基于独立分类网络的开集识别研究

徐雪松，付瑜彬，于波

（华东交通大学电气与自动化工程学院，江西 南昌 330013）

摘　要：目前，大多数的研究在解决图像分类模型的开集识别问题时，依赖于大量的标注样本来刻画已知类别的特征分布，这可能会导致两个问题：标注样本不足时，模型的训练容易陷入局部极值；根据已知类别特征差异构建的决策边界仅反映已知类间的特征分布，缺乏开集泛化性。本文提出了一种分离式的独立分类网络结构，每个类别都包含独立的线性特征层，特征层中设计的神经元节点能够在有限的数据样本下更准确地捕获类别特征。同时，在模型训练时，文中引入了一类无需标注的负样本，使得模型在构建决策边界时不仅依赖于已知类别的特征差异，在不增加额外标注样本的情况下，增加模型决策边界的开集泛化性。实验结果表明，在 FDS 和 Imagenet-Crop 数据集上采用独立分类网络并融合开集自适应训练的算法比现有开集识别算法具有更优的开集识别性能。

关键词：深度学习；开集识别；图像分类；迁移学习

中图分类号：TP391　　**文献标志码：**A

本文引用格式：徐雪松，付瑜彬，于波.基于独立分类网络的开集识别研究[J].华东交通大学学报，2023，xx(x)：x-x.

Research on open set recognition based on independent Classification Network

Xu XueSong, Fu YuBin, Yu Bo

（School of Electrical & Automation Engineering, East China Jiaotong University, Nanchang 330013, China）

Abstract: At present, most studies rely on a large number of annotated samples to describe the feature distribution of known categories when solving the open set recognition problem of image classification models, which may lead to two problems: when the annotated samples are insufficient, the model training is easy to fall into the local extreme value; The decision boundary constructed according to the feature difference of known classes only reflects the feature distribution among known classes and lacks the generalization of open sets. In this paper, we propose a separate independent classification network structure, in which each category contains an independent linear feature layer. The neural nodes designed in the feature layer can capture the category features more accurately under limited data samples. At the same time, a class of negative samples without labeling is introduced in the model training, so that the model not only relies on the feature difference of the known categories when constructing the decision boundary, but also increases the open set generalization of the model decision boundary without adding additional labeled samples. The experimental results show that the open set recognition algorithm based on independent classification network and adaptive open set training on FDS and Imagenet-Crop data sets has better open set recognition performance than the existing open set recognition algorithm.

Key words: Deep learning; Open set recognition; Image classification; Transfer learning

Citation format: XU X S，FU Y B，YU B，et al. Research on open set recognition based on independent Classification Network [J]. Journal of East China Jiaotong University，2023，xx(x)：x-x

---

收稿日期：2023−03−06

基金项目：国家自然科学基金资助项目(1763012)

在传统图像分类模型的设计中，通常遵循“类内距离最小化、类间距离最大化”的设计思想，这要求事先知晓各类别样本特征的分布情况。然而，在实际工程应用中，常常无法准确预知样本的分布。例如，在自动驾驶汽车中，可能会遇到完全未知的场景。这时使用预先设计好的分类器进行分类会导致较高的错误率。如何使预先设计的分类器在面对未知类别样本时仍然能够保持较高的分类准确性，即所谓开集识别问题(Open Set Recognition, OSR)[1]一直是实际工程应用中的难题。

在早期的研究中，一些学者就考虑过 OSR 问题。其中，文献[2]中通过在算法中设计“拒判”规则，以区分已知和未知样本。然而，直到 2013 年，Scheirer 等人[1]才首次将 OSR 问题进行了系统性的总结。自那时起，OSR 问题引起了越来越多学者的关注。最近的一些研究中，Chuanxing 和 Mahdavi 等人[3,4]对以往的开集识别模型进行了总结，将现有的模型算法分为生成式和判别式。

判别式模型的基本思路是：通过正则化收缩已知类在特征空间的决策边界，并在特征空间内给未知类分配区域，以此缓解分类模型的过度泛化，降低模型的 OSR 风险。Scheirer 等人[1]提出了一种基于支持向量机的 OSR 算法，该方法通过在模型中额外设计一个超平面来区分已知和未知类。类似的做法还有文献[5]，超平面的存在压缩了已知类的特征空间，但已知类空特征间依旧是无界的，开集风险依旧存在。Zhang 等人[6]利用极值理论对数据的尾部分布进行重新建模，提高了模型对尾部分布数据的处理能力，但这种基于统计学的方式仅在模型的预测阶段使用，并未参与模型的训练。文献[7]首次在深度神经网络中引入 OSR 问题，证明了 Softmax 激活函数不利于 OSR 任务，并提出了 OpenMax 算法作为深度神经网络的第一个开集解决方案。文献[8]在文本分类任务中提出了使用采用 K-Sigmoid 方法代替神经网络的 Softmax 层，将 K 分类任务转换成 K+1 个二分类任务，并额外设计了一个未知类。然而，由于未知类的无限性和不可预见性，模型依旧无法有效的形成收缩的特征边界。总的来说，判别式模型通过收缩已知类的边界，给未知类在特征空间分配了空间，在一定程度上缓解了分类模型过度泛化的问题。

生成式模型则显式地建模了已知类在特征空间分布，旨在通过约束已知类在特征空间内的分布来缩小模型的决策边界，以此提高模型对开集数据的识别性。文献[9,10]提出使用生成对抗网络(Generative Adversarial Networks，GAN)来生成实例样本，并对生成样本进行概率估计来扩充已知类的分布，使分类器能够更加准确地区分已知类和未知类。文献[11,12]采用自动编码器技术对输入样本进行重建，通过对比重建误差来进行区分已知类和未知类。然而，在训练数据样本数量不足或类别单一化的情况下，代表已知类的特征分布本身十分稀疏，这会导致即使输入样本属于已知类，其重构误差仍然很大。最近，基于原型网络[13]的 ORS 算法得到了广泛关注，文献[14–17]通过在特征空间内对已知类原型设计空间约束，限定未知类原型在特征空间的分布区域，以此来降低模型的 OSR 风险。生成式模型在特征空间内更好地建模了已知类的决策边界，但在一般情况下，生成式模型的训练需要更加丰富的数据样本，在样本匮乏的条件下，生成式模型的闭集精度要低于判别式模型。

