# 基于坐标重要性池化和解耦类别对齐蒸馏的图像分类算法

刘 颖[1,2],薛家昊[1*],张伟东[1,2],许志杰[2,3]
(1. 西安邮电大学图像与信息处理研究所,陕西西安 710121;2. 无线通信与信息处理技术国际联合研究中心,陕西西安 710121;
3. 英国哈德斯菲尔德大学,西约克郡 HD13DH)

**摘 要:** 为提高卷积神经网络图像分类精度的同时实现网络轻量化,本文提出一种基于坐标重要性池化和解耦类别对齐蒸馏的图像分类算法. 首先,设计一种坐标重要性池化模块并将其嵌入ResNet34,充分利用图像像素的位置信息,以增强其判别重要性特征的能力;其次,采用BlurPool缓解在下采样过程中移位等变性丢失对网络性能的影响,以此构建教师网络;最后,构造一种解耦类别对齐蒸馏算法,分别考虑目标类和非目标类的知识并引入类别之间的关联信息,以高效地将分类知识从教师网络迁移到轻量级MobileNetV3学生网络. 在不同数据集上的实验结果表明,本文提出的教师网络有效提高了分类性能,且蒸馏训练后的学生网络明显优于其他同量级网络,实现了更优越的综合性能,能够更好地应用于计算和内存资源受限的实际场景.

**关键词:** 图像分类;轻量化;知识蒸馏;ResNet34;坐标重要性池化
**基金项目:** 国家自然科学基金(No.62106195)
**中图分类号:** TP391 **文献标识码:** A **文章编号:** 0372-2112(2025)03-0962-12
**电子学报URL:**http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20240754

# Image Classification Algorithm Based on Coordinate Importance Pooling and Decoupled Class Alignment Distillation

LIU Ying[1,2], XUE Jia-hao[1*], ZHANG Wei-dong[1,2], XU Zhi-jie[2,3]
(1. *Center for Image and Information Processing*, *Xi'an University of Posts and Telecommunications*, *Xi'an*, *Shaanxi* 710121, *China*;
2. *International Joint-Research Center for Wireless Communication and Information Processing*, *Xi'an*, *Shaanxi* 710121, *China*;
3. *University of Huddersfield*, *West Yorkshire* HD13DH, *United Kingdom of Great Britain and Northern Ireland*)

**Abstract:** An image classification algorithm based on coordinate importance pooling and decoupled class alignment distillation is proposed to improve the image classification accuracy of convolutional neural networks while achieving network lightweighting. Firstly, a coordinate importance pooling module is designed and embedded it into ResNet34, in order to fully utilize the positional information of image pixels to enhance the ability to discriminate important features. Secondly, BlurPool is used to mitigate the impact on network performance due to shift equivariance during down-sampling, and to construct the teacher network. Finally, the decoupled class alignment distillation algorithm was constructed to efficiently migrate image classification knowledge from the teacher network to the lightweight MobileNetV3 network, which considers the knowledge of target and non-target class separately and introduces correlation information between the class. The experimental results on different datasets showed that the proposed teacher network effectively improves the classification performance, and the distillation-trained student network achieves superior overall performance than other networks of the same magnitude, making it better applicable to practical scenarios with limited computational and storage power.

**Key words:** image classification; lightweight; knowledge distillation; ResNet34; coordinate importance pooling
**Foundation Item(s):** National Natural Science Foundation of China (No.62106195)

收稿日期:2024-08-13;修回日期:2025-01-13;责任编辑:朱梅玉
*通讯作者:薛家昊

## 1　引言

作为计算机视觉的一个重要研究领域,图像分类的应用十分广泛,已覆盖医疗、安全、农业等与人们日常生活相关的诸多方面. 卷积神经网络(Convolutional Neural Network,CNN)作为深度学习的突出技术之一,具有自主学习特征、分类效果好等优势,能够克服传统方法的许多限制,在图像分类中得到了广泛应用[1,2].

随着神经网络研究的深入和互联网时代的到来,学者们发现,以往通过加深网络来提高分类性能的方法会导致参数量和计算量急剧增加. 例如,一个152层的ResNet中有超过6 000万个参数[3],这导致其在一些对即时性要求较高的应用场景中难以给出实时分类结果,而且在资源有限的设备上部署也会受到限制[4]. 为此,研究人员提出一系列解决方法,如轻量化网络设计[5~10]、神经网络架构搜索[11~13]、模型压缩[14~16]等,通过将复杂网络轻量化使其得到更广泛的应用. 作为模型压缩的代表方法,知识蒸馏(Knowledge Distillation,KD)因强大的实用性而备受关注,该方法可以将复杂大型网络的知识转移到小型轻量级网络中,使轻量级网络的性能逼近大型网络,同时大幅降低对资源的需求[17].

KD的概念最早由Buciluǎ等人[18]提出,通过训练带有伪数据标记的强分类器的压缩模型再现原始较大模型的输出,后由Hinton等人[19]推广,引起学术界的广泛关注. KD的重点是通过教师-学生(teacher-student)网络来实现知识迁移的过程,而教师网络作为知识的源头和指导学生网络训练的角色,其具有强大的泛化能力,直接影响KD的效果和学生网络的性能提升. 刘立波等人[20]提出在教师网络中的每组Block后与SimAM注意力机制相结合,以进一步提高特征提取能力. 李大湘等人[21]将通道注意力和空间注意力相结合,构造一个新的双注意力模块,以将教师网络中的"注意力知识"迁移到学生网络中. Li等人[22]设计一种辅助分类器来捕获跨层语义信息,帮助网络学习更丰富的知识. 上述方法虽然可以增强教师网络的特征学习能力,但它们均忽略了池化过程中存在潜在的信息损失,若出现特征采样不精细和信息丢失的问题,会对分类结果造成影响. 研究表明,当使用强大的教师网络时,即当教师网络与学生网络之间有较大差距时,蒸馏后的性能可能会下降[23,24]. 因此,后来的研究工作尝试改进蒸馏算法以改善知识迁移的有效性,如多阶段蒸馏[25]、知识编码[26]、对比学习表征[27]、加入引导策略[28]等,通过考虑样本之间的关联信息选择性迁移重要知识,剔除冗余信息,这样虽然能够有效提高蒸馏效果,进而提升学生网络的性能,但其忽略了不同类型的知识在传递时存在相互影响,在知识迁移的灵活性方面存在限制,且未利用类别相关性信息进一步蒸馏知识,降低了学习效果.

针对上述问题,本文综合设计了坐标重要性池化模块(Coordinate Importance Pooling,CIP),并结合BlurPool采样策略对教师网络ResNet34进行改进. 同时,通过构造解耦类别对齐蒸馏(Decoupled Class Alignment Distillation,DCAD)算法进行知识迁移,以指导学生网络获取更丰富的图像信息,从而更准确地进行图像分类. 综上所述,本文主要贡献如下:

(1)针对传统网络对图像特征采样不精细、存在丢失重要信息的问题,设计了坐标重要性池化模块代替教师网络的池化层. 该模块通过图像每个像素的坐标信息生成特征权重,并通过一个固定的放大系数动态调节重要性权重,以确定重要特征,进而提高模型的特征判别能力. 引入BlurPool作为网络的子采样方式,缓解移位等变性的丢失对网络性能的影响,以此构建改进后的教师网络E-ResNet34.

(2)设计DCAD算法,将蒸馏过程分为目标类和非目标类知识蒸馏,并引入类别之间的关联信息进行类别对齐,进一步增强知识迁移的高效性和灵活性,提高分类准确性.

(3)在Caltech101、BIRDS 525 SPECIES和CIIP-CSID数据集上使用本文算法进行图像分类,并与其他算法进行对比,以验证本文算法的有效性.

