arXiv:2301.00622v1 [cs.CV] 2 Jan 2023

基于空地双视图图像证据融合的可信遥感场景分类

KUN ZHAO、高倩、郝思远、孙杰、周利建*

摘要—多视图（多源、多模态、多视角等）数据在遥感任务中的应用日益广泛，因为与单一来源数据相比，它们能够提供更多有用信息。然而，不同视图的数据质量各不相同，限制了多视图数据的潜在优势；若依据其可信度进行融合，便可最大限度地发挥这些优势。尽管近期基于深度学习的模型能够自适应地学习数据权重，但由于缺乏对多视图数据可信度进行显式量化的研究，这些模型难以解释，进而影响其在下游遥感任务中的性能和灵活性。为弥补这一不足，本文将证据深度学习理论引入空地双视图遥感场景分类任务，以建模每个视图的可信度。具体而言，本文采用证据理论计算不确定性值，用以描述每个视图的决策风险。基于不确定性，本文提出一种新颖的决策级融合策略，确保决策风险较低的视图获得更高权重，从而使分类结果更加可信。我们的方法在两个经典的公共空地双视图遥感图像数据集上取得了最先进的性能，证明了所提方法的有效性。

索引词—多视图数据融合、遥感场景分类、不确定性估计、证据学习、狄利克雷分布。

一、引言

遥EMOTE 感场景分类是遥感与卫星图像分析领域的一项重要研究课题，旨在依据遥感图像的不同语义特征，将场景图像划分为离散且有意义的土地利用和土地覆盖类别。它广泛应用于多种实际场景，包括地理图像检索 [1]–[3]，自然灾害检测 [4]–[6]，植被制图 [7]–[9]以及地理空间目标检测 [10]–[12]。

过去几十年来，研究人员在利用单一数据源设计高效分类器方面取得了巨大进展，例如高光谱 [13]，多光谱 [14]，合成孔径雷达 [15]，高分辨率图像 [16]。尽管遥感数据更容易获取，但由于俯视图像缺乏多样而详细的信息，仅使用俯视图像进行遥感场景分类 [17]仍被视为一项具有挑战性的任务。值得庆幸的是，随着各种社交媒体平台（如微博、Flickr 等）和地图软件（如谷歌地图、必应地图等）的兴起，从各种来源收集地理标记数据变得越来越容易。Antoniou 等人 [18] 将带有地理标记的社交媒体图像用作输入数据，并证明这些图像能够提供有关土地利用和土地覆盖的重要信息。Pei 等人 [19] 利用来自手机的整合数据，在网格层面开展土地利用分类。Zhang 等人 [20] 和 Kang 等人 [21]利用街景图像对城市土地功能区进行分类。这些研究表明，地面的地理标记数据能够为土地利用和土地覆盖分类提供有用信息。

将俯视图像与其他地理标记数据融合以进行遥感场景分类的研究日益受到关注。Hu 等人 [22] 和 Liu 等人 [23] 将卫星图像与兴趣点（POI）及其他社交媒体数据相结合，对城市地块进行分类，证明了社交媒体数据有助于提升场景分类性能。Jia 等人 [24] 采用支持向量机，分别对从遥感图像和移动电话位置数据中提取的特征进行分类，再通过决策级融合整合两种分类结果，以实现城市土地利用分类。Tu 等人 [25] 提出了一种数据融合框架，用于融合遥感图像与人类感知数据，并采用层次聚类方法识别城市功能区。Hong 等人 [26]针对高光谱-多光谱对应关系提出了公共子空间学习（CoSpace），将遥感场景分类问题转化为跨模态学习问题。

在用于遥感场景分类的多源数据中，空地双视图图像因其地理覆盖范围广、易于获取而被频繁使用 [27]。Cao 等人 [28] 从稀疏分布的街景图像中提取语义特征，使其与航空图像的空间分辨率相匹配，随后通过深度神经网络融合这些特征，以对土地利用和土地覆盖进行分类。Srivastava 等人 [29] 采用端到端学习，对单模态特征提取与融合进行学习，从遥感图像、街景图像以及 OpenStreetMap 的标注中预测土地覆盖标签。对于场景分类，决策级融合优于直接进行数据级融合 [30]。由于空地双视图图像由具有不同分布的异构特征表示，Deng 等人 [31] 提出了一种半监督流形正则化多核学习（SMRMKL）算法来解决这些问题。

然而 此外，多源数据的使用也带来了 关于

通讯作者：周利建。KUN ZHAO、高倩、郝思远、孙杰和周利建就职于青岛理工大学信息与控制工程学院（中国青岛，邮编266520）。电子邮箱：sterling1982@163.com、q17863936190@163.com、lemonbananan@163.com、sunjie1979@qut.edu.cn、zhoulijian@qut.edu.cn

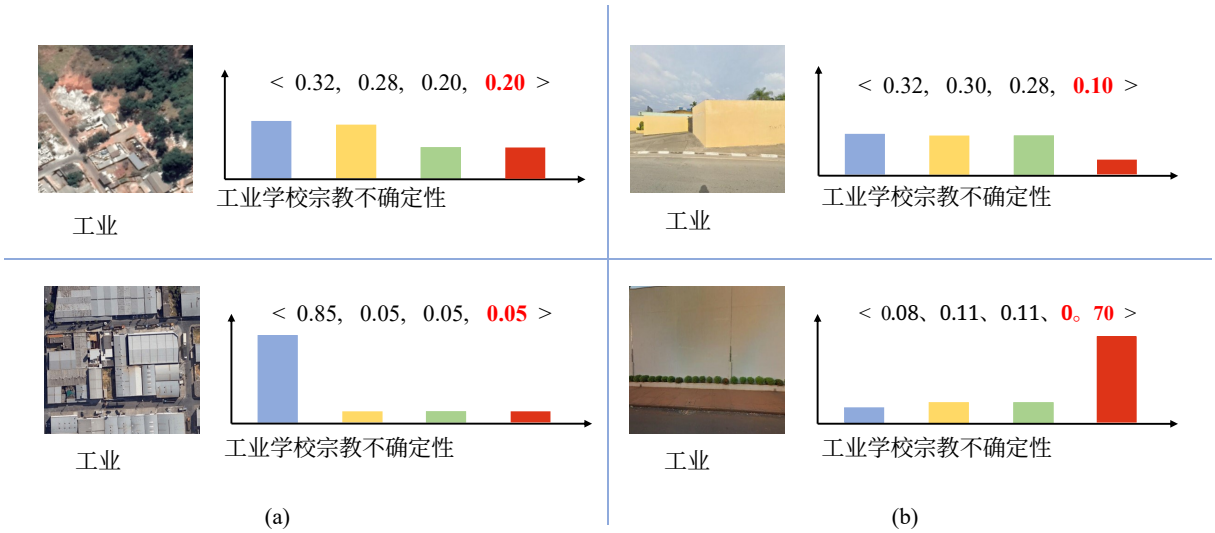


图1. 哪幅图像更可信？(a) 航空图像和(b) 地面图像中的不同类型样本不确定性。每个直方图中的前三个柱表示各类别的可能概率。每个直方图中最后一个红色柱表示可能的不确定性值。

样本不确定性的增加 [32]。特别是，作为遥感多模态数据的一种特殊类型，由于航空图像存在类间相似性和类内多样性，以及地面图像质量较差（例如严重遮挡、室内视角等），空地双视图图像的样本不确定性问题更加严重。例如，图 1（a）展示了工业区的两幅航空图像。底部图像显然更可信，因为它清晰地描绘了一个工业区。而顶部图像应具有更高程度的不确定性，因为它看起来也像是学校区域。样本不确定性的量化（红色区间）有助于描述航空图像中普遍存在的类间相似性和类内多样性，从而促进更好的分类。图 1（b）展示了另一种常见情况。在这种情况下，顶部的地面视角图像所描绘的场景可以被看作三个类别中的任意一个，而底部图像则完全没有有用信息。显然，顶部图像包含更多有用信息，因此也更可信。然而，如果没有对样本不确定性进行量化（红色区间），这两幅图像将在场景分类中发挥相同的作用。遗憾的是，迄今为止，对多视图遥感数据融合中的样本不确定性研究仍然不足。

本文基于证据深度学习提出了一种新颖的融合方法，其出发点是对样本不确定性进行显式建模 [33]，用于空地双视图图像的遥感场景分类。本文的主要贡献如下。

- 一种**基于证据理论的样本不确定性量化（TEB-SUQ）**方法应用于航空和地面图像两个视图，以衡量二者融合过程中的决策风险。
- 本文提出一种**证据融合** 策略，在决策级融合空地双视图图像，用于遥感场景分类。与现有的其他决策级融合网络不同，所提出的策略不仅关注分类概率还关注每个视角的决策风险。因此，最终结果将更多地取决于决策风险较低的视图。
- 本文设计了一种更简洁的损失函数，即**倒数损失**，用于同时约束单个视图的不确定性及其融合结果的不确定性。该损失函数不仅可用于训练端到端的空地双视图遥感场景分类模型，还可用于训练无需特征提取的融合分类器。

II. 相关工作

A. 遥感多视图数据融合

多视图学习的目标是利用或关联来自不同视图（例如不同来源、模态或视角）的数据来构建模型 [34]。多视图学习的融合方法可在三个层级上实现 [35]，即**数据级、特征级**和**决策级**。**数据级** 融合策略通常融合来自多个数据源 [36] 或多分辨率图像 [37],等的原始或预处理数据。多尺度小波 [38]，、二维经验模态分解 [39] 和拉普拉斯金字塔 [40] 被用于融合多光谱图像数据。另一方面，**特征级** 融合会结合从多视图数据中提取的多个中间特征。融合后的特征随后用于下游任务，例如通过面向对象的多尺度分割，将其分类为土地利用和土地覆盖类别 [41]。Wang 等人 [42]提出了一种多层注意力融合网络，该网络以层次化方式融合并分类高光谱图像和激光测距数据（激光雷达）的特征。**决策级融合**采用不同的融合规则来汇总多个分类器 [43], [44],的预测结果，其中每个分类器均由一个独立模型得到。Yu 等人 [45] 将航空图像及从图像中提取的两个图像块输入三个具有不同感受野的卷积神经网络，然后使用概率融合模型进行最终分类。Yang 等

人。 [46]提出了一种超高分辨率遥感图像场景分类方法：去除原始图像中的部分光谱，将图像输入卷积神经网络（CNN），并结合网络输出的预测结果。Zhang 等人 [47]提出了一种梯度增强随机卷积网络（GBRCN）结构，可在决策级融合中成功结合多个深度神经网络进行场景分类。由于每个视图的特征都是单独学习的，彼此互不影响，因此决策级融合是三个层级中最灵活的。由于可以忽略不同视图中数据结构的差异，这种方法非常适合空地双视图图像的遥感场景分类任务。然而，融合质量在很大程度上取决于各视图决策结果的可信度。因此，量化各视图预测结果的不确定性尤为重要。