解决 OSR 问题的核心是准确描述各类别特征真实的边缘分布，并缩小已知类别的决策边界。判别式模型依赖于对照组样本来确定各类别特征的分布，因此需要有足够数量的对照组样本，才能准确刻画每个类别的边缘特征。生成式模型通过生成方法学习每个类别自身的特征分布，因此需要足够丰富的自身样本才能训练出可靠的生成模型。在实际应用中，有时可能无法获得足够数量的标注样本。此时，由于自身样本特征稀疏，无论是判别式模型还是生成式模型的训练都面临较大困难。

卷积神经网络(Convolutional Neural Network -CNN)和 Vision Transformer(ViT)[18]被广泛应用于图像分类任务。然而，传统 CNN 和 ViT 模型结构在常规训练方式下效果并不理想。本文以 CNN 和 ViT 模型结构为基础，对模型结构和训练策略进行了优化，提出了一种适用于 OSR 问题的模型结构：独立分类网络开集识别(**I**ndependent **C**lassifiers for **O**pen-set **R**ecognition，ICOR)模型，以及一种新的模型训练策略：开集自适应训练。全文主要创新点有以下几个方面：

1) 传统的分类模型通常在输出端使用 Softmax 函数，其本质上将整个特征空间划分给了已知类，没有给未知类保留分类区域。本文提出了一种独立分类网络结构，在分类形式上将 M 分类任务转换为 M 个单分类任务，降低了由 Softmax

函数归一化所导致的开集识别风险。独立分类网络为未知类别提供了分类区域，并通过独立的线性层捕获更多类别特征，帮助模型学习更完备的特征。

2)分类模型在训练中通过对照其它类别的特征来确定自身的特征分布，以此构建类别间的决策边界。由于传统模型在训练过程中遵循闭集设定，这使得模型形成的决策边界具有固有的闭集特性，缺乏对未知类别的先验知识。本文提出了一种开集自适应训练策略，在训练中增加了一类未标注的开集数据集作为所有已知类数据的负样本，使类别间决策边界的划分不仅限于已知类之间的特征差异。通过对比已知类和负样本的特征差异来使模型学习更完备的已知类的特征，以此收缩已知类别的决策边界，降低模型的OSR风险。

3)ICOR 模型结构可以实现根据实际任务差异和应用需求，仅通过少量数据样本就可实现模型的训练，有效地降低了在研究过程中模型对标注图像的依赖。同时，本文不仅在多种 CNN 结构的模型上验证了本文方案的可行性，还在纯注意力机制的 Vision Transformer 模型上进行了验证。

1　本文方案

本文对传统模型的末端分类器的全连接结构进行重新设计，在网络末端使用 Sigmoid 激活函数代替传统算法中的 Softmax 激活函数，降低了由 Softmax 函数归一化机制所带来的开集识别风险，ICOR 算法的总体框架如图 1 所示：

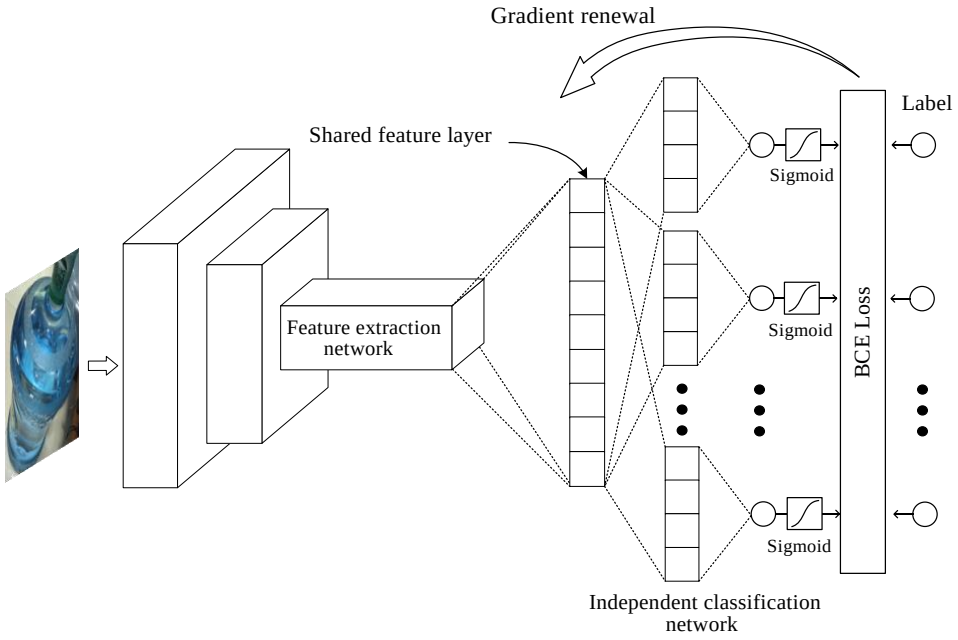


图 1　ICOR 模型总体框架

Fig. 1　Overall framework of ICOR mode

1.1　特征提取网络

现有的 OSR 模型通常需要依赖大量的数据样本来学习更完备的特征信息。然而，实际工程应用中数据样本的获取和标注往往是个难题，为了弥补这一不足，本文在对 ICOR 模型的特征提取网络(Feature extraction network)权重参数初始化时，本文采用了迁移学习的策略[19]，以提高模型的基础泛化能力。此外，ICOR 模型的特征提取网络直接迁移了传统模型的特征提取网络，这种方法减少了在不同模型框架上部署本文算法的工作量，提高了算法的灵活性。为了验证 ICOR 模型结构的普适性，在实验中，本文采用了三种具有代表性的特征提取网络：经典的 ResNet50 深度模型[20]、轻量化的 MobileNet 模型[21]和基于注意力机制模块的 ViT 模型[18]。

