E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1006-8961(2024)12-3756-14

论文引用格式:Zhang R, Chen Y, Wang J B, Li Y and Zhang X. 2024. Self-knowledge distillation for fine-grained image classification. Journal of Image and Graphics, 29(12):3756-3769(张睿, 陈瑶, 王家宝, 李阳, 张旭. 2024. 细粒度图像分类的自知识蒸馏学习. 中国图象图形学报, 29(12):3756-3769)[DOI:10.11834/jig.230846]

细粒度图像分类的自知识蒸馏学习

张睿[1],陈瑶[1],王家宝[1*],李阳[1],张旭[1,2]

1. 陆军工程大学指挥控制工程学院,南京　210007; 2. 江苏经贸职业技术学院,南京　211168

摘　要:**目的**　在无教师模型指导的条件下,自知识蒸馏方法可以让模型从自身学习知识来提升性能,但该类方法在解决细粒度图像分类任务时,因缺乏对图像判别性区域特征的有效提取导致蒸馏效果不理想。为了解决该问题,提出了一种融合高效通道注意力的细粒度图像分类自知识蒸馏学习方法。**方法**　首先,引入高效通道注意力(efficient channel attention, ECA)模块,设计了ECA残差模块并构建ECA-ResNet18(residual network)轻量级骨干网,用以更好地提取图像判别性区域的多尺度特征;其次,构建了高效通道注意力加权双向特征金字塔ECA-BiFPN(bidirectional feature pyramid network)模块,用以融合不同尺度的特征,构建更加鲁棒的跨尺度特征;最后,提出了一种多级特征知识蒸馏损失,用以跨尺度特征对多尺度特征的蒸馏学习。**结果**　在Caltech-UCSD Birds 200、Stanford Cars和FGVC-Aircraft 3个公开数据集上,所提方法分别取得了76.04%、91.11%和87.64%的分类精度,与已有15种自知识蒸馏方法中最佳方法的分类精度相比,分别提高了2.63%、1.56%和3.66%。**结论**　所提方法具有高效提取图像判别性区域特征的能力,能获得更好的细粒度图像分类精度,其轻量化的网络模型适合于面向嵌入式设备的边缘计算应用。

关键词:细粒度图像分类;通道注意力;知识蒸馏(KD);自知识蒸馏(SKD);特征融合;卷积神经网络(CNN);轻量级模型

Self-knowledge distillation for fine-grained image classification

Zhang Rui[1], Chen Yao[1], Wang Jiabao[1*], Li Yang[1], Zhang Xu[1,2]

1. Command and Control Engineering College, Army Engineering University of PLA, Nanjing　210007, China;

2. Jiangsu Vocational Institute of Commerce, Nanjing　211168, China

Abstract: **Objective** Fine-grained image classification aims to classify a super-category into multiple sub-categories. This task is more challenging than general image classification due to the subtle inter-class differences and large intra-class variations. The attention mechanism enables the model to focus on the key areas of the input image and the discriminative regional features of the image, which are particularly useful for fine-grained image classification tasks. The attention-based classification model also shows high interpretability. To improve the focus of this model on the image discriminative region, attention-based methods have been applied in fine-grained image classification. Although the current attention-based fine-grained image classification models achieve high classification accuracy, they do not adequately consider the number of model parameters and computational volume. As a result, they cannot be easily deployed on low-resource devices, thus greatly limiting their practical application. The concept of knowledge distillation involves transferring knowledge from a high-accuracy, high-parameter, and computationally expensive large teacher model to a low-parameter and computationally

收稿日期:2023-12-12;**修回日期**:2024-02-06;**预印本日期**:2024-02-13

*通信作者:王家宝　jiabao_1108@163.com

基金项目:江苏省自然科学基金项目(BK20200581)

Supported by:Natural Science Foundation of Jiangsu Province, China(BK20200581)

