郑州大学，中国郑州[3]计算机科学与技术学院，西安交通大学，中国西安 xuejx@zhejianglab.cn、xuecwu@gmail.com

arXiv:2601.00278v1 [cs.CV] 1 Jan 2026

摘要——**由于地面目标出现的固有不均衡性，长尾分布在遥感领域普遍存在。然而，一个关键挑战在很大程度上仍被忽视，即如何将困难的尾部数据样本与含噪的模糊样本解耦。传统方法通常无差别地强调全部低置信度样本，导致在噪声数据上过拟合。为弥合这一差距，我们以证据深度学习为基础，提出了一种模型无关的不确定性感知框架，称为 DUAL。该框架能够将预测不确定性动态解耦为认知不确定性（EU）和偶然不确定性（AU）。具体而言，我们将 EU 作为样本稀缺性的指标，用于指导针对难学习尾部样本的重加权策略；同时利用 AU 量化数据模糊性，并采用自适应标签平滑机制来抑制噪声的影响。在多个数据集和各种骨干网络上的大量实验表明，我们的框架具有良好的有效性和泛化能力，性能超过了 TGN 和 SADE 等强基线方法。消融研究进一步揭示了我们设计中关键选择的重要作用。**

索引词—**不确定性、长尾数据、遥感、图像分类**

I. 引言

在遥感图像场景中 [1], [2],土地覆盖类别通常呈现显著的不平衡分布，通常称为长尾分布：少数高频类别拥有大量样本，而大多数稀有类别可用于学习的样本数量有限。这种分布给深度学习模型的表征能力带来了重大挑战 [3], ，使模型容易对头部类别过拟合，而在尾部类别上的表现不佳。

近年来，关于长尾分布的研究主要聚焦于重采样 [4], 损失重加权 [5], 和对数几率调整 [6]等方法。重采样通过对尾部类别过采样或对头部类别欠采样来平衡类别分布。类别平衡损失 [4]和焦点损失 [7], 等损失重加权方法，会根据类别频率或样本难度调整损失权重，以提升尾部类别性能。尽管这些方法在一定程度上有效，但遥感领域仍有一个关键挑战在很大程度上被忽视：如何区分难以学习的尾部样本与含噪模糊样本。与自然图像不同，遥感图像常受传感器分辨率变化、云层遮挡、土地覆盖重叠和光照条件变化的影响，从而产生固有噪声或歧义。

传统方法通常仅依靠类别频率或预测逻辑值来评估样本的重要性。因此，它们会不加区分地强调所有难学习样本，导致模型对噪声数据过拟合，而非减轻其负面影响。

因此，核心研究挑战随之产生：如何鼓励模型从稀有样本中学习，同时抑制噪声数据的影响？为应对这一核心挑战，我们引入不确定性估计，以更好地区分稀有样本与噪声数据。不确定性通常分为两种类型：认知不确定性（EU）和偶然不确定性（AU） [3]。认知不确定性反映了模型知识的缺乏，通常源于输入空间中模型训练不充分或无法作出自信预测的区域。相比之下，偶然不确定性刻画数据中固有的噪声或模糊性，这些因素无法仅通过收集更多数据来减少。

这两种不确定性恰好对应长尾遥感中的两个关键问题：模型对尾部类别学习不足（EU），以及部分样本的质量下降或语义模糊性（AU）。因此，与仅依赖类别频率或 logits 的方法相比，不确定性能够提供更加细粒度的训练信号。一方面，认知不确定性可作为模型当前学习不足样本的指标，帮助识别哪些实例值得优先训练。另一方面，偶然不确定性能够评估每个样本的质量，从而动态调整监督强度，避免对噪声或语义模糊数据过拟合。

基于这一洞见，我们提出了一种双重不确定性

*共同贡献。
†通讯作者。

感知长尾学习框架，称为 **DUAL**。首先，我们采用证据深度学习（EDL） [8]动态建模不确定性，并在训练过程中将认知不确定性与偶然不确定性区分开来。随后，我们进行基于认知不确定性的样本重加权，根据模型当前的认知状态重新分配权重，以突出需要更多学习的样本；此外，还提出一种偶然不确定性引导的动态标签平滑方法，根据估计的偶然不确定性调整监督强度。

总而言之，本文的主要贡献有三点：

- **一种将困难与噪声解耦的新型不确定性感知框架**。我们指出了现有方法在将困难尾部样本与噪声模糊样本解耦方面的关键局限。为此，我们提出**DUAL**，一种基于 EDL 的模型无关框架，通过动态地将预测不确定性解耦为 EU 和 AU，识别低置信度的来源。