迁移学习策略的适用性受到许多因素的限制，包括源任务与目标任务的相似度、数据分布的差异等。因此，在进行权重参数迁移时，本文选择了与 ICOR 模型相同类型的图像分类模型，使用在 ImageNet1K 数据集（ResNet50、MobileNet 模型）[20,21]和 JFT-3B 数据集[18]（ViT 模型）上预训练的模型参数初始化 ICOR 模型的特征提取网络。输入图像经过特征提取后的对应表达式如下：

$$V_k = G(\theta_G, I_k) \tag{1}$$

式(1)中的 I_k 表示输入图像，θ_G 为特征提取网络 $G(\cdot)$ 的模型参数。对于输入图像 I_k，经过特征提取网络后得到共享特征向量 V_k (Share feature layer)。其中，$G(\cdot)$ 决定特征向量 V_k 的输出维度，θ_G 决定特征向量 V_k 的特征属性。

1.2 独立分类网络

传统的图像分类网络模型通常使用 Softmax 交叉熵函数来构建分类损失，Softmax 函数的归一化机制能加快类间特征的分离，加速模型的训练。但这也直接导致了两个问题：Softmax 函数在进行归一化时将整个特征空间划分给了已知类，没有给未知类别分配分类区域；Softmax 的归一化机制会导致模型过于关注目标类而忽略非目标类别的信息，降低了模型对非目标类的学习权重，这容易导致模型陷入局部优化，使模型构建决策边界的特征单一化，造成模型的过度拟合。

在 ICOR 模型中，本文对传统分类模型末端的全连接层结构进行了优化，引入了如图 1 所示的分离式独立分类网络结构(Independent classification network)。每个独立分类网络结构包含了若干个独立神经元节点，与共享特征层之间使用全连接结构连接，激活函数采用 GELU。同时，在输出端使用 Sigmoid 激活函数对分类器的输出进行独立归一化到 0 到 1 之间。第 i 个类别的预测概率可以表达为：

$$P(y_i \mid V_k) = \mathrm{Sigmoid}(F_i(\varphi_i \mid V_k)) \tag{2}$$

式(2)中，V_k 表示特征提取网络输出的特征向量，所有独立分类器共享该特征向量。$F_i(\cdot)$ 表示属于类别 y_i 独立分类器，其中 φ_i 为该独立分类器的模型参数。独立分类网络，将 N 分类任务转换成了 N 个单分类任务，每个独立网络层只需要独立判断输入样本属于本类别的概率，这可以使得已知类别在特征空间内的特征分布形成闭合的区域，从而实现已知类和未知类在特征空间上的分离，给未知类别在特征空间内保留了分类区域。此外，针对每一个独立分类器的预测输出，均使用二值交叉熵函数进行单独构建误差损失，对应的损失计算为：

$$L = \begin{cases} -\dfrac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{M}\log(P(y_i \mid x_k)), & t_k = 1 \\ -\dfrac{1}{N}\sum_{k=1}^{N}\sum_{i=1}^{M}\log(1 - P(y_i \mid x_k)), & t_k = 0 \end{cases} \tag{3}$$

式(3)中，N 表示迭代中的批次数量，M 表示已知类别的数量。在模型训练中，每个独立的分类网络会产生独立的损失输出。对于单个独立分类器 $F_i(\cdot)$ 而言，只有当输入样本属于第 i 类时，才被视为该类别的正样本，其余的则为负样本。因此，ICOR 模型对正负样本的学习权重是相同的，通过增加模型对负样本特征的学习，可以促使模型缩小正样本类在特征空间内的决策边界。同时，独立分类网络中的独立神经元节点，可以帮助独立分类网络捕获更多属于本类别的特征，帮助模型收缩决策边界，降低模型的开集识别风险。

1.3 开集自适应训练

神经网络的训练是一个随机优化的过程，模型根据不同任务学习到具有良好区分度的特征来完成任务需求。但在常规的监督分类任务中，模型分类特征的学习过程是不可控的，因为模型只是通过标签来训练，而标签只提供了有限的信息，并且受到数据分布和标签分配等因素的制约。因此，模型可能会拟合一些仅在闭集训练集下有利于分类的特征，这些特征在开集情况下可能并不具有泛化能力。此外，在实际工程应用中，往往无法有效收集到足够的数据样本，样本不足时模型可能会过度拟合训练集，使得其特征学习偏向于局部优化，从而忽略一些泛化性更好的特征。降低 OSR 算法在研究过程中对标注样本的依赖性，实现在少量的标签数据的情况下尽可能的收缩模型的决策边界，是 OSR 问题研究过程中的重点也是难点，对实际工程应用具有重要的实际价值。

本文提出了一种如图 2 所示的开集自适应训练策略。模型在训练时，不仅使用了已标注的样本图像，还添加了额外的负样本图像。尽管负样本图像未进行细致的类别标注，但在模型训练中仍然可以提供额外的对比特征，使模型在构建已知类的决策边界时，不局限于已知类间的特征差异，这些额外的特征差异信息可以从侧面反映已知类别的边缘特征分布。当模型用于构建决策边界的特征足够丰富时，模型在面对全新的开集数据时，紧凑的决策边界能赋予模型更好的鲁棒性。开集适应性训练的优势在于，数据样本十分易得，可以直接对大型数据集进行随机采样。

