

带噪声标签细粒度图像分类方法研究综述

王 浩, 姚俊萍*, 李晓军, 阙小翔

(火箭军工程大学, 陕西 西安 710025)

摘 要: 开展带噪声标签细粒度图像分类方法综述, 在细粒度图像分类问题定义及方法类型梳理的基础上, 对噪声标签问题进行描述, 总结了标签噪声的来源, 从影响因素依赖关系和实例依赖关系两个角度对标签噪声进行分类, 分析了标签噪声对细粒度图像分类任务的影响, 从基于显式噪声模型方法、基于隐式噪声模型方法两个维度梳理了细粒度图像分类的研究进展, 提出了未来值得关注的研究主题。

关键词: 细粒度图像分类; 噪声标签; 深度学习; 噪声鲁棒

中图分类号: TP391

文献标志码: A

文章编号: 2097-2741(2024)05-096-13

Review on Fine-Grained Image Classification with Noise Labels

WANG Hao, YAO Junping, LI Xiaojun, QUE Xiaoxiang

(Rocket Force University of Engineering, Xi'an 710025, Shaanxi)

ABSTRACT: A review of classification methods of fine-grained images with noise labels was conducted. Firstly, based on the definition of fine-grained image classification problems and the classification of method types, the problems of noise labels were described, as well as the sources of label noise were summarized. Then, label noise was classified from the perspectives of influence factor dependency and instance dependency, and the impact of label noise on fine-grained image classification was analyzed. Furthermore, the progress of fine-grained image classification was discussed from two dimensions: explicit noise model based methods and implicit noise model based methods. Thereby, future noteworthy research topics were proposed.

KEYWORDS: fine-grained image classification; noise label; deep learning; noise robust

细粒度图像分类 (Fine-Grained Image Classification, FGIC) 是区分在视觉上相似且属于同一基本类别的从属类别, 如舰艇种类、飞机型号^[1]、鸟类品种等^[2]。FGIC 具有类内方差大、类间差异小、类别数量大、含噪声标签可能性高等特点, 在生物物种识别^[3]、疾病诊断^[4]、智能交通^[5]、目标检测等领域发挥着重要作用。细粒度图像数据集需要经过大量的标注, 尤其需要领域专家精确

标记, 否则会产生标签噪声^[6]。为了解决这一问题, 采用深度学习技术作为直接从数据中学习特征表示的强大方法, 在 FGIC 中取得了显著突破^[7]。但基于深度学习的 FGIC 问题需要在大规模数据集上训练模型, 同时提升不断增大的数据集可用性。然而要收集并准确标记足够大的数据集是很困难的。常用的解决方案是使用非专家人工、众包形式或无监督的自动化标注系统来标记大规

收稿日期: 2024-04-03

基金项目: 国家社科基金 (2023-SKJJ-B-063)

第一作者: 王浩 (1996—), 男, 硕士研究生, 主要从事智能信息处理研究。E-mail: wanghao18811506308@163.com

***通信作者:** 姚俊萍 (1978—), 女, 教授, 主要从事信息系统与数据工程研究。E-mail: junpingy200225@163.com

引用格式: 王浩, 姚俊萍, 李晓军, 等. 带噪声标签细粒度图像分类方法研究综述[J]. 火箭军工程大学学报, 2024, 38 (5): 96-108.

模数据集。

然而使用这种方法收集到的数据集会存在很高的标签噪声, 因此适用性有限。标签噪声的可能来源主要分为4类: 1) 标注过程信息不足; 2) 数据辨识度低; 3) 人工标注主观性; 4) 数据编码或通信问题导致的标签噪声。一些研究已成功利用自动化挖掘细粒度图像数据库, 如医院图像存档系统、零售百货系统等来构建大型训练数据集。然而, 系统存储的历史数据可能不包括全部所需的标签或图像。此外, 此类数据集中的标签噪声预计要高于专家标记的数据集中的标签噪声。

区别于传统FGIC问题, 带噪声标签FGIC问题面临更复杂的数据集以及标签噪声对分类方法的负面影响。尽管基于深度学习的方法可以学习对象的整个结构、附加细节和局部特征, 以达到最先进的性能, 但研究表明, 深度学习模型可能在带噪声标签的情况下过度拟合甚至拟合完全随机的噪声^[8]。作为数据集采集和标注过程的自然结果, 噪声标签严重降低了深度神经网络的泛化性能。因此, 在带噪声标签深度学习 (Learning with Noisy Label, LNL) 的研究领域中, 如何提高FGIC的准确性成为该领域的重要研究方向之一。尤其在FGIC方法中处理噪声标签 (LNL-FGIC) 是非常重要的^[9], 对提高分类模型精度、增强分类模型鲁棒性, 降低FGIC标注成本等具有显著的现实意义。综上, 本文围绕带噪声标签FGIC方法开展综述。

1 带噪声标签细粒度图像分类问题及描述

现实数据集可能包含噪声, 其被描述为由非系统误差组成, 即使对专家来说, 收集正确注释的数据集也并不总是可行的。许多研究表明噪声产生过拟合现象, 诱导分类器性能产生不利影响。因此, 有必要明确噪声标签的定义、来源和分类问题, 并研究消除噪声或减少其后果的技术。

1.1 标签噪声概念及分类

数据噪声通常分为两类: 特征噪声和标签噪声^[6]。特征噪声影响特征的观测值, 例如通过在测量过程中对某些特征添加高斯噪声。标签噪声^[10]是指用于训练的目标标签与相应实例本身的真实标签之间的偏差, 例如设定标签时错误设置标签所属类别。两种噪声都会影响系统性能, 但标签噪声比特征噪声危害更大^[11]。例如当有多个

特征时, 标签对于每个数据都是唯一的, 并且每个特征的重要性各不相同。

1.1.1 标签噪声概念

图像分类任务涉及从训练数据建模来预测新样本的类别, 每个训练样本都与一个观察到的标签相关联, 该标签通常对应于样本的真实类别, 但这可能会受到噪声过程的影响而产生标签噪声。经典监督学习^[12]由一个输入数据集 $S = \{(x_1, y_1), \dots, (x_N, y_N)\} \in (X, Y)^N$ 根据未知分布 D , 在 (X, Y) 上绘制而成的输入数据集, X 表示观察到的实例特征, Y 表示真实标签, N 表示标签数量。学习目标是在函数 $\mathcal{F}$ 中找到最佳的映射函数 $f: X \rightarrow Y$, 其中每个函数都用 θ 进行参数化为 f_θ 。

损失函数 l 记为 $l: R \times Y \rightarrow R^+$, R 代表经验风险。给定一个例子 $(x_i, y_i) \in (X, Y)$, $l(f_\theta(x_i), y_i)$ 评估分类器预测的好坏。然后, 对于任意分类器 f , 经验风险 $R_{l,D}(f_\theta)$ 的定义为

$$R_{l,D}(f_\theta) = E_D[l(f_\theta(x_i), y_i)] \quad (1)$$

式中: E_D 表示分布 D 上的期望。

由于对分布有完备的知识一般是不可行的, 因此, 用 $\hat{R}_{l,D}(f_\theta)$ 作为近似使用经验风险。

$$\hat{R}_{l,D}(f_\theta) = \frac{1}{N} \sum_{i=1}^N l(f_\theta(x_i), y_i) \quad (2)$$