- **一种由不确定性驱动的动态引导训练框架**。我们在**DUAL**中设计了一种不确定性感知机制，用于处理尾部样本和噪声样本。具体而言，**DUAL**利用 EU 表示样本稀缺性，以对困难尾部样本进行重加权，同时利用 AU 衡量数据模糊性，通过自适应标签平滑抑制噪声。

- **在多个长尾遥感基准数据集上的广泛验证**。 实验表明，**DUAL** 能够持续提升总体准确率，并显著提高尾部类别性能，证实了其有效性和实用性。

II. 相关工作

A. 长尾学习

长尾数据是许多现实场景中的常见挑战，包括类别分布严重不均衡的遥感图像分析。在这些数据集中，少数类别（称为头部类别）占据了大多数样本，而许多其他类别（称为尾部类别）的样本数量则不足。这种不均衡给机器学习模型带来了重大挑战，因为标准训练方法往往会导致模型对头部类别过拟合，而在尾部类别上的表现不佳。重采样 [4] 和类别敏感学习[7] 是应对长尾数据的主要方法。重采样通过调整样本的采样权重来平衡数据分布；类别敏感学习则通过调整模型的损失函数来应对数据分布的不均衡。然而，这些方法通常仅依赖类别频率或预测逻辑值。因此，它们往往不加区分地强调所有尾部类别样本或难学习样本，而忽视了数据本身固有的质量问题。由于无法识别不适合用于训练的样本（例如，因噪声或模糊性所致），这些方法往往会导致过拟合，而非减轻低质量样本带来的负面影响。

B. 不确定性估计

近年来，深度神经网络在各个领域都取得了显著成功 [9]–[12]。然而，随着这些模型日益广泛地应用于现实场景，其预测结果的可靠性已成为一个关键问题。深度学习中的不确定性通常分为模型不确定性（认知不确定性），源于数据有限所导致的知识缺口；以及数据不确定性（偶然不确定性），由固有噪声引起。用于估计这些不确定性的早期方法 [3] 包括贝叶斯神经网络（BNN）、深度集成和蒙特卡洛（MC）Dropout。然而，这些方法往往因需要多次前向传播（深度集成、MC Dropout）而产生较高的计算成本，或面临收敛困难（BNN），因此难以应用于大规模遥感任务。相比之下，EDL [8]提供了一种确定性且高效的替代方案。通过将预测分布建模为狄利克雷分布，EDL能够在单次前向传播中同时量化预测结果、认知不确定性和偶然不确定性。更重要的是，EDL能够在训练过程中动态估计不确定性，无需改变骨干架构或进行昂贵的采样，因此非常适合用于优化模型训练过程。

III. 方法

如图 1 所示，**DUAL** 由三个关键组成部分构成：（1）通过 EDL 进行不确定性估计；（2）从模型预测中分离认知不确定性和偶然不确定性；（3）基于认知不确定性的样本重加权，以解决尾部类别学习不足的问题；以及基于偶然不确定性的动态标签平滑，以降低模糊样本的影响。

A. 证据深度学习

在 EDL 中，需要确定狄利克雷分布的参数以评估不确定性。证据是指从输入中获取、用于支持分类的指标，并且与狄利克雷分布的参数密切相关。根据邓普斯特-谢弗证据理论（DST）[13], ，在 K 类分类问题中，模型尝试为每个类别分配一个信念分布，并为整个框架分配总体不确定性。因此，对于每个输入，都有 K+1 个非负信念分布值，其总和为 1，如公式1所示。

$$u_i + \sum_{j=1}^{K} b_{ij} = 1, \tag{1}$$

其中， u_i 和 b_{ij} 分别表示 i^{th} 样本的总体不确定性和 k^{th} 类的概率。

对于输入 i^{th}，将狄利克雷分布 $\alpha = [\alpha_1, \cdots, \alpha_K]$ 的参数与不确定性相关联。不确定性 u 的计算方式如下：