efficient small student model to enhance the performance of the latter and to reduce the cost of model learning. To further reduce the model learning cost, researchers have proposed the self-knowledge distillation method that, unlike traditional knowledge distillation methods, enables models to improve their performance by utilizing their own knowledge instead of relying on teacher networks. However, this method falls short in addressing fine-grained image classification tasks due to its ineffective extraction of discriminative region features from images, which results in unsatisfactory distillation outcomes. To address this issue, we propose a self-knowledge distillation learning method for fine-grained image classification by fusing efficient channel attention (ECASKD). **Method** The proposed method embeds an efficient channel attention mechanism into the structure of the self-knowledge distillation framework to effectively extract the discriminative regional features of images. The framework mainly consists of a self-knowledge distillation network with a lightweight backbone and a self-teacher subnetwork and a joint loss with classification loss, knowledge distillation loss, and multi-layer feature-based knowledge distillation loss. First, we introduce the efficient channel attention (ECA) module, propose the ECA-Residual block, and construct the ECA-Residual Network18 (ECA-ResNet18) lightweight backbone to improve the extraction of multiscale features in discriminative regions of the input image. Compared with the residual module in the original ResNet18, the ECA-Residual block introduces the ECA module after each batch normalization operation. This module consists of two ECA-Residual blocks to form a stage of the ECA-ResNet18 backbone network, enhance the network's focus on discriminative regions of the image, and facilitate the extraction of multiscale features. Unlike ResNet18, which is commonly used in self-knowledge distillation methods, the proposed backbone is based on the ECA-Residual module, which can significantly enhance the ability of the model to extract multi-scale features while maintaining lightweight and highly efficient computational performance. Second, considering the differences in the importance of different scale features output from the backbone network, we design and propose the efficient channel attention bidirectional feature pyramid network (ECA-BiFPN) block that assigns weights to the channels during the feature fusion process to differentiate the contribution of features from various channels to the fine-grained image classification task. Finally, we propose a multi-layer feature-based knowledge distillation loss to enhance the backbone network's learning from the self-teacher subnetwork and to focus on discriminative regions. **Result** Our proposed method achieves classification accuracies of 76.04%, 91.11%, and 87.64% on three publicly available datasets, namely, Caltech-UCSD Birds 200 (CUB), Stanford Cars (CAR), and FGVC-Aircraft (AIR). To ensure a comprehensive and objective evaluation, we compared the proposed ECASKD method with 15 other methods, including data-augmentation, auxiliary-network, and attention-based methods. Compared with data-augmentation-based methods, ECASKD improves the accuracy by 3.89%, 1.94%, and 4.69% on CUB, CAR, and AIR, respectively, with respect to the state-of-the-art (SOTA) method. Compared to the auxiliary network-based method, ECASKD improves the accuracy by 6.17%, 4.93%, and 7.81% on CUB, CAR, and AIR, respectively, with respect to SOTA method. Compared to the joint auxiliary network and data augmentation methods, ECASKD improves the accuracy by 2.63%, 1.56%, and 3.66% on CUB, CAR, and AIR, respectively, with respect to SOTA method. In sum, ECASKD demonstrates a better fine-grained image classification performance compared with the joint auxiliary network and data augmentation methods even without data augmentation. Compared with the attention-based self-knowledge distillation method, ECASKD improves about 23.28%, 8.17%, and 14.02% on CUB, CAR and AIR, respectively, with respect to SOTA method. In sum, the ECASKD method outperforms all three types of self-knowledge distillation methods and demonstrates a better fine-grained image classification performance. We also compare this method with four mainstream modeling methods in terms of the number of parameters, floating-point operations (FLOPs), and TOP-1 classification accuracy. Compared with ResNet18, the ECA-ResNet18 backbone used in the proposed method significantly improves the classification accuracy with an increase of only 0.4 M parameters and 0.2 G FLOPs. Compared with the larger-scale ResNet50, the performance of the proposed method is less than one-half of that of ResNet50 in terms of number of parameters and computation, but its classification accuracy on the CAR dataset differs from ResNet50 by only 0.6%. Compared with the larger ViT-Base and Swin-Transformer-B, the proposed method is about one-eighth of both in terms of number of parameters and computation, and its classification accuracies on the CAR and AIR datasets are 3.7% and 5.3% lower than the optimal Swin-Transformer-B. These results demonstrate that the classification accuracy of the proposed method is significantly improved with only a small increase in model complexity. **Conclusion** The proposed self-knowledge distillation fine-grained image

classification method achieves good performance results with 11.9 M parameters and 2.0 G FLOPs, and its lightweight network model is suitable for edge computing applications for embedded devices.

Key words: fine-grained image classification; channel attention; knowledge distillation (KD); self-knowledge distillation (SKD); feature fusion; convolutional neural network (CNN); lightweight model

0 引 言

细粒度图像分类,又称为子类别图像分类,其目的是对粗粒度的大类别进行更加细致的子类划分,如区分鸟的种类、车的款式及狗的品种等,是近年来计算机视觉、模式识别等领域一个非常热门的研究课题(魏秀参 等,2022)。由于子类别间细微的类间差异和较大的类内变化,较之普通的图像分类任务,细粒度图像分类难度更大(Wei 等,2022)。随着深度学习的迅速发展,各种卷积神经网络(convolutional neural network, CNN)(Krizhevsky 等,2012;He 等,2016;Hu 等人,2018)和 Transformer(Vaswani 等,2017;Dosovitskiy 等,2021;Liu 等,2021)应用于细粒度图像分类任务,凭借其越来越深的网络层数和精妙的设计结构不断降低模型的错误率(江铃燚 等,2023)。然而性能越好的深度学习模型往往需要越多的资源,使其在物联网、移动互联网等低资源设备的应用中受到限制。

为了解决这一问题,基于知识蒸馏(knowledge distillation, KD)(Hinton 等,2015)的模型压缩受到研究者的广泛关注。知识蒸馏的主要思想是将精度高、参数多和计算量大的教师大模型中的"知识"迁移到参数少、计算量小的学生小模型中,以此来提升学生小模型的性能(Wang 和 Yoon, 2022;司兆峰和齐洪钢,2023)。主要方法包括基于类预测结果的知识蒸馏(Hinton 等, 2015; Mirzadeh 等, 2020; Zhao 等,2022)、基于特征的知识蒸馏(Romero 等,2015;Tian 等,2020;Chen 等,2021)和基于注意力的知识蒸馏(Zagoruyko 和 Komodakis, 2017; Guo 等, 2023)。但教师大模型在知识蒸馏之前需要耗费大量计算资源进行预训练,且已有预训练大模型与具体任务之间还存在着跨域适配问题(Gou 等,2021)。因此,是否可以在无教师模型的条件下提升学生小模型的学习效果,是一个非常值得研究的问题。

与传统的知识蒸馏不同,自知识蒸馏(self-knowledge distillation, SKD)允许网络从自身学习知识,而不需要任何额外教师大模型的指导。自知识蒸馏是单个网络被同时用做教师模型和学生模型,让网络模型在自我学习的过程中通过知识蒸馏去提升性能(黄震华 等,2022)。现有自知识蒸馏方法主要分为基于辅助网络的方法、基于数据增强的方法和基于注意力的方法。前两种方法虽然有着较高的分类性能表现,但它们在迁移的知识类型上大多选择类预测结果或者特征图,忽视了对图像判别性区域的重点关注。注意力机制能够使模型关注输入图像的重点区域,理论上相较于前两者会更加关注到图像判别性区域特征,更适合细粒度图像分类任务,且基于注意力的自知识蒸馏模型有着更高的可解释性(Guo 等,2023)。Hou 等人(2019)提出的自注意力蒸馏(self-attention distillation, SAD)模型虽然验证了采用基于分层和自顶向下注意力图自知识蒸馏方法的有效性,并成功将其应用到车道线检测任务中,但该方法所使用的注意力图并未使用注意力机制,而是直接来自于模型本身的网络深层特征,并只采用深层特征蒸馏浅层特征的方式,缺乏对图像判别性区域特征的有效提取。与基于辅助网络的方法和基于数据增强的方法相比,直接将 SAD 方法应用于细粒度图像分类任务中,所得分类精度较低。