学习分类器的各种方法可以被看作是最大限度地降低受网络参数影响的经验风险, θ^* 为使 $\hat{R}_{l,D}(f_\theta)$ 达到最小值时的参数 θ 的取值, 有

$$\theta^* = \arg \min_{\theta} \hat{R}_{l,D}(f_\theta) \quad (3)$$

当标签噪声存在时, 数据集变成 $S_n = \{(x_1, \tilde{y}_1), \dots, (x_N, \tilde{y}_N)\} \in (X, Y)^N$, 根据噪声分布 D_n 从 (X, Y) 中抽取。然后, 风险最小化结果为

$$\theta_n^* = \arg \min_{\theta} \hat{R}_{l,D_n}(f_\theta) \quad (4)$$

因此, 通过最小化经验风险得到的参数与期望的最优分类器参数不同。

经典的监督学习是在给定分布 D 的情况下寻找最优的估计器参数 θ^* 。然而, 在有噪声的标签设置中, 任务仍然是在分布 D_n 上寻找 θ^* 。因此, 在标签噪声存在的情况下, 经典的风险最小化是不充分的, 因为它会导致 θ_n^* 。许多文献提出了经典风险最小化方法的变体并对它们进行评估。在这里, 由于噪声标签计算的损失, 风险最小化过程不再具有噪声容忍性。深度学习可以很容易地记忆损坏的标签, 并相应地在看不见的数据上退

化它们的泛化^[13-14]。因此,减少噪声标签的不利影响对实现深度学习的抗噪声训练至关重要。

1.1.2 标签噪声分类

标签噪声可能受到以下 3 个因素影响:数据特征、数据的真实标签和标注者特征。根据这些因素的依赖关系,Frénay 等^[10]将标签噪声分为 3 个子类:完全随机噪声(Noisy Completely at Random, NCAR),随机噪声(Noisy at Random, NAR)和非随机噪声(Noisy not at Random, NN-AR)。如图 1 所示。

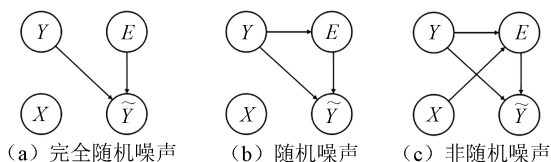


图 1 影响因素依赖关系的标签噪声分类

Fig. 1 Label noise classification based on the dependency of influence factors

图 1 中模型的依存关系复杂性从左到右依次增加,箭头表示统计相依性。其中: $\tilde{Y}$ 表示相应的噪声标签;随机变量 E 表示该实例特征是否受到标签噪声带来的错误影响。

另外, Song 等^[8]考虑数据标签破坏概率受到数据特征或类标签之间依赖关系的影响,将标签噪声分为两个子类:实例无关的标签噪声(Instance-Independent Label Noise, IILN)、实例依赖的标签噪声(Instance-Dependent Label Noise, IDLN)。考虑在实际细粒度图像分类任务中,真实世界的噪声模式通常更依赖于实例,这给噪声模式带来了很大的挑战,尤其是在存在类间不平衡的情况下。因此,本文遵循相同的分类法。

1) 实例无关的标签噪声。一种典型的标签噪声建模方法假设,当给定真实标签时,损坏过程与数据特征是有条件独立的,即真实标签被噪声转移矩阵 $T \in [0, 1]^{c \times c}$ 破坏,其中: $T_{ij} = P(\tilde{y} = j | y = i)$ 是真实标签 i 被翻转为损坏标签 j 的概率; $\tilde{y}$ 是噪声标签样本集合。如果

$$\forall i=j T_{ij} = 1 - \tau \wedge \forall i \neq j T_{ij} = (\tau / (c - 1))$$

则该噪声被称为噪声率 $\tau \in [0, 1]$ 的对称(均匀)噪声,其中一个真标签以相同的概率被翻转成其他标签。与对称噪声相反,如果

$$\forall i=j T_{ij} = 1 - \tau \wedge \exists i \neq j, i \neq k, j \neq k T_{ij} > T_{ik}$$

则该噪声称为非对称(标签相关)噪声,其中真实标签更有可能被错误标记为特定的标签。例如,

“狗”更容易与“猫”混淆,而不是“鱼”。

2) 实例依赖的标签噪声。为了噪声建模更真实,假设损坏概率同时依赖于数据特征和类标签。因此,将破坏概率定义为 $\rho_{ij}(x) = P(\tilde{y} = j | y = i, x)$ 。与实例无关标签噪声不同,示例 x 的数据特征也会影响 x 被错误标记的概率。在这种情况下,标签错误的概率依赖于每个具体的实例。比如,某些猫的图片可能由于其特定的特征(如颜色或姿态)更容易被错误地标记为狗。

现有的深度学习算法大多处理与实例无关的噪声,实例依赖噪声由于其复杂的建模尚未被广泛研究。Zhao 等^[15]提出了一种两阶段清洁样品识别方法来解决带实例依赖噪声标签的问题。Chen 等^[16]证实了标签噪声应该取决于实例,并提出了与实例相关标签噪声的参数化模型。Xie 等^[17]通过识别基于模糊实例的噪声标签,将部分多标签学习(Partial Multi-label Learning, PMLNI)方法扩展到多实例多标签学习,对来自各种实际任务的多个数据集的泛化束缚和广泛的实验进行了理论分析,证明了所提方法有效。

1.1.3 标签噪声影响

标签噪声广泛存在于现实数据集,对分类性能带来不利影响。但不可否认,人工标签噪声也有其潜在作用。例如:可以在统计研究中添加标签噪声以保护图像的数据隐私^[18]。为了提高分类器效果可加入标签噪声^[19],本文主要关注标签噪声对细粒度图像分类任务的负面影响。

1) 分类性能恶化。标签噪声会导致分类方法预测性能降低,例如,分类时经常使用的单元感知器、线性分类器、K-最近邻分类器都会受到标签噪声的影响^[20]。同时标签噪声也会影响常用的决策树、支持向量机等方法^[21-22]。深度学习方法中,分类任务需要大量的数据标签训练模型,但标注者可能在数据集采集和标注过程产生噪声标签,模型会在带噪声标签的情况下过度拟合甚至产生完全随机噪声,进而影响深度神经网络(Deep Neural Networks, DNN)的泛化性能。Shorten 等^[23]证明了深度神经网络可以很容易地用任意比例的损坏标签来拟合整个训练数据集,最终导致在测试数据集上的泛化性很差。

目前,正则化技术^[24]到了广泛应用,但它们并没有完全克服过拟合问题。例如,在对称噪声为 40% 的 CIFAR-100 数据集上,使用标准训练方法训练 WideResNet-16-8 时尽管相关正则化技术

都被激活,但在干净和噪声数据上训练的模型之间的测试精度差距仍然很大^[8]。此外,标签噪声导致的精度下降比输入噪声等其他噪声的危害更大^[6]。因此,在有噪声标签的情况下获得良好的泛化能力是一个关键的挑战。

2) 影响学习需求和模型复杂性。标签噪声会影响特征数或学习模型的复杂性。例如,在标签噪声的情况下,决策树的大小可能会使深度模型过于复杂^[25]。另外,标签噪声需要进行多次训练才会达到指定的性能指标^[26],标签噪声会使决策树的节点数量和支持向量机中支持向量的数量增加^[27]。此外,Angluin等^[28]讨论了在合取范式命题公式存在标签噪声的情况下,可能近似正确学习(Probably Approximately Correct Identification, PAC)的可行性。因此,在许多情况下需要通过降噪产生更容易理解的模型。