B. 遥感中的空地双视图图像融合

与其他常用的多视图遥感图像（例如卫星图像与激光雷达图像、多光谱图像等）不同，匹配的航空图像与地面图像之间差异过大，无法直接在数据级进行融合。因此，通常采用特征级和决策级的融合方法。空地双视图图像融合最初应用于图像地理定位。Lin 等人 [48]利用四种人工设计特征，并加入土地覆盖特征作为第三种模态，将来自必应地图的高分辨率正射影像（HRO）与来自 Panoramio 的街景图像进行匹配。为扩展这一方法，他们使用孪生卷积神经网络，在 Google街景（GSV）图像与 45 度倾斜航空影像 [49]之间学习深度特征。Workman 等人 [50]通过端到端深度网络融合航空图像和 GSV 图像，该网络针对三种不同的分类问题输出航空图像的像素级标注：土地利用、建筑功能和建筑年代。Zhang 等人 [20] 融合机载激光探测与测距（LiDAR）数据、HRO 和 GSV 图像，用于土地利用分类。在其研究中，选取了十三种地块特征作为随机森林分类器的输入变量。Cao 等人 [28] 使用来自必应地图和 GSV 的图像，结合由 SegNet [51]演化而来的双流编码器和单解码器架构，实现土地利用分割。Hoffmann 等人 [52] 使用双流卷积神经网络模型进行建筑功能分类。他们通过融合航空图像和地面视图图像的深度特征，为俯视图像预测四个类别标签，即商业、住宅、公共和工业。其模型借助决策级融合策略，显著提高了精确率。Srivastava 等人 [29]将其早期工作 [53]扩展为一种多模态策略，利用空地双视图的互补性。他们基于双流卷积神经网络模型，采用典型相关分析（CCA）处理航空图像缺失的情况。

C. 不确定性估计

为了更全面地描述结果，我们不仅致力于提高分类准确率，还致力于对预测结果进行风险评估。不确定性估计旨在输出预测结果的不确定性并降低过度自信。关于不确定性估计，已有大量研究。Jiang 等人 [54]提出了“信任分数”这一术语，用作置信度的度量，其置信度准则是样本到最近类别的距离与样本到预测类别的距离之比。然而，该方法的计算开销很大。另一个缺点是，局部距离计算 [55]在高维空间中并无意义，还可能对结果产生负面影响。近年来，贝叶斯方法在神经网络不确定性估计方面受到了广泛关注。通过对模型参数进行分布建模并对这些参数进行边缘化，贝叶斯神经网络（BNN） [56] 对预测分布中的不确定性进行了建模。真实类别概率（TCP） [57] 提出了一个额外模块来推导输出的不确定性；这是一种新准则，也有助于预测失败。证据深度学习（EDL） [33]采用主观逻辑对输出概率的不确定性进行建模，并在类别概率上建立狄利克雷分布。深度集成 [58]通过对多个与训练数据一致的模型的预测结果取平均，来捕捉“模型不确定性”。确定性不确定性量化（DUQ） [59] 通过质心的相对位置计算径向基函数（RBF）距离，以表征不确定性。基于高斯过程映射的深度置信网络，即深度高斯过程 [60]，利用非参数核函数对样本不确定性进行建模。提出不确定性感知注意力（UA）机制 [61]，以捕捉异方差不确定性或实例特定不确定性，从而更准确地校准预测不确定性。随着不确定性相关研究不断增多，人们发现不确定性可以降低，使网络做出的决策更加确定 [56]。遗憾的是，以上所有方法都专注于估计单视图数据的不确定性，尽管利用不确定性融合多视图能够提升性能和可信度。

III. 方法

本节介绍用于航空—地面双视图图像遥感场景分类的**证据融合网络（EFN）**。首先简要介绍网络的整体架构；其次，详细说明如何利用证据深度学习获取样本不确定性，以及如何进行证据融合；最后，介绍用于训练该网络的**倒数损失**。

A. 概述

近期研究 [29], [30] 表明，在空地双视图图像遥感场景分类任务中，采用深度学习框架的决策级融合通常表现最佳。然而，由于缺乏对其可信度的度量，多视图决策（即输入样本的推断结果）在融合过程中发挥着同等重要的作用，尽管这些样本具有

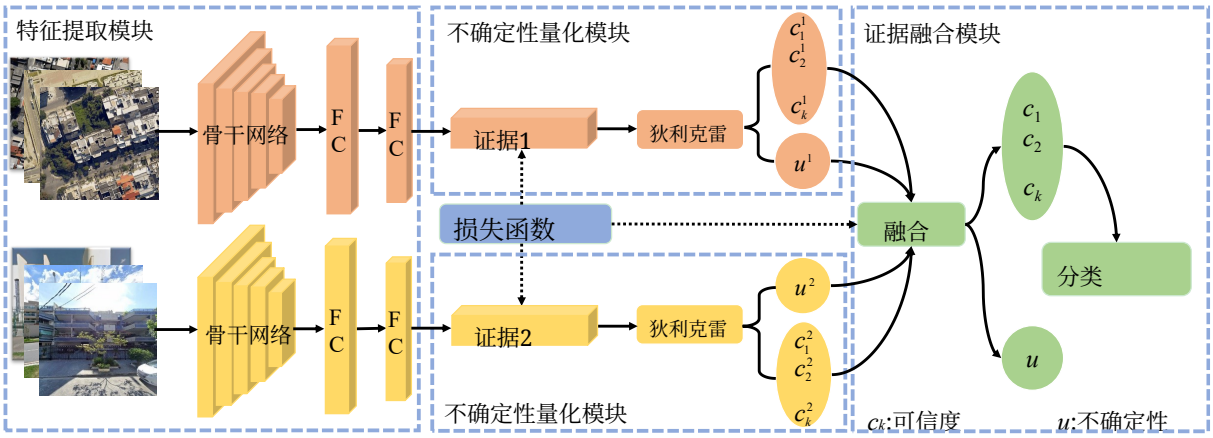


图 2. 所提出的空地双视图图像证据融合网络（EFN）整体框架。

不确定性并不相同。一方面，尽管空中视角图像的质量相对较高，但类内多样性和类间相似性问题更为突出。另一方面，尽管地面视图图像提供了更多可区分的细节，但常因遮挡、视角变化幅度大以及定位精度低等问题而质量欠佳。若不考虑样本不确定性，融合过程就会受到大量“过度自信”结果的误导。所提出的证据融合网络（EFN）通过明确量化样本不确定性来评估各视角的决策风险，并在融合时根据风险为各视角的决策分配不同权重。因此，最终分类会更多地依赖决策风险较低的视角，从而更加可信。EFN 的总体框架如图2所示。该框架可进一步分为三个模块：特征提取模块（FEM）、不确定性量化模块（UQM）和证据融合模块（EFM）。

在每个视图中，使用训练好的骨干模型提取一维特征向量，然后将其输入带有非负激活函数的全连接层，以获得用于计算不确定性的“证据”。随后，将证据分别映射为狄利克雷分布的浓度参数，用于获得可信度和不确定性（更多详情见第III-B 节）。最后，采用加权决策级融合来融合航拍与地面双视图图像的可信度和不确定性（更多详情见第III-C 节）。这种方法充分利用了各视图的不确定信息，在最终决策中可以降低决策风险较高的视图所占的权重。

所提出的 FEM 分为两个子网络，每个子网络包含一个骨干网络和两个额外的全连接层。两个子网络的骨干网络可以相同，也可以不同，因为所提出的方法独立于骨干网络。正是这种灵活性使所提出的方法适用于任何网络结构。最后，末端的 Softmax算子
将最后一个全连接层替换为非负激活函数（例如 ReLU）。这一简单的替换操作使所提出的方法具有极强的可移植性。

B. 基于证据学习的不确定性估计

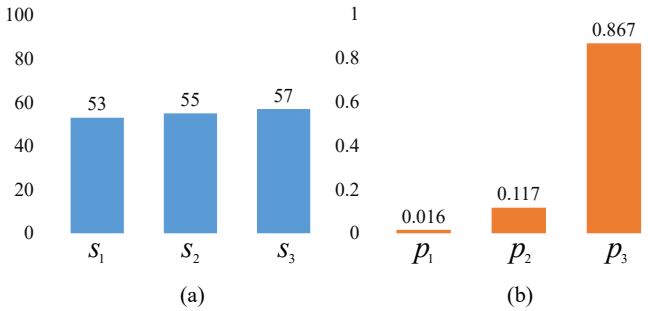


图3. softmax 导致的“过度自信”：(a) 为特征提取后三个类别的可能得分，(b) 为经 softmax 映射得到的对应概率。

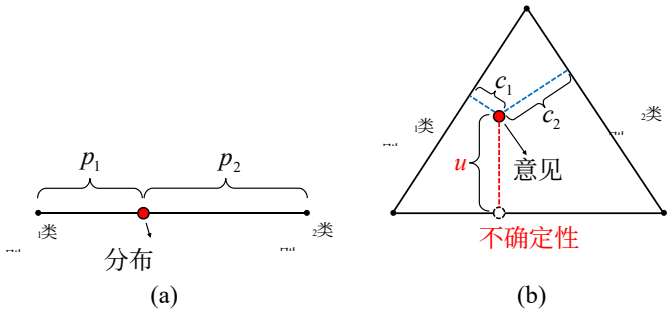


图4. 二分类问题示例：(a) 为由 softmax 生成的样本类别概率分布；(b) 为通过不确定性估计操作生成的同一样本的“意见”，其中 u 是不确定性，c_i 是类别 i 的可信度。

众所周知，使用 softmax 算子将输出的连续激活值转换为离散的类别概率，是分类网络的黄金标准。然而，近期研究表明，softmax 算子存在诸多缺陷。首先，它可能导致“过度自信”问题 [62]，，即通过放大不同类别预测分数之间的差异，迫使模型将不确定样本归入某个确定类别。