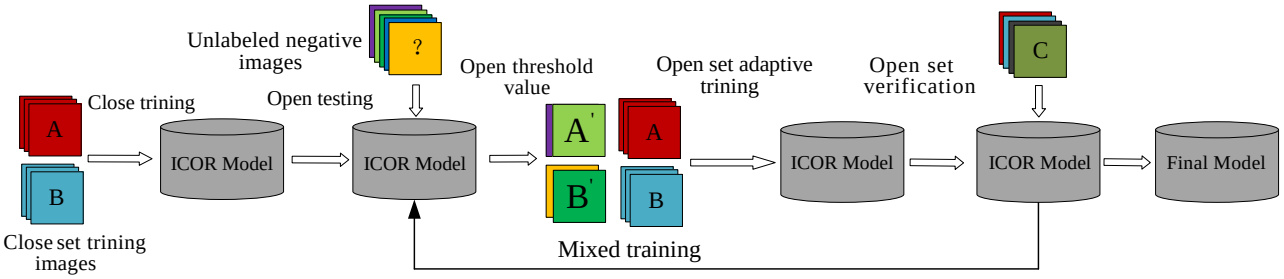


图 2　开集适应性训练

Fig. 2　Open set adaptive training

图 2 中，A 和 B 表示的是训练集中闭集类对应的图像，实际问题可能还存在其它类，本文仅以 A 和 B 类为例进行描述说明。'?'表示未经过标注的未知类别的任意图像，经过处理后 $(A,B)\notin$'?'。根据图 2 描述的开集自适应训练的流程框图，训练的步骤包括：

步骤 1：构建如图 1 所示的 ICOR 模型，加载特征提取网络的预训练权重，并在闭集条件下训练模型(Close trining)，直至模型收敛，保存 ICOR 模型。

步骤 2：使用步骤 1 中保存的模型对未标注的负样本图像'?'进行预测(Open testing)，并将预测值高于设定开集阈值的负样本图像保存。图 2 中“A'”表示图像'?'的预测值高于设定的开集阈值(Open threshold value)，且对应的预测类别为“A”。

步骤 3：将步骤 2 中保存的图像(“A'”和“B'”)与闭集图像进行混合，并用于模型的开集自适应训练(Open set adaptive trining)，直至模型收敛，并保存模型。其中，图像“A'”和“B'”对于任何一个独立分类网络的真实标签均设置为'0'，表示不属于任何闭集已知类别。

步骤 4：使用验证集图像对步骤 3 保存的模型进行验证(Open set verification)，计算模型的开集和闭集精度。其中，开集验证集‘C’中包含了已知类和未知类图像。

步骤 5：重复步骤 2 到 4，直到模型的开集和闭集精度达到预期要求后，开集自适应训练结束，保存最终模型。

在开集测试中，当输入样本的预测值小于设定的开集阈值时，被判定为开集类；反之，为闭集类。在步骤 2 中，尽管负样本图像可能被识别为已知类，但负样本图像仍然在特征空间内有其所属的分类区域，在步骤 3 中通过标签信息来不断的迭代学习，负样本图像能逐步地压缩已知类在特征空间内的区域，达到收缩已知类别决策边界的目的。本文在步骤 2 中使用步骤 1 中训练好的模型对未知类图像进行自动筛选，从而使得获取的负样本图像对模型的开集性能改善更具有针对性。

2　实验与分析

为了验证本文提出的 ICOR 模型的有效性，本节中设计了多组对比实验。实验中，模型的学习率为 0.001，学习动量为 0.5，权重衰减率为 0.00005。实验平台服务器 CPU 为 W-2133 CPU@3.60GHZ，GPU 为 GeForce GTX 1080Ti，搭载 Pytorch 深度学习框架，迭代次数根据不同的实验进行确定。

2.1　数据集介绍

正如 1.3 节所分析的，解决模型 OSR 问题的必要措施是帮助模型学习到更加完备的类别特征，以缩小模型决策边界的范围。然而，模型学习完备的特征需要覆盖某类目标在各种情况下的数据，实际应用中，这面临数据采集和标注的困难[22]。因此，研究在少量样本条件下的 OSR 问题对解决实际工程应用问题具有极大的实际价值。为此，本文针对性地收集了两个小型的数据集。

2.1.1　小型数据集 1-FDS

为了使实验更贴合实际工程应用中所面临的数据样本匮乏条件，本文制作了小型数据集(Few Data Set--FDS)，数据集的部分展示如图 3 所示：

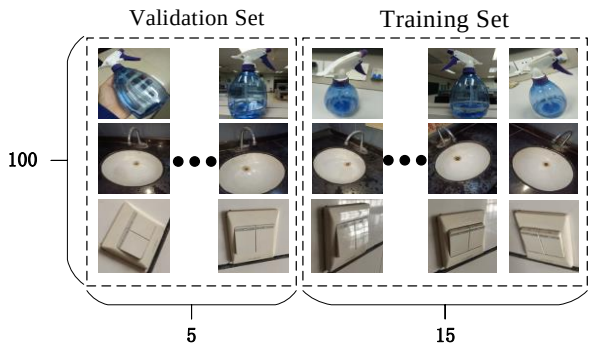


图 3　部分数据集展示

Fig. 3　Partial presentation of the data set

FDS 数据集包含 100 类实际环境中的物品，每

类物品从不同角度采集 20 张图像。实验中，按照 1:9 来随机划分闭集和开集类别，每个类别按照 1:3 随机划分训练集和验证集。同时，为了便于 ICOR 模型的开集自适应训练，本文收集了一个不做类别标注的未知类数据集(Unknown Data Set--UDS)，如图 4 所示，一共包含 5000 张图像，涵盖各类生活用品、服饰、地标等。其中，UDS 数据集与 FDS 数据集之间在类别上不存在交集。

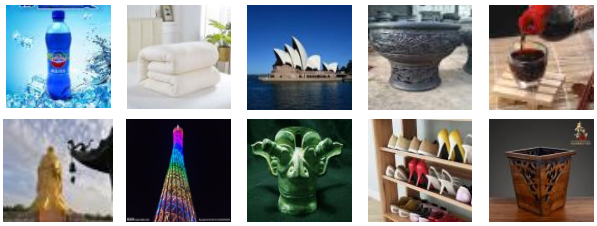


图 4　未知类数据集

Fig. 4　Unknown data set

2.1.2　数据集 2-Imagenet-Crop

FDS 数据集图像中物品的背景较为纯净，背景信息对模型训练的干扰较小。然而，在某些较为复杂的情况下，数据集图像中物品的背景往往复杂多样，背景信息会对模型的训练造成极大的影响。为此，本文还基于 Imagenet1k 数据集制作了一个比 FDS 数据更加复杂的数据集，以增加实验的对比性。如图 5 所示，Imagenet1k 数据集包含了 1000 个物品类别，涵盖了日常生活用品、交通工具、乐器等。为了满足本文设定的少量数据样本的实验条件，每次试验随机从 Imagenet1K 的 1000 个类别中选取了 10 个类别作为闭集类别用于模型的闭集训练，在剩余的 990 类中，划分 400 类作为开集类别用于模型开集测试，590 类作为未知类别，该数据集命名为 Imagenet-Crop。