$$u_i = \frac{K}{\sum_{j=1}^{K} \alpha_{ij}} = \frac{K}{S_i}. \tag{2}$$

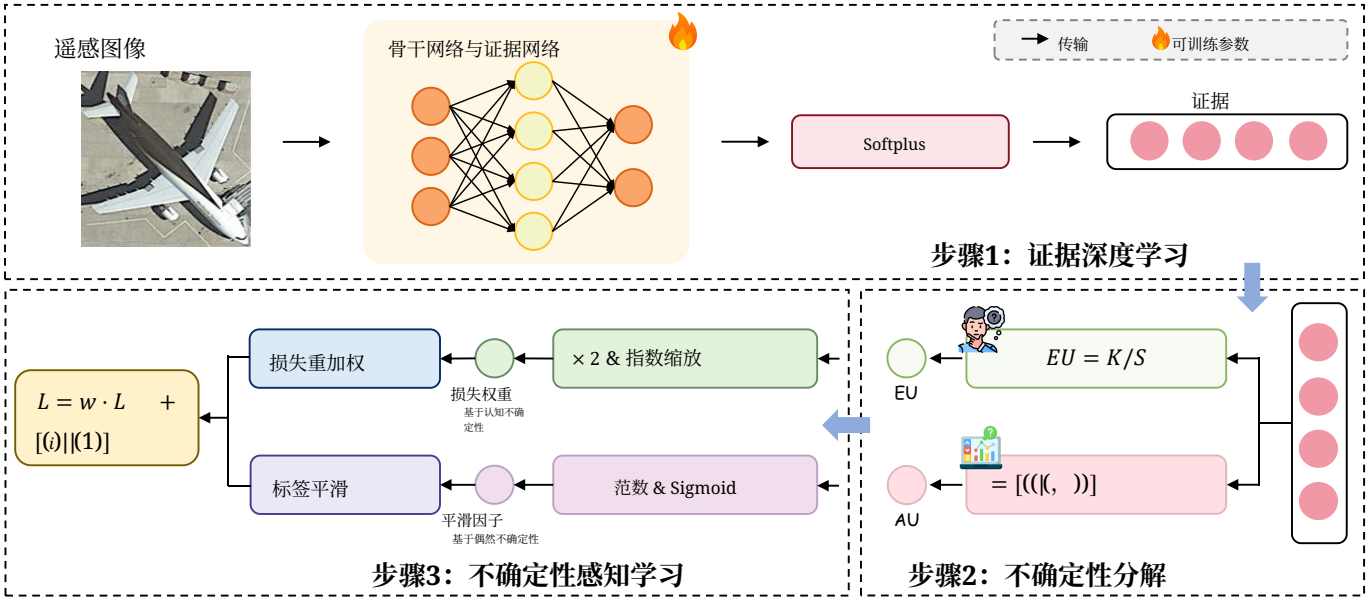

图 1. 我们提出的 DUAL 框架概览。该流程包含三个阶段：(1) 证据深度学习，根据骨干网络预测类别级证据；(2) 不确定性分解，将预测不确定性分解为认知不确定性和偶然不确定性；(3) 不确定性感知学习，其中认知不确定性作为样本稀缺的指标，用于重新加权困难尾部样本，而偶然不确定性则量化数据模糊性，以指导自适应标签平滑，实现噪声抑制。

在此语境中，$S_i = \sum_{j=1}^{K} \alpha_{ij}$ 是狄利克雷分布的强度，可将其视为证据的总量。

对于 i^{th} 输入， j^{th} 类别的预测概率 p_{ij} 是相应狄利克雷分布的均值，计算如下：

$$p_{ij} = \frac{\alpha_{ij}}{S_i}. \tag{3}$$

对于传统的基于深度神经网络的分类器，通常使用交叉熵损失：

$$L_{ce} = -\sum_{j=1}^{K} y_{ij} log(p_{ij}), \tag{4}$$

其中， p_{ij} 是j^{th} 类别中 i^{th} 样本的预测概率。

对于本章中的模型，狄利克雷分布的参数 α_i 可通过证据神经网络获得。对公式 4进行简单修改后，即可得到调整后的交叉熵损失，即，

$$\begin{aligned} L_{ace} &= \int [\sum_{j=1}^{K} -y_{ij} log(p_{ij})] \frac{1}{B(\alpha_{ij})} \prod_{j=1}^{K} p_{ij}^{\alpha_{ij}-1} dp_i \\ &= \sum_{j=1}^{K} y_{ij} (\psi(S_i) - \psi(\alpha_{ij})), \end{aligned} \tag{5}$$

其中 $\psi(\cdot)$ 表示双伽马函数，且式5 是由 α_i确定的交叉熵损失函数的积分。

尽管上述损失函数能够确保每个样本的正确标签比其他类别的标签产生更多证据，但它并不能保证错误标签产生的证据更少。因此，我们希望模型中错误标签对应的证据逐步降低至接近 0。为此，引入以下 KL散度项：

$$\begin{aligned} KL[D(p_i|\tilde{\alpha}_i)||D(p_i|1)] &= log(\frac{\Gamma(\sum_{j=1}^{K} \tilde{\alpha}_{ij})}{\Gamma(K) \prod_{j=1}^{K} \Gamma(\tilde{\alpha}_{ij})}) \\ &+ \sum_{j=1}^{K} (\tilde{\alpha}_{ij} - 1)[\psi(\tilde{\alpha}_{ij}) - \psi(\sum_{j=1}^{K} \tilde{\alpha}_{ij})], \end{aligned} \tag{6}$$