3) 观察频率失真。已知带有噪声标签的训练会损害预测模型的性能,因为后者可能被迫记忆错误的标签,而不是学习通用的模式。在医学分析领域,通常通过医学测试进行诊断,估计对应症状在人群中的流行程度。然而,标签噪声会影响医学检测结果的观察频率,从而可能导致不正确的结论^[29]。评估分类器的错误率对于模型选择和模型评估也很重要。在这种背景下,Xie等^[30]评估带有嘈杂标签的多标签分类器,确定了3种常见的现实世界标签噪声场景,说明当测试样本被污染时,标签噪声会对估计的错误率产生重要影响。

4) 相关特征识别困难。Shanthini等^[31]研究表明无论特征噪声是熟悉1个字符还是熟悉所有字符,标签噪声都比特征噪声具有更大的破坏性。例如,Broeker等^[32]证明,标签噪声的后果在微阵列数据的特征选择中很重要。在一个实验中,只有1个错误标记的样本导致大约20%的未识别的歧视性基因。在微阵列数据中,只有很少的数据是可用的。类似地,相关研究表明^[33],标签噪声降低了特征排名的稳定性,逻辑回归中特征选择对标签噪声具有敏感性。

1.2 噪声标签细粒度图像分类数学模型

1.2.1 细粒度图像模型

1) 定位-分类子网络方法。定位-分类子网络模型思路是:定位捕捉细粒度图像的判别语义部分,接着构建与关键部位相对应的局部特征向量,然后与全局图像(鸟的整体图像)特征结合以表示细粒度对象,接下来由分类子网络执行细微的

识别任务。其中,为解决信息区域定位这一挑战,Hanselmann等^[34]通过由3个专门为FGIC问题设计的高效且轻量级的组件来增强骨干卷积神经网络(Convolutional Neural Network, CNN)的性能(例如ResNet网络),这一高效的边界框估计量提高了对信息区域的定位效果。通过使用掩码进行弱监督目标检测和实例分割来提取粗略的目标实例。考虑到这些区域的组合具有更强的区域代表性、更好的区分能力,一些工作试图从联合部分特征学习方面着手。Lam等^[35]提出的长短期记忆网络(Long Short-Term Memory, LSTM),Wang等^[36]提出了交叉图传播(Criss-cross Graph Propagation, CGP)子网络来学习区域相关性。

在基于卷积神经网络的方法中:Yang等^[37]提出了自监督机制NTS-Net模型这一多智能体协作学习方案;Rodríguez等^[38]提出的机制由注意力模块和注意力门构成,根据局部信息,从不同抽象层次的主要CNN特征图激活中产生类别预测;Chen等^[39]关注有助于准确分类的特征图之间的关系,增加了对特征图的进一步处理,提出了一种高效的自监督注意力过滤和多尺度特征网络(Self-supervised Attention Filtering and Multi-scale Features Network, SAMFN)来提高解决FGIC问题的有效性。另外,定位-分类子网络方法中还包括滤波器输出作为局部检测器^[40]、结合深度强化学习方法^[41]、DF-GMM(Discriminative Feature-oriented Gaussian Mixture Model)框架^[42]等多种方法。

2) 端到端特征编码方法。端到端特征编码方法通过开发用于细粒度识别的强大深度学习模型,旨在统一但有细微差异的图像表示。本文主要介绍执行高阶特征交互、优化损失函数和辅助任务的方法。高阶特征交互旨在让深度卷积网络学习到具有判别性的深度特征,其中代表技术是基于协方差矩阵^[43]的表示及双线性CNN^[44]。双线性CNN架构以平移不变的方式对局部成对特征交互进行建模,这对于细粒度分类特别有用。为了规范化双线性表示并解决上述问题,Ma等^[45]提出了高效地学习对象任意子群之间的高阶交互来进行细粒度的视频理解。最近,Du等^[46]提出了一种由深度网络模块和双线性池化模块组成的DeepBP模型,实现多级信息集成,在细粒度数据集上取得了较好的结果。

另外,一些学者尝试为细粒度量身设计特定

损失函数,可以直接影响学习到的分类器和特征。Xie 等^[47]提出的 LG-CNN 由两个流损失函数联合训练,用于指导共享权重的更新。Yan 等^[48]提出了一个成对损失函数,代替传统三元组损失驱动,使行人再识别模型能够学习细粒度的特征。Chang 等^[49]提出了一种互信道损失 (MC-Loss) 的单一损失函数,MC-Loss 主要研究单个特征通道,而不是像传统那样从一个综合的特征图开始。因此,最终结果是一组特征通道,每个特征通道反映了特定类别的不同局部判别区域。MC-Loss 可以进行端到端的训练,不需要任何边界框或部件标注,通过关注各种部分级别的区域,鼓励特征通道更具区分性,不需要对细粒度对象的部分定位进行任何特定的网络修改。

3) 利用外部信息方法。针对 FGIC 任务,大量标记良好的数据集是很必要的。但由于细粒度图像标记中存在的需要领域专家和对象差异细微等困难,因此在细粒度分类中使用一些外部但成本较小的信息,以进一步提高准确率。Guo 等^[50]提出了一种基于 Webly 监督细粒度视觉分类端到端的方法,将不确定的动态损失校正和全局样本选择相结合,以解决标签噪声问题。Cai 等^[51]提出了一种 MSDeJOR (Multi-Scale Training with Decoupled Joint Optimization and Refurbishment) 模型,将样本选择与训练过程解耦,来解决当二者联合时无法胜任标签噪声存在的情况,该模型可以显式地遗忘那些被识别为噪声的样本。Zhang 等^[52]提出了一种简单而有效的纯网络监督 FGIC 方法,来处理训练过程中的噪声图像和困难样本。

零样本学习和元学习方法的思路是知识迁移,通常将具有良好标记训练数据的辅助类别的知识迁移到测试类别。其中,零样本学习利用辅助信息以类语义描述的形式泛化到未见(目标)类中。Shermin 等^[53]提出的零样本学习方法集成了嵌入学习网络和特征融合网络,实现了对目标类进行迁移学习。Huynh 等^[54]开发了一个针对 FGIC 任务的零样本学习框架,在对注意力模型、属性嵌入技术进行优化后,新设计的自校准损失解决了测试过程中可见类偏见的挑战。Jiang 等^[55]运用组合零样本学习 (Compositional Zero-Shot Learning, CZSL) 将不同组合下的相同属性(或对象)视为细粒度类,并提出了一种新颖的细粒度原始表示学习框架,来处理 CZSL 中基元概念的属性内或对象内视觉多样性,有助于学习更可靠的基元分

类器。

在细粒度分类中,可利用多模态数据来建立联合表示或嵌入,以整合多模态信息。这种方法可以提高细粒度分类的准确率。Xu 等^[56]提出了一种改进的融合多模态信息的细粒度图像分类方法,包括多模态数据预处理、多模态特征提取、多时相特征融合和决策校正模块。Zanzotto 等^[57]提出了 Human-in-the-loop 方法,将人类知识的互补优势与计算机视觉算法相结合。