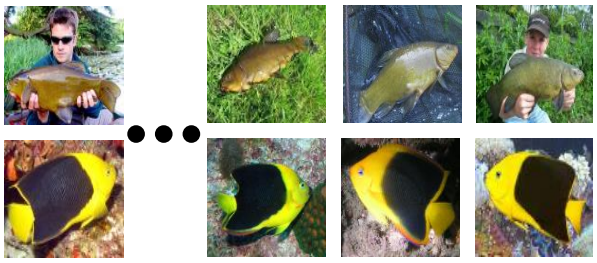


图 5　Imagenet-Crop 数据集

Fig. 5　Imagenet-Crop data set

由于本文聚焦于研究模型的 OSR 性能，在实验中适当的增加了 Imagenet-Crop 数据集中闭集图像的数量，以保证模型的闭集精度。实验中，在闭集类别对应的训练集内随机抽取 100 张图像，按照 2:8 的比例划分训练集和测试集。开集类和未知类选取的图像均是对应类别中 Imagenet1K 数据集的测试集图像，其中开集类包括 20000 张图像，未知类包括 29500 张图像。

2.2　实验设计

为了评估 ICOR 模型和开集自适应训练策略的有效性、泛化性及鲁棒性能。本文设计了以经典 Resnet50(Res50)模型、轻量化 MobilenetV2(MobV2)模型以及 Vision Transformer(ViT)等三种特征提取网络的 ICOR 模型，并设计同等实验条件下关键参数的消融实验、模型开放度实验以及与现有 OSR 算法的对比实验。

2.3　消融实验

为了分别测试 ICOR 模型结构和开集自适应训练对模型 OSR 性能改善的有效性，本节设计了消融实验进行对比分析。实验中，在 FDS 数据集和 Imagenet-Crop 数据集的闭集训练集上进行模型训练，在开集类别上进行开集性能测试，多次实验统计取平均值。

2.3.1　独立分类网络对模型开集性能的影响

在 OSR 问题当中，通常采用 AUROC 来测评模型的 OSR 性能[14–17]，然而 AUROC 值只考虑到模型分辨对已知类和未知类的能力，并未考虑到模型对已知类的正确分类能力，这使得 AUROC 对模型的 OSR 性能的评估并不直观。本文采用“开集精度(Open Accuracy-OA)”来评估模型识别已知类和未知类的能力，并通过“闭集精度(Close Accuracy-CA)”来评估模型对闭集类别的正确分类能力。对于 OA 和 CA 的定义为：

$$\begin{cases} OA = \dfrac{\sum_{i=1}^{P_O} \mathrm{II}(y_i^* < od)}{P_O} \\ CA = \dfrac{\sum_{i=1}^{P_C} \mathrm{II}(y_i^* >= od \ \&\ y_i^* = cls_i)}{P_C} \end{cases} \tag{4}$$

式(4)中，P_O 与 P_C 分别为测试集中开集和闭集样本的总数；$y_i^* = \arg\max(y_1^*, y_2^*, \ldots y_M^*)$，即对于输入样本 x_i，y^*_i 表示 M 个独立分类网络中对应预测输出的最大值。II 为指示函数，当逻辑为真，函数结果为 1。cls_i 为输入样本 x_i 的正确标签。od 为开集拒判阈值，由于 ICOR 模型中每一个独立分类网络独通过 Sigmoid 函数激活输出，可以看作是逻辑回归的二分类任务，故设定开集拒判阈值为 0.5[23]。

为了平衡模型性能和计算资源的消耗，本文通过多次实验的先验知识确定每个独立分类网络的神经元节点个数为共享特征层的十分之一。在 FDS

和 Imagenet-Crop 数据集上，训练次数为 200 次，采取解冻训练模式，即不对网络层进行冻结，训练所有模型的所有网络层。实验结果如表 1 所示：

表 1　开集和闭集精度对比

Tab.1　Comparison of open set and closed set

Model order	Model name	FDS		Imagenet-Crop	
		Closed	Open	Close	Open
1	Res50+Softmax	0.999	0.474	0.999	0.471
2	Res50+Sigmoid	0.996	0.824	0.984	0.744
3	**Res50+ICOR**	**0.995**	**0.940**	**0.971**	**0.882**
4	MobV2+Softmax	0.982	0.399	0.998	0.294
5	MobV2+Sigmoid	0.980	0.707	0.994	0.692
6	**MobV2+ICOR**	**0.986**	**0.971**	**0.998**	**0.936**
7	ViT+Softmax	0.999	0.345	0.999	0.331
8	ViT+Sigmoid	0.994	0.826	0.995	0.437
9	**ViT+ICOR**	**0.999**	**0.890**	**0.999**	**0.792**

表 1 中，1、4、7 三组使用 Softmax 函数的模型在闭集条件下模型具有良好的识别精度，表明模型已经成功收敛。然而，这三组模型的最高 OSR 精度不超过 0.474(1 组)，特别是在相对复杂的 Imagenet-Crop 数据集上，同组实验模型的 OSR 风险更大。此外，从 Sigmoid 组模型对应的表 1 中的 2、5、8 组实验结果可以看出，在 FDS 数据集上，这三组模型的 OSR 精度分别提高至 0.824、0.707 和 0.826，表明 Sigmoid 函数对模型的 OSR 性能有所提高。然而，在 Imagenet-Crop 数据集上，这些模型仍然表现出极高的 OSR 风险。例如，第 8 组的 OSR 精度仅为 0.437，相对于 Softmax 组的 0.331 仅有 10%的提升。