其中， $\tilde{\alpha}_i = y_i + (1 - y_i) \odot \alpha_i$ 是经过调整的狄利克雷分布参数，用于避免正确标签的证据为零； $\Gamma(\cdot)$是伽马函数。

因此，给定每个样本 i的狄利克雷分布参数 α_i ，该样本的特异性损失为：

$$L_{EDL} = L_{ace} + \lambda_t KL[D(p_i|\tilde{\alpha}_i)||D(p_i|1)], \tag{7}$$

其中， $\lambda_t > 0$ 是平衡因子。在实验中，随着训练的进行，可以逐渐增大 λ_t，以防止网络在训练初期过于关注 KL散度项，否则可能导致网络无法充分优化参数以输出均匀分布。

B. 认知不确定性与偶然不确定性

预测不确定性（PU）可分解为两部分：认知不确定性（EU）和偶然不确定性（AU）。认知不确定性源于模型参数的不确定性，通常与训练数据不足或模型存在知识缺口有关。相比之下，偶然不确定性反映了数据的内在噪声，即使模型实现了完美拟合，也无法消除

表 I DUAL 与其他方法在 DOTA、DIOR 和 FGSC-23 测试数据集上的性能比较。该表列出了头部和尾部类别的平均 Top-1 准确率（%），其中“全部”表示总体准确率（%）。最佳数值以粗体标出。TLC、BKD 和 LAL 方法采用 ResNet32 作为骨干网络，其他方法则采用 ResNet50。请注意，我们以粗体 标出最佳性能，并以下划线标出第二佳性能。

方法	DOTA			DIOR			FGSC-23		
	头部 (↑)	尾部 (↑)	全部 (↑)	头部 (↑)	尾部 (↑)	全部 (↑)	头部 (↑)	尾部(↑)	全部 (↑)
SADE [14]	94.27	88.67	93.57	88.68	86.90	88.40	68.70	76.08	70.79
RIDE [15]	85.10	78.15	81.54	88.33	83.19	87.57	42.68	57.22	52.27
ResLT [16]	94.75	81.74	94.97	78.81	81.05	72.95	64.38	63.19	62.55
LDMLR [17]	87.33	80.92	92.95	80.59	79.82	86.60	52.54	51.01	51.82
TLC [18]	88.25	78.99	88.24	82.27	76.39	82.70	29.00	68.50	44.12
BKD [19]	85.16	55.96	84.74	75.31	61.43	75.71	65.41	72.05	65.94
LAL [20]	93.60	68.12	92.77	82.80	76.32	84.73	54.16	49.62	53.45
T2FTS [21]	86.96	87.80	87.29	-	-	-	75.70	71.46	73.46
EME [22]	90.32	**89.32**	89.92	-	-	-	73.71	73.81	73.77
TGN [23]	95.56	81.49	96.10	**91.81**	84.46	90.68	72.24	68.93	71.76
DUAL	**97.13**	89.18	**96.66**	90.54	**87.47**	**91.07**	**78.21**	**82.72**	**79.98**

这种分解对于不确定性建模至关重要，预测不确定性（PU）可以表示为认知不确定性（EU）与偶然不确定性（AU）之和：

$$\mathrm{PU} = \mathrm{EU} + \mathrm{AU}. \tag{8}$$

为了量化PU、认知不确定性和偶然不确定性，可以使用预测分布熵度量。输入 x 的不确定性可以通过其预测分布 $p\,(y \mid x)$ 的熵来近似：

$$\mathrm{PU} = H(p(y \mid x)) = -\sum_{j=1}^{K} p_{ij} \log p_{ij}, \tag{9}$$

其中， $H\,(\cdot)$ 表示熵函数。

偶然不确定性通过对使用采样的模型参数 θ 得到的多个预测的熵取期望来估计，即，

$$\begin{aligned} \mathrm{AU} &= \mathbb{E}\big[H(p(y \mid x, \theta))\big] \\ &= \sum_{j=1}^{K} p_{ij}\big[\psi(S_i + 1) - \psi(\alpha_{ij} + 1)\big]. \end{aligned} \tag{10}$$