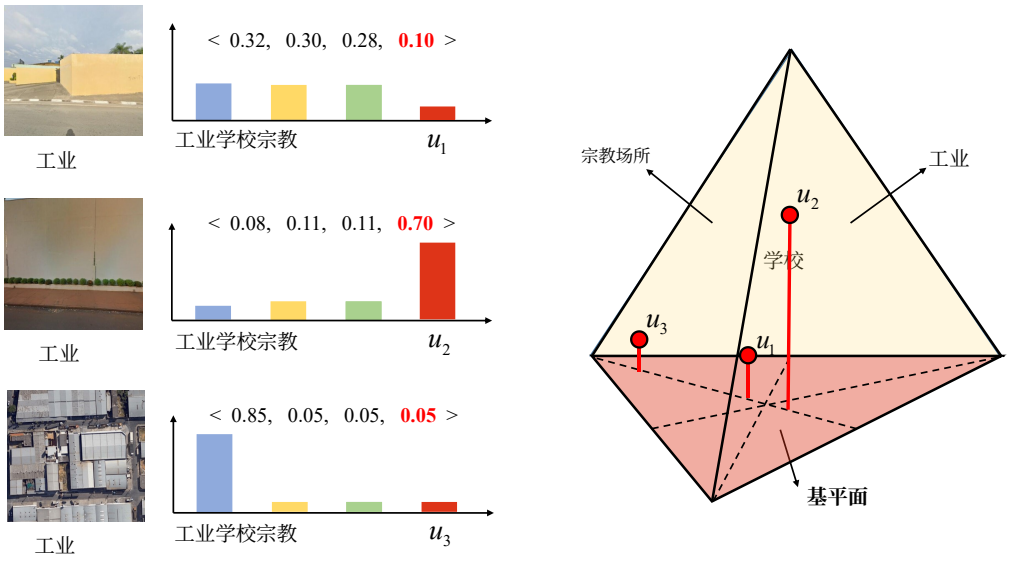


图 5. 所提出的不确定性估计操作的一个多分类案例：(a) 展示了三个样本的“意见”（两个地面视图图像和一个航空图像）；(b) 为包含示例意见的意见四面体。

以图3为例，图3(a)中三个类别的分数非常接近，表明模型无法确定输入样本所属的类别。然而，经过图3(b)所示的 Softmax映射后，各类别概率之间的差异大幅增加，使模型产生结果非常确定的错觉。

其次，softmax 只能为样本的类别概率提供点估计，而无法提供相关的不确定性，这往往会导致不可靠的结论 [59]。当输入图像中没有相关特征时，模型会被诱导甚至被迫在最终预测中给出一个结果，却没有任何证据支持 [63]。图4展示了在一个二分类问题案例中，softmax 与不确定性估计操作的区别。在图4(a) 中，softmax 生成了一个类别概率分布，其中 p_i 是类别 i的概率，并且我们有 $p_1 + p_2 = 1$ 和 $p_1 < p_2$，这意味着分类器更倾向于将样本分类为类别 2。然而，我们并不知道这一预测是否可信。要评估样本的可信度，需要对不确定性进行明确建模。

为了解决 softmax 在场景分类中的缺点，使用非负激活函数（如 ReLU）替代 Softmax算子，输出样本属于每个类别的“证据”。基于这些证据，可以通过证据的邓普斯特-谢弗理论（DST）描述样本不确定性 [64]，该理论在构建“意见空间”时将“基平面”（见图 5（b））纳入考虑，从而能够明确表达“无知”，即缺乏关于样本真实性的证据，并进一步将样本的整体可信度分配给可能的类别标签。在图 4（b）中，通过不确定性估计操作生成了同一样本的一个“意见”（等边三角形中的红点），其中 c_i 是类别的可信度

类别 i 其数值等于“意见”到代表类别 i的边的距离，而 u 是样本不确定性，其数值等于“意见”到基平面的距离。众所周知，等边三角形中任意一点到其三条边的距离之和等于该等边三角形的高。当将等边三角形的高设为 1 时，我们有 $c_1 + c_2 + u = 1$ 和 $p_1 < p_2 < u$，这意味着分类器更有可能发现该样本不可靠，而不是做出二元决策。

图5 展示了包含三个类别的示例，这些示例已在图1中展示。在这种情况下，图4 中的意见等边三角形增加了一个维度，成为意见四面体 [65]其中每个类别都表示为一个侧平面。每个类别的可信度等于意见点到相应侧平面的距离。意见点的垂直高度表示样本不确定性。

更一般地说，在包含 K 个互斥单元素集（例如类别标签）的空间中，DST 可以为每个单元素集提供一个可信度以及总体不确定性。样本的预测通常称为“意见”。 K 侧平面表示可能的类别，其可信度值等于这些平面到意见点的距离。意见的不确定性被解释为对样本真实性的度量。这些 K 可信度值 c_k 和不确定性 u 均为非负数，且总和为一，即

$$u + \sum_{k=1}^{K} c_k = 1, \tag{1}$$

其中， $u \geq 0$和 $c_k \geq 0$分别表示 $k = 1$、2、 $\cdots$、 K的总体不确定性和 k类的可信度。对于每个视图，可以利用各类别的证据计算其可信度。样本的非负证据集可表示为**证据** $= [e_1, e_2, \cdots, e_K]$。令 e_k表示

该样本中第 k类的证据，因此，可信度c_k 和不确定性 u可计算如下：

$$c_k = \frac{e_k}{\sum_{k=1}^{K}(e_k+1)}, \tag{2}$$

$$u = \frac{K}{\sum_{k=1}^{K}(e_k+1)}. \tag{3}$$

从上述公式可以看出，样本不确定性与总体证据量成反比。当尚未学习到任何证据时，每个类别的可信度均为 0，样本不确定性为 1。随着可用于分类的证据越来越多，可信度会增加，而样本不确定性则会降低。

DST 的概念可进一步形式化为狄利克雷分布 [65]，该分布基于概率“范围”，而非单个概率值。正是这一“范围”在量化不确定性方面发挥着核心作用：当某一类别的概率范围过宽时，意味着分类器对其意见并不确定；相反，当范围收缩为一个离散的概率值时，则意味着分类器对预测非常确定。换言之，借助狄利克雷分布（亦称“分布的分布”），样本不确定性可以解释为一阶概率的二阶概率。而狄利克雷分布的浓度参数能够精确描述这一“范围”。相比之下，标准神经网络分类器的输出是针对每个样本的可能类别所作的离散概率分配，这种输出“过于确定”，会导致过度自信。

具体而言，为了通过 DST 计算样本不确定性，每个意见都对应一个浓度参数为

$$\alpha_k = e_k + 1. \tag{4}$$

也就是说，可以分别利用以下公式，根据相应的狄利克雷分布轻松获得可信度和不确定性：

$$c_k = \frac{\alpha_k - 1}{\alpha_0}, \tag{5}$$

$$u = \frac{K}{\alpha_0}, \tag{6}$$

$$\alpha_0 = \sum_{k=1}^{K}\alpha_k. \tag{7}$$

对于一个 K 分类问题，浓度参数为 $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_K]$，狄利克雷分布定义为

$$D(\mathbf{p}|\boldsymbol{\alpha}) = \begin{cases} \frac{1}{B(\boldsymbol{\alpha})}\prod_{k=1}^{K} p_k^{\alpha_k - 1} & \mathbf{p} \in S_K \\ 0 & otherwise \end{cases}, \tag{8}$$

其中，$B(\boldsymbol{\alpha})$ 是 K 维多元贝塔函数，S_K 是 K维单位单纯形，

$$S_K = \{\mathbf{p}|\sum_{k=1}^{K} p_k = 1 \quad and \quad 0 \le p_1, p_2, \cdots, p_K \le 1\}. \tag{9}$$

输入场景图像时，如果观察到它与某个 K 类别相似，狄利克雷分布的相应浓度参数就会增加。随后，狄利克雷分布将根据新样本进行更新。因此，狄利克雷分布的浓度参数与每个类别的证据相关。至此，本节中的**TEB-SUQ** 已准备就绪，可用于航拍和地面图像的两个视图，以衡量它们融合过程中的决策风险。

C. 证据融合

对于场景分类，双视图图像的融合确实可以利用不同视图的互补信息。然而，由于第 I 节中提到的航拍视图图像和地面视图图像存在的问题（见图 1），直接融合双视图的 softmax 输出会忽略各视图的决策风险，从而导致结果不可靠。正如第III-B节所述，不确定性估计是量化意见可信度的关键。利用两个视图中的样本不确定性，可以最大限度地发挥双视图图像的潜在优势，并使分类结果更加可靠。

为了降低最终分类中决策风险较高的视图所占的权重，本节基于式**证据融合**提出了一种新颖的决策级融合策略，2 和式3 该策略将各视图的样本不确定性作为其决策风险，用于分配最终预测中的权重。

经过 UQM 后，可得到两个视图的意见 $\mathcal{O}^1 = \{c_1^1, c_2^1, \ldots, c_K^1, u^1\}$和 $\mathcal{O}^2 = \{c_1^2, c_2^2, \ldots, c_K^2, u^2\}$，其中上标表示不同视图的编号。最终决策意见 $\mathcal{O} = \{c_1, c_2, \ldots, c_K, u\}$ 计算如下：

$$c_k = \frac{1}{\lambda}[c_k^1 c_k^2 + (1-u^1)c_k^1 + (1-u^2)c_k^2], \tag{10}$$

$$u = \frac{1}{\lambda}u^1 u^2, \tag{11}$$

$$\lambda = u^1 u^2 + (1-u^1)^2 + (1-u^2)^2 + \sum_{k=1}^{K} c_k^1 c_k^2, \tag{12}$$

其中，λ 是用于执行归一化的缩放因子，可确保融合后式1 仍然成立。

最终决策意见由两个视图的意见融合而成。该融合策略旨在涵盖三种一般情况。

- 当两个视图的决策风险都较低（ u^1 和 u^2都较小），也就是说，两个视图的可信度都较高时，最终预测的可信度也应较高（ c 会较大）。
- 当两个视图中一个的决策风险较高、另一个较低时（只有 u^1 或 u^2 较大），最终预测应更多地依赖决策风险较低的视图。

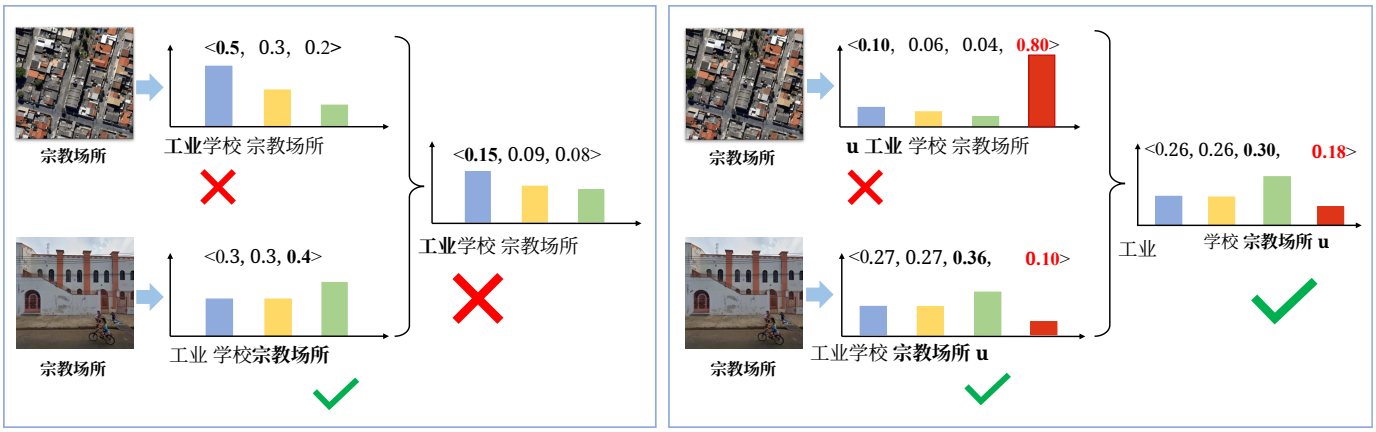


图 6. 在空地双视图的决策上采用不同融合策略时得到的不同预测结果：(a) 对两个视图的 Softmax 输出进行乘法融合；(b) 对 UQM 输出的意见采用所提出的证据融合。所有预测均取最高分对应的类别标签，无论是单视图预测还是融合后的预测。