表 1 中的 3、6、9 组对比实验结果表明，在两个数据集下，ICOR 模型均实现了最高的 OSR 精度。在 Imagenet-Crop 数据集上，最低的 OSR 精度也达到了 0.792(9 组)，高于 Sigmoid 组最高的 0.744(2 组)。此外，通过比较 Softmax 和 Sigmoid 组的闭集精度可以发现，使用 Sigmoid 函数可以提高模型的 OSR 精度，但同时也可能降低模型的闭集识别精度。而本文提出的 ICOR 模型在同样的实验条件下，既保证了较高的闭集精度，同时对模型的 OSR 性能具有良好的改善效果。

2.3.2　开集自适应训练对模型开集性能的影响

为了评估开集自适应训练策略对模型 OSR 性能的影响，本文在三组模型的 Sigmoid 和 ICOR 组的闭集训练基础上进行了开集自适应训练。具体而言，开集自适应训练的过程如图 2 所示，在开集测试时，选择模型在闭集验证集中有 95%以上样本被分类正确的预测置信度作为开集阈值。模型训练分为两个阶段：首先进行 100 次的闭集迭代训练，然后进行 100 次的交替闭集和开集自适应训练。实验结果如表 2 所示：

表 2　开集自适应训练试验对比

Tab.2　Comparison of Open Set Adaptive Training

Model order	Model name	FDS		Imagenet-Crop	
		Close	Open	Close	Open
1	Res50+Sigmoid	0.996	0.824	0.984	0.744
2	**Res50+Sigmoid(K+)**	**0.999**	**0.949**	**0.995**	**0.969**
3	Res50+ICOR	0.995	0.940	0.971	0.882
4	**Res50+ICOR(K+)**	**0.996**	**0.988**	**0.968**	**0.992**
5	MobV2+Sigmoid	0.979	0.707	0.994	0.692
6	**MobV2+Sigmoid(K+)**	**0.980**	**0.940**	**0.951**	**0.972**
7	MobV2+ICOR	0.986	0.971	0.998	0.936
8	**MobV2+ICOR(K+)**	**0.999**	**0.988**	**0.978**	**0.989**
9	ViT+Sigmoid	0.993	0.826	0.999	0.437
10	**ViT+Sigmoid(K+)**	**0.995**	**0.959**	**0.988**	**0.933**
11	ViT+ICOR	0.999	0.890	0.999	0.792
12	**ViT+ICOR(K+)**	**0.994**	**0.981**	**0.993**	**0.974**

表 2 中，K+表示模型经过开集自适应训练后的结果。2、6、10 组的对比实验结果表明，在经过开集自适应训练后，Sigmoid 和 ICOR 组模型在两个数据集上的 OSR 精度均提升至 0.9 以上。因此，可以证明开集自适应训练方式能够有效提高模型的 OSR 性能。此外，2、4、6、8、10、12 组实验结果表明，在相同的特征提取网络下，ICOR 模型相较于直接使用 Sigmoid 方法更具优势，能更有效地提升模型的开集性能。这也表明独立分类网络的设计能够更准确地捕获类别特征，使得模型形成更加收缩的决策边界。

2.4　开放度测试实验

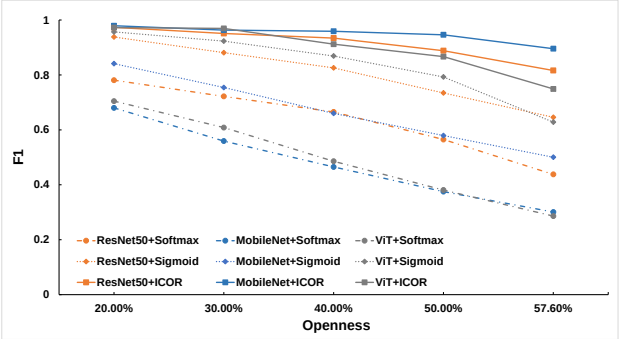
在实际应用环境中，如果场景足够复杂，模型出现开集识别问题是不可避免的。为了从有限的未知类别中估计模型在更加开放的环境中的开集识别性能的变化趋势，以评价模型的鲁棒性，本文设计了开放度测试实验来进一步评估模型的开集性能。数据集的开放度(openness)[1]是指在测试过程中出现未知类别的比例，其定义如式(5):

$$\text{openness}=1-\sqrt{\frac{2C_T}{C_E+C_R}} \tag{5}$$

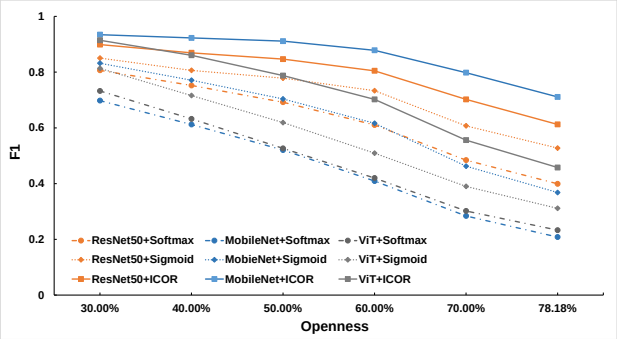
式子(5)中，C_T 表示在训练中的已知类样本类别数，C_E 表示验证集中包含的已知类的类别数量，C_R 表示验证集中所有的类别数，即已知类别和未知类别数之和。本文的实验中，$C_E=C_T$，即验证集

包含了所有训练类别。根据开放度定义，FDS 和 Imagenet-Crop 数据集对应的最大开放度分别为 0.574 和 0.782。

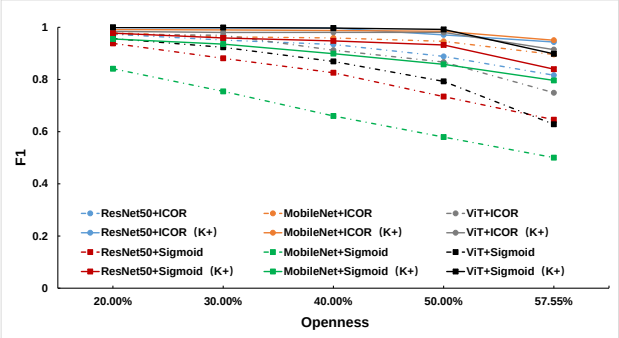
F1 值结合了准确率和召回率两个指标，常被用于评价二分类任务模型的综合性能，开集识别问题可以被视为一个简化的二分类任务[23,24]，即模型是否准确识别已知类和未知类。多次试验，记录模型在不同开放度条件下的 F1 分数的变化趋势，实验结果如图 6 和图 7 所示：