认知不确定性可通过认知不确定性与偶然不确定性之差计算得出：

$$\mathrm{EU} = H(p(y \mid x)) - \mathbb{E}[H(p(y \mid x, \theta))]. \tag{11}$$

在证据深度学习框架中，狄利克雷分布参数 $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_K]$ 可用于 K/S 作为认知不确定性的度量，其中 K 是类别数量，而 $S = \sum_{c=1}^{K} \alpha_c$。图2 显示，$K/S$ 与基于熵的认知不确定性高度相关。与熵相比，K/S产生的取值范围分布更加均匀，因此更适合作为损失函数中的权重。因此，我们选择 K/S 作为认知不确定性的度量，以指导不确定性感知优化。

C. 不确定性感知长尾学习

为解决长尾遥感分类中样本可学习性存在异质性这一挑战，我们利用认知不确定性和偶然不确定性，通过动态重加权和标签平滑来优化训练过程。以下将介绍这些机制的详细内容以及最终的损失函数。

基于认知不确定性的样本重加权。我们利用认知不确定性在训练过程中动态调整样本权重，重点关注认知不确定性较高的样本，以加强对代表性不足类别的学习。具体而言，对于样本 i，其训练权重 w_i 计算如下：

$$w_i = (2 \times \mathrm{EU}_i)^{\sigma}, \tag{12}$$

其中， σ 是指数缩放因子（通常为 $\sigma \in [1, 5]$），用于放大认知不确定性之间的差异，使认知不确定性较高的样本获得更大的权重，同时降低置信度较高样本的权重。乘以 2 是为了将认知不确定性的区间设为 [0, 2],，使其经过指数缩放后不会趋近于 0。这种方法促使模型优先处理不确定性较高的尾部样本。

基于偶然不确定性的动态标签平滑。偶然不确定性反映了数据中固有的噪声或模糊性，例如遥感图像中的云层遮挡或混合土地覆盖。为减轻此类样本的负面影响，我们引入了一种基于偶然不确定性的动态标签平滑机制。传统的标签平滑会将独热标签 y_i 修改为：

$$\tilde{y}_i = (1 - \epsilon) y_i + \frac{\epsilon}{K}, \tag{13}$$

其中， ϵ 是固定的平滑因子， K 是预定义类别的总数

我们进一步将 ϵ_i 调整为适应每个样本的偶然不确定性值：

$$\tilde{\epsilon}_i = \mathrm{sigmoid}(\mathrm{AU}_i) \cdot \epsilon, \tag{14}$$

其中，Sigmoid 函数将偶然不确定性映射到 $[0,1]$。偶然不确定性较高的样本会获得更大的平滑因子，从而生成更柔和的标签分布，并降低过拟合风险。

最终损失函数。 通过结合基于认知不确定性的重新加权和 AU 驱动的动态标签平滑，我们设计了最终损失函数如下：

$$L = w_i \cdot L_{ace} + \lambda_t KL[D(p_i|\tilde{\alpha}_i)||D(p_i|1)]. \quad (15)$$

该组合损失函数促使模型关注尾部类别（通过 EU），并抑制噪声样本的影响（通过 AU），从而提升长尾遥感分类的性能和鲁棒性。

IV. 实验

数据集。 我们在三个遥感基准数据集上评估所提出的框架：DIOR [2]，DOTA [1]，和 FGSC-23[24]，，这些数据集涵盖了复杂背景下的大规模目标检测和细粒度分类。三个数据集的详细信息如下所述：

- DIOR 是一个大规模光学遥感目标检测基准数据集，包含 20 个类别和 192,465 个标注实例。该数据集具有较高的类间差异以及显著的类内外观变化。

- DOTA 包含 2,806 幅航空图像，其类别与 DIOR 大体重叠，但具有更复杂的背景和尺度变化。

- FGSC-23 专注于包含 23 个类别的细粒度船舶分类。由于各子类之间具有高度相似性，其长尾分布更具挑战性。

在下文中 [23], 我们采用头部–尾部划分协议。类别不平衡程度通过不平衡比率（IR）进行量化： IR =最大值（N^c_{real}）/最小值（N^c_{real}），其中 N^c_{real}表示类别 c 的样本数。详细统计数据和类别汇总于表 II。

实现细节。 我们使用 ResNet-50 作为骨干网络，并采用 ImageNet 预训练权重进行初始化。全部实验均使用 PyTorch 2.1，在配备 CUDA 12.2 和 cuDNN 加速的 NVIDIA RTX A6000 GPU（48GB）上进行。训练共进行 100 个 epoch，批大小为 64，采用 Adam 优化器（$\beta_1 = 0.9$，$\beta_2 = 0.999$），学习率从 1×10^{-3} to 1×10^{-6} 进行余弦学习率衰减。权重衰减设置为 1×10^{-4}。数据增强策略包括随机裁剪、水平/垂直翻转和归一化。不确定性感知模块的超参数通过经验设置为 $\sigma = 3$ 和 $\lambda = 0.2$。我们采用总体准确率、平均类别准确率、头部类别平均准确率和尾部类别平均准确率来评估分类性能。