- 当两个视图的决策风险都较高时（u^1 和 u^2都较大），两个视图的可信度都较低，因此最终预测的可信度也应较低（c会较小）。

从两个视图融合得到最终决策意见 $\mathcal{O}$ 后，可根据以下公式计算第 k 类的融合证据 e_k：

$$e_k = \frac{K \cdot c_k}{u}, \tag{13}$$

以及 α_k 狄利克雷分布中 k类的浓度参数可通过公式4 更新，用于计算训练步骤中的损失；而 e_k最大的类别是测试步骤中的最终预测标签。

图 6 展示了一个示例，以说明所提出的证据融合如何发挥作用。图 6(a) 展示了对空地双视图图像的 Soft max 输出进行乘法融合的结果。最终预测同等依赖于两个视图的决策。忽略决策风险会导致分类结果错误。图 6(b) 展示了对 UQM 输出的意见进行所提出的证据融合。根据 TEB-SUQ 计算出的各视图的不确定性（决策风险），最终预测会更多地依赖风险较低视图的决策，从而获得正确的分类结果。也就是说，所提出的方法能够获得更可信的分类结果。

D. 损失函数

如第 III-A 节所述，在 FEM 中，softmax 层被非负输出激活层（例如 ReLU、Softplus 等）取代，以为狄利克雷分布预测提供证据向量。为训练所提出的模型，我们设计了以下损失函数，并将其应用于每个视图：

$$L(\boldsymbol{\alpha}) = \sum_{i=1}^{N} (L_{pc}(\boldsymbol{\alpha}_i) + L_{nc}(\boldsymbol{\alpha}_i)), \tag{14}$$

$$L_{pc}(\boldsymbol{\alpha}_i) = \sum_{k=1}^{K} y_{ik} \left[\psi(\alpha_{i0}) - \psi(\alpha_{ik})\right], \tag{15}$$

$$L_{nc}(\boldsymbol{\alpha}_i) = \sum_{k=1}^{K} (1 - y_{ik}) \frac{1}{\psi(\alpha_{i0}) - \psi(\alpha_{ik})}, \tag{16}$$

其中 $L_{pc}(\boldsymbol{\alpha}_i)$ 和 $L_{nc}(\boldsymbol{\alpha}_i)$ 分别是第 i 个样本的正类损失和负类损失，$\psi(\cdot)$ 是 digamma 函数，该函数在$(0, +\infty)$ 内单调递增，α_{i0} 与公式 7一致，K 是类别数，而 N 是样本数。

如式 14 所示，所提出的损失函数清晰地区分了正类和负类的惩罚项，易于解释。此外，式 15 和式 16表明，正类损失与负类损失互为倒数。因此，本文将所提出的损失函数称为“倒数损失”。更具体地说，由于 $\alpha_{ik} = e_{ik}+1$ 和 $e_{ik} > 0$，因此 $\psi(\alpha_{ik})$ 随α_{ik}单调递增，而 $L_{pc}(\boldsymbol{\alpha}_i)$ 随 α_{ik}单调递减，这确保每个样本的正类产生更多证据。相反，$L_{nc}(\boldsymbol{\alpha}_i)$ 随 α_{ik} 单调递增，以确保每个样本的负类产生更少证据。接下来，以 $L_{pc}(\boldsymbol{\alpha}_i)$ 为例，推导所提出的倒数损失与狄利克雷分布之间的关系。

对于多类分类任务，最常用的是交叉熵损失函数（交叉熵损失）。对于特定样本，其交叉熵损失可通过以下方式计算：

$$L_{ce} = -\sum_{k=1}^{K} y_k \log p_k, \tag{17}$$

其中，p_k 是第 k类的预测概率，y_k 是其类别标签，正类取值为 $y_k = 1$，负类取值为 $y_k = 0$。在所提出的模型中，softmax 算子被非负激活函数取代。因此，无法直接使用交叉熵损失。由于证据

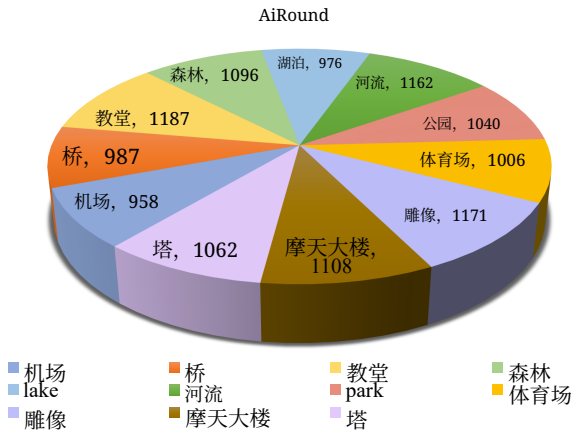


图7。AiRound数据集的类别分布。



图8. AiRound数据集的样本。

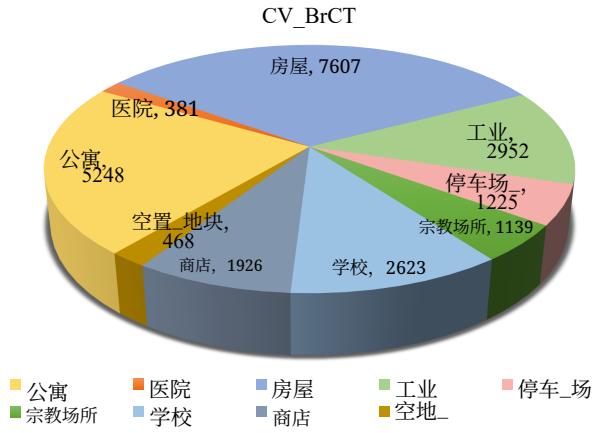


图9. CV-BrCT数据集的类别分布。



图10. CV-BrCT数据集的样本。

该样本 e 由 FEM 得到的结果可映射为狄利克雷分布 $\boldsymbol{\alpha}$ 的浓度参数$D(\mathbf{p}|\boldsymbol{\alpha})$ 利用式 4，交叉熵损失的贝叶斯风险 $L_{cer}(\boldsymbol{\alpha})$ 可计算为

$$\begin{aligned} L_{cer}(\boldsymbol{\alpha}) &= \int L_{ce} D(\mathbf{p}|\boldsymbol{\alpha}) d\mathbf{p} \\ &= \int (-\sum_{k=1}^{K} y_k \log p_k) D(\mathbf{p}|\boldsymbol{\alpha}) d\mathbf{p}. \end{aligned} \tag{18}$$

由于 p_k是一个 $D(\mathbf{p}|\boldsymbol{\alpha})$ 随机变量，函数 $\log p_k$ 是狄利克雷分布的充分统计量。因此，可以使用指数族微分恒等式，得到 $\log p_k$ [66] 的解析表达式：

$$\begin{aligned} \mathbb{E}_{D(\mathbf{p}|\boldsymbol{\alpha})}(\log p_k) &= \int (\log p_k) D(\mathbf{p}|\boldsymbol{\alpha}) d\mathbf{p} \\ &= \psi(\alpha_k) - \psi(\alpha_0). \end{aligned} \tag{19}$$

在这方面，公式18 可进一步展开为

$$\begin{aligned} L_{cer}(\boldsymbol{\alpha}) &= \int (-\sum_{k=1}^{K} y_k \log p_k) D(\mathbf{p}|\boldsymbol{\alpha}) d\mathbf{p} \\ &= -\sum_{k=1}^{K} y_k \int (\log p_k) D(\mathbf{p}|\boldsymbol{\alpha}) d\mathbf{p} \\ &= -\sum_{k=1}^{K} y_k \left[\psi(\alpha_k) - \psi(\alpha_0)\right] \\ &= \sum_{k=1}^{K} y_k \left[\psi(\alpha_0) - \psi(\alpha_k)\right]. \end{aligned} \tag{20}$$

鉴于公式 20 只能惩罚正类别， L_{cer}可写为 L_{pc}。最终得到 $L_{pc}(\boldsymbol{\alpha}_i)$ 在公式 15 中适用于所有样本的情形。

为了确保两个视图都能为场景分类提供合理的意见，从而改进融合后的整体意见，最终采用多视图全局互惠损失：

$$L_{global} = L^1 + L^2 + L_{fused}, \tag{21}$$

其中 L^1， L^2 分别表示第一视图和第二视图的倒数损失（式14），而 L_{fused} 则根据式14，利用融合参数并结合式10、式11、式13和式4得到。

空中视角

u = 0.2258



证据

机场	桥	教堂	森林	lake	park
9.7799e-12	2.7680e-18	1.1594e-06	9.9307e-15	2.8358e-07	5.9592e-12
河流	摩天大楼	体育场	雕像	塔	
1.4766e-12	2.6596e-21	**3.1589e+01**	6.1345e+00	4.5939e-06	

地面视图

u = 0.5428



证据

机场	桥	教堂	森林	lake	park
1.6814e-01	**3.2422e+00**	1.6782e+00	4.0884e-03	6.2256e-03	7.3449e-05
河流	摩天大楼	体育场	雕像	塔	
3.0051e+00	9.7359e-04	1.1220e+00	1.1253e-02	2.8245e-02	

图 11. TEB-SUQ 计算的 AiRound 数据集中类别为“体育场”的第 9360855 号样本的不确定性和证据。加粗的证据值对应预测结果。

IV. 实验

A. 数据集描述与实验设置

本文通过在两个公共空地双视角图像数据集上进行实验，验证了所提出方法的性能 [30]。

AiRound 数据集收集了来自世界各地的地标。它由来自三种不同视图的图像组成：哨兵-2图像、高分辨率 RGB 航空图像和地面图像。哨兵-2图像的尺寸为 224×224 像素。航空图像的尺寸为 500×500像素。地面图像通过两种不同的方式获取，即 Google Places 数据库和 Google Images。因此，它们的尺寸各不相同。其标签取自公开可用的 OpenStreetMap 数据。如图 7所示，AiRound 中有 11 个不同的土地利用类别，共包含 11,753 组图像。我们的实验使用航空图像和地面图像。图8 展示了若干样本。

CV-BrCT数据集包含9个类别的23,569对图像：公寓、医院、房屋、工业、停车场、宗教场所、学校、商店和空地。每对图像由一张航空视角图像和一张地面视角图像组成，两者均为 500×500 RGB图像。类别分布和若干样本分别如图9和图10所示。