## 2　相关研究

### 2.1　池化层

在深度学习迅速发展的推动下,尤其是CNN在图像识别和处理任务中的成功应用,池化层已成为神经网络模型中的一个关键组成部分,其能够在保留特征图主要信息的同时减小特征图参数量,使模型在分类时更关注关键特征,还能防止过拟合以提高网络泛化能力[29]. 在许多经典的神经网络架构中,池化层的使用已形成标准化模式,如最大池化和平均池化. 此外,研究人员提出多种实现池化层的方法,旨在优化其在图像分类任务中的应用.

空间金字塔池化[30]将图像按照从较细到较粗的级别划分为多个部分,以聚合来自不同尺度的局部特征. Xu等人[31]提出一种边缘感知池化模块,以保留更多的边缘结构信息. Wijaya等人[32]提出可学习池化方法,根据特征与可学习参数之间的相关性进行信息聚合,以最小的信息损失产生具有代表性的特征. Zhao等人[33]提出T-Max-Avg池化方法,引入$K$个具有最高表示能力的像素,并结合一个可学习参数$T$,使其能够根据关键特征信息计算像素的最大值和平均值,以有效捕获和表示输入数据中的判别性信息,从而提高分类性能. 与以往研究工作不同,本文利用图像每个像素的坐标信

息来生成重要性权重，并通过固定系数动态调节输入特征的重要性权重以自适应地确定重要特征，旨在克服传统池化中产生的潜在信息损失，提升判别性特征的提取能力和分类效果.

### 2.2 知识蒸馏

KD作为一种有效的网络轻量化方法，因其不需要修改网络结构和参数而备受研究者关注. 通过将复杂教师网络中的知识提取到较小的学生网络中，使它们在性能上更加接近. 在此背景下，Kim等人[25]从教师网络提取多阶段知识，即不同层级的知识，以引导学生网络生成与教师网络类似的最终表示；Li等人[26]将每个样本上的知识价值编码为一个隐变量，并以此建立一个期望最大化框架交替执行教师知识集的浓缩和学生网络的蒸馏，这样学生网络就能逐渐获得更精细、更关键的知识表征；Sharma等人[27]在教师-学生网络之间应用对比学习的方式使学生网络学习到更接近教师网络的判别特征，以缩小学生和教师之间的能力差距；Zhang等人[28]提出Top-$K$引导策略选择性迁移教师网络的知识，抑制教师网络输出中可能存在的不确定性问题或错误，以确保将更可靠的知识转移到学生网络中. 虽然KD方法取得了较好的成果，可在教师网络指导下提升学生网络的性能，但其并未考虑到不同类型知识之间的耦合关系会影响知识迁移，且忽略了类别相关性这部分知识的重要性，降低了蒸馏效果. 本文引入解耦蒸馏的方法对不同类型的知识进行解耦处理，并提出在蒸馏时进行类别对齐，以更好地理解和吸收类别间的相关性，减小教师和学生在类别上的预测差异，从而更好地应用于图像分类任务.

## 3 本文算法描述

由图1可知，整体算法结构主要由3个部分组成，即大型教师网络、轻量级学生网络与蒸馏函数，旨在通过改进教师网络来提供更丰富且准确的知识，并使用设计的DCAD函数将教师网络中的知识高效迁移到学生网络中，使其在参数量较小的情况下性能接近教师网络. 选取经典的卷积神经网络ResNet34并对其进行改进以作为KD的教师网络，学生网络则选择更适用于有限资源的轻量级网络MobileNetV3. 其中，ResNet34主要由16个残差模块(Basicblock)组成，每块残差结构包含2个3×3卷积层，且这些残差模块以3∶4∶6∶3的比例排列分布在4个不同的阶段，使网络可以更加容易地学习恒等映射，从而避免梯度消失和梯度爆炸的问题，有助于教师网络生成更高质量的知识. MobileNetV3采用深度可分离卷积方法和具有线性瓶颈的残差块(Bottleneck)，每个瓶颈块均包含深度卷积、逐点卷积、激活函数，同时在一些瓶颈块中引入SE(Squeeze-and-Excitation)注意力机制以增强特征表示.

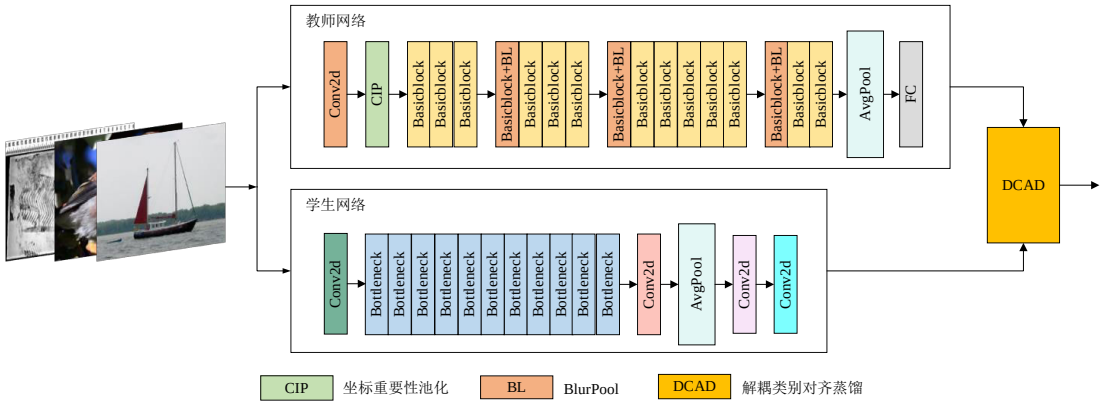



图1 整体算法结构

### 3.1 教师网络构建

目前，CNN通常利用池化层来减小特征映射以减少内存消耗，并提取区域中最重要的特征信息. 池化通常存在以下问题：(1)易丢失一些重要的判别性信息，如纹理、边缘信息等；(2)网络模型中的移位等变性可能会丢失. 因此，本节提出一种坐标重要性池化模块以强化判别性特征提取能力，并引入BlurPool来缓解移位等变性丢失对网络性能的影响，以此来改进并构建教师网络，从而提高分类性能.

#### 3.1.1 坐标重要性池化模块

针对常见的池化(如最大池化、平均池化)可能会丢失重要特征的问题，Gao等人[34]提出一种根据特征的重要性进行池化的方法，以便更好地保留重要信息. 相关研究表明，学习池化的特征是有效的方法[35]，在关注特征重要性的同时也不能忽略位置信息，在图像分类任务中位置信息对于捕获目标特征至关重要. 基于此，设计坐标重要性池化模块(CIP)，将位置信息应用到学习重要性特征中，以自动增强判别特征.

该模块可分为2个步骤:(1)在水平和垂直2个空间范围内对输入图像进行特征提取,得到每个输入的水平和垂直特征图,并通过logit模块中的卷积自适应生成特征权重,然后将两者在局部卷积窗口内进行归一化操作;(2)通过固定的放大系数动态调节重要权重,自适应地确定重要特征,以提高模型的特征判别能力. 该模块具体结构如图2所示.

对于输入特征图$I$,该模块首先通过所设计的坐标模块(CoordModule)获得图像中每个像素的坐标信息,分别包含水平方向$x$和垂直方向$y$上的坐标信息,CoordModule结构如图3所示. 通过在普通卷积的基础上添加2个通道(分别表示每个像素的$x$、$y$坐标),将坐标信息与卷积后的特征图相结合,为卷积操作提供位置信息,使其在后续对图像进行局部运算时增加图像局部信息与整体信息的关联性. 在这个过程中,2个坐标经历了线性变换,并在[-1,1]内进行归一化处理.