A. 主要结果

我们在三个遥感长尾分类数据集：DOTA、DIOR 和 FGSC-23 上，将 **DUAL** 与先进方法进行了性能比较。评估指标包括总体准确率（Top-1 准确率）、头部类别的平均准确率和尾部类别的平均准确率。基线包括通用的长尾分类方法以及面向遥感的专用方法。实验结果如表I所示。

表 II DOTA、DIOR 和 FGSC-23 数据集的统计信息，其中不平衡比率和尺度范围分别表示训练数据集中的不平衡比率和尺度分布范围。

数据集	类别数	训练样本	测试样本	不平衡比率
DOTA [1]	15	98,906	28,853	86
DIOR [2]	20	68,025	124,440	54
FGSC-23 [24]	23	3,256	825	25

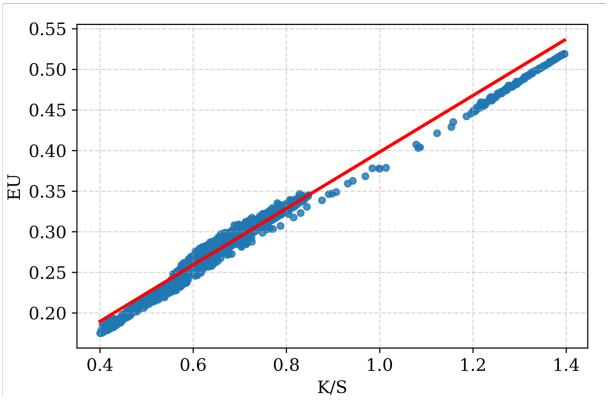

图 2. 所提出的 K/S 指标与 FGSC-23数据集上基于熵的认知不确定性之间的相关性。

表 III DUAL 在三个遥感基准数据集上使用不同骨干网络的性能（%）。

骨干网络	DOTA	DIOR	FGSC-23
EfficientNet-B0	96.38	91.45	77.55
MobileNetV2	96.09	89.83	79.00
ResNet-18	95.72	89.26	78.52

在DOTA数据集上，我们的方法在头部类别上的平均准确率为97.13%，在尾部类别上的平均准确率为89.18%，总体准确率为96.66%，分别比TGN提高了1.57%、7.69%和0.56%。在DIOR数据集上，我们的方法在尾部类别上的平均准确率达到87.47%，比TGN提高了3.01%，进一步表明我们的方法能够有效提升尾部类别学习效果。在FGSC-23数据集上，我们的方法在头部类别上的平均准确率为78.21%，在尾部类别上的平均准确率为82.72%，总体准确率为79.98%，分别比TGN显著提高了5.97%、13.79%和8.22%，充分凸显了其在各种细粒度场景中的卓越鲁棒性。

B. 不同认知不确定性之间的关系

为验证我们方法的核心假设，我们分析了所提出的K/S指标与基于熵的认知不确定性之间的关系，以及认知不确定性与类别分布之间的相关性。图2 展示了FGSC-23数据集上K/S指标与基于熵的认知不确定性的散点图，结果表明二者具有很强的

表 IV 不同指数缩放因子下在 FGSC-23 上的性能 σ，其中 ϵ 固定为 0.2。

σ	准确率(%)	平均准确率(%)
1	78.64	79.65
2	78.40	78.74
3	**79.98**	**81.35**
4	78.88	79.75
5	**79.98**	79.75
6	74.64	73.78

相关性，斯皮尔曼系数为0.99。这表明K/S指标能够有效近似基于熵的认知不确定性。然而，基于熵的认知不确定性值集中在较窄的范围内，限制了其动态调整训练权重的灵敏度。相比之下，K/S指标在 $[0, +\infty]$上呈现出更均匀的分布，因此更适合用于损失加权，以提升尾部类别学习。

C. 骨干网络泛化能力

为验证**DUAL**的泛化能力，我们在三种常见骨干网络上评估其性能：EfficientNet- B0、MobileNetV2 和 ResNet-18。如表 III所示，我们的方法在 DOTA、DIOR 和 FGSC-23 数据集上的这三种骨干网络中均始终表现出色。这表明我们方法的有效性并不依赖于特定的网络架构，并且可以灵活地集成到各种骨干网络模型中。