对于这两个数据集，我们从每个类别中随机选取80%的样本作为训练/验证集，其余20%作为测试集，从而形成一个数据划分。除第IV-B1 节外，所有测试结果均为10次划分的平均值。随后，按照9:1的比例将训练/验证集随机划分为训练集和验证集。

所有模型均使用 PyTorch 在配备 GTX 1080Ti 显卡的计算机上进行模拟。训练期间的具体设置如下：批量大小：128；特征提取和可信融合的训练轮次均为 200，并保存验证集数据上表现最佳的所有模型；学习率调度：初始值为 0.01 的余弦衰减；优化器：特征提取使用随机梯度下降，可信融合使用 Adam；权重衰减：0.1；随机梯度下降的动量：0.9。所有特征提取模型均通过微调官方发布的预训练模型进行训练。

为定量评估各模型的性能，我们采用分类准确率（Acc）和 F1 分数（F1）作为评价指标。

B. 消融研究

1）不确定性估计有效性的验证：图11 展示了AiRound中一个体育场类别的研究案例。通过TEB-SUQ，获得了两个不同视图图像的不确定性和证据。航空图像中大于1.000的证据值为31.589（体育场）和6.135（雕像），呈现出良好的集中性。因此，其不确定性仅约为0.226。相比之下，地面图像中大于1.000的证据值为3.242（桥）、3.005（河流）、1.678（教堂）和1.122（体育场），分布更加分散。因此，其不确定性约为0.543，这表明模型对其预测结果的置信度不到一半。从图像中可以看出，上述结论符合直观判断。

从统计角度来看，图12 展示了 AiRound 和 CV-BrCT 测试集中各视图样本的不确定性分布。图中清晰地显示，AiRound 的样本在不确定性较低的区域分布得更密集，且峰值更高。这一优势在地面视图中尤为明显（图12（b））。换言之，经过 TEB-SUQ 计算后，AiRound 的质量高于 CV-BrCT，尤其是在地面视图方面。以下事实支持这一结论。首先，CV-BrCT 的样本数是 AiRound 的两倍以上。其次，与 AiRound 的均匀分布不同，CV-BrCT 的类别分布呈现典型的长尾分布（图9）。众所周知，样本数增加以及类别分布不平衡都会导致数据质量下降。最后但同样重要的是，这两个数据集采集地面视图图像所采用的方法有所不同 [30]。AiRound 的地面视图图像主要来自 Google Places 数据库，这是一个广为人知的高质量数据集。而 CV-BrCT 的地面视图图像则全部通过 Google Images 搜索引擎获取，无法保证图像质量。

为了进一步验证 TEB-SUQ 在测量样本不确定性方面的有效性，进行了主观评估

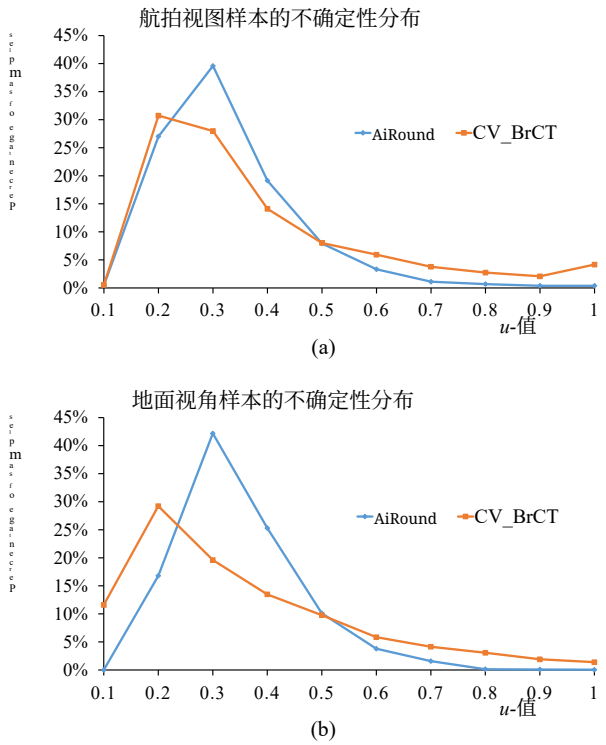


图12. AiRound 和 CV BrCT 测试集中的 (a) 航空视角和 (b) 地面视角样本的不确定性分布。

样本可信度。两个数据集中的所有测试样本（AiRound 的 2,347 对和 CV-BrCT 的 4,830 对）均由青岛市城乡建设研究院的 9 位城市规划专家根据图像内容是否与标签相符进行评估。9 位专家的投票决定每个测试样本的最终结论（可信或不确定）。对于 TEB-SUQ， u_{th} 设定适当的阈值，以区分可信样本和不确定样本。将使用 $u \leq 0.4$ (n_k 计算得到的 k类样本数与专家投票生成的样本数（m_k 对于 k类）进行比较，结果如两个数据集各视图中每个类别的图 13 所示。13。表I 展示了 TEB-SUQ 相对于各视图主观评估的平均误差 $\bar{\sigma}$，使用

$$\bar{\sigma} = \frac{1}{K}\sum_{k=1}^{K}\frac{|n_k - m_k|}{m_k}, \tag{22}$$

其中 K 为类别总数。

表 I TEB-SUQ 相对于主观评估的平均误差。

视图*	A-*AiRound*	G-*AiRound*	A-*CV-BrCT*	G-*CV-BrCT*
$\bar{\sigma}$	0.065	0.078	0.183	0.090

***A-** 表示航拍视图，**G-**表示地面视图。

从表I可以看出，除 CV-BrCT 的空中视角外，其他视图中预测的可信样本数量与主观评估结果之间的相对误差均

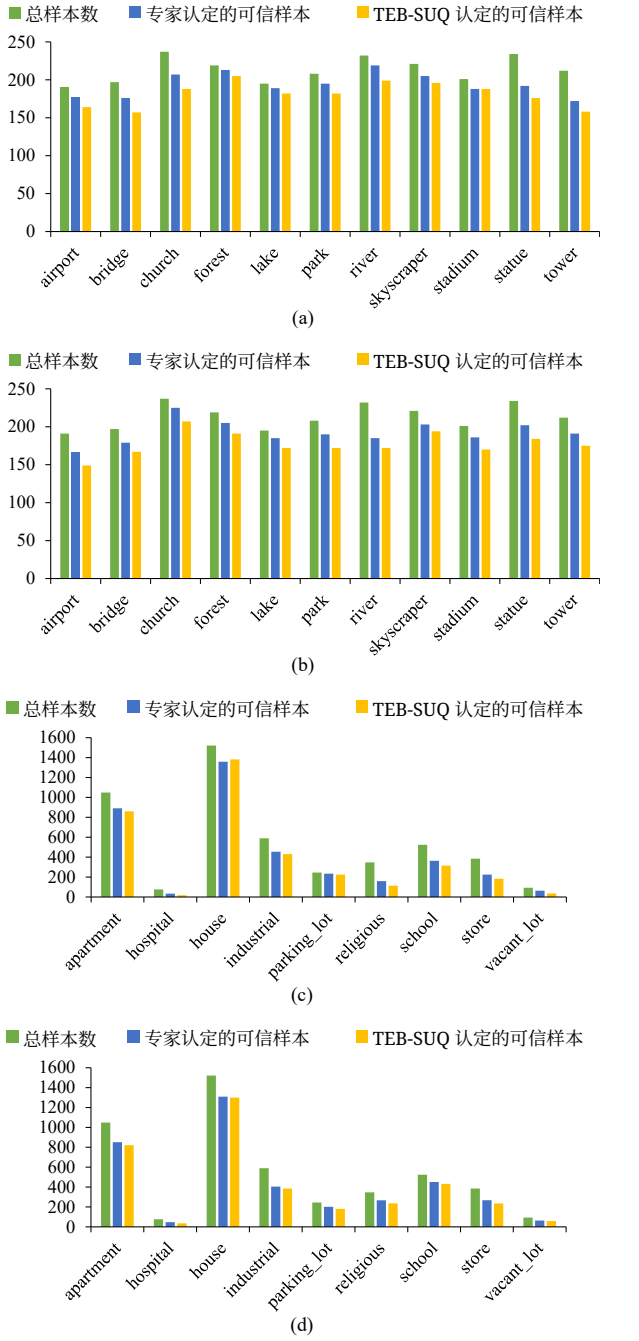


图 13. 使用 $u \leq 0.4$，通过专家投票和 TEB-SUQ 生成的 AiRound 各类别可信样本数量：(a) 航拍视图和 (b) 地面视图；CV-BrCT 的 (c) 航拍视图和 (d) 地面视图。

小于 0.1，证明了所提出的样本不确定性估计方法有效。图14展示了 CV-BrCT 数据集中更多样本不确定性的案例，这些样本的类别是随机选取的。

2）倒数损失有效性的验证：广泛用于证据深度学习的损失函数（简称 EDL 损失）由一项类交叉熵损失和一项两个狄利克雷分布之间的 Kullback-Leibler（KL）散度组成，并带有一个额外的退火系数 [33]。相比之下，它在形式上过于复杂，难以训练

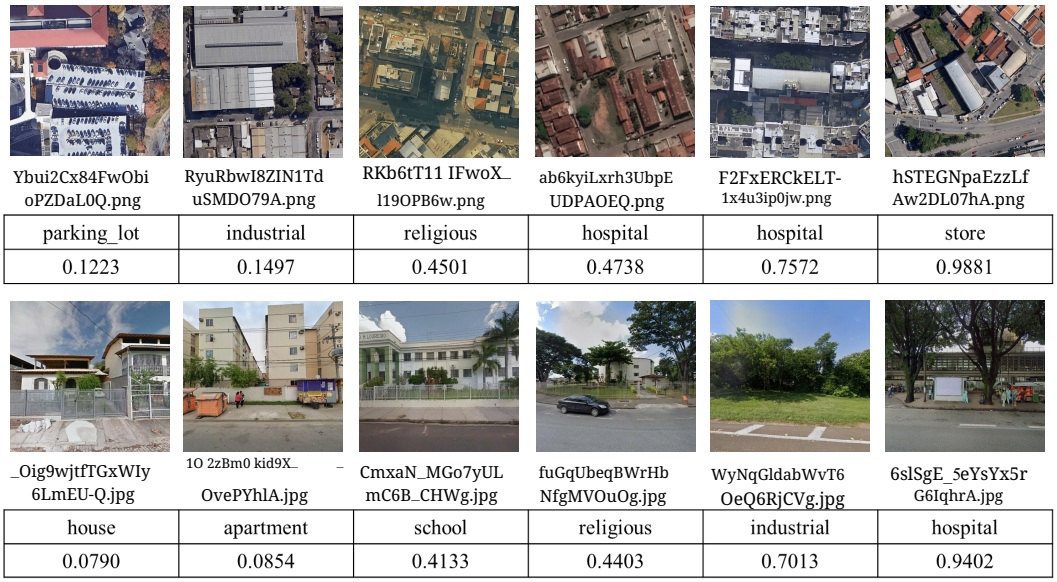


图 14. TEB-SUQ 计算的 CV-BrCT 中样本的不确定性。

，而所提出的倒数损失（式14 至式16）。图15展示了在相同设置下使用 EDL 损失和倒数损失（式21）时的训练损失和验证损失。使用 EDL 损失训练时出现了明显的过拟合。改用倒数损失后，过拟合现象显著改善。EDL 损失和所提出的倒数损失的场景分类性能见表II。所有结果均使用 VGG-11 [67] 作为骨干网络，并采用相同的训练设置和融合策略。EDL 中的退火系数设为 50。显然，基于倒数损失优化训练过程后，测试集上的性能有所提升。