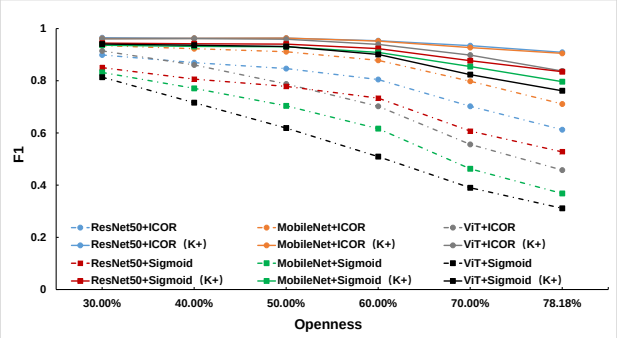
(a) Closed set openness experiment (FDS)



(b) Closed set openness experiment (Imagenet-Crop)



(c) Open adaptive training openness experiment (FDS)



(d) Open adaptive training openness experiment (Imagenet-Crop)

图 6 模型开放度实验结果

Fig. 6　Experimental results of model openness

图 6 中，(a)和(b)展示了三组模型闭集条件下，在 FDS 和 Imagenet-Crop 数据集上进行开放度实验的结果。实验结果表明，随着数据集开放度的增加，各组模型的综合 F1 分数逐渐降低，这与预期实验趋势一致。开放度的增加，意味着模型在测试过程中需要检测更多未知类别。其中，Sigmoid 组的折线图位于 Softmax 组上方，这表明使用 Sigmoid 比使用 Softmax 具有更优的开集识别性能，但随着开放度的增加，Sigmoid 组的 F1 分数迅速下降，表明 Sigmoid 仅在低开放度条件下表现出良好的鲁棒性。相比之下，ICOR 模型的折线位于最上方，并且随着开放度的增加，F1 分数下降的速度要较为缓慢，表明相对于其它两组模型，ICOR 模型具有更优的开集鲁棒性。

在图 6 中，(c)和图(d)展示了三组模型经过开集自适应训练后在 FDS 和 Imagenet-Crop 数据集上进行开放度实验的结果。在相同的开放度条件下，ICOR 组的综合 F1 分数高于 Sigmoid 组。Sigmoid 组的折线在开放度较小时呈水平状态，但在开放度大于 50%和 60%时，折线的斜率急剧降低，模型的 F1 分数急剧下降。这表明 Sigmoid 组在开集适应性训练后，在低开放度下能够保持较高的开集精度，但随着开放度的增加，仍然存在严重的开集识别风险。实验中，ICOR 组的折线斜率始终保持着缓慢的下降趋势，模型表现出更优的鲁棒性。这也证明了，本文提出的开集自适应训练对模型开集识别性能具有更好的改善效果。

2.5 算法对比实验

目前，在解决 OSR 的问题上已经有了一些优秀的算法，为了验证 ICOR 算法对比其它 OSR 算法(AMPFL[14]、SLCPL[15]、ARPL[16]、GCPL[17])在少量标注数据样本条件下对现有模型 OSR 性能的改善的优势，本文在 FDS 和 Imagenet-Crop 数据集上进行了对比实验。为了公平起见，所有算法使用同一特征提取网络，并使用迁移自 Imagenet1K 预训练模型的参数进行多次随机试验。现有算法的迭代次数均为 200 次，实验结果如表 3 所示：

表 3　ICOR 及开集适应性训练与现有 OSR 算法性能对比试验

Tab.3　Comparing the performance of ICOR and open set adaptive training with existing OSR algorithms

Model name		Model order	FDS		Imagenet-Crop		Model name		Model order	FDS		Imagenet-Crop	
			Close	Open	Close	Open				Close	Open	Close	Open
	Res50	1	0.999	0.474	0.999	0.471		Res50	13	0.992	0.334	0.999	0.610
Softmax	MobV2	2	0.982	0.399	0.998	0.294	ARPL	MobV2	14	0.970	0.752	0.986	0.547
	ViT	3	0.999	0.345	0.999	0.331		ViT	15	0.999	0.648	0.978	0.462
	Res50	4	0.996	0.824	0.984	0.744		Res50	16	0.999	0.395	0.999	0.674
Sigmoid	MobV2	5	0.980	0.707	0.994	0.692	GCPL	MobV2	17	0.962	0.721	0.994	0.527
	ViT	6	0.994	0.826	0.999	0.437		ViT	18	0.998	0.624	0.998	0.506
	Res50	7	0.999	0.321	0.998	0.558		**Res50**	**19**	**0.995**	**0.940**	**0.971**	**0.882**
AMPFL	MobV2	8	0.909	0.727	0.995	0.511	**ICOR**	**MobV2**	**20**	**0.994**	**0.928**	**0.997**	**0.898**
	ViT	9	0.999	0.638	0.999	0.476		**ViT**	**21**	**0.999**	**0.890**	**0.999**	**0.792**
	Res50	10	0.992	0.377	0.998	0.587		**Res50**	**22**	**0.996**	**0.988**	**0.968**	**0.991**
SLCPL	MobV2	11	0.962	0.781	0.989	0.493	**ICOR(K+)**	**MobV2**	**23**	**0.997**	**0.976**	**0.979**	**0.982**
	ViT	12	0.999	0.749	0.996	0.340		**ViT**	**24**	**0.995**	**0.981**	**0.993**	**0.974**