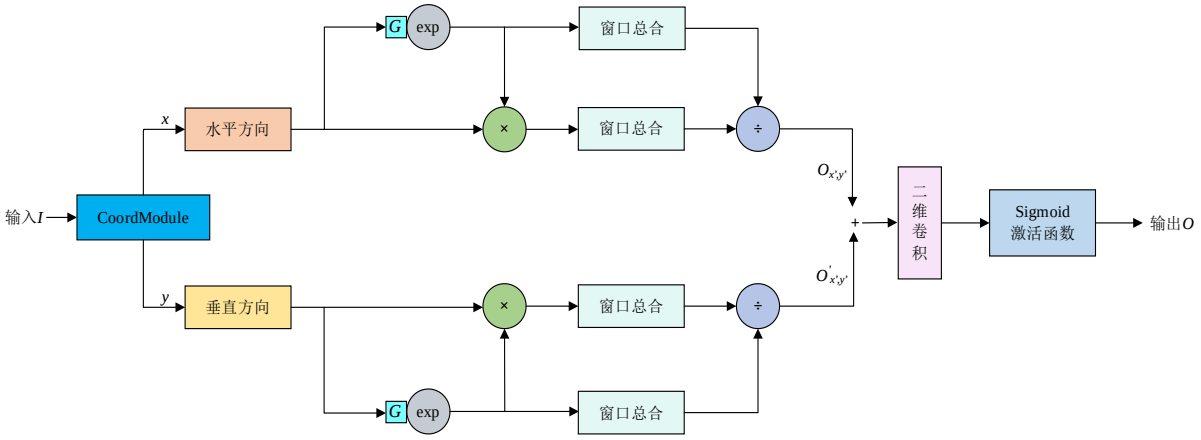



图2　坐标重要性池化模块结构

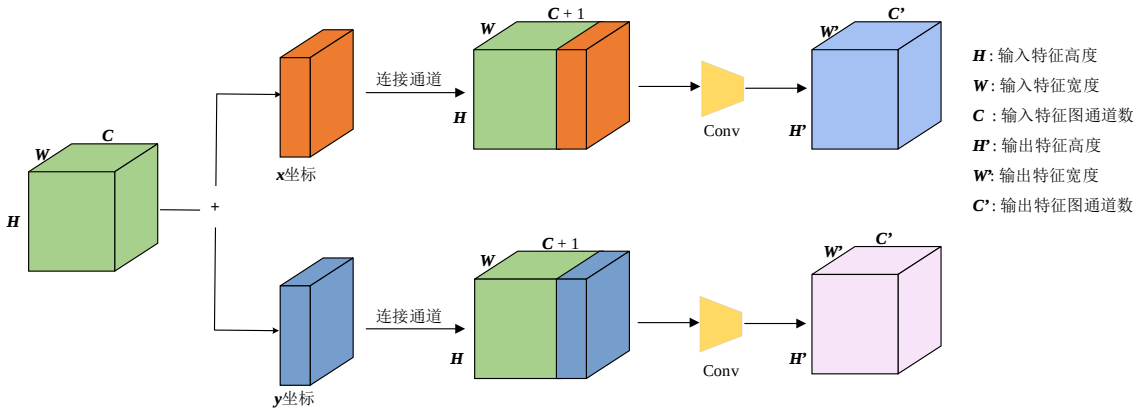



图3　CoordModule结构

其次,logit模块$G$通过由坐标模块获得的包含坐标信息的特征图来生成对应不同方向上的重要性权重,使网络在后续池化中更倾向于对重要性高的区域进行下采样. 为了保证重要性权重为正值,使用logit模块$G$对输出结果进行指数运算和局部归一化,见式(1).

$$F(I)=\exp\left[G(I)\right] \tag{1}$$

其中,$F(I)$为与输入特征图$I$尺寸相同的重要性权重图. 具体地,logit模块$G$由一个$1\times1$卷积层组成,其目的是高效获取空间信息,通过该卷积与坐标信息结合来生成水平和垂直方向上的特征权重,然后将仿射实例归一化(affine instance normalization)作为归一化形式,以适应不同方向上的输入分布,从而在后续计算中正确突出重要特征.

再次,通过将平均池化加权后的特征与权值相除可以分别得到水平和垂直方向上归一化后的输出$O_{x',y'}$和$O'_{x',y'}$,计算公式分别为

$$O_{x',y'}=\frac{\sum_{(\Delta x,\Delta y)\in\Omega}F(I)_{x+\Delta x,y}I_{x+\Delta x,y}}{\sum_{(\Delta x,\Delta y)\in\Omega}F(I)_{x+\Delta x,y}} \tag{2}$$

$$O'_{x',y'}=\frac{\sum_{(\Delta x,\Delta y)\in\Omega}F(I)_{x,y+\Delta y}I_{x,y+\Delta y}}{\sum_{(\Delta x,\Delta y)\in\Omega}F(I)_{x,y+\Delta y}} \tag{3}$$

其中,$(x,y)$为输入特征图对应滑动窗口的位置,$(x',y')$

为对应的输出位置，$(\Delta x,\Delta y)$为滑动窗口内的相对位置，$\Omega$为$(\Delta x,\Delta y)$组成的采样点集合．因此，在滑动窗口中激活值的线性加权过程可以看作池化过程，窗口总和即为特定窗口中所有激活值的和，用以聚合重要信息．在本文设计中，窗口大小为3．

最后，分别沿着2个空间方向聚合特征，使用1×1卷积对特征进行变换，并将Sigmoid激活函数作为一种调节器，通过固定的放大系数动态调节输入特征的重要性，使模型在不同特征上应用不同的权重，增强网络对不同特征的敏感度，从而获得最终的输出特征图，进一步提升模型表现．为提供足够大的范围以有效捕捉特征的重要性，在本文实验中放大系数设置为12．由于所设计的CIP模块的重要性权重为Softmax形式，可将其看作池化设计的注意力方法，仅通过水平和垂直方向卷积生成的特征来计算其权重，并未使用键-查询(key-query)方案．

### 3.1.2 BlurPool

常用的下采样方法通常忽略了奈奎斯特采样定理导致移位等变性丢失，即图像输入特征出现小移位或平移会使得输出剧烈变化．移位等变性(shift equivariance)公式如下：

$$\mathrm{Shift}_{\Delta h,\Delta\omega}\left(\tilde{F}(X)\right)=\tilde{F}\left(\mathrm{Shift}_{\Delta h,\Delta\omega}(X)\right),\forall(\Delta h,\Delta\omega) \tag{4}$$

其中，$(\Delta h,\Delta\omega)$为移位量，$\tilde{F}$为移位函数，则对于$L$层的网络可表示为$\tilde{F}_L(X)\in\mathbf{R}^{H_L\times W_L\times C_L}$，$H\times W$为图像$X$的分辨率，$C$为通道数．若输入移位等于输出移位，则意味着移位和特征提取可以交换，即网络中的移位操作不会影响特征提取结果．因此，在图像分类中移位等变性是一个需要重要考虑的因素．

为了缓解下采样过程中忽略采样定理对网络性能造成的影响，教师网络将使用BlurPool[36]作为下采样方式．网络在采样时，输出结果可能会随着输入的振荡产生较大的波动．BlurPool通过在ResNet34的Basicblock中的下采样部分引入一个核为$m\times m$的抗锯齿滤波器来平滑之后的值，用$\mathrm{Blur}_m$来表示，计算公式如下：

$$\mathrm{BlurPool}_{m,s}=\mathrm{downsample}_s\circ\mathrm{Blur}_m \tag{5}$$

其中，$s$为步长，$\circ$为连接关系．将抗锯齿和下采样合并成一个操作，并在Basicblock中根据步长进行下采样，以平滑图像的边缘和细节，如图4所示，这样可以有效减轻锯齿状边缘效应，从而在很大程度上缓解移位等变性的丢失．