表 II 不同损失函数的性能（%） ±标准差 。

Loss	AiRound		CV-BrCT	
函数	Acc	F1	Acc	F1
EDL损失	90.23±0.32	90.64±0.29	86.98±0.17	81.56±0.34
倒数损失	**92.16±0.31**	**92.49±0.25**	**88.21±0.26**	**83.57±0.29**

3）证据融合策略有效性的验证：如第III-A节所述，所提出的方法与骨干网络无关，因此可以灵活地与不同的深度特征提取网络搭配使用。在本实验中，采用三种最常见的骨干网络，对比使用不同融合策略前后在场景分类任务上的性能。在表III的单视图列中，“-s”表示传统深度学习（即 Softmax 深度学习）方法，该方法在训练阶段使用 Softmax 层和交叉熵损失，在测试阶段则将 Softmax 算子计算出的最大概率所对应的类别作为预测结果。“-e”表示第III-B 节和第III-D 节提出的证据深度学习方法，其中 Softmax 层被 Softplus 取代层，并使用所提出的倒数损失进行训练。测试时，将最大证据值所对应的类别作为预测结果。在融合策略列中，采用两种常见的决策层融合策略（求和和乘积） [30]作为基线方法，与所提出的证据融合方法进行比较。最佳结果以粗体突出显示。

从表III中可以得出以下观察结果。首先，所有融合结果均优于任何单视图结果。这表明，来自多个视角的信息能够显著提高分类准确率。其次，通过证据深度学习获得的两个单视图的性能略逊于通过Softmax深度学习获得的相应单视图。这是因为，当某些样本的不确定性较高时，证据深度学习会更加注重尽可能准确地估计不确定性值，这可能导致这些样本的分类准确率有所下降。然而，从样本质量来看，这些高不确定性样本的类别标签缺乏实际语义。它们的预测结果正确与否也无关紧要。量化它们的不确定性更有意义。最后但同样重要的是，本文提出的证据融合策略优于另外两种基线融合方法，这证明了证据融合在多视图遥感场景分类任务中的有效性。

C. 数据级、特征级和决策级不同融合方法的对比实验

在第 IV-B节中，分别验证了本文三项创新贡献（即 TEB-SUQ、证据融合和倒数损失）的有效性。本节将更多多视图融合方法与所提出的方法进行比较，以评估其在

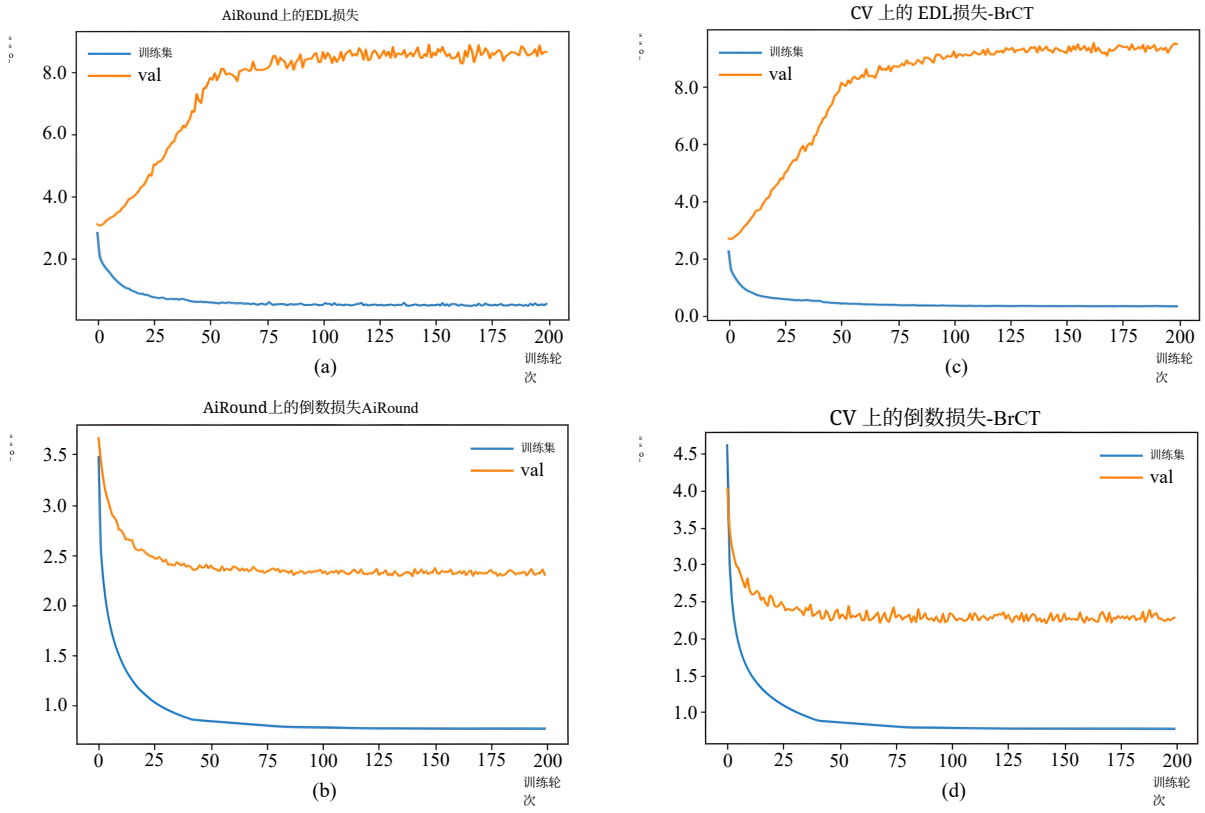


图 15. AiRound 上使用 (a) EDL损失和 (b) 倒数损失的训练集与验证集损失；CV-BrCT 上使用 (c) EDL损失和 (d) 倒数损失的训练集与验证集损失。

表 III 分类准确率 (%) ±标准差 决策级融合前后。

数据集	骨干网络	单视图（softmax）		单视图（证据）		决策级融合策略		
		空中-s	地面-s	空中-e	地面-e	求和 [30]	乘积 [30]	证据（所提出的方法）
Airound	AlexNet [68]	76.96±0.52	71.35±0.24	76.04±0.46	70.96±0.21	84.02±0.47	86.74±0.25	**88.12±0.23**
	VGG-11 [67]	82.75±0.61	77.10±0.28	82.64±0.49	76.99±0.17	87.75±0.38	90.41±0.27	**92.16±0.31**
	ResNet-18 [69]	80.93±0.49	76.68±0.19	80.83±0.52	76.36±0.22	88.02±0.25	89.56±0.24	**91.02±0.35**
CV-BrCT	AlexNet [68]	84.63±0.24	68.01±0.12	84.37±0.10	66.36±0.25	85.26±0.45	86.52±0.25	**88.02±0.28**
	VGG-11 [67]	87.11±0.42	71.43±0.22	87.06±0.38	70.15±0.29	86.70±0.58	87.21±0.22	**88.21±0.26**
	ResNet-18 [69]	86.74±0.38	70.96±0.25	86.18±0.29	70.86±0.24	85.59±0.62	86.83±0.18	**87.95±0.19**

航拍—地面双视图遥感场景分类任务上的整体性能。如第 II-A 节所述，现有的多视图融合方法大致可分为数据级、特征级和决策级。本实验选取一种数据级融合方法（六通道 [70]）、两种特征级融合方法（特征拼接 [30] 和 CILM [71]），以及四种决策级融合方法 [30]（最大值、最小值、求和和乘积），与所提出的证据融合方法进行比较。下面对这些方法作简要介绍。

- 六通道 [70]：该方法将配对的空中视角和地面视图图像的RGB通道拼接成一个六通道图像，作为卷积神经网络的输入。
- 特征拼接 [30]：该方法使用孪生式卷积神经网络，在第一个卷积层（其卷积核数量加倍）之前拼接中间特征向量。
- *CILM* [71]：该方法将对比学习的损失函数与交叉熵损失结合，使两个子网络提取的特征能够在不共享任何权重的情况下进行融合。
- 最大值 [30]：每个视角分别采用独立的深度神经网络获得其预测结果，该结果由一个类别标签及其概率组成。最终预测为各视角预测的类别概率中最大值所对应的类别标签。
- 最小值 [30]：每个视角分别采用独立的深度神经网络获得其预测结果，该结果由一个类别标签及其概率组成。最终预测为各视角预测的类别概率中最小值所对应的类别标签。
- 求和 [30]：每个视图都使用一个独立的深度神经网络，生成一个包含各类别概率的向量。

表IV 不同融合策略下的分类准确率（%） ±标准差

数据集	骨干网络	数据级	特征级		决策级				
		六通道 [70]	拼接 [30]	CILM [71]	最大值 [30]	最小值 [30]	求和 [30]	乘积 [30]	证据（本文提出）
Airound	AlexNet [68]	70.19±0.23	82.52±0.32	83.49±0.17	84.86±0.36	85.52±0.23	84.02±0.47	86.74±0.25	**88.12±0.23**
	VGG-11 [67]	72.34±0.21	84.69±0.41	85.72±0.19	88.17±0.34	89.56±0.25	87.75±0.38	90.41±0.27	**92.16±0.31**
	Inception [72]	71.76±0.24	83.91±0.45	85.05±0.15	88.39±0.33	89.12±0.27	88.05±0.30	90.02±0.14	**91.41±0.18**
	ResNet-18 [69]	71.29±0.26	83.56±0.39	84.72±0.21	88.21±0.38	88.42±0.22	88.02±0.25	89.56±0.24	**91.02±0.35**
	DenseNet [73]	71.57±0.25	83.72±0.42	84.91±0.20	89.96±0.35	90.41±0.24	89.88±0.31	91.16±0.17	**92.16±0.19**
CV-BrCT	AlexNet [68]	71.92±0.26	81.86±0.42	83.10±0.20	85.52±0.24	86.02±0.21	85.26±0.45	86.52±0.25	**88.02±0.28**
	VGG-11 [67]	73.46±0.24	83.25±0.39	84.32±0.19	86.74±0.39	86.95±0.27	86.70±0.58	87.21±0.22	**88.21±0.26**
	Inception [72]	75.26±0.25	84.65±0.38	85.22±0.16	86.70±0.41	86.95±0.28	86.24±0.35	87.02±0.21	**88.21±0.23**
	ResNet-18 [69]	73.25±0.28	83.28±0.41	84.31±0.17	85.84±0.43	86.24±0.25	85.59±0.62	86.83±0.18	**87.95±0.19**
	DenseNet [73]	74.19±0.27	84.24±0.40	85.19±0.22	86.95±0.38	87.02±0.25	86.85±0.39	87.54±0.17	**88.34±0.20**