表 3 中 7 到 18 组的实验结果表明，针对 OSR 问题，AMPFL、SLCPL、ARPL 和 GCPL 等现有算法对模型的 OSR 性能改善有限，特别是在 ResNet50 特征提取网络框架下(第 7、10、13 和 16 组)，尽管这些算法保持着较高的闭集精度，但存在极大的 OSR 风险，甚至高于传统的 softmax 和 sigmoid(第 1、4 组)。此外，ICOR 算法在同等特征提取网络和实验条件下，在未进行开集自适应训练时(第 19、20、21 组)，实现了试验中最高的 OSR 精度；在进行开集自适应训练后，ICOR 模型的 OSR 精度均在 0.97 以上，相比于现有 OSR 算法，ICOR(K+)在 FDS 和 Imagenet-Crop 数据集上的表现(22、23、24 组)均有明显的优势。

同时，AMPFL、SLCPL、ARPL、GCPL 等 OSR 算法，随着特征提取网络的改变，模型的开集识别精度也会发生明显变化，这表明特征提取网络的类型对传统开集识别算法的开集识别精度具有很大影响，而本文算法随着特征提取网络的变化，开集识别精度却一直十分稳定，表现出了更好的鲁棒性。综上所示，ICOR 算法和开集自适应训练策略能更有效地降低模型在数据样本匮乏条件下的 OSR 风险。

3　结论

本文针对传统模型存在的 OSR 问题，设计了 ICOR 模型和开集自适应训练策略，通过实验可以得出以下结论：

1)本文对 ICOR 模型和开集自适应训练策略进行了消融实验，实验证明，ICOR 模型结构和开集自适应训练策略均能有效改善传统模型 OSR 性能。

2)本文设计了开放度实验，实验证明，本文的 ICOR 模型及开集自适应训练策略随着开放度的增加，表现出更高的鲁棒性能。

3)相比于本文测试的其它现有 OSR 算法，ICOR 模型和开集自适应训练具有更好的普适性，同时还能更有效地降低模型的 OSR 风险。

参考文献

[1] Scheirer W J， Anderson D R R， Sapkota A，et al. Toward Open Set Recognition[J]. Pattern Analysis & Machine Intelligence IEEE Transactions on, 2013, 35(7):1757-1772.[J].

[2] Ming Y， Wegkamp M H . Classification Methods with Reject Option Based on Convex Risk Minimization[J]. Journal of Machine Learning Research, 2010, 11(1):111-130.[J].

[3] Geng C， Huang S J， Chen S . Recent Advances in Open Set Recognition: A Survey[J]. IEEE Trans Pattern

Anal Mach Intell, 2021(10).[J].

[4] Mahdavi A， Carvalho M . A Survey on Open Set Recognition[J]. 2021.[J].

[5] Cevikalp H， Triggs B， Franc V . Face and Landmark Detection by Using Cascades of Classifiers[J]. IEEE, 2013.[J].

[6] Zhang H， Patel V M . Sparse Representation-Based Open Set Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, PP(8):1-1.[J].

[7] Bendale A， Boult T . Towards Open Set Deep Networks[J]. IEEE, 2016.[J].

[8] Shu L， Xu H， Liu B . DOC: Deep Open Classification of Text Documents[J]. 2017.[J].

[9] Ge Z Y， Demyanov S， Chen Z，et al. Generative OpenMax for Multi-Class Open Set Classification:, 10.5244/C.31.42[P]. 2017.[J].

[10] Neal L, Olson M, Fern X, et al. Open set learning with counterfactual images[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 613-628.[J].

[11] Zhang Y， Lee K， Lee H . Augmenting Supervised Neural Networks with Unsupervised Objectives for Large-scale Image Classification[J]. JMLR.org, 2016.[J].

[12] Rasmus A， Valpola H， Honkala M，et al. Semi-Supervised Learning with Ladder Networks[J]. Computer Science, 2015, 9 Suppl 1(1):1-9.[J].

[13] Kuncheva L I， Bezdek J C . Nearest prototype classification: clustering, genetic algorithms, or random search [J]. IEEE Press, 1998.[J].

[14] Xia Z， Wang P， Dong G，et al. Adversarial Motorial Prototype Framework for Open Set Recognition[J]. 2021.[J].

[15] Xia Z， Dong G， Wang P，et al. Spatial Location Constraint Prototype Loss for Open Set Recognition[J]. 2021.[J].

[16] Chen G， Peng P， Wang X，et al. Adversarial Reciprocal Points Learning for Open Set Recognition[J]. 2021.[J].

[17] Yang H M， Zhang X Y， Yin F，et al. Convolutional Prototype Network for Open Set Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, PP(99):1-1.[J].

[18] Dosovitskiy A， Beyer L， Kolesnikov A，et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale[J]. 2020.[J].

[19] MD Zeiler, R Fergus. Visualizing and Understanding Convolutional Networks[C]// ECCV 2014. 2014.[J].

[20] He K， Zhang X， Ren S，et al. Deep Residual Learning for Image Recognition[J]. IEEE, 2016.[J].

[21] Howard A G， Zhu M， Chen B，et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications[J]. 2017.[J].

[22] 杨丰萍,彭云帆,李远征.基于深度学习的小样本绝缘子自爆检测研究[J].华东交通大学学报,2022,39(02):110-117.DOI:10.16749/j.cnki.jecjtu.2022 0314.010.

YANG F P，PENG Y F，LI Y Z . Research on insulator self -explosion detection with small sample based on deep learning[J]. Journal of East China Jiaotong University，2022，39（2）：110-117.[J].

[23] Sun X， Yang Z， Zhang C，et al. Conditional Gaussian Distribution Learning for Open Set Recognition[J]. 2020.[J].

[24] Vendramini M， Oliveira H， Machado A，et al. Opening Deep Neural Networks with Generative Models[J]. 2021.[J].

第一作者：徐雪松（1970—），男，教授，博士，硕士生导师，研究方向为移动机器人视觉导航与控制、无人机控制、计算机视觉、模式识别。E-mail：cedarxu@163.com

通信作者：付瑜彬（1996—），男，硕士研究生，研究方向为深度学习、开集识别。E-mail：1668875496@qq.com