D. 超参数分析

EU-Based Reweighting 中 σ 的影响。我们研究了缩放因子 σ 对 FGSC-23 的影响，并将 ϵ 固定为 0.2。如表 IV 所示，性能在 $\sigma = 3$ 时达到峰值，总体准确率（79.98%）和平均类别准确率（81.35%）均为最高。较小的 σ（例如 1）无法充分强调不确定样本，而过大的 σ（例如 6）则会过度抑制样本权重，导致头部类别欠拟合。这表明，适中的 σ 能有效平衡头部类别和尾部类别的学习。AU-Based Label Smoothing 中 ϵ 的影响。我们进一步考察了参数 ϵ 的影响，并将 σ 固定为 3。较高的 ϵ 会增强对偶然不确定性较高的样本的平滑强度，旨在减少模型对噪声或模糊输入的过拟合。如表 V 所示，性能在 $\epsilon = 0.2$ 时达到峰值，表现最佳。当 ϵ 过小时（例如 0.1），平滑效果有限，模型的抗噪鲁棒性随之降低。相反，当 ϵ 增大至超过 0.2 时，两项准确率指标均逐渐下降。这些结果表明，适度的平滑能够在数据模糊性存在时提升泛化能力，同时不牺牲判别能力。

E. 消融实验

为验证我们提出的不确定性感知长尾学习框架中各组成部分的作用，我们在FGSC-23数据集上进行了消融实验。具体而言，我们逐步整合证据深度学习（EDL）、基于认知不确定性的重加权和基于偶然不确定性的标签平滑，然后评估总体准确率和平均类别准确率。详细结果汇总于表VI。

表V 在不同 ϵ 下 FGSC-23 上的性能，其中 σ 固定为 3。

ϵ	准确率（%）	平均准确率（%）
0.1	77.55	79.23
0.2	**79.98**	**81.35**
0.3	79.49	78.59
0.4	79.13	78.62

表VI 消融实验。**EDL**。表示证据深度学习。**认知不确定性**。表示基于认知不确定性的重新加权，而**偶然不确定性**。表示基于偶然不确定性的标签平滑。

EDL.	EU.	AU.	准确率（%）	平均准确率（%）
✓	✗	✗	72.33	68.24
✓	✓	✗	76.58	75.32
✓	✓	✓	**79.98**	**81.35**

V. 结论与讨论

本文提出**DUAL**，用于分解认知不确定性与偶然不确定性。通过结合基于认知不确定性的样本重加权与 AU 驱动的动态标签平滑，我们的方法显著提升了长尾遥感分类的性能。在 DOTA、DIOR 和 FGSC-23 上进行的大量实验表明，该方法效果显著，尾部类别性能得到明显提升。消融研究验证了**DUAL**各组成部分的必要性。

未来，我们将探索自适应超参数调度，以提升模型在动态噪声场景下的鲁棒性，推动其在复杂现实环境中的广泛应用。

参考文献

[1] G.-S. Xia、X. Bai、J. Ding、Z. Zhu、S. Belongie、J. Luo、M. Datcu、M. Pelillo 和 L. Zhang，《Dota：用于航空图像目标检测的大规模数据集》，载于 *IEEE*计算机视觉与模式识别会议论文集，2018，3974 3983 1, 5 第 – 页。K.

[2] Li、G. Wan、G. Cheng、L. Meng 和 J. Han，《光学遥感图像中的目标检测：综述与新基准》1, 5 *ISPRS*摄影测量与遥感学报，第159卷，307，2020年 296页。–。

[3] Y. Zhang、B. Kang、B. Hooi、S. Yan 和 J. Feng，《深度长尾学习：综述》，*IEEE*模式分析与机器智能汇刊，第 45， 9， 10 795 10 816, 2023 1, 2 卷，第 期，第 – 页。Y.

[4] Cui、M. Jia、T.-Y. Lin、Y. Song 和 S. Belongie，《基于样本有效数的类别平衡损失》，载于 *IEEE/CVF*计算机视觉与模式识别会议论文集，2019年，第9268–9277 1, 2 页。K. Cao、C. Wei、A. Gaidon、N. Arechiga 和 T. Ma，《利用标

[5] 签分布感知边际损失学习不平衡数据集》，神经信息处理系统进展，第 32, 2019 1 卷。M. Chen、Y. Du、

[6] W. Jiang、B. Zhang、S. Feng、Y. Xin 和 C. Wang，《针对长尾噪声数据学习的稳健对数几率调整》，载于 *AAAI*人工智能会议论文集，第39卷，第15期，2025年，第15 830–15 838页。1