图16. 单视图、乘积融合和所提出的证据融合方法的预测示例。

融合向量是单视图向量的和。最终预测结果是融合向量中最大元素对应的类别标签。

- 乘积 [30]：每个视图均采用独立的深度神经网络，为每个类别生成一个包含其概率的向量。对单视图向量进行逐元素乘法，得到融合向量。最终预测结果为融合向量中最大元素对应的类别标签。

表IV 展示了采用不同骨干网络时上述融合方法的性能。可以得出以下观察结果。首先，数据级融合类别的方法表现最差，而决策级融合类别的方法则全面胜出。这可能是因为空中视角的视觉特征与地面视角的视觉特征差异极大。该方法的性能

随着参与融合的特征变得更加抽象，融合效果也会提升。这也解释了为何在两种特征级融合方法中，CILM 的表现优于特征拼接：CILM 缺少共享权重结构，因此在形式上更接近决策级融合。其次，在决策级融合方法中，求和与最大值的表现几乎相同，且二者的表现都略逊于最小值。乍看之下，这一结果似乎有悖直觉。事实上，这证实了第 III-B 节（见图3 和图4）中提到的 softmax 所导致的过度自信问题：被高估的预测更有可能是错误的。最后但同样重要的是，乘积方法和所提出的证据融合在所有融合方法中脱颖而出，且后者优于前者。事实上，式10 可视为对乘积方法的改进。引入样本不确定性打破了融合中各视图的平等性：具有较低 u 值的视图会

获得更高的自适应权重。

最后，在两个数据集的所有骨干网络上，所提出的证据融合方法均优于最佳决策级融合方法1。26% ±0.27，优于最佳特征级融合方法4。96% ±1.46，并优于数据级融合方法17。04%±2.68。预测结果示例见图16。

V. 结论

多视图数据（例如空地双视图图像）日益广泛地应用于场景分类等各种遥感任务。然而，随着视图数量的增加，原始单视图中的数据质量问题变得愈发突出，使融合效果远不及预期。深度学习模型容易受到数据质量的影响，尤其是在处理海量数据时。李飞飞团队提出了“可信人工智能”这一概念 [74]，，并着重关注数据质量。这个问题在空地多视图图像数据中也十分明显。空中视角图像的质量相对较高。然而，由于拍摄尺度的视野范围较大，类别标签可能难以准确描述图像中包含的所有场景类别，导致大量样本类别内差异较大、类别间相似度较高。另一方面，地面视图中存在许多低质量样本，例如目标过大、严重遮挡以及室内拍摄。面对这些挑战，现有融合方法往往难以发挥作用。

本文引入不确定性理论，尝试量化样本的可信度。具体而言，通过设计的狄利克雷分布，将深度神经网络（DNN）输出的类别得分映射到“意见空间”（见图5），从而获得输入样本的总体不确定性值。在此基础上，提出一种决策级多视图融合策略，为决策风险较低的视图赋予较高权重。此外，本文还提出了一种新颖的损失函数——倒数损失。与证据深度学习中常用的损失函数相比，倒数损失更加简洁、更易训练，且测试性能更优。基于上述三项创新，本文提出的证据融合方法在空地双视图图像遥感场景分类任务的两个经典数据集上取得了最佳性能。

在多视图任务中关注数据质量是一个新的研究方向，仍有许多工作有待开展。首先，可供遥感任务使用的公共多视图数据集很少。亟需公开大规模、实例级对齐的遥感多视图数据集，以支持相关研究。此外，目前缺乏对样本不确定性估计进行有效客观评估的方法。遥感领域尚未出现带有样本质量标注的数据集。最后，还需要进一步探索更明确的样本不确定性表征方法。

致谢

作者谨向来自青岛城市与乡村建设研究院 BIM研究中心、城市规划领域的专家 Ruxuan Bi、Zhiwei He 和 Mengshuo Fan 致谢，感谢他们为样本可信度的主观评估提供专业指导。感谢参与主观评估的 Chunting Zhao、Xiayue Wang、Qi Liu、Mingdong Ding、Jinhong Guo 和 Zhengnan Fang。本研究得到国家自然科学基金委员会（项目编号：62171247）资助。

参考文献

[1] S. Roy、E. Sangineto、B. Demir 和 N. Sebe，“基于度量学习的深度哈希网络：用于遥感图像的基于内容检索”，《*IEEE*地球科学与遥感快报》，卷 18，第 2 期，第 226–230 页，2020年。[2] Y. Wang、L. Zhang、X. Tong、L. Zhang、Z. Zhang、H. Liu、X. Xing 和 P. T. Mathiopoulos，“用于遥感图像检索的三层图学习方法”，《*IEEE*地球科学与遥感汇刊》，卷 54，第 10 期，第 6020–6034 页，2016年。[3] X.-Y. Tong、G.-S. Xia、F. Hu、Y. Zhong、M. Datcu 和 L. Zhang，“利用深度特征进行遥感图像检索：系统性研究”，《*IEEE*大数据汇刊》，卷 6，第 3 期，第 507–521 页，2019 年。[4] T. R. Martha 、 N. Kerle 、C. J. Van Westen、V. Jetten 和 K. V. Kumar，“面向对象图像分析中用于知识驱动滑坡检测的分割优化与数据驱动阈值化”，《*IEEE*地球科学与遥感汇刊》，卷 49，第 12 期，第 4928–4943 页，2011年。[5] G. Cheng、L. Guo、T. Zhao、J. Han、H. Li 和 J. Fang，“基于视觉词袋和概率潜在语义分析的场景分类方法自动检测遥感影像中的滑坡”，《国际遥感杂志》，卷 34，第 1 期，第 45–59 页，2013年。[6] A. Stumpf 和 N. Kerle，“利用随机森林进行滑坡面向对象制图”，《遥感环境》，卷 115，第 10 期，第 2564–2577 页，2011年。[7] M. Kim、M. Madden 和 T. A. Warner，“利用多光谱 IKONOS 影像的对象特定纹理测度绘制森林类型图”，《摄影测量工程与遥感》，卷 75，第 7 期，第 819–829 页，2009年。[8] A. Capolupo、L. Kooistra 和 L. Boccia，“一种利用历史及当代摄影测量航空照片检测农业梯田景观的新方法”，《国际应用地球观测与地理信息杂志》，卷 73，第 800–810 页，2018。[9] N. B. Mishra 和 K. A. Crews，“干旱稀树草原生态系统中的植被形态类型制图：结合分层面向对象图像分析与随机森林”，《国际遥感杂志》，卷 35，第 3 期，第 1175–1198 页，2014年。[10] B. Zhao、Y. Zhong、G.-S. Xia 和 L. Zhang，“用于高空间分辨率遥感影像的狄利克雷衍生多主题场景分类模型”，《*IEEE*地球科学与遥感汇刊》，卷 54，第 4 期，第 2108–2123 页，2015年。[11] X. Yao、J. Han、G. Cheng、X. Qian 和 L. Guo，“通过弱监督学习对高分辨率卫星图像进行语义标注”，《*IEEE*地球科学与遥感汇刊》，卷 54，第 6 期，第 3660–3671 页，2016年。[12] H. Wu、B. Liu、W. Su、W. Zhang 和 J. Sun，“用于场景级土地利用分类的层次编码向量”，《遥感》，卷 8，第 5 期，第 436 页，2016年。[13] R. Liu、X. Ning、W. Cai 和 G. Li，“结合协方差池化的多尺度密集交叉注意力机制，用于高光谱图像场景分类”，《移动信息系统》，卷 2021，2021年。[14] J. Jiang、F. Liu、Y. Xu、H. Huang、等，“使用双通道卷积神经网络进行多光谱 RGB-NIR 图像分类”，《*IEEE Access*》，卷 7，第 20607–20613 页，2019年。[15] L. Chen、X. Cui、Z. Li、Z. Yuan、J. Xing、X. Xing 和 Z. Jia，“一种基于空间统计建模和特征再校准的 SAR 场景分类新深度学习算法”，《传感器》，卷 19，第 11 期，第 2479 页，2019年。[16] Q. Wang、S. Liu、J. Chanussot 和 X. Li，“利用循环注意力进行超高分辨率遥感图像场景分类”，《*IEEE*地球科学与遥感汇刊》，卷 57，第 2 期，第 1155–1167 页，2018。[17] Y. Zhu 和 S. Newsam，“将卷积神经网络应用于地面图像的土地利用分类”，载于第*23*届*SIGSPATIAL*地理信息系统进展国际会议论文集，第 1–4 页，2015年。