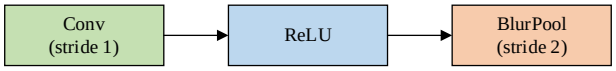



图4 BlurPool下采样过程

## 3.2 解耦类别对齐蒸馏

为了提高分类准确率，避免教师和学生网络架构悬殊时导致的蒸馏效果不佳的问题，同时更好地增强类别区分能力，本文将教师和学生网络的分类预测分为2个部分，即目标类知识蒸馏(Target Class Knowledge Distillation，TCKD)和非目标类知识蒸馏(Non-target Class Knowledge Distillation，NCKD)，并在输出部分进行类别对齐以考虑类别之间的关系．目标类知识是学生模型主要关注和学习的类别，通常是教师模型在预测过程中产生的与训练数据直接相关的信息，包括训练样本的“困难”知识；非目标类知识则包含大量的暗知识，两者进行解耦时应分别考虑各部分的知识，从而提高蒸馏效率．同时，在输出部分加入类别对齐函数来吸收类别相关性知识，以减小教师与学生在类别上的预测差异，进而增强预测．算法原理如图5所示．

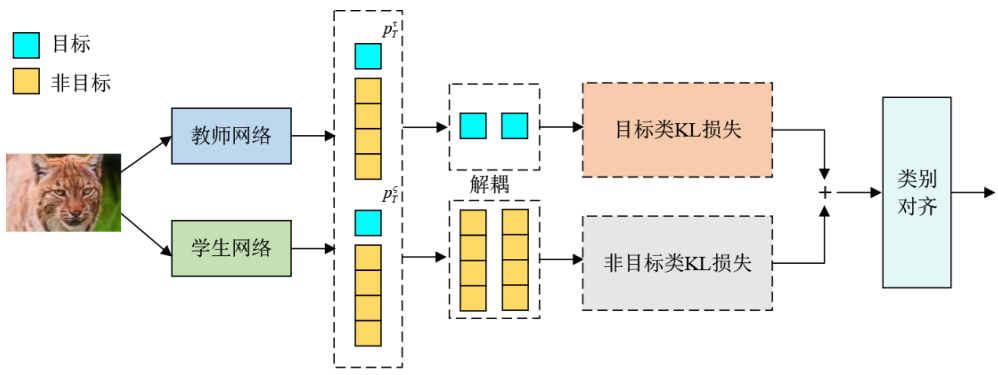



图5 解耦类别对齐蒸馏原理

对于第$t$类的训练样本，分类概率可以表示为$p=\left[p_1,p_2,\cdots,p_t,\cdots,p_n\right]\in\mathbf{R}^{1\times n}$，其中$n$为类别的数量，$\mathbf{R}^{1\times n}$为教师和学生网络的logit输出空间．$p$中的每个元素都可以通过Softmax函数得到，表达式为

$$p_i=\frac{\exp\left(z_i\right)}{\sum_{j=1}^{n}\exp\left(z_j\right)} \tag{6}$$

其中，$z_i$为第$i$类的logit输出．

对于解耦目标类和非目标类，定义 $b=\left[p_T,p_N\right]\in\mathbf{R}^{1\times 2}$，其中 $b$ 为目标类和非目标类的二值概率，$p_T$ 为目标类的概率，$p_N$ 为非目标类的概率，其计算公式分别为

$$p_T=\frac{\exp\left(z_T\right)}{\sum_{j=1}^{n}\exp\left(z_j\right)} \tag{7}$$

$$p_N=\frac{\sum_{k=1,k\neq T}^{n}\exp\left(z_k\right)}{\sum_{j=1}^{n}\exp\left(z_j\right)} \tag{8}$$

对非目标类之间的概率进行建模，即 $\tilde{p}=\left[\tilde{p}_1,\tilde{p}_2,\cdots,\tilde{p}_{t-1},\tilde{p}_{t+1}\cdots,\tilde{p}_n\right]\in\mathbf{R}^{1\times(n-1)}$，每个元素的概率可表示为

$$\tilde{p}_i=\frac{\exp\left(z_i\right)}{\sum_{j=1,j\neq t}^{n}\exp\left(z_j\right)} \tag{9}$$

经典KD算法使用KL散度表示的损失计算公式为

$$L_{\mathrm{KD}}=\mathrm{KL}\left(p^{\tau}\middle\| p^{\varsigma}\right)=p_T^{\tau}\log\left(\frac{p_T^{\tau}}{p_T^{\varsigma}}\right)+\sum_{i=1,i\neq T}^{n}p_i^{\tau}\log\left(\frac{p_i^{\tau}}{p_i^{\varsigma}}\right) \tag{10}$$

其中，$\tau$ 和 $\varsigma$ 分别为教师和学生. 根据式(6)~式(9)可以推导出 $p_i=\tilde{p}\times p_N$，则蒸馏损失公式为

$$L_{\mathrm{KD}}=p_T^{\tau}\log\left(\frac{p_T^{\tau}}{p_T^{\varsigma}}\right)+p_N^{\tau}\log\left(\frac{p_N^{\tau}}{p_N^{\varsigma}}\right)+p_N^{\tau}\sum_{i=1,i\neq T}^{n}\tilde{p}_i^{\tau}\log\left(\frac{\tilde{p}_i^{\tau}}{\tilde{p}_i^{\varsigma}}\right) \tag{11}$$

式(11)可变为

$$L_{\mathrm{KD}}=\mathrm{KL}\left(b^{\tau}\middle\| b^{\varsigma}\right)+\left(1-p_T^{\tau}\right)\mathrm{KL}\left(\tilde{p}^{\tau}\middle\|\tilde{p}^{\varsigma}\right) \tag{12}$$

其中，$\mathrm{KL}\left(b^{\tau}\middle\| b^{\varsigma}\right)$ 为目标类教师和学生二值概率之间的相似度，可表示为 $S_{\mathrm{TCKD}}$；$\mathrm{KL}\left(\tilde{p}^{\tau}\middle\|\tilde{p}^{\varsigma}\right)$ 为教师和学生在非目标类中的概率相似度，可表示为 $S_{\mathrm{NCKD}}$.

考虑到 $\left(1-p_T^{\tau}\right)$ 较小时NCKD的效果会受到抑制，引入2个超参数 $\alpha$ 和 $\beta$，分别作为TCKD和NCKD的权重，以控制它们在损失函数中的贡献度. 解耦后的蒸馏损失函数可表示为

$$L_{\mathrm{KD}}=\alpha S_{\mathrm{TCKD}}+\beta S_{\mathrm{NCKD}} \tag{13}$$

可以通过调整参数 $\alpha$ 和 $\beta$ 来平衡TCKD和NCKD的重要性，使其更有效和灵活地发挥作用. 针对学生网络可能过于依赖训练数据标签且参数较少，能力有限，在特定类别上存在预测差异的问题(尤其是在样本较少的情况下)，本文提出在输出部分进行类别对齐使网络预测可以描述类别之间的关系，以更好地学习教师网络的知识. 类别之间的关系可以通过预测一批数据进行建模，表达式为

$$\boldsymbol{M}^k=p_k^T p_k \tag{14}$$

$$M_{ab}^k=\sum_{i=1}^{N}p_{i,a,k}\cdot p_{i,b,k} \tag{15}$$

其中，$\boldsymbol{M}^k$ 为 $A\times A$ 矩阵；$M_{ab}^k$ 为一批输入同时被分类到第 $a$ 类和第 $b$ 类的概率；$N$ 为类别数；$p_k$ 为通过温度超参数 $T_k$ 得到的预测概率值；$p_{i,a,k}$ 为第 $i$ 个输入在第 $a$ 个类别上的概率；$p_{i,b,k}$ 为第 $i$ 个输入在第 $b$ 个类别上的概率.