针对以上问题,本文提出了一种融合高效通道注意力的细粒度图像分类自知识蒸馏学习方法(efficient channel attention based self-knowledge distillation, ECASKD),并应用于 ResNet18 网络,提高了自知识蒸馏细粒度图像分类性能。主要贡献包括:1)提出了一种融合高效通道注意力的细粒度图像分类自知识蒸馏学习方法,将高效通道注意力机制嵌入到自知识蒸馏框架结构中,实现了图像判别性区域特征的有效提取;2)设计了 ECA 残差模块并构建 ECA-ResNet18 骨干网和 ECA-BiFPN 模块,实现不同尺度特征的有效提取和融合,并提出了一种多级特征知识蒸馏损失,指导骨干网对图像判别性区域的重点关注;3)在 Caltech-UCSD Birds 200(Wah 等,2011)、Stanford Cars(Krause 等,2013)和 FGVC-Aircraft(Maji 等,2013) 3 个公开细粒度图像数据集上,实验结果

表明ECASKD的分类精度优于同类自知识蒸馏方法，验证了所提方法的有效性。

1 相关工作

1.1 基于注意力的细粒度图像分类

注意力机制模仿人类的视觉系统使神经网络聚焦于图像需要重点关注的区域，已经被证明是增强模型性能的一种重要手段。Hu等人（2018）设计了“压缩—激励”（squeeze-and-excitation，SE）模块，通过全局平均池化对通道特征进行“压缩”，再通过“激励”步骤学习通道间非线性关系，实现对各通道特征权重的自适应调整。以SE模块为基础的压缩—激励网络（squeeze-and-excitation network，SENet）获得了2017年ImageNet大赛图像分类任务的冠军。此后，基于通道注意力的研究相继出现，通过改进或扩展SE模块以获得更好的模型性能。Woo等人（2018）提出了卷积块注意力模块（convolutional block attention module，CBAM），在采用通道注意力机制的基础上引入空间注意力机制，自适应调整通道权重的同时强化图像空间维度上的重要位置信息。

为了能使模型更好地关注到图像判别性区域，各种基于注意力的方法相继应用于细粒度图像分类任务研究。现有基于注意力的细粒度图像分类模型虽然有着较高的分类精度，但缺乏对模型参数量和计算量的考虑，难以部署到低资源设备上，落地应用受到极大限制。Wang等人（2020）提出了一种高效通道注意力（efficient channel attention，ECA）机制，在不降低维度的情况下进行逐通道全局平均池化，获取特征图的空间信息；通过一维卷积操作，获取每个通道及其近邻通道的跨通道全局信息。由于ECA仅通过执行一维卷积操作来捕获全局上下文信息，所以其能够在几乎不增加参数量和模型计算复杂度的同时较大幅度地提高模型性能。因此，本文探索引入高效通道注意力来改进小模型的细粒度图像分类性能。

1.2 经典知识蒸馏方法

为压缩模型的参数量、减少计算量并保持较好的精度性能，Hinton等人（2015）提出了知识蒸馏（knowledge distillation，KD）。KD作为一种迁移学习方法，其目标是通过迁移教师大网络的“暗知识”来提高学生小网络的表现。在此之后，Zagoruyko和Komodakis（2017）首次提出了基于注意力的知识蒸馏方法，该方法将注意力图定义为模型最关注的输入区域的空间注意力图，通过计算特征映射绝对值的和来获得注意力映射，将教师模型所得注意力图迁移（attention transfer，AT）到学生模型上以指导学生模型需要关注的区域。Guo等人（2023）证明了仅需通过转移类激活图（class activation map，CAM）（Zhou等，2016）就可以获得识别输入图像类判别性区域的能力，并在此基础上提出了基于类注意力迁移的知识蒸馏（class attention transfer based knowledge distillation，CAT-KD），该方法通过迁移教师类激活图的方式指导学生模型，具有很高的可解释性以及良好的精度。以上知识蒸馏方法需要教师大模型的支撑，且教师大模型本身学习也需要大量的计算和存储资源。

1.3 自知识蒸馏方法

为了进一步减少模型学习代价，研究者们提出了自知识蒸馏方法。与传统的知识蒸馏方法不同，自知识蒸馏在训练过程中逐步从学生网络本身提取知识以提高自身性能。

由于没有额外教师大模型指导，一个很自然的想法就是构造辅助网络来生成额外的知识（Zhang等，2021）。Ji等人（2021）将设计修改后的加权双向特征金字塔网络（bidirectional feature pyramid network，BiFPN）（Tan等，2020）作为辅助的“自教师”，以学生分类器的输出特征作为“自教师”网络的输入，“自教师”网络的输出特征为学生分类器提供特征指导。同时，Wang等人（2023）和Cho等人（2023）通过引入或提出辅助网络分支辅助骨干网进行优化，提升了模型的学习效果。

以上方法虽然展现了较好的图像分类性能，但忽视了对细粒度图像判别性区域的重点关注。基于注意力的自知识蒸馏模型SAD（Hou等，2019）在车道线检测任务中的表现验证了基于注意力自知识蒸馏方法的有效性，但直接将其运用到细粒度图像分类任务中所得精度表现并不优秀，主要原因也是对细粒度图像判别性区域的重点关注不够。为此，引入ECA注意力机制强化对细粒度图像判别性区域的关注，提出了融合高效通道注意力的细粒度图像分类自知识蒸馏学习方法。

2 ECASKD方法

2.1 方法框架与流程

图1展示了所提融合高效通道注意力的细粒度图像分类自知识蒸馏学习方法的框架。该框架主要包括由轻量级骨干网和自教师子网组成的自知识蒸馏网络，以及由分类损失L_{CE}、知识蒸馏损失L_{KD}和多级特征知识蒸馏损失L_{FD}组成的联合损失。