1.2.2 带噪声标签的细粒度图像分类模型

考虑不同分类依据,现有基于深度学习处理噪声标签的分类归纳方法有多种。Nigam 等^[13]简要介绍了用于管理噪声标签的深度学习算法,其范围仅限于损失函数和样本选择两类。相关研究用嘈杂的标签总结了鲁棒学习的基本组成部分,他们更关注输入数据、目标函数和优化策略。还可以根据处理的标签类型进行分类,例如分为单标签和多标签。然而多标签问题常被当做单标签问题的拓展^[58],如 Tarekegn 等^[59]在使用单个正标签进行注释的图像数据集中,训练多标签分类器。Song 等^[8]针对噪声标签问题,将深度学习方法分为 5 类:(1) 鲁棒结构;(2) 鲁棒正则化;(3) 鲁棒损失函数;(4) 损失调整;(5) 样本选择。

另外值得关注的是,Algan 等^[12]将处理噪声标签的深度学习方法分为两大类:基于噪声模型的方法 (Noise Model Based Methods) 和无噪声模型的方法 (Noise Model Free Methods)。前者旨在估计噪声结构,并使用该信息来避免噪声标签的不利影响,优点是将噪声估计和分类任务解耦^[60],因此,更容易在现有分类算法上实现。然而,这类方法的一个基本任务是防止系统偏离现实。后一种方法旨在开发固有的噪声鲁棒策略,而无需对噪声结构进行显式建模,因此,主要关注通过规范网络训练程序来避免过度拟合。该分类标准涵盖了较为广泛全面的带噪声标签深度学习方法,而且是面向图像分类任务划分的分类标准,这对于本文所研究的 FGIC 问题很有参考价值,因此本文将按照这一分类标准开展综述。

2 基于显式噪声模型的 FGIC 方法

基于显式噪声模型的方法旨在对噪声结构进行建模,以便在训练期间以使用这些信息来避免噪声标签的不利影响。通常,这类方法旨在通过

忽略噪声或去处理来自噪声样本的信息来提取数据集的无噪声信息。此外, 一些方法试图通过校正有噪声的标签来优化数据集, 以提高分类器的数据集质量, 这类方法的性能很大程度上依赖估计潜在噪声的准确性。LNL 模型的优点在于: 1) 将分类任务和标签噪声估计解耦, 有利于其使用现有的分类算法; 2) 在拥有关于噪声结构的先验知识的情况下, 将额外信息插入系统中, 可以很容易地运行基于噪声模型的方法。本节中的算法同时试图找到潜在的噪声结构。

2.1 噪声通道方法

如图2所示, 噪声可以被建模为基础分类器之上的噪声通道。噪声通道适应噪声的特性, 使基础分类器在训练过程中采用无噪声的梯度进行训练。噪声通道估计中的一个常见问题是可伸缩性。在标签噪声学习中, 噪声转移矩阵表示干净标签翻转为噪声标签的概率, 在构建统计一致分类器中起着核心作用。随着类别的数量增加, 噪声转移矩阵的大小呈指数级增加, 使得难以计算。一类方法是显式地计算噪声转移矩阵, 然后使用该矩阵训练基础分类器。Hendrycks 等^[61]提出一种损失校正技术来利用可信数据, 方法是基于预处理模型的混淆矩阵来构造噪声转移矩阵。

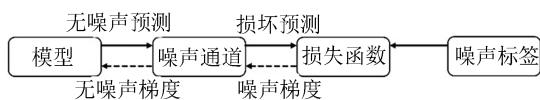


图2 噪声通道的常规设置

Fig. 2 General setup of noise channel

噪声忽略通常由噪声转移矩阵与类间损失组成^[62], 同时通过计算预测值和由概率转移矩阵生成的累积标签间的相关度, 提高分类器的分类精度。在基本分类器的训练期间, 噪声转移矩阵被增量地估计, 相关研究用于迭代训练网络以匹配给定的分布, 并在给定模型预测的情况下估计噪声转移矩阵。常用方法是利用锚点(属于特定类的数据点极大概率属于特定的类)来学习噪声转移矩阵, Xia 等^[63]提出了一种不用锚点的转移修正方法来有效学习噪声转移矩阵。首先利用与锚点相似的数据点对其进行初始化, 然后通过添加松弛变量来修改初始化矩阵, 该变量可以通过使用噪声数据与分类器一起进行学习和验证, 进而迭代地更新矩阵以最小化训练期间的损失。

2.2 噪声标签清理

标签清理算法是一种识别可疑标签并纠正为

真实标签的方法, 依赖于特征提取器将数据映射到特征域以研究噪声。这种方法通常用于区分信息量大的样本和带有噪声标签的样本。当存在干净标签的情况下, 将无噪声标签结构融合到有噪声标签中进行校正, 即根据干净数据上训练的预测网络对噪声标签重新标记。Gao 等^[64]将重新标记方法建模为一个动态决策过程, 提出了一种基于强化学习的标签增强算法, 有助于打破噪声结构, 使噪声在标签空间中更均匀地分布, 而不是保留有噪声的标签。Ortego 等^[65]提出了一种噪声检测策略, 该策略提出了一个合适的损失, 并以半监督的方式进行训练, 有助于重新标记噪声标签。在实际情况中, 无噪声数据并不总是可用的, 因此, 通常使用额外的正则化器来逐步估计更干净的后验标签分布。Zheng 等^[66]提出了一种基于噪声分类器预测的标签修正算法, 校正后的标签与真正的贝叶斯最优分类器一致。具体思路是部署置信策略, 其中标签由网络输出或给定噪声标签确定, 确定标准取决于模型预测的置信度。

2.3 数据集修剪

通过删除尽可能少的样本以防止不必要的数数据丢失, 有效减少噪声标签重标记的巨大成本。第一种选择是完全去除噪声样本, 并在修剪后的数据集上训练分类器。Sharma 等^[67]提出使用马尔可夫随机场进行无监督标签降噪, 首先在噪声数据上训练网络, 并使用该模型提取特征向量, 然后使用基于提取特征的 K-means 算法对数据进行聚类并去除异常值。另一种选择是只去除有噪声数据的标签, 并将数据集转换为两个子集(标记和未标记的数据), 接着可以在得到的数据集上使用半监督学习算法。Sachdeva 等^[68]提出了一种基于期望最大化框架训练算法 ScanMix, 探索了语义聚类和半监督学习。分步估计了潜在变量以根据训练图像的外观和分类结果对训练图像进行聚类, 优化了半监督学习分类并通过语义聚类学习有效的特征表示。通过利用自监督训练整合来自所有训练数据的信息, 然后采用 K 近邻方法来去除部分分布不均匀的样本, 并通过这种迭代的算法来寻找干净和有噪声的样本, 并以半监督的方式训练模型。

2.4 样本选择

通过选择正确的实例来指导网络, 旨在对分类器的输入流进行操作, 并提升分类器在带噪声标签情况下的分类精度。Guo 等^[50]提出了一种端

到端的方法,通过结合不确定的动态损失校正和全局样本选择来解决标签噪声问题。具体来说,利用深度神经网络来预测所有样本,记录最近几个时期的预测,计算整个时期全局样本选择的不确定动态损失。即保证给定样本的标签和模型预测一致,可以在训练集中使用给定样本。