对类别相关性进行量化后，可以通过损失计算迫使学生网络从教师网络中吸收这部分知识，计算公式为

$$L_{\mathrm{class}}=\frac{1}{A}\sum_{k=1}^{K}\left\|\boldsymbol{M}_{\mathrm{tea}}^k-\boldsymbol{M}_{\mathrm{stu}}^k\right\|_2^2 \tag{16}$$

其中，$L_{\mathrm{class}}$ 为类别对齐损失，$\boldsymbol{M}_{\mathrm{tea}}^k$ 和 $\boldsymbol{M}_{\mathrm{stu}}^k$ 分别为教师和学生网络用 $T_k$ 计算得到的类别相关矩阵. DCAD的损失函数为

$$L_{\mathrm{DCAD}}=L_{\mathrm{KD}}+L_{\mathrm{class}} \tag{17}$$

通过蒸馏算法不仅可以独立研究TCKD和NCKD的效果，还能更有效地学习目标类特征，且不受非目标类的噪声影响，使知识在传递过程中更精准，增强知识迁移的有效性和灵活性. 同时，该算法能够增强学生网络在不同类别上的区分能力，如果预测之间存在较大差异，网络会更加谨慎，避免过度自信做出错误的分类，从而提高学生网络的鲁棒性和其他性能.

## 4　实验结果与分析

### 4.1　数据集

本实验分别在公开数据集Caltech101[37]、BIRDS 525 SPECIES[38]和自建数据集CIIP-CSID[39]上测试并验证本文算法的性能，各数据集示例如图6所示.

Caltech101是加州理工学院发布的物体图像数据集，类别范围广泛，有动物、车辆、乐器、日常物品等，包含来自101个对象类别和1个背景类别的9 144张图像，每一类图像个数为40~800张. 本次实验随机选取6 454张图像作为训练集、1 827张图像作为验证集、863张图像作为测试集.

BIRDS 525 SPECIES为Kaggle公开鸟类的数据集，该数据集规模大，类别多样，涵盖525种鸟类，共计89 885张图像，其中包含84 635张训练图像、2 625张验证图像和2 625张测试图像，并且每个物种至少有130张训练图像.

CIIP-CSID是西安邮电大学图像与信息处理研究所(Center for Image and Information Processing，CIIP)依托公安部门收集的刑侦现场勘查图像，包含生物物证、血迹、车辆、现场远景、指纹等17类图像. 由于该数据集图像均来自真实现场案例，且在不同环境、姿态及拍

(a) Caltech101

(b) BIRDS 525 SPECIES

(c) CIIP-CSID

图6 3个数据集示例

摄条件下采集,增加了图像分类的挑战性,因此很适合作为验证算法有效性的数据集. 在实验过程中,选取该数据集中17个大类中的19 359张图像,并使用70%的数据用于训练,20%的数据用于验证,10%的数据用于测试.

### 4.2 实验方法与评价指标

在训练教师和学生网络中设置相同的训练参数,epoch和batch size分别设置为160和32,初始学习率为0.01,所有网络统一使用随机梯度下降(Stochastic Gradient Descent,SGD)优化器,且每30个epoch后的学习率衰减为原来的一半. 在分类任务中,使用准确率Accuracy、精确率Precision、召回率Recall、$F_1$分数和参数量5种指标评价实验结果. 考虑到样本在不同标签上往往分布不均匀,各类标签对于总评价指标的贡献存在差异,因此在加权平均(weighted avg)规则下根据各标签样本数$N'$来选择不同的权重值,其中Weighted Precision、Weighted Recall和Weighted $F_1$分别表示不同类别间贡献差异下所有类别的综合精确率、召回率和$F_1$分数,表达式如下:

$$\text{Weighted Precision}=\frac{1}{N'}\sum_{i=1}^{n}P_i\cdot N_i \tag{18}$$

$$\text{Weighted Recall}=\frac{1}{N'}\sum_{i=1}^{n}R_i\cdot N_i \tag{19}$$

$$\text{Weighted } F_1=\frac{1}{N'}\sum_{i=1}^{n}F_{1,i}\cdot N_i \tag{20}$$

### 4.3 对比实验

为验证所提算法应用于图像分类的有效性,分别在Caltech101、BIRDS 525 SPECIES和CIIP-CSID数据集上与其他主流算法进行对比,包括经典的卷积神经网络Vgg16、ResNet34,轻量级网络ShuffleNetV2、MobileNetV2、MobileOne等,以及MobileViT、RepViT等基于Transformer的视觉转换器(Vision Transformer,ViT)模型. 不同算法的分类结果如表1所示,其中改进后的教师网络命名为E-ResNet34,蒸馏后的学生网络命名为D-MobileNetV3.

由表1可知,在网络性能方面,3个数据集上E-ResNet34的准确率分别达到79.95%、99.41%、90.84%,较ResNet34分别提高1.85、1.01和1.21个百分点;通过蒸馏训练后的D-MobileNetV3准确率分别达到79.37%、98.17%、90.41%,相较于E-ResNet34分别减少0.58、1.24和0.43个百分点,但相较于原网络分别提高5.44、4.19和1.88个百分点,准确率显著提升,表明提出的DCAD算法可以有效引导学生网络学习教师的知识. 在网络规格方面,改进后的E-ResNet34相较于ResNet34不仅实现了精度提升,而且未额外增加参数量,在CIIP-CSID数据集上表现优异,准确率显著高于其他CNN网络,说明E-ResNet34能够成功捕捉复杂环境下的关键信息,尤其是在处理复杂背景和多样性特征时表现出色,进一步表明其在图像分类任务中具有良好的泛化性.