2.1.1 轻量级骨干网

轻量级骨干网负责对输入的图像进行特征提取，输出多尺度特征。该骨干网选用ResNet18作为基准，实现小的参数量和计算量。为了提升模型对图像判别性区域特征提取能力，引入高效通道注意力，设计了基于高效通道注意力的残差模块——ECA残差模块，并用其替换ResNet18（He等，2016）中的残差模块，构建了ECA-ResNet18轻量级骨干网。ECA-ResNet18骨干网与ResNet18架构一致，包括5个降低尺度、扩充通道数的操作，形成不同大小的多尺度特征。综合考虑计算代价和已有研究（Ji等，2021）经验，选择ECA-ResNet18后4个阶段输出特征构建多尺度特征。

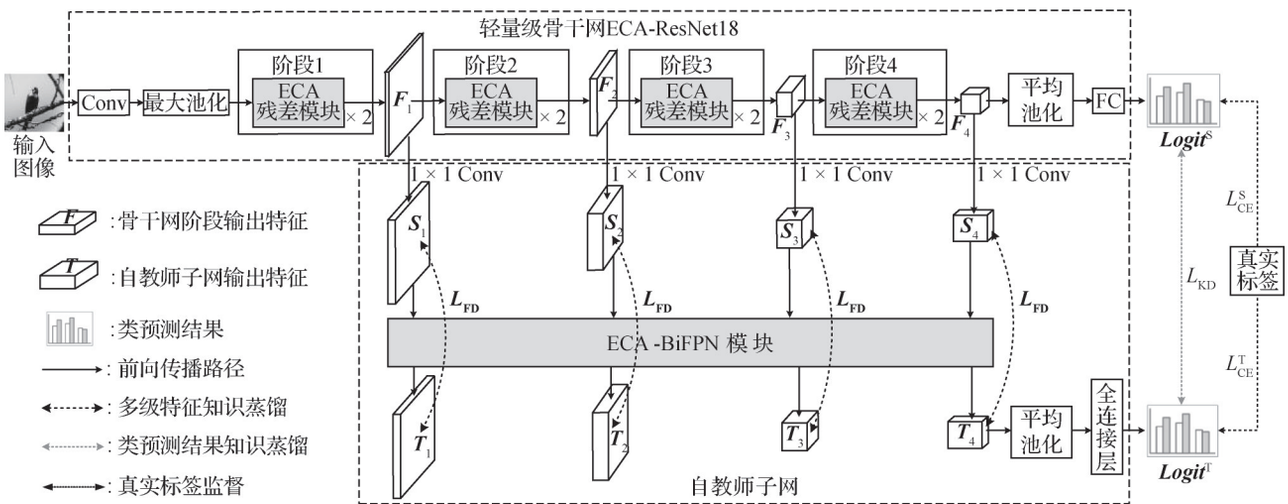


图1 所提ECASKD框架图

Fig. 1 Framework of the proposed ECASKD

给定一幅图像$\boldsymbol{x}$，ECA-ResNet18轻量级骨干网输出多尺度特征$\boldsymbol{F}=\{\boldsymbol{F}_i\}_{i=1}^4$，其中，$\boldsymbol{F}_i\in\mathbf{R}^{C_i^{\mathrm{F}}\times H_i^{\mathrm{F}}\times W_i^{\mathrm{F}}}$是第$i$个阶段输出的特征图，$C_i^{\mathrm{F}}$、$H_i^{\mathrm{F}}$和$W_i^{\mathrm{F}}$分别表示特征图$\boldsymbol{F}_i$的通道数、高和宽。关于ECA残差模块的详细过程见2.2.2节。

2.1.2 自教师子网

自教师子网负责对输入的多尺度特征进行特征融合，输出更加鲁棒的跨尺度特征，用于指导骨干网更好地学习特征。为了统一通道数便于融合，特征图$\boldsymbol{F}_i$先通过1×1卷积（convolution，Conv）以扩充通道数，具体过程为

$$\boldsymbol{S}_i=\boldsymbol{F}_i\otimes\boldsymbol{w}_i \tag{1}$$

式中，$\boldsymbol{S}_i$为输出结果，$\boldsymbol{F}_i$是第i个阶段输出的特征图，$\otimes$表示卷积操作，$\boldsymbol{w}_i\in\mathbf{R}^{2C_i^{\mathrm{F}}\times C_i^{\mathrm{F}}\times 1\times 1}$表示1×1卷积核参数矩阵。然后，将$\boldsymbol{S}=\{\boldsymbol{S}_i\}_{i=1}^4$输入ECA-BiFPN模块进行自适应加权融合，获得更丰富的融合特征$\boldsymbol{T}=\{\boldsymbol{T}_i\}_{i=1}^4$，ECA-BiFPN模块的详细融合过程见2.2.3节。

2.1.3 联合损失

联合损失由分类损失、知识蒸馏损失和多级特征知识蒸馏损失组成，计算为

$$L=L_{CE}^{\mathrm{S}}+L_{CE}^{\mathrm{T}}+aL_{KD}+\beta L_{FD} \tag{2}$$

式中，超参数a和β用于平衡不同损失。骨干网分类损失L_{CE}^{S}和自教师子网分类损失L_{CE}^{T}采用交叉熵损失函数，以真实标签监督网络学习，计算为

$$L_{CE}^{\mathrm{S}}=CorssEntropy(f_{\mathrm{S}}(\boldsymbol{x},\boldsymbol{w}_{\mathrm{S}}),\boldsymbol{y}) \tag{3}$$

式中，$CorssEntropy(\cdot)$表示交叉熵函数，$\boldsymbol{x}$表示输入的图像，$f_{\mathrm{S}}(\cdot)$表示参数为$\boldsymbol{w}_{\mathrm{S}}$的ECA-ResNet18网络，$\boldsymbol{y}$表示由图像真实标签构建的one-hot向量，$L_{CE}^{\mathrm{T}}$可由类似公式得到。