2.5 样本重要性加权

与样本选择类似,通过根据实例的估计噪声水平为其分配权重,可以使训练效果更加精准。最简单的思路,尽可能对干净数据增加权重。然而,这不仅无法充分利用信息,而且干净的数据并不总是可用的。孪生神经网络^[11]通过学习区分干净和有噪声数据的判别特征来检测有噪声的标签,迭代地从干净样本中提取噪声样本,对噪声样本重新计算每个迭代加权因子,并在整个数据集上训练基础分类器。还有一类研究使用元学习范式来确定权重因子,在每次迭代中,对给定的小批量加权因子执行梯度下降步骤,以最大限度地减少干净验证数据的损失。Shu 等^[60]采用了类似方法,但使用多层感知器来估计加权函数,而不是隐式计算加权因子。

2.6 贴标机质量评估

针对多个注释者标记数据集产生的标签噪声,建立模型并使用贴标机特征来提高标注性能。其中两个关键因素是噪声贴标机特征和真实标签,可以用期望最大化对两者进行估计。若噪声与实例相关,则可以使用噪声转移矩阵对贴标机进行建模。Sun 等^[69]在元学习场景中训练标签纠正处理,其中元网络生成一个整流向量,通过嘈杂的标签促进鲁棒学习,从标签输入和预测向量的输入中估计整体向量的分布,无缝集成到分类网络的优化中,进而提高标注性能。

整体而言,基于噪声模型的方法在很大程度上依赖于噪声结构的精确估计,因此需要更好的估计量,这就需要对潜在噪声进行估计。难点在于防止模型过多偏离现实,虽然对噪声结构的假设在处理特定领域噪声时效果显著,但会损害对于不同设置的适用性和泛化性。

3 基于隐式噪声模型的 FGIC 方法

基于隐式噪声模型的方法旨在开发固有的噪声鲁棒策略,而无需对噪声结构进行显式建模。此类方法假设分类器对噪声不太敏感,并且过拟

合会导致性能下降,因此,主要关注通过规范网络训练程序来避免过度拟合。它们通常依赖于分类器的内部噪声容忍度,并旨在通过正则化不需要的噪声数据记忆来提高性能。

3.1 鲁棒损失方法

旨在设计损失函数,降低噪声标签对性能的影响。若一个分类器在有噪声和无噪声数据环境中,均能达到相同的分类精度,则可认为损失函数是噪声鲁棒的。这类方法以相同方式处理有噪声数据和干净数据,防止了在数据分布上利用任何先验信息。Manwani 等^[70]从理论上证明了损失函数的充分条件,使得该函数的风险最小化为二元分类的噪声容限。随后,使用深度学习扩展了多类分类的充分条件^[71]。具体而言,如果损失函数 ℓ 满足噪声率 $\tau < ((c-1)/c)$, c 是指类别数量,且满足

$$\sum_{j=1}^c \ell((x; \Theta), y=j) = C \forall x \in X \forall f \quad (6)$$

式中: Θ 是神经网络的一个参数; C 是一个常数。此条件保证在指定假设下,在噪声数据上与在无噪声数据上训练的分类器具有相同的误分类概率。在分类任务中,分类交叉熵 (Categorical Cross Entropy, CCE)^[72] 损失是应用最广泛的损失函数,因为它具有快速收敛和高度泛化能力。然而,在存在噪声标签的情况下,鲁棒平均绝对误差 (Mean Absolute Error, MAE)^[8] 损失比 CCE 损失实现了更好的泛化,因为只有 MAE 损失满足上述条件。MAE 损失的局限性是当涉及复杂数据时其泛化性能会显著下降。广义交叉熵 (Generalized Cross Entropy, GCE) 可以兼顾 MAE 和 CCE 损失的优点。GCE 作为更通用的一类噪声鲁棒损失函数,更改损失函数只需对现有架构和算法进行最少的干预,因此可以迅速推广应用。Lü 等^[73]提出的课程损失 (Curriculum Loss, CL) 是 0-1 损失函数的替代损失,它提供了一个更严格的上限,并可以自适应地选择样本进行模型训练。其中噪声修剪课程损失 (Noise Pruned Curriculum Loss, NPCL) 将 CL 扩展到解决标签损坏问题。主动被动损失 (Active Passive Loss, APL)^[74] 是两种鲁棒损失函数的组合,一种是使属于给定类别的概率最大化的主动损失,另一种是使属于其他类别的概率最小化的被动损失。值得关注的是,以上基于损失函数鲁棒性的方法在理论上得到了很好的验证,然而,它们仅

在标注数据或类别数量较少时表现良好。此外,损失函数的修改优化增加了收敛的训练时间。

3.2 元学习方法

旨在消除网络架构、损失函数、优化器算法等必要条件,它不仅能学习任务所需的复杂函数,还能学习“学习”本身。自动重新加权^[75]是一种元学习算法,可根据训练实例的梯度方向学习其权重。它在训练数据集中加入一个小的干净验证数据集,并对小批量实例的后向损失进行重新加权,从而使更新后的梯度最小化为该干净验证数据集的损失。Shu等^[60]提出的元加权网络将加权函数参数化为只有1个隐藏层的多层感知器网络。元目标的定义是更新其参数,使其最小化为一个小的纯净数据集的经验风险。每次迭代时,目标网络的参数都由通过元目标更新的权重函数指导。Xu等^[76]提出了一种元权重学习方法,用于最小化不同训练任务间的性能不平等,引入的端到端训练方法可以有效地学习权重和初始化网络模型。然而,由于元学习的每个训练循环都需要嵌套梯度计算循环,因此比传统训练过程更消耗时间和计算成本。

3.3 正则化方法

可防止DNN过度拟合噪声标签,其将噪声数据导致的性能下降视为对噪声的过度拟合,广泛使用的技术包括权重衰减、dropout^[24]、对抗训练^[77]、混合^[78]、标签平滑^[79]。正则化方法在通常噪声比率的数据上运行良好,但单独使用并不能充分提高测试精度,然而当噪声较大时泛化能力较差。Menon等^[80]提出了一种基于复合损失的梯度裁剪,它是标准梯度裁剪的一种变体,用于标签噪声的鲁棒性。鲁棒早期学习^[81]对关键参数和非关键参数进行分类,分别用于拟合干净标签和噪声标签。然后,仅对非关键的更新规则进行惩罚。Wei等^[82]利用开集辅助数据,通过给开集样本分配随机标签,从标签集中均匀采样,防止噪声标签的过拟合。以上改进方法与经典方法一起使用,可进一步提高FGIC对标签噪声的鲁棒性。

总体而言,属于隐式噪声模型的方法将噪声数据视为异常,因此,在避免过拟合和异常检测方面是相似的。尽管这一假设对随机噪声可能非常有效,但在更复杂和结构化的噪声情况下就失去了有效性。由于噪声建模没有显式地从分类任务中解耦,所提出的方法在一般意义上是嵌入到现有算法中的,这阻止了它们快速部署到现有系