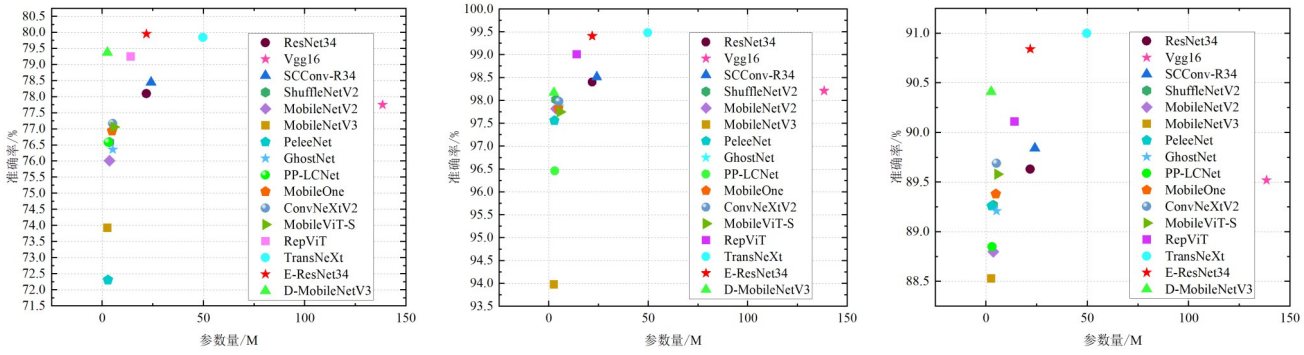
**表1 不同算法的分类结果**

| Algorithm | Model | Params/M | Accuracy/% | | |
|---|---|---|---|---|---|
| | | | Caltech101 | BIRDS 525 SPECIES | CIIP-CSID |
| CNN | ResNet34[3] | 21.8 | 78.10 | 98.40 | 89.63 |
| | Vgg16[40] | 138.4 | 77.75 | 98.21 | 89.52 |
| | SCConv-R34[41] | 24.1 | 78.45 | 98.51 | 89.84 |
| Lightweight CNN | ShuffleNetV2[6] | 3.5 | 76.59 | 98.02 | 89.27 |
| | MobileNetV2[8] | 3.5 | 76.01 | 97.82 | 88.80 |
| | MobileNetV3[9] | 2.5 | 73.93 | 93.98 | 88.53 |
| | PeleeNet[42] | 2.8 | 72.31 | 97.56 | 89.26 |
| | GhostNet[43] | 5.2 | 76.36 | 97.87 | 89.21 |
| | PP-LCNet[44] | 3.0 | 76.59 | 96.46 | 88.85 |
| | MobileOne[45] | 4.8 | 76.94 | 97.83 | 89.38 |
| | ConvNeXtV2[46] | 5.2 | 77.17 | 97.98 | 89.69 |
| ViT | MobileViT-S[47] | 5.6 | 77.06 | 97.75 | 89.58 |
| | RepViT[48] | 14.0 | 79.25 | 99.01 | 90.11 |
| | TransNeXt[49] | 49.7 | 79.84 | 99.48 | 91.00 |
| Ours | E-ResNet34 | 21.8 | 79.95 | 99.41 | 90.84 |
| | D-MobileNetV3 | 2.5 | 79.37 | 98.17 | 90.41 |

所有模型中,E-ResNet34在Caltech101数据集中准确率最高;在BIRDS 525 SPECIES和CIIP-CSID数据集

中,相较于TransNeXt,其准确率分别降低0.07个百分点和0.16个百分点,但优于其他模型.这是由于在Caltech101数据集中,E-ResNet34加入了坐标重要性池化模块,能够增强网络对空间位置的感知,特别是对图像中有明显空间结构或大尺度特征的物体具有明显优势,可以帮助网络更有效地聚焦于重要区域.在另外2组数据集中,E-ResNet34的准确率略低于TransNeXt,原因是这2组数据集包含鸟类和现场勘查图像,可能只存在细微差异,且类内的多样性较大.E-ResNet34通过加入坐标重要性池化模块并引入BlurPool帮助网络关注细粒度的局部信息,并平滑图像的边缘和细节.TransNeXt较E-ResNet34增加了一种像素聚焦注意力机制,采用双路径设计逐像素聚合细粒度和粗粒度特征,模仿人眼在观察过程中从局部到全局的连续运动,因此TransNeXt对细微差异更为敏感.另外,E-ResNet34不仅在模型参数量方面压缩至TransNeXt的43.86%,而且在BIRDS 525 SPECIES和CIIP-CSID数据集上表现出相似的分类结果.综上所述,所提出的E-ResNet34更能兼具分类性能和模型复杂度.

图7为本文所提模型与其他模型的性能对比,直观地展现了模型参数量和准确率的表现.由图7可知,D-MobileNetV3相较于E-ResNet34,在微小精度损失的前提下大幅减少网络的参数量;相较于ViT模型,虽然D-MobileNetV3的参数量比RepViT和TransNeXt分别降低82%和95%,但三者准确度十分接近.例如,在Caltech101数据集上,D-MobileNetV3在精度上比TransNeXt略低0.47个百分点,但参数量仅为其1/20,表现出更优的综合性能,从而证明了本文算法的有效性.本文算法不仅增强了D-MobileNetV3的分类能力,还使其在实际应用场景中具有更强的竞争力和实用性.




(a) Caltech101　　(b) BIRDS 525 SPECIES　　(c) CIIP-CSID

图7　在不同数据集上的性能对比

表2进一步对精确率Precision、召回率Recall和$F_1$分数3个分类指标进行展示,以便更全面地评估模型性能.由表2可知,本文提出的E-ResNet34在各数据集上的精确率、召回率和$F_1$分数远优于其他CNN,表明模型识别准确且全面.相较于MobileViT-S和RepViT,E-ResNet34表现出较好的性能,尤其是在Caltech101数据集上,E-ResNet34相比其他2种模型在$F_1$分数上分别具有4.22个百分点和1.26个百分点的性能增益.其主要原因是MobileViT-S和RepViT虽然通过各种设计降低了模型复杂度,但仍需要大量的训练数据,而Caltech101数据集的训练样本数较少,导致其无法充分发挥优势.实验结果进一步说明E-ResNet34能够更好地适应数据量较小的情况,具有广泛的应用价值,尤其在数据稀缺的场景中能够显著提升模型性能.此外,D-MobileNetV3的精确率和召回率相近,且在$F_1$分数方面表现良好,说明该模型具有良好的分类性能和泛化性.尽管D-MobileNetV3的精确率不及TransNeXt,但在3个数据集上分别达到79.12%、98.48%和90.19%,结合参数量优势,该模型的综合性能最佳.

由表1和表2的实验结果可知,本文算法在模型参数量和分类性能上表现更好.这是由于其在特征提取时充分利用图像像素的坐标信息,增强了模型对重要性区域的关注,且提取的信息保有量高,有利于提高分类准确度;它在蒸馏时分别考虑目标类和非目标类知识,有利于提高蒸馏效率;同时,它还进行了类别对齐以更有效地利用教师网络知识,极大地提升了分类准确性.蒸馏后的学生网络D-MobileNetV3在保证参数量尽可能小的前提下拥有优异的性能,更适合部署于储存资源有限和硬件水平低的场景.

### 4.4　消融实验

为了验证教师网络每个改进步骤的有效性,本文分别在3个数据集上对坐标重要性池化和BlurPool进行测试,探索其对网络分类效果的作用.消融实验结果如表3所示.

由表3可知,在3个数据集上,以ResNet34为基线网络,加入坐标重要性池化模块CIP,利用水平和垂直

表2 各数据集上的具体对比结果 单位:%

| Model | Caltech101 | | | BIRDS 525 SPECIES | | | CIIP-CSID | | |
|---|---|---|---|---|---|---|---|---|---|
| | Precision | Recall | $F_1$ | Precision | Recall | $F_1$ | Precision | Recall | $F_1$ |
| ResNet34[3] | 78.18 | 78.10 | 77.23 | 98.65 | 98.40 | 98.38 | 89.37 | 89.63 | 89.44 |
| Vgg16[40] | 77.83 | 77.75 | 76.43 | 98.51 | 98.21 | 98.17 | 88.92 | 89.53 | 89.09 |
| SCConv-R34[41] | 79.53 | 78.45 | 77.57 | 98.75 | 98.51 | 98.50 | 89.44 | 89.84 | 89.54 |
| ShuffleNetV2[6] | 78.02 | 76.59 | 75.73 | 98.35 | 98.02 | 98.00 | 88.57 | 89.28 | 88.75 |
| MobileNetV2[8] | 77.06 | 76.01 | 75.08 | 98.22 | 97.83 | 97.80 | 88.71 | 88.80 | 88.51 |
| MobileNetV3[9] | 73.99 | 73.94 | 72.93 | 94.93 | 93.98 | 93.89 | 88.28 | 88.54 | 88.13 |
| PeleeNet[42] | 72.56 | 72.32 | 71.14 | 98.00 | 97.56 | 97.55 | 88.76 | 89.27 | 88.88 |
| GhostNet[43] | 77.13 | 76.36 | 75.75 | 98.20 | 97.87 | 97.80 | 88.97 | 89.21 | 88.86 |
| PP-LCNet[44] | 78.22 | 76.59 | 76.02 | 97.04 | 96.46 | 96.36 | 88.37 | 88.85 | 88.37 |
| MobileOne[45] | 77.34 | 76.94 | 76.05 | 98.20 | 97.83 | 97.80 | 89.09 | 89.38 | 89.06 |
| ConvNeXtV2[46] | 78.63 | 77.17 | 76.45 | 98.30 | 97.98 | 97.95 | 89.21 | 89.70 | 89.25 |
| MobileViT-S[47] | 76.51 | 77.05 | 75.15 | 98.15 | 97.75 | 97.73 | 89.19 | 89.58 | 89.30 |
| RepViT[48] | 79.58 | 79.26 | 78.11 | 99.16 | 99.01 | 99.00 | 89.73 | 90.10 | 89.74 |
| TransNeXt[49] | 79.47 | 79.83 | 78.47 | 99.73 | 99.48 | 99.44 | 90.93 | 91.00 | 90.71 |
| E-ResNet34 | 80.92 | 79.96 | 79.37 | 99.65 | 99.42 | 99.39 | 90.51 | 90.84 | 90.53 |
| D-MobileNetV3 | 79.12 | 79.37 | 78.12 | 98.48 | 98.16 | 98.14 | 90.19 | 90.42 | 90.17 |