知识蒸馏损失的计算过程为

$$L_{\mathrm{KD}} = KL\left(\sigma\left(\frac{f_{\mathrm{T}}(\boldsymbol{x},\boldsymbol{w}_{\mathrm{T}})}{K}\right)\middle\|\,\sigma\left(\frac{f_{\mathrm{S}}(\boldsymbol{x},\boldsymbol{w}_{\mathrm{S}})}{K}\right)\right) \tag{4}$$

式中，$KL(\cdot)$代表KL(Kullback-Leibler)散度函数，$\sigma(\cdot)$表示softmax函数，$f_{\mathrm{T}}(\cdot)$表示参数为$\boldsymbol{w}_{\mathrm{T}}$的自教师子网，$\boldsymbol{Logit}^{\mathrm{T}} = f_{\mathrm{T}}(\boldsymbol{x},\boldsymbol{w}_{\mathrm{T}})$，$f_{\mathrm{S}}(\cdot)$表示参数为$\boldsymbol{w}_{\mathrm{S}}$的骨干网，$\boldsymbol{Logit}^{\mathrm{S}} = f_{\mathrm{S}}(\boldsymbol{x},\boldsymbol{w}_{\mathrm{S}})$，$K$代表KL散度函数的温度。该损失是基于类预测结果的知识蒸馏，能有效利用“自教师”网络输出的高质量“软标签”监督学习。

此外，为了提升对细粒度识别图像的判别性区域特征提取能力，提出了多级特征知识蒸馏损失，用于跨尺度特征对多尺度特征的蒸馏学习。该损失可以让骨干网更好地学习到跨尺度特征，具体计算过程见2.3节。

2.2 融合高效通道注意力的自知识蒸馏网络

所提自知识蒸馏网络包括轻量级骨干网和自教师子网。其中，轻量级骨干网主要由ECA残差模块组成；自教师子网主要是ECA-BiFPN模块。两者均由高效通道注意力模块构建而成。

2.2.1 高效通道注意力模块

高效通道注意力模块通过对原特征图进行注意力加权，能实现对细粒度图像判别性区域的关注，该模块的主要实现过程如图2所示。

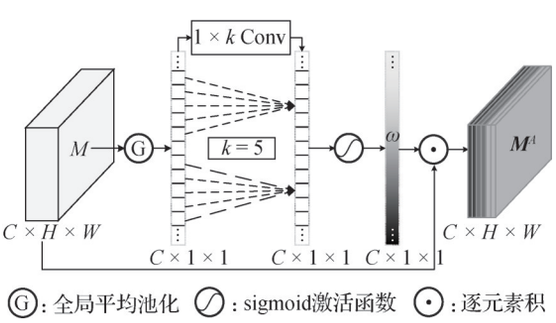


图2 ECA模块结构

Fig. 2 Structure of ECA module

首先，给定输入的特征图$\boldsymbol{M} \in \mathbf{R}^{C\times H\times W}$，执行全局平均池化操作，以获得基于全局空间信息的特征表达。然后，采用卷积核大小为k、填充(padding)为$\lfloor (k-1)/2 \rfloor$的一维卷积操作($1\times k$ Conv)，将当前通道与其k个邻近通道交互。接着，经过sigmoid激活函数得到各个通道的权重值$\boldsymbol{\omega}_i \in \boldsymbol{\omega}(i = 1, 2, \cdots, C)$，再将$\boldsymbol{\omega}_i$与原始输入特征图$\boldsymbol{M}$对应元素相乘得到最终输出特征图$\boldsymbol{M}^A \in \mathbf{R}^{C\times H\times W}$，具体为

$$\boldsymbol{M}^A = ECA(\boldsymbol{M}) = expand(\boldsymbol{\omega},\boldsymbol{M})\odot\boldsymbol{M} \tag{5}$$

式中，$expand(\boldsymbol{\omega},\boldsymbol{M})$表示将$\boldsymbol{\omega}$的尺寸扩展到与$\boldsymbol{M}$相同的尺寸大小，$\odot$表示逐元素积。更多权重及参数设置可详见ECA的相关文献(Wang等，2020)。

2.2.2 ECA残差模块

为了提升骨干网提取多尺度特征的能力，使其更好地聚焦于有效区域特征，设计了ECA残差模块结构，如图3所示。图中点状线框表示有下采样操作的ECA残差模块，下采样操作采用步长为2的3×3卷积实现；而虚线框表示无下采样操作的ECA残差模块，由两个步长为1的3×3卷积构成。

借鉴ResNet18的网络结构设计，ECA-ResNet18网络的第1阶段为两个无下采样操作的ECA残差模块，后3个阶段为有下采样操作的ECA残差模块和无下采样操作的ECA残差模块。相较于ResNet18中的残差模块，ECA残差模块在每个批归一化(batch normalization，BN)操作后引入ECA模块，并由两个ECA残差模块构成ECA-ResNet18骨干网的一个阶段，实现骨干网对多尺度特征的提取并增强其对图像判别性区域的关注。与自知识蒸馏方法中广泛应用的ResNet18相比，所提出的骨干网基于ECA残差模块，可以在保持轻量化和较高计算效率的情况下大幅提升模型提取多尺度特征的能力。