统中。此外,属于元学习的算法在计算上代价高昂,因为它们需要训练循环的多次迭代。

4 研究展望

综合已有研究以及图像处理领域工程实践需求的发展,以下研究内容需进一步关注。

1) 针对细粒度图像的局部和全局特征提取。信息区域定位和细粒度特征学习是细粒度类别分类的难点。近年来深度学习已成为判别性特征学习的一种有效方法,并广泛部署在实际问题中。基于注意力机制的深度学习方法善于捕捉细粒度图像的全局特征,但其在提取特征时缺乏全面性,限制了它在计算机视觉任务中的表现^[83]。现有多尺度注意力机制使注意力在图像(特征图)的作用范围不同,捕捉从大到小、从全局到局部的特征,最终通过特征融合操作融合各分支的特征。目前,已有研究将Vision Transformer (ViT)^[84]运用在FGIC任务中,其自注意力机制可以提取不同像素块的判别性标记信息,然而其深层的分类标记同样会忽略层与层之间的局部特征。Zhang等^[85]提出了包含3个组件的ACC-ViT方法,其中的CTI组件集成了不同ViT层中类标记捕获的互补信息,取得了良好的实验结果。未来研究可以探索更加有效的特征提取方法,以提高细粒度图像分类的准确性和鲁棒性。同时,加深了解网络的哪些部分受到标签噪声的高度影响,可能有助于解决标签噪声对特征提取的不利影响。

2) 实例依赖噪声的深度学习方法。FGIC介于语义级图像分类与实例级图像分类之间,需关注同一类别内相似的子类别,识别其固有的细微类间差异,因此,细粒度图像中的噪声标签往往与实例紧密相关。然而,现有的深度学习算法大多处理与实例无关的噪声,实例依赖噪声由于其复杂的建模尚未被广泛研究。未来应从现有的噪声建模方式和算法出发,分析推理出实例依赖噪声可能潜在的建模方式,并根据FGIC问题具体任务构建鲁棒的FGIC深度学习方法,以便解决深度学习在真实场景中实际应用的问题。

3) 噪声鲁棒的FGIC方法研究。现实场景下的标签噪声污染程度往往更高,当带噪声数据的规模接近或大于干净数据规模时,算法的效率会显著下降。同时,由于提升模型性能的关键是要处理好复杂样本,然而复杂样本无论标签是干净

的还是噪声的,都会产生相对较大的损失,目前的方法都无法精确地分类复杂样本。模型训练过程中操作处理也可能会受环境因素影响而发生异常,往往会严重影响系统性能。因此,如何建立高效准确的噪声鲁棒模型,更好地适应复杂噪声标签数据环境,为基于深度学习的FGIC方法拓展现实数据集分类任务适用场景,未来需要进一步探索研究。

4) FGIC的轻量化模型。深度学习方法在具体FGIC任务中显示出巨大的潜力^[86],但泛化性和轻量化方面仍需加强。针对当前Vision transformer模型需要大量训练数据且模型参数大的情况,Yu等^[87]提出的Mix-ViT模型可以从输入数据的内在结构中获得更多知识,因此能够在不需要大量训练数据的情况下学习判别表示。在车辆类型分类任务中,Sun等^[88]提出了一种使用轻量级卷积神经网络方法有效降低了模型的复杂度。Boukerche等^[89]提出了一种轻量级递归注意力单元来增强卷积神经网络的特征提取能力。然而鲁棒性提升与泛化性增强之间存在矛盾,例如很多噪声建模方法具有鲁棒性,但其没有显式地从分类任务中解耦,所提出的方法在一般意义上是嵌入到现有算法中的,这影响了它们快速部署到其他领域分类任务的泛化性。因此,如何处理这个矛盾是现有深度学习在其他分类任务应用面临的重要问题。

参考文献:

- [1] MAJI S, RAHTU E, KANNALA J, et al. Fine-grained visual classification of aircraft [EB/OL]. 2013[2024-08-12]. <https://doi.org/10.48550/arXiv.1306.5151>.
- [2] VAN HORN G, BRANSON S, FARRELL R, et al. Building a bird recognition app and large scale dataset with citizen scientists: The fine print in fine-grained dataset collection[C]//Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2015: 595-604.
- [3] ZHANG Y S, CHEN L, YUAN Y. Multimodal fine-grained transformer model for pest recognition [J]. Electronics, 2023, 12(12): 2620.
- [4] CAO Q, DU N, YU L, et al. Practical fine-grained learning based anomaly classification for ECG image[J]. Artificial Intelligence in Medicine, 2021, 119: 102130.
- [5] YIN J H, WU A C, ZHENG W S. Fine-grained person re-identification[J]. International Journal of Computer Vision, 2020, 128(6): 1654-1672.
- [6] PATEL D, SASTRY P S. Adaptive sample selection for robust learning under label noise. [C]//2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Piscataway, NJ, USA: IEEE, 2023: 3921-3931.
- [7] WEI X S, SONG Y Z, AODHA O M, et al. Fine-grained image analysis with deep learning: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(12): 8927-8948.
- [8] SONG H, KIM M, PARK D, et al. Learning from noisy labels with deep neural networks: A survey[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(11): 8135-8153.
- [9] XU Z, HUANG S L, ZHANG Y, et al. Webly-supervised fine-grained visual categorization via deep domain adaptation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(5): 1100-1113.
- [10] FRÉNEY B, VERLEYSEN M. Classification in the presence of label noise: A survey[J]. IEEE Transactions on Neural Networks and Learning Systems, 2014, 25(5): 845-869.
- [11] JIA S, JIANG S G, LIN Z J, et al. A semisupervised Siamese network for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5516417.
- [12] ALGAN G, ULUSOY I. Image classification with deep learning in the presence of noisy labels: A survey[J]. Knowledge-Based Systems, 2021, 215: 106771.
- [13] NIGAM N, DUTTA T, GUPTA H P. Impact of noisy labels in learning techniques: A survey [C]//Proceedings of the International Conference on Data Intelligence and Security (ICDIS). Piscataway, NJ, USA: IEEE, 2020: 403-411.
- [14] ZHANG J, WU X D, SHENG V S. Learning from crowdsourced labeled data: A survey[J]. Artificial Intelligence Review, 2016, 46(4): 543-576.
- [15] ZHAO G L, LI G B, QIN Y P, et al. Centrality and consistency: Two-stage clean samples identification for learning with instance-dependent noisy labels[C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2022: 21-37.