[7] T.-Y. Ross 和 G. Doll´ar，《用于密集目标检测的焦点损失》，载于 IEEE计算机视觉与模式识别会议论文集，2017，2980 2988 1, 2 第 – 页。[8] M. Sensoy、L. Kaplan 和 M. Kandemir，《用于量化分类不确定性的证据深度学习》，神经信息处理系统进展，第31, 2018 2 卷。X. Wu、H. Sun、Y. Wang、J. Nie、J. Zhang、Y. Wang、J. Xue 和 L. He，《Avf-mae++：通过高效视听自监督学习扩展情感视频面部掩码自编码器》，载于计算机视觉与模式识别会议论文集，2025年，第9142–9153页。2[9] X. Wu、J. Xue、X. Yin、Y. Shi、L. Fu、D. Huang、Y. Wang、J. Zhang、J. Nie 和 J. Wang，《用于高效情感视频面部分析的可扩展视听掩码自编码器》，智能计算，2025 2 。[10] X. Wu、J. Liu、D. Huang、X. Li、Y. Wang、C. Chen、L. Ma、X. Cao 和 J. Xue，《Vic-bench：使用自由形式中间状态表示对多模态大语言模型的视觉交错思维链能力进行基准测试》，arXiv预印本 arXiv: *2505 14404*, 2025 2 。[11] J. Xue、J. Wang、X. Liu、Q. Zhang 和 X. Wu，《情感视频内容分析：十年回顾与新视角》，大数据挖掘与分析，第8, 1, 118 144, 2024 2 卷，第 期，第 – 页。[12] L. Fidon、M. Aertsen、F. Kofler、A. Bink、A. L. David、T. Deprest、D. Emam、F. Guffens、A. Jakab、G. Kasprian 等，《一种用于可信人工智能的 Dempster-Shafer 方法及其在胎儿脑部磁共振成像分割中的应用》，IEEE模式分析与机器智能汇刊，第46, 5, 3784 3795, 2024 2 卷，第 期，第 – 页。[13] Y. Zhang、B. Hooi、L. Hong 和 J. Feng，《面向与测试无关的长尾识别的多样化专家自监督聚合》，神经信息处理系统进展，第35卷，34 090, 2022年 34 077 4，第 – 页。[14] X. Wang、L. Lian、Z. Miao、Z. Liu 和 S. X. Yu，《通过路由多样化分布感知专家实现长尾识别》，arXiv预印本 arXiv: *2010 01809*, 2020 4 。[15] J. Cui、S. Liu、Z. Tian、Z. Zhong 和 J. Jia，《ResLT：面向长尾识别的残差学习》，IEEE模式分析与机器智能汇刊，第45, 3, 3695 3706, 2022 4 卷，第 期，第 – 页。[16] P. Han、C. Ye、J. Zhou、J. Zhang、J. Hong 和 X. Li，《用于长尾识别的基于潜变量的扩散模型》，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2024, 2639 2648 4 第 – 页。[17] B. Li、Z. Han、H. Li、H. Fu 和 C. Zhang，《可信长尾分类》，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2022, 6970 6979 4 第 – 页。[18] S. Zhang、C. Chen、X. Hu 和 S. Peng，《面向长尾学习的平衡知识蒸馏》，Neurocomputing，第527, 36 46, 2023 4 卷，第 – 页。[19] A. K. Menon、S. Jayasumana、A. S. Rawat、H. Jain、A. Veit 和 S. Kumar，《通过对数几率调整进行长尾学习》，arXiv预印本 arXiv: *2007 07314*, 2020 4 。[20] W. Zhao、J. Liu、Y. Liu、F. Zhao、Y. He 和 H. Lu，《先教教师，再教学生：通过分层蒸馏提升航空图像中的长尾目标识别》，IEEE地球科学与遥感汇刊，第60, 1 12, 2022 4 卷，第 – 页。[21] Y. Bai、S. Shao、S. Zhao、W. Liu、D. Tao 和 B. Liu，《EME：面向长尾遥感图像分类的基于能量的多专家模型》，IEEE地球科学与遥感汇刊，第62卷，第1 12, 2024 4 –页。[22] H. Tang、W. Zhao、G. Hu、Y. Xiao、Y. Li 和 H. Wang，《用于长尾遥感目标分类的文本引导多样化图像合成》，IEEE地球科学与遥感汇刊，2024 4, 5 。[23] X. Zhang、Y. Lv、L. Yao、W. Xiong 和 C. Fu，《光学遥感图像细粒度船舶分类的新基准与属性引导的多层级特征表示网络》，IEEE应用地球观测与遥感选题期刊，第13卷，2020年，第1271–1285页。5[24]