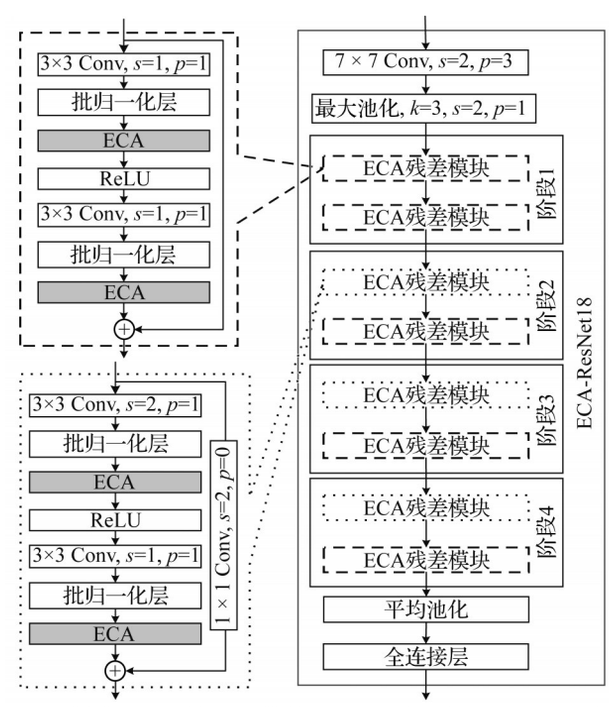


图3 ECA残差模块和ECA-ResNet18结构

Fig. 3 Structure of ECA-Residual block and ECA-ResNet18

2.2.3 ECA-BiFPN模块

考虑到骨干网输出的不同尺度特征的重要性差异，本文设计了ECA-BiFPN模块，在特征融合过程中对通道赋予权重，从而区分不同通道的特征对细粒度图像分类任务的贡献度。ECA-BiFPN流程如图4所示。图4中，$\boldsymbol{P}_i$和$\boldsymbol{Q}_i$ $(i=1,2,3,4)$是主要的中间计算结果，是计算的核心，下面将围绕两者展开描述。首先，对第$i+1$ $(i=1,2,3)$层特征$\boldsymbol{P}_{i+1}\in\mathbf{R}^{C_{i+1}\times H_{i+1}\times W_{i+1}}$进行上采样(upsample)，并用一个$1\times1$卷积缩减通道数；然后，采用ECA模块进行通道加权，所得特征图记做$\boldsymbol{P}_{i+1}^{\mathrm{U_A}}\in\mathbf{R}^{C_i\times H_i\times W_i}$，具体为

$$\boldsymbol{P}_{i+1}^{\mathrm{U\_A}}=ECA(\boldsymbol{P}_{i+1}^{\mathrm{U}}\otimes\boldsymbol{W}_{i+1}^{\mathrm{U}}) \tag{6}$$

式中，$\boldsymbol{P}_{i+1}^{\mathrm{U}}\in\mathbf{R}^{C_{i+1}\times H_i\times W_i}$表示对$\boldsymbol{P}_{i+1}$采用双线性插值上采样操作所得的特征图，$\boldsymbol{W}_{i+1}^{\mathrm{U}}\in\mathbf{R}^{C_i\times C_{i+1}\times1\times1}$表示$1\times1$卷积核参数。

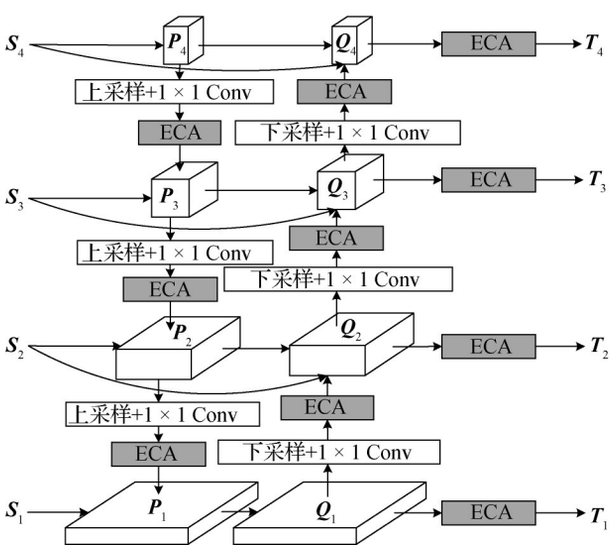


图4 ECA-BiFPN模块结构图

Fig. 4 Structure of ECA-BiFPN block

其次，将$\boldsymbol{P}_{i+1}^{\mathrm{U_A}}$与特征$\boldsymbol{S}_i$加权融合，得到自顶向下的融合特征$\boldsymbol{P}_i\in\mathbf{R}^{C_i\times H_i\times W_i}$，具体为

$$\boldsymbol{P}_i=DW(\boldsymbol{\lambda}_{i,1}^{\mathrm{P}}\times\boldsymbol{S}_i+\boldsymbol{\lambda}_{i,2}^{\mathrm{P}}\times\boldsymbol{P}_{i+1}^{\mathrm{U\_A}}) \tag{7}$$

式中，$DW(\cdot)$表示深度可分离卷积(Chollet, 2017)，$\boldsymbol{\lambda}_{i,1}^{\mathrm{P}}$和$\boldsymbol{\lambda}_{i,2}^{\mathrm{P}}$分别为$\boldsymbol{S}_i$和$\boldsymbol{P}_{i+1}^{\mathrm{U_A}}$的权重。特别地，位于最深层的$\boldsymbol{P}_4=\boldsymbol{S}_4$。

进一步地，对第$i-1$ $(i=2,3,4)$层特征$\boldsymbol{Q}_{i-1}\in\mathbf{R}^{C_{i-1}\times H_{i-1}\times W_{i-1}}$进行下采样(downsample)并用一个$1\times1$卷积扩充通道数；然后，采用ECA模块进行通道加权，所得特征图记做$\boldsymbol{Q}_{i-1}^{\mathrm{D_A}}\in\mathbf{R}^{C_i\times H_i\times W_i}$，具体为

$$\boldsymbol{Q}_{i-1}^{\mathrm{D\_A}}=ECA(\boldsymbol{Q}_{i-1}^{\mathrm{D}}\otimes\boldsymbol{W}_{i-1}^{\mathrm{D}}) \tag{8}$$

式中，$\boldsymbol{Q}_{i-1}^{\mathrm{D}}\in\mathbf{R}^{C_{i-1}\times H_i\times W_i}$表示对$\boldsymbol{Q}_{i-1}$采用$2\times2$大小的最大池化下采样操作所得的特征图，$\boldsymbol{W}_{i-1}^{\mathrm{D}}\in\mathbf{R}^{C_i\times C_{i-1}\times1\times1}$表示$1\times1$卷积核参数。