表3 教师网络消融实验结果 单位: %

| 基线 | CIP | BlurPool | Accuracy | | |
|---|---|---|---|---|---|
| | | | Caltech101 | BIRDS 525 SPECIES | CIIP-CSID |
| √ | | | 78.10 | 98.40 | 89.63 |
| √ | √ | | 79.14 | 99.01 | 90.26 |
| √ | √ | √ | 79.95 | 99.41 | 90.84 |

方向的位置信息来指导特征的加权和选择,其网络分类准确率分别达到79.14%、99.01%和90.26%,相较于基线网络分别提高1.04、0.61和0.63个百分点,这说明坐标信息有助于充分提取图像特征并更好地保留重要特征,用于整体的图像分类. 引入BlurPool的抗混叠能力后,准确率进一步提高0.81、0.40和0.58个百分点,分别达到79.95%、99.41%、90.84%. 测试结果表明,改进后的E-ResNet34中的2个关键点均可提高网络的分类性能.

为了研究DCAD算法对学生网络分类性能的影响,采用D-MobileNetV3作为学生网络,并使用KD、解耦知识蒸馏(Decoupled Knowledge Distillation,DKD)和本文方法进行蒸馏消融实验对比,在各数据集上得到的实验结果如表4所示. 由表4可知,本文提出的DCAD算法使得学生网络的分类准确率在3个数据集上分别达到79.37%、98.17%和90.41%,相较于其他2种蒸馏算法精度明显提高,证明了该蒸馏算法的有效性. 相较于只进行解耦处理的DKD算法,DCAD在3个数据集上的准确率分别提升0.88、0.57和0.41个百分点,验证了所设计的DCAD能够指导学生网络更好地学习知识,尤其能通过学习类别上的细微差异来提升整体分类性能. 通过整体的消融实验结果可以证明,本文改进方法均能提升教师与学生网络的分类准确率.

表4 学生网络蒸馏消融实验结果 单位: %

| Model | Accuracy | | |
|---|---|---|---|
| | Caltech101 | BIRDS 525 SPECIES | CIIP-CSID |
| 基线[9] | 73.93 | 93.98 | 88.53 |
| +KD[19] | 77.98 | 95.77 | 89.58 |
| +DKD[50] | 78.49 | 97.60 | 90.00 |
| +DCAD | 79.37 | 98.17 | 90.41 |

## 4.5 可视化实验

为了更直观地验证D-MobileNetV3的有效性,本文使用GradCAM生成可视化激活图对基线网络(MobileNetV3)和D-MobileNetV3进行可视化分析,如图8所示,其中数据样本来自Caltech101数据集. 图8中,第1行为数据样本原图,第2行为基线网络的激活图,第3行为D-MobileNetV3的激活图,激活图中颜色越鲜艳的区域表示越具有辨别力,是网络对该图像进行分类时最感兴趣的区域. 由图8可知,相较于基线网络,D-MobileNetV3注意力聚焦的感兴趣区域更精准且更完整,例如第1列的可视化图,基线网络的注意区域仅在帆船桅杆部分,而D-MobileNetV3则将注意力集中在重要的主体区域,进而可以更准确地对图像进行分类,验证了D-MobileNetV3卓越的图像分类性能.

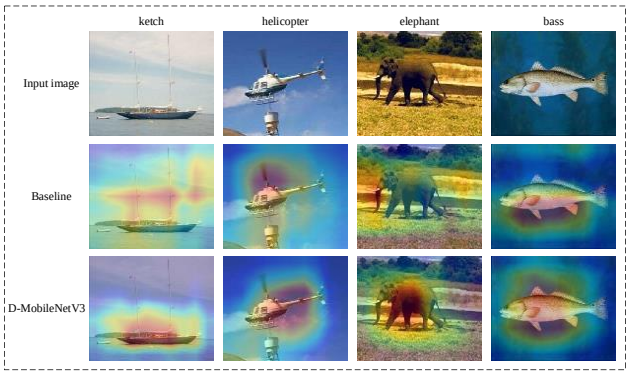



图8　使用GradCAM进行可视化对比

## 5　结论

为了提高卷积神经网络在图像分类任务中的精度,同时实现网络轻量化,本文提出一种基于坐标重要性池化和解耦类别对齐蒸馏的图像分类算法. 选取基于坐标重要性池化模块和BlurPool改进后的E-ResNet34作为教师网络,并选取轻量级MobileNetV3作为学生网络. 通过构造一种解耦类别对齐蒸馏算法,将教师和学生网络的分类预测分为目标类知识蒸馏与非目标类知识蒸馏2个部分,并在输出部分进行类别对齐,使网络预测可以描述类别之间的关系,进而使学生网络可以更好地学习教师网络的知识. 通过蒸馏训练完成后的学生网络D-MobileNetV3被用于图像分类,提高分类精度的同时实现了网络轻量化. 在不同数据集上的实验结果表明,本文算法E-ResNet34在准确率、精确率、召回率和$F_1$分数方面的综合性能优于其他主流模型,且未额外增加参数量;蒸馏后的学生网络D-MobileNetV3在较低的参数量下具有出色的分类性能,更利于在资源受限的场景中部署.

### 参考文献


[1] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.

[2] WANG S X, VELDHUIS R, BRUNE C, et al. What do neural networks learn in image classification? A frequency shortcut perspective[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2023: 1433-1442.

[3] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 770-778.

[4] 葛道辉, 李洪升, 张亮, 等. 轻量级神经网络架构综述[J]. 软件学报, 2020, 31(9): 2627-2653.
GE D H, LI H S, ZHANG L, et al. Survey of lightweight neural network[J]. Journal of Software, 2020, 31(9): 2627-2653. (in Chinese)

[5] IANDOLA F N, HAN S, MOSKEWICZ M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size[EB/OL]. (2016-02-24)[2024-08-13]. https://arxiv.org/abs/1602.07360v4.

[6] MA N N, ZHANG X Y, ZHENG H T, et al. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design[M]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 122-138.

[7] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17) [2024-08-13]. https://arxiv.org/ abs/1704.04861v1.

[8] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 4510-4520.

[9] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2019: 1314-1324.

[10] SHEN H, WANG Z W, ZHANG J X, et al. L-Net: A lightweight convolutional neural network for devices with low computing power[J]. Information Sciences, 2024, 660: 120131.

[11] TAN M X, CHEN B, PANG R M, et al. MnasNet: Platform-aware neural architecture search for mobile[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 2820-2828.

[12] TAN M X, LE Q V. EfficientNet: Rethinking model scaling for convolutional neural networks[EB/OL]. (2020-09-11)[2024-08-13]. https://arxiv.org/abs/1905.11946v5.