- [16] CHEN P F, YE J J, CHEN G Y, et al. Beyond class-conditional assumption: A primary attempt to combat instance-dependent label noise[C]//Proceedings of the Conference on Artificial Intelligence. Washington, D.C., USA: AAAI, 2021, 35(13): 11442-11450.
- [17] XIE M K, HUANG S J. Partial multi-label learning with noisy label identification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(7): 3676-3687.
- [18] TIAN Q, KUANG K, JIANG K L, et al. Confounder GAN: Protecting image data privacy with causal confounder[J]. Advances in Neural Information Processing Systems, 2022, 35: 32789-32800.
- [19] JU L, WANG X, WANG L, et al. Improving medical images classification with label noise using dual-uncertainty estimation[J]. IEEE Transactions on Medical Imaging, 2022, 41(6): 1533-1546.
- [20] YANG W J, LI C Q, JIANG L X. Learning from crowds with robust support vector machines [J]. Science China Information Sciences, 2022, 66(3): 132103.
- [21] LAXMI S, GUPTA S K. Intuitionistic fuzzy proximal support vector machines for pattern classification[J]. Neural Processing Letters, 2020, 51(3): 2701-2735.
- [22] LI W W, LU Y Q, CHEN L, et al. Label distribution learning with noisy labels via three-way decisions[J]. International Journal of Approximate Reasoning, 2022, 150: 19-34.
- [23] SHORTEN C, KHOSHGOFTAAR T M. A survey on Image Data Augmentation for Deep Learning[J]. Journal of Big Data, 2019, 6(1): 60.
- [24] XIE J Y, MA Z Y, LEI J J, et al. Advanced dropout: A model-free methodology for Bayesian dropout optimization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(9): 4605-4625.
- [25] JIA Q R, LI X H, YU L, et al. Learning from training dynamics: Identifying mislabeled data beyond manually designed features[C]//Proceedings of the Conference on Artificial Intelligence. Washington, D.C., USA: AAAI, 2023, 37(7): 8041-8049.
- [26] ALBERY G F, KIRKPATRICK L, FIRTH J A, et al. Unifying spatial and social network analysis in disease ecology[J]. The Journal of Animal Ecology, 2021, 90(1): 45-61.
- [27] MÜLLER N M, MARKERT K. Identifying mislabeled instances in classification datasets[C]//Proceedings of the International Joint Conference on Neural Networks (IJCNN). Piscataway, NJ, USA: IEEE, 2019: 1-8.
- [28] ANGLUIN D, LAIRD P. Learning from noisy examples[J]. Machine Learning, 1988, 2(4): 343-370.
- [29] DONEGAN C, CHUN Y W, GRIFFITH D A. Modeling community health with areal data: Bayesian inference with survey standard errors and spatial structure[J]. International Journal of Environmental Research and Public Health, 2021, 18(13): 6856.
- [30] XIE M K, HUANG S J. Partial multi-label learning with noisy label identification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(7): 3676-3687.
- [31] SHANTHINI A, VINODHINI G, CHANDRAS-EKARAN R M, et al. RETRACTED ARTICLE: A taxonomy on impact of label noise and feature noise using machine learning techniques[J]. Soft Computing, 2019, 23(18): 8597-8607.
- [32] BROEKER C D, ORTIZ M M O, MURILLO M S, et al. Integrative multi-omic sequencing reveals the MMTV-Myc mouse model mimics human breast cancer heterogeneity[J]. Breast Cancer Research, 2023, 25(1): 120.
- [33] YIN N, SHEN L, WANG M Z, et al. OMG: Towards effective graph classification against label noise[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(12): 12873-12886.
- [34] HANSELMANN H, NEY H. ELoPE: Fine-grained visual classification with efficient localization, pooling and embedding[C]//2020 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway, NJ, USA: IEEE, 2020: 1236-1245.
- [35] LAM M, MAHASSENI B, TODOROVIC S. Fine-grained recognition as HSnet search for informative image parts[C]//Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 6497-6506.
- [36] WANG Z H, WANG S J, LI H J, et al. Graph-propagation based correlation learning for weakly supervised fine-grained image classification [C]//Proceedings of the Conference on Artificial Intelligence. Washington, D.C., USA: AAAI, 2020, 34(7): 12289-12296.

- [37] YANG Z, LUO T G, WANG D, et al. Learning to navigate for fine-grained classification[C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2018: 438-454.
- [38] RODRÍGUEZ P, VELAZQUEZ D, CUCURULL G, et al. Pay attention to the activations: A modular attention mechanism for fine-grained image recognition[J]. IEEE Transactions on Multimedia, 2020, 22(2): 502-514.
- [39] CHEN H Y, CHENG L L, HUANG G H, et al. Fine-grained visual classification with multi-scale features based on self-supervised attention filtering mechanism[J]. Applied Intelligence, 2022, 52(13): 15673-15689.
- [40] HUANG Z X, LI Y. Interpretable and accurate fine-grained recognition via region grouping [C]//Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 8659-8669.
- [41] HE X T, PENG Y X, ZHAO J J. Which and how many regions to gaze: Focus discriminative regions for fine-grained visual categorization[J]. International Journal of Computer Vision, 2019, 127(9): 1235-1255.
- [42] WANG Z H, WANG S J, YANG S H, et al. Weakly supervised fine-grained image classification via Gaussian mixture model oriented discriminative learning[C]//Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 9746-9755.
- [43] WANG Q L, XIE J T, ZUO W M, et al. Deep CNNs meet global covariance pooling: Better representation and generalization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(8): 2582-2597.
- [44] HOSSAIN S, UMER S, ROUT R K, et al. Fine-grained image analysis for facial expression recognition using deep convolutional neural networks with bilinear pooling[J]. Applied Soft Computing, 2023, 134(5): 109997.
- [45] MA C Y, KADAV A, MELVIN I, et al. Attend and interact: Higher-order object interactions for video understanding[C]//Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2018: 6790-6800.
- [46] DU Y N, RUI T, LI H W, et al. DeepBP: A bilinear model integrating multi-order statistics for fine-grained recognition[J]. Computers and Electrical Engineering, 2023, 105: 108432.
- [47] XIE G S, ZHANG X Y, YANG W H, et al. LG-CNN: From local parts to global discrimination for fine-grained recognition[J]. Pattern Recognition, 2017, 71: 118-131.
- [48] YAN C, PANG G S, BAI X, et al. Beyond triplet loss: Person re-identification with fine-grained difference-aware pairwise loss[J]. IEEE Transactions on Multimedia, 2022, 24: 1665-1677.
- [49] CHANG D L, DING Y F, XIE J Y, et al. The devil is in the channels: Mutual-channel loss for fine-grained image classification[J]. IEEE Transactions on Image Processing, 2020, 29: 4683-4695.
- [50] GUO J, DING M Y, WANG Q, et al. An uncertainly dynamic loss correction and global sample selection method for Webly supervised fine-grained visual classification[J]. Circuits, Systems, and Signal Processing, 2022, 41(6): 3265-3281.
- [51] CAI Z H, XIE G S, HUANG X G, et al. Robust learning from noisy web data for fine-Grained recognition[J]. Pattern Recognition, 2023, 134: 109063.
- [52] ZHANG C Y, YAO Y Z, ZHANG J C, et al. Web-supervised network for fine-grained visual classification[C]//Proceedings of the International Conference on Multimedia and Expo (ICME). Piscataway, NJ, USA: IEEE, 2020: 1-6.
- [53] SHERMIN T, TENG S W, SOHEL F, et al. Integrated generalized zero-shot learning for fine-grained classification[J]. Pattern Recognition, 2022, 122: 108246.
- [54] HUYNH D, ELHAMIFAR E. Fine-grained generalized zero-shot learning via dense attribute-based attention[C]//Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 4482-4492.
- [55] JIANG H, YANG X S, CHEN C F, et al. Fine-grained primitive representation learning for compositional zero-shot classification[C]//Proceedings of the International Conference on Multimedia and Expo (ICME). Piscataway, NJ, USA: IEEE, 2023: 456-461.
- [56] XU J, ZHANG X Q, ZHAO C M, et al. Improving fine-grained image classification with multi-modal information[J]. IEEE Transactions on Multimedia, 2024, 26: 2082-2095.