接着，将$\boldsymbol{Q}_{i-1}^{\mathrm{D_A}}$与自顶向下融合特征$\boldsymbol{P}_i$和特征$\boldsymbol{S}_i$加权融合，得到自底向上融合特征$\boldsymbol{Q}_i\in\mathbf{R}^{C_i\times H_i\times W_i}$，具体为

$$\boldsymbol{Q}_i=DW(\boldsymbol{\lambda}_{i,1}^{\mathrm{Q}}\times\boldsymbol{Q}_{i-1}^{\mathrm{D\_A}}+\boldsymbol{\lambda}_{i,2}^{\mathrm{Q}}\times\boldsymbol{P}_i+\boldsymbol{\lambda}_{i,3}^{\mathrm{Q}}\times\boldsymbol{S}_i) \tag{9}$$

式中，$DW(\cdot)$表示深度可分离卷积(Chollet, 2017)，$\boldsymbol{\lambda}_{i,1}^{\mathrm{Q}}$、$\boldsymbol{\lambda}_{i,2}^{\mathrm{Q}}$和$\boldsymbol{\lambda}_{i,3}^{\mathrm{Q}}$分别为$\boldsymbol{Q}_{i-1}^{\mathrm{D_A}}$、$\boldsymbol{P}_i$和$\boldsymbol{S}_i$的权重。特别地，位于最浅层的$\boldsymbol{Q}_1=\boldsymbol{P}_1$。

最后，在$\boldsymbol{Q}_i$的输出后添加ECA模块，将$\boldsymbol{Q}_i$作为ECA模块的输入得到最终输出特征$\boldsymbol{T}_i$，生成在特征融合的基础上关注网络判别性区域的跨尺度特征。

2.3 多级特征知识蒸馏损失函数

为了使骨干网能够从自教师子网中更好地学习到关注判别性区域的能力，在CAM(Zhou等，2016)和CAT-KD(Guo等，2023)的基础上，本文提出了多级特征知识蒸馏损失(multi-layer feature based knowledge distillation，MFKD)，计算过程如图5所示。

多级特征知识蒸馏损失函数为

$$L_{\mathrm{FD}}=\sum_{1\leqslant i\leqslant4}\sum_{1\leqslant j\leqslant C_i}\frac{1}{C_i}\left\|\frac{f(\boldsymbol{T}_i^j)}{\left\|f(\boldsymbol{T}_i^j)\right\|_2}-\frac{f(\boldsymbol{S}_i^j)}{\left\|f(\boldsymbol{S}_i^j)\right\|_2}\right\|_2^2 \tag{10}$$

式中，C_i表示特征$\boldsymbol{S}_i$和$\boldsymbol{T}_i$的通道数，$\boldsymbol{S}_i^j$、$\boldsymbol{T}_i^j\in\mathbf{R}^{H_i\times W_i}$分别表示$\boldsymbol{S}_i$、$\boldsymbol{T}_i$在通道$j$中的特征，$\|\cdot\|_2$表示L2距离，$f(\cdot)$表示自适应平均池化操作，计算过程为

$$f(\boldsymbol{S}_i)=\frac{W^p\times H^p}{W_i\times H_i}\sum_{x,y}\sum_{1\leqslant j\leqslant C_i}a_i^j(x,y) \tag{11}$$

式中，W^p、H^p分别表示经自适应平均池化操作后注

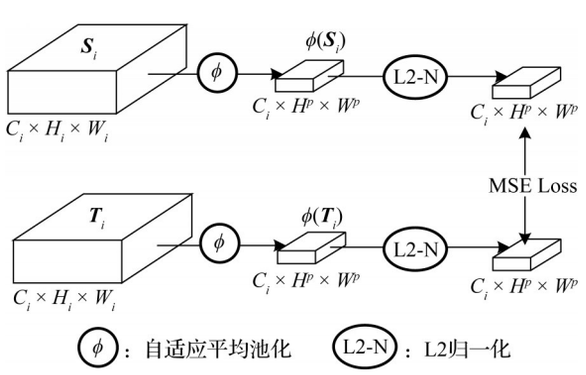


图5 多级特征知识蒸馏中单层损失计算过程

Fig. 5 The process of calculation of single layer loss in MFKD

意力图的宽和高，$a_i^j(x,y)$表示特征图在通道j中于空间位置(x,y)处的激活情况，$f(\boldsymbol{T}_i)$的计算方法与$f(\boldsymbol{S}_i)$类似。

3 实　验

3.1 数据集

实验采用3个公开的细粒度图像分类数据集，分别是CUB（caltech-UCSD birds 200）（Wah等，2011）、CAR（Stanford cars）（Krause等，2013）和AIR（FGVC-aircraft）（Maji等，2013）。其中，CUB是一个包含200种不同类型的鸟类数据集，共有11 788幅图像，其中5 994幅为训练图像，5 794幅为测试图像；CAR数据集由196类不同车型共16 185幅图像组成，其中8 144幅为训练图像，8 041幅为测试图像；AIR数据集包含100种不同型号的飞机，每种100幅图像，共10 000幅，其中6 667幅为训练图像，3 333幅为测试图像。

3.2 参数设置

借鉴方法FRSKD（feature refinement via self-knowledge distillation）（Ji等，2021）的实验设置，对CAR和AIR数据集采用的批处理大小为8，对CUB数据集的批处理大小为16。

在训练阶段，图像先随机剪裁到尺寸为224×224像素，再进行水平翻转。采用随机梯度下降算法进行参数更新，初始学习率为0.1，权重衰减为0.000 1。学习总迭代次数为300，采用余弦模拟退火策略调整学习率。在超参数方面，α取1，β取值细节可详见3.3.3超参数分析。

在测试阶段，仅使用骨干网而不采用任何辅助网络。先将输入图像尺寸缩放至256×256像素，再中心裁剪到224×224像素。

实验采用图像分类任务中常用的Top-1分类精度作为评测指标。

3.3 实验结果与分析

3.3.1 结果与分析

为了更客观全面地评测ECASKD，将其与基于数据增强的方法、基于辅助网络的方法和基于注意力的方法等共15种方法进行对比，各方法在3个数据集上的结果如表1所示。所有对比方法均基于