[13] PENG C, LI Y Y, SHANG R H, et al. ReCNAS: Resource-constrained neural architecture search based on differentiable annealing and dynamic pruning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(2): 2805-2819.

[14] VADERA S, AMEEN S. Methods for pruning deep neural networks[J]. IEEE Access, 2022, 10: 63280-63300.

[15] ROKH B, AZARPEYVAND A, KHANTEYMOORI A. A comprehensive survey on model quantization for deep neural networks in image classification[EB/OL]. (2023-10-23)[2024-08-13]. https://arxiv.org/abs/2205.07877v5.

[16] SAINATH T N, KINGSBURY B, SINDHWANI V, et al. Low-rank matrix factorization for Deep Neural Network

training with high-dimensional output targets[C]//2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2013: 6655-6659.

[17] 黄震华, 杨顺志, 林威, 等. 知识蒸馏研究综述[J]. 计算机学报, 2022, 45(3): 624-653.
HUANG Z H, YANG S Z, LIN W, et al. Knowledge distillation: A survey[J]. Chinese Journal of Computers, 2022, 45(3): 624-653. (in Chinese)

[18] BUCILUǍ C, CARUANA R, NICULESCU-MIZIL A. Model compression[C]//Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2006: 535-541.

[19] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[EB/OL]. (2015-03-09)[2024-08-13]. https://arxiv.org/abs/1503.02531v1.

[20] 刘立波, 郗思宇, 邓箴. 结合改进ConvNeXt网络与知识蒸馏的天气识别[J]. 光学 精密工程, 2023, 31(14): 2123-2134.
LIU L B, XI S Y, DENG Z. Weather recognition combining improved ConvNeXt models with knowledge distillation[J]. Optics and Precision Engineering, 2023, 31(14): 2123-2134. (in Chinese)

[21] 李大湘, 南艺璇, 刘颖. 面向遥感图像场景分类的双知识蒸馏模型[J]. 电子与信息学报, 2023, 45(10): 3558-3567.
LI D X, NAN Y X, LIU Y. A double knowledge distillation model for remote sensing image scene classification[J]. Journal of Electronics & Information Technology, 2023, 45(10): 3558-3567. (in Chinese)

[22] LI S H, SHAO M W, GUO Z H, et al. Improving knowledge distillation *via* pseudo-multi-teacher network[J]. Machine Vision and Applications, 2023, 34(2): 33.

[23] MIRZADEH S I, FARAJTABAR M, LI A, et al. Improved knowledge distillation *via* teacher assistant[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(4): 5191-5198.

[24] HUANG T, YOU S, WANG F, et al. Knowledge distillation from a stronger teacher[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. New York: ACM, 2022: 33716-33727.

[25] KIM J, JUNG J, KANG U. Compressing deep graph convolution network with multi-staged knowledge distillation[J]. PLoS One, 2021, 16(8): e0256187.

[26] LI C X, LIN M B, DING Z Y, et al. Knowledge Condensation Distillation[M]//Computer Vision-ECCV 2022. Cham: Springer Nature Switzerland, 2022: 19-35.

[27] SHARMA S, LODHI S S, CHANDRA J. SCL-IKD: Intermediate knowledge distillation via supervised contrastive representation learning[J]. Applied Intelligence, 2023, 53(23): 28520-28541.

[28] ZHANG J, TAO Z, GUO K H, et al. Hybrid mix-up contrastive knowledge distillation[J]. Information Sciences, 2024, 660: 120107.

[29] GHOLAMALINEZHAD H, KHOSRAVI H. Pooling methods in deep neural networks, a review[EB/OL]. (2020-09-16)[2024-08-13]. https://arxiv.org/abs/2009.07485v1.

[30] HE K M, ZHANG X Y, REN S Q, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[M]//Computer Vision-ECCV 2014. Cham: Springer International Publishing, 2014: 346-361.

[31] XU L J, YAN S Z, CHEN X, et al. Motion recognition algorithm based on deep edge-aware pyramid pooling network in human-computer interaction[J]. IEEE Access, 2019, 7: 163806-163813.

[32] WIJAYA K T, PAEK D H, KONG S H. Advanced feature learning on point clouds using multi-resolution features and learnable pooling[J]. Remote Sensing, 2024, 16(11): 1835.

[33] ZHAO L, ZHANG Z L. A improved pooling method for convolutional neural networks[J]. Scientific Reports, 2024, 14(1): 1589.

[34] GAO Z T, WANG L M, WU G S. LIP: Local importance-based pooling[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2019: 3355-3364.

[35] WANG L, GAO C Q, LIU J, et al. A novel learning-based frame pooling method for event detection[J]. Signal Processing, 2017, 140: 45-52.

[36] ZHANG R. Making convolutional networks shift-invariant again[C]//36th International Conference on Machine Learning. New York: PMLR, 2019: 12712-12722.

[37] LI F F, FERGUS R, PERONA P. One-shot learning of object categories[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(4): 594-611.

[38] GPIOSENKA G. BIRDS 525 SPECIES[EB/OL]. (2023-04-20) [2024-08-13]. https://www. kaggle. com/datasets/gpiosenka/100-bird-species.

[39] 刘颖, 胡丹, 范九伦. 现勘图像检索综述[J]. 电子学报, 2018, 46(3): 761-768.
LIU Y, HU D, FAN J L. A survey of crime scene investigation image retrieval[J]. Acta Electronica Sinica, 2018,

46(3): 761-768. (in Chinese)
[40] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-08-10)[2024-08-13]. https://arxiv.org/abs/1409.1556v6.
[41] LI J F, WEN Y, HE L H. SCConv: Spatial and channel reconstruction convolution for feature redundancy[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2023: 6153-6162.
[42] WANG R J, LI X, LING C X. Pelee: A real-time object detection system on mobile devices[C]//NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: ACM, 2018: 1967-1976.
[43] HAN K, WANG Y H, TIAN Q, et al. GhostNet: More features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 1580-1589.
[44] CUI C, GAO T Q, WEI S Y, et al. PP-LCNet: A lightweight CPU convolutional neural network[EB/OL]. (2021-09-17)[2024-08-13]. https://arxiv.org/abs/2109.15099v1.
[45] VASU P K A, GABRIEL J, ZHU J, et al. MobileOne: An improved one millisecond mobile backbone[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2023: 7907-7917.
[46] WOO S, DEBNATH S, HU R H, et al. ConvNeXt V2: Co-designing and scaling ConvNets with masked autoencoders[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2023: 16133-16142.
[47] MEHTA S, RASTEGARI M. MobileVIT: Light-weight, general-purpose, and mobile-friendly vision transformer[EB/OL]. (2021-10-05)[2024-08-13]. https://arxiv.org/abs/2110.02178v2.
[48] WANG A, CHEN H, LIN Z J, et al. Rep ViT: Revisiting mobile CNN from ViT perspective[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2024: 15909-15920.
[49] SHI D. TransNeXt: Robust foveal visual perception for vision transformers[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2024: 17773-17783.
[50] ZHAO B R, CUI Q, SONG R J, et al. Decoupled knowledge distillation[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 11943-11952.

## 作者简介

**刘　颖**　女,西安邮电大学通信与信息工程学院教授. 主要研究方向为图像处理与模式识别.
E-mail: liuying_ciip@163.com

**薛家昊**　男,西安邮电大学通信与信息工程学院硕士研究生. 主要研究方向为图像分类. E-mail: xuejiahao0803@163.com

**张伟东**　男,西安邮电大学通信与信息工程学院副教授. 主要研究方向为室内场景理解. E-mail: chluzhre@126.com

**许志杰**　男,英国哈德斯菲尔德大学(University of Huddersfield)工程与计算机学院教授. 主要研究方向为图形图像处理.
E-mail: z.xu@hud.ac.uk