- [57] ZANZOTTO F M. Viewpoint: Human-in-the-loop artificial intelligence[J]. *Journal of Artificial Intelligence Research*, 2019, 64: 243-252.
- [58] HU Y Y, DONALD C, GIACAMAN N. Can multi-label classifiers help identify subjectivity? A deep learning approach to classifying cognitive presence in MOOCs[J]. *International Journal of Artificial Intelligence in Education*, 2023, 33(4): 781-816.
- [59] TAREKEGN A N, GIACOBINI M, MICHALAK K. A review of methods for imbalanced multi-label classification[J]. *Pattern Recognition*, 2021, 118: 107965.
- [60] SHU J, XIE Q, YI L X, et al. Meta-Weight-Net: Learning an explicit mapping for sample weighting [EB/OL]. 2019[2024-08-12]. <https://doi.org/10.48550/arXiv.1902.07379>.
- [61] HENDRYCKS D, MAZEIKA M, WILSON D, et al. Using trusted data to train deep networks on labels corrupted by severe noise [EB/OL]. 2018[2024-08-12]. <https://doi.org/10.48550/arXiv.1802.05300>.
- [62] MATTU S R, CHOCKALINGAM A. Learning-based channel estimation and phase noise compensation in doubly-selective channels[J]. *IEEE Communications Letters*, 2022, 26(5): 1052-1056.
- [63] XIA X B, LIU T L, WANG N N, et al. Are anchor points really indispensable in label-noise learning? [C]//*Proceedings of the 33rd International Conference on Neural Information Processing Systems (NeurIPS)*. Piscataway, NJ, USA: IEEE, 2019: 6838-6849.
- [64] GAO Y B, ZHANG Y, GENG X. Label enhancement for label distribution learning via prior knowledge[C]//*Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI)*. California: International Joint Conference on Artificial Intelligence Organization, 2020: 3223-3229.
- [65] ORTEGO D, ARAZO E, ALBERT P, et al. Towards robust learning with different label noise distributions[C]//*Proceedings of the 25th International Conference on Pattern Recognition (ICPR)*. Piscataway, NJ, USA: IEEE, 2021: 7020-7027.
- [66] ZHENG S Z, WU P X, GOSWAMI A, et al. Error-bounded correction of noisy labels[C]//*Proceedings of the 37th International Conference on Machine Learning (ICML)*. San Diego, CA, USA: ICML, 2020: 11447-11457.
- [67] SHARMA K, DONMEZ P, LUO E M, et al. NoiseRank: Unsupervised label noise reduction with dependence models[C]//*Proceedings of the Conference on European Conference on Computer Vision (ECCV)*. Piscataway, NJ, USA: IEEE, 2020: 737-753.
- [68] SACHDEVA R, CORDEIRO F R, BELAGI-ANNIS V, et al. ScanMix: Learning from severe label noise via semantic clustering and semi-supervised learning[J]. *Pattern Recognition*, 2023, 134: 109121.
- [69] SUN H L, GUO C H, WEI Q, et al. Learning to rectify for robust learning with noisy labels[J]. *Pattern Recognition*, 2022, 124: 108467.
- [70] MANWANI N, SASTRY P S. Noise tolerance under risk minimization[J]. *IEEE Transactions on Cybernetics*, 2013, 43(3): 1146-1151.
- [71] QIN Z, ZHANG Z W, LI Y, et al. Making deep neural networks robust to label noise: Cross-training with a novel loss function[J]. *IEEE Access*, 2019, 7: 130893-130902.
- [72] LI P L, HE X H, CHENG X J, et al. An improved categorical cross entropy for remote sensing image classification based on noisy labels[J]. *Expert Systems with Applications*, 2022, 205: 117296.
- [73] LYU Y M, TSANG I W. Curriculum loss: Robust learning and generalization against label corruption[C]//*Proceedings of the 7th International Conference on Learning Representations (ICLR)*. New Orleans, LA, USA: Computational and Biological Learning Society, 2019: 1-22.
- [74] MA X J, HUANG H X, WANG Y S, et al. Normalized loss functions for deep learning with noisy labels[C]//*Proceedings of the 37th International Conference on Machine Learning (ICML)*. San Diego, CA, USA: ICML, 2020: 6543-6553.
- [75] REN M Y, ZENG W Y, YANG B, et al. Learning to reweight examples for robust deep learning[C]//*Proceedings of the 35th International Conference on Machine Learning (ICML)*. San Diego, CA, USA: ICML, 2018: 4334-4343.
- [76] XU Z X, CHEN X L, TANG W, et al. Meta weight learning via model-agnostic meta-learning[J]. *Neurocomputing*, 2021, 432: 124-132.
- [77] XIAO J, BI X J. Model-guided generative adversarial networks for unsupervised fine-grained image generation[J]. *IEEE Transactions on Multimedia*, 2024, 26: 1188-1199.

- [78] TIAN R L, SHI H M. A transfer-based few-shot classification approach via masked manifold mixup and fuzzy memory contrastive learning[J]. *Neural Computing and Applications*, 2023, 35(14): 10069–10082.
- [79] HUANG H, LIU Y, XIONG Q J, et al. A fatty liver diseases classification network based on adaptive coordination attention with label smoothing [J]. *Biomedical Signal Processing and Control*, 2023, 86: 105267.
- [80] MENON A K, RAWAT A S, REDDI S J, et al. Can gradient clipping mitigate label noise? [C/OL]//*Proceedings of the 8th International Conference on Learning Representations (ICLR)*. New Orleans, LA, USA: Computational and Biological Learning Society, 2020. (2023-05-06) [2024-04-03]<https://openreview.net/group?id=ICLR.cc/2020/Conference>.
- [81] XIA X B, LIU T L, HAN B, et al. Robust early-learning: Hindering the memorization of noisy labels[C/OL]//*Proceedings of the 9th International Conference on Learning Representations (ICLR)*. New Orleans, LA, USA: Computational and Biological Learning Society, 2021. (2023-05-06) [2024-04-03]https://openreview.net/forum?id=EqI5b1_hTE4.
- [82] WEI H X, TAO L, XIE R, et al. Openset label noise can improve robustness against inherent label noise[C]//*Proceedings of the 35th International Conference on Neural Information Processing Systems (NeurIPS)*. Piscataway, NJ, USA: IEEE, 2021, 34: 7978–7992.
- [83] XIE J, PANG Y W, NIE J, et al. Latent feature pyramid network for object detection[J]. *IEEE Transactions on Multimedia*, 2023, 25: 2153–2163.
- [84] HE J, CHEN J N, LIU S, et al. TransFG: A transformer architecture for fine-grained recognition [C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. Washington, D.C., USA: AAAI, 2022, 36(1): 852–860.
- [85] ZHANG Z C, CHEN Z D, WANG Y X, et al. A vision transformer for fine-grained classification by reducing noise and enhancing discriminative information[J]. *Pattern Recognition*, 2024, 145: 109979.
- [86] SU Y, LI X J, YAO J P, et al. A spectral spatial feature rotation-based ensemble method for imbalanced hyperspectral image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 5515918.
- [87] YU X H, WANG J, ZHAO Y, et al. Mix-ViT: Mixing attentive vision transformer for ultra-fine-grained visual categorization[J]. *Pattern Recognition*, 2023, 135: 109131.
- [88] SUN W, ZHANG G C, ZHANG X R, et al. Fine-grained vehicle type classification using lightweight convolutional neural network with feature optimization and joint learning strategy[J]. *Multimedia Tools and Applications*, 2021, 80(20): 30803–30816.
- [89] BOUKERCHE A, MA X R. A novel smart lightweight visual attention model for fine-grained vehicle recognition[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(8): 13846–13862.