[18] V. Antoniou、C. C. Fonte、L. See、J. Estima、J. J. Arsanjani、F. Lupia、M. Minghini、G. Foody 和 S. Fritz，“探讨带地理标签的照片作为土地覆盖输入数据来源的可行性”，ISPRS《国际地理信息杂志》，卷 5，第 5 期，第 64 页，2016年。[19] T. Pei、S. Sobolevsky、C. Ratti、S.-L. Shaw、T. Li 和 C. Zhou，“基于聚合手机数据的土地利用分类新视角”，《国际地理信息科学杂志》，卷 28，第 9 期，第 1988–2007 页，2014年。[20] W. Zhang、W. Li、C. Zhang、D. M. Hanink、X. Li 和 W. Wang，“利用机载激光雷达、高分辨率正射影像和 Google 街景对特大城市进行基于地块的城市土地利用分类”，《计算机、环境与城市系统》，卷 64，第 215–228 页，2017年。[21] J. Kang、M. K¨orner、Y. Wang、H. Taubenb¨ock 和 X. X. Zhu，“利用街景图像进行建筑实例分类”，ISPRS《摄影测量与遥感杂志》，卷 145，第 44–59 页，2018。[22] T. Hu、J. Yang、X. Li 和 P. Gong，“利用 Landsat图像和开放社交数据绘制城市土地利用图”，《遥感》，卷 8，第 2 期，第 151 页，2016年。[23] X. Liu、J. He、Y. Yao、J. Zhang、H. Liang、H. Wang 和 Y. Hong，“整合遥感与社交媒体数据进行城市土地利用分类”，《国际地理信息科学杂志》，卷 31，第 8 期，第 1675–1696 页，2017年。[24] Y. Jia、Y. Ge、F. Ling、X. Guo、J. Wang、L. Wang、Y. Chen 和 X. Li，“结合遥感图像与手机定位数据绘制城市土地利用图”，《遥感》，卷 10，446，2018 3，第 p. .。[25] W. Tu、Z. Hu、L. Li、J. Cao、J. Jiang、Q. Li 和 Q. Li，“结合遥感图像与人类感知数据描绘城市功能区”，《遥感》，卷 10，第 1 期，第 141 页，2018年。[26] D. Hong、N. Yokoya、J. Chanussot 和 X. X. Zhu，“Cospace：从高光谱-多光谱对应关系中学习公共子空间”，《IEEE地球科学与遥感汇刊》，卷 57，第 7 期，第 4349–4359 页，2019 年。[27] X. Li 、 C. Zhang 和 W. Li，“基于 Google街景图像的建筑街区级城市土地利用信息检索”，《GIScience与遥感》，卷 54，第 6 期，第 819–835 页，2017年。[28] R. Cao、J. Zhu、W. Tu、Q. Li、J. Cao、B. Liu、Q. Zhang 和 G. Qiu，“融合航空与街景图像进行城市土地利用分类”，《遥感》，卷 10，第 10 期，第 1553 页，2018年。[29] S. Srivastava、J. E. Vargas-Munoz 和 D. Tuia，“从空中与地面视角理解城市土地利用：一种深度学习多模态解决方案”，《遥感环境》，卷 228，第 129–143 页，2019年。[30] G. Machado、E. Ferreira、K. Nogueira、H. Oliveira、M. Brito、P. H. T. Gama 和 J. A. dos Santos，“AiRound 和 CV-BrCT：用于场景分类的新型多视图数据集”，《IEEE应用地球观测与遥感精选主题杂志》，卷 14，503，2020 488pp. – 。[31] Z. Deng、H. Sun 和 S. Zhou，“利用异构特征学习进行场景分类的半监督地面到航空适应”，《ISPRS国际地理信息杂志》，卷 7，182，2018 5，第 p. .。[32] Z. Han、C. Zhang、H. Fu 和 J. T. Zhou，“可信多视图分类”，arXiv预印本 arXiv:2102.02051，2021年。[33] 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》，卷 31，2018。[34] T. Baltruˇsaitis、C. Ahuja 和 L.-P. Morency，“多模态机器学习：综述与分类法”，《IEEE模式分析与机器智能汇刊》，卷 41，第 2 期，第 423–443 页，2018年。[35] D. Ramachandram 和 G. W. Taylor，“深度多模态学习：近期进展与趋势综述”，《IEEE信号处理杂志》，卷 34，第 6 期，第 96–108 页，2017年。[36] T.-M. Tu、P. S. Huang、C.-L. Hung 和 C.-P. Chang，“一种用于 IKONOS影像的快速强度-色调-饱和度融合技术及光谱调整”，《IEEE地球科学与遥感快报》，卷 1，第 4 期，第 309–312 页，2004年。[37] K. Amolins、Y. Zhang 和 P. Dare，“基于小波的图像融合技术——介绍、综述与比较”，《ISPRS摄影测量与遥感杂志》，卷 62，263，2007 4，249 第 pp. – 。[38] J. Nunez、X. Otazu、O. Fors、A. Prades、V. Pala 和 R. Arbiol，“采用加性小波分解的基于多分辨率的图像融合”，《IEEE地球科学与遥感汇刊》，卷 37，第 3 期，第 1204–1211 页，1999年。[39] Z. Liu、P. Song、J. Zhang 和 J. Wang，“用于融合多光谱与全色图像的二维经验模态分解”，《国际地理信息杂志》，卷 28，第 4 期，第 1067–1080 页，2007年。[40] B. Aiazzi、L. Alparone、S. Baronti 和 A. Garzelli，“基于过采样多分辨率分析的高空间和光谱分辨率图像上下文驱动融合”，《IEEE地球科学与遥感汇刊》，卷 40，第 10 期，第 2300–2312 页，2002年。[41] R. Zhang、X. Tang、S. You、K. Duan、H. Xiang 和 H. Luo，“面向多云山区土地利用/土地覆盖（lulc）分类的光学与 SAR 遥感图像新型特征级融合框架”，《应用科学》，卷 10，第 8 期，页码 2928，2020年。[42] X. Wang、Y. Feng、R. Song、Z. Mu 和 C. Song，“用于高光谱和激光雷达数据融合分类的多注意力层级密集融合网络”，《信息融合》，卷 82，第 1–18 页，2022年。[43] G. Ye、D. Liu、I.-H. Jhuo 和 S.-F. Chang，“基于秩最小化的稳健后期融合”，载于《2012年IEEE计算机视觉与模式识别会议》，第 3021–3028 页，电气与电子工程师学会，2012年。[44] H. Gunes 和 M. Piccardi，“基于面部和身体的情感识别：早期融合与后期融合”，载于《2005年IEEE系统、人类与控制论国际会议》，卷 4，第 3437–3443 页，电气与电子工程师学会，2005年。[45] Y. Yu 和 F. Liu，“基于深度卷积神经网络多层级融合的航空场景分类”，《IEEE地球科学与遥感快报》，卷 15，第 2 期，第 287–291 页，2018。[46] N. Yang、H. Tang、H. Sun 和 X. Yang，“Dropband：一种简单有效的方法，可提高甚高分辨率遥感图像卷积神经网络的场景分类准确率”，《IEEE地球科学与遥感快报》，卷 15，2018 年第 2 期，第 261、257 页。[47] F. Zhang、B. Du 和 L. Zhang，“基于梯度提升随机卷积网络框架的场景分类”，《IEEE地球科学与遥感汇刊》，卷 54，第 3 期，第 1793–1802 页，2015年。[48] T.-Y. Lin、S. Belongie 和 J. Hays，“跨视角图像地理定位”，载于《IEEE计算机视觉与模式识别会议论文集》，第 891–898 页，2013年。[49] T.-Y. Lin、Y. Cui、S. Belongie 和 J. Hays，“学习用于地面到航空视角地理定位的深度表征”，载于《IEEE计算机视觉与模式识别会议论文集》，第 5007–5015 页，2015年。[50] S. Workman、M. Zhai、D. J. Crandall 和 N. Jacobs，“近距离与遥感的统一模型”，载于《IEEE国际计算机视觉会议论文集》，第 2688–2697 页，2017年。[51] V. Badrinarayanan、A. Kendall 和 R. Cipolla，“Segnet：用于图像分割的深度卷积编码器-解码器架构”，《IEEE模式分析与机器智能汇刊》，卷 39，第 12 期，第 2481–2495 页，2017年。[52] E. J. Hoffmann、Y. Wang、M. Werner、J. Kang 和 X. X. Zhu，“基于航空与街景图像的建筑类型分类模型融合”，《遥感》，卷 11，第 11 期，页码 1259，2019年。[53] S. Srivastava、J. E. Vargas-Mu˜noz、D. Swinkels 和 D. Tuia，“利用卷积神经网络从地面照片进行多标签建筑功能分类”，载于《第 2 届 ACM SIGSPATIAL地理知识发现人工智能国际研讨会论文集》，第 43–46 页，2018。[54] H. Jiang、B. Kim、M. Guan 和 M. Gupta，“是否信任分类器”，《神经信息处理系统进展》，卷 31，2018。[55] K. Beyer、J. Goldstein、R. Ramakrishnan 和 U. Shaft，“何时‘最近邻’才有意义？”，载于《国际数据库理论会议》，第 217–235 页，Springer，1999年。[56] C. Blundell、J. Cornebise、K. Kavukcuoglu 和 D. Wierstra，“神经网络中的权重不确定性”，载于《国际机器学习会议》，第 1613–1622 页，PMLR，2015年。[57] C. Corbi`ere、N. Thome、A. Bar-Hen、M. Cord 和 P. P´erez，“通过学习模型置信度应对失效预测”，《神经信息处理系统进展》，卷 32，2019年。[58] B. Lakshminarayanan、A. Pritzel 和 C. Blundell，“使用深度集成进行简单且可扩展的预测不确定性估计”，《神经信息处理系统进展》，卷 30，2017年。[59] J. Van Amersfoort、L. Smith、Y. W. Teh 和 Y. Gal，“使用单个深度确定性神经网络进行不确定性估计”，载于《国际机器学习会议》，第 9690–9700 页，PMLR，2020年。[60] A. Damianou 和 N. D. Lawrence，“深度高斯过程”，载于《人工智能与统计》，第 207–215 页，PMLR，2013年。[61] J. Heo、H. B. Lee、S. Kim、J. Lee、K. J. Kim、E. Yang 和 S. J. Hwang，“用于可靠解释与预测的不确定性感知注意力”，《神经信息处理系统进展》

[62] J. Moon、J. Kim、Y. Shin 和 S. Hwang，“深度神经网络的置信度感知学习”，载于国际机器学习会议，第 7034–7044 页，PMLR，2020 年。[63] L. Neumann、A. Zisserman 和 A. Vedaldi，“松弛 soft max：用于安全行人检测的高效置信度自动校准”，2018。[64] R. R. Yager 和 L. Liu，邓普斯特-谢弗信任函数理论经典著作，第 219 卷。Springer，2008年。[65] A. Jøsang，主观逻辑，第 3 卷。Springer，2016年。[66] J. Lin，“关于狄利克雷分布”，女王大学数学与统计系，2016年。[67] K. Simonyan 和 A. Zisserman，“用于大规模图像识别的极深卷积网络”，*arXiv* 预印本 *arXiv:1409.1556*，2014。[68] A. Krizhevsky、I. Sutskever 和 G. E. Hinton，“使用深度卷积神经网络进行 ImageNet 分类”，*ACM* 通讯，第 60 卷，第 6 期，第 84–90 页，2017年。[69] K. He、X. Zhang、S. Ren 和 J. Sun，“用于图像识别的深度残差学习”，载于*IEEE*计算机视觉与模式识别会议论文集，第 770–778 页，2016年。[70] N. N. Vo 和 J. Hays，“利用俯视影像定位街景并确定其朝向”，载于欧洲计算机视觉会议，第 494–509 页，Springer，2016年。[71] W. Geng、W. Zhou 和 S. Jin，“基于互补信息学习模型的多视图城市场景分类”，《摄影测量工程与遥感》，第 88 卷，第 1 期，第 65–72 页，2022年。[72] C. Szegedy、V. Vanhoucke、S. Ioffe、J. Shlens 和 Z. Wojna，“重新思考计算机视觉的 Inception 架构”，载于*IEEE*计算机视觉与模式识别会议论文集，第 2818–2826 页，2016年。[73] G. Huang、Z. Liu、L. Van Der Maaten 和 K. Q. Weinberger，“密集连接卷积网络”，载于*IEEE*计算机视觉与模式识别会议论文集，第 4700–4708 页，2017年。[74] W. Liang、G. A. Tadesse、D. Ho、L. Fei-Fei、M. Zaharia、C. Zhang 和 J. Zou，“为可信人工智能创建数据的进展、挑战与机遇”，《自然·机器智能》，第 4 卷，677，2022 8，669no. pp. –。