表1　自知识蒸馏模型细粒度图像分类精度对比

Table 1　Comparison of self-knowledge distillation models' fine-grained image classification accuracy

/%

类型	方法	CUB	CAR	AIR
	Baseline（He等，2016）	57.48	83.50	77.07
基于数据增强的方法	DDGSD（Xu和Liu，2019）	58.49	85.04	74.91
	TF-KD（Yuan等，2020）	57.44	83.59	76.76
	SLA-SD（Lee等，2020）	62.54	81.91*	75.58*
	CS-KD（Yun等，2020）	66.72	86.87	80.92
	CS-KD+Mixup（Yang等，2022）	69.29	87.10	81.13
	MixSKD（Yang等，2022）	72.15	89.17	82.95
基于辅助网络的方法	BYOT（Zhang等，2019）	58.66	85.36	79.32
	DKS（Sun等，2019）	63.72	86.13	79.69
	FRSKD（Ji等，2021）	65.39	84.73	78.85
	ART-KD（Cho等，2023）	67.23	/	/
	DRG+DSR（Wang等，2023）	69.87*	86.18*	79.83*
联合辅助网络和数据增强的方法	FRSKD+SLA（Ji等，2021）	67.80	<u>89.55</u>*	<u>83.98</u>*
	FRSKD+Mixup（Yang等，2022）	67.98	86.25	79.97
	ART-KD+SLA（Cho等，2023）	<u>73.41</u>	/	/
基于注意力的方法	SAD（Hou等，2019）	55.51	82.94	73.62
	ECASKD（本文）	**76.04**	**91.11**	**87.64**

注：加粗、下划线字体表示各列最优、次优结果。“*”表示该方法在数据集上自测的结果，“/”表示该方法未报告结果且未公开代码。

ResNet18骨干网,其中结果数据来自原文献报告的结果,对于SLA-SD(self-supervised label augmentation based self-distillation)(Lee等,2020)、FRSKD+SLA(self-supervised label augmentation)(Ji等,2021)和DRG+DSR(distillation with reverse guidance adds distillation with shape-wise regularization)(Wang等,2023)方法中标注"*"的数据是本文利用其官方代码进行评测后得到的结果。

从表1可以发现,所提ECASKD方法在CUB、CAR和AIR 3个数据集的细粒度图像分类任务中,分别获得76.04%、91.11%和87.64%的Top-1分类精度。对比不同类的方法:1)与基于数据增强的方法相比,ECASKD较最先进的MixSKD(self- knowledge distillation from image mixture)(Yang等,2022)在CUB、CAR和AIR 3个数据集的细粒度图像分类任务中Top-1分类精度分别提升了3.89%、1.94%和4.69%。2)与基于辅助网络的方法相比,ECASKD较最先进方法DRG+DSR(Wang等,2023)在CUB、CAR和AIR 3个数据集的细粒度图像分类任务中Top-1分类精度分别提升了6.17%、4.93%和7.81%。3)与联合辅助网络和数据增强的方法相比,ECASKD较其中最优方法在CUB、CAR和AIR 3个数据集的细粒度图像分类任务中Top-1分类精度分别提高2.63%、1.56%和3.66%。可见ECASKD即使不采用数据增强(Lee等,2020;Zhang等,2018),也有着比联合辅助网络和数据增强方法更好的细粒度图像分类性能。4)与基于注意力的自知识蒸馏方法SAD(Hou等,2019)相比,ECASKD在CUB、CAR和AIR 3个细粒度图像分类任务中Top-1分类精度分别提升了20.53%、8.17%和14.02%。综上所述,在CUB、CAR和AIR 3个数据集的细粒度图像分类任务中,本文所提方法的Top-1分类精度均优于其他15种自知识蒸馏方法,获得了更好的细粒度图像分类性能。

3.3.2 模块消融实验

表2展示了所提方法的消融实验的结果,其中第1行表示未采用任何注意力机制和自知识蒸馏学习方法的ResNet18。

表2 消融实验结果

Table 2 Results of ablation experiments

/%

残差模块	ECA残差模块	BiFPN模块	ECA-BiFPN模块	AT	MFKD	CUB	CAR	AIR
√	–	–	–	–	–	57.48	83.50	77.07
√	–	√	–	√	–	65.39	84.73	78.85
	√	√	–	√	–	72.08	88.02	85.84
√	–	–	√	√	–	74.75	89.23	86.62
√	–	√	–	–	√	74.39	87.97	86.35
–	√	–	√	√	–	74.94	90.88	87.49
–	√	–	√	–	√	**76.04**	**91.11**	**87.64**

注:加粗字体表示各列最优结果。"√"表示使用该模块,"–"表示未使用该模块。

对比第2、3行的实验结果显示,仅将原ResNet18中残差模块替换为ECA残差模块所得模型在3个数据集上分类精度分别提高了6.69%、3.29%和6.99%,验证了所设计ECA残差模块的有效性。

对比第2、4行的实验结果显示,仅将BiFPN模块替换为ECA-BiFPN模块所得模型在3个数据集上分类精度分别提高了9.36%、4.50%和7.77%,验证了所设计ECA-BiFPN模块的有效性。

对比第2、5行的实验结果显示,仅将AT(Zagoruyko等,2017)替换为多级特征知识蒸馏损失之后的所得方法在3个数据集上分类精度分别提高了9.00%、3.24%和7.50%,验证了所提多级特征知识蒸馏损失的有效性。

对比第2、6行的实验结果显示,同时采用ECA残差模块和ECA-BiFPN模块的方法较均不采用两者的方法在3个数据集上分类精度分别提高了9.55%、6.15%和8.64%,验证了引入ECA模块的有

效性。

对比第2、7行的实验结果显示，同时采用ECA残差模块、ECA-BiFPN模块和多级特征知识蒸馏的ECASKD方法在3个数据集上分类精度均达到了最优，较均不采用上述3种方法的模型分别提高了10.65%、6.38%和8.79%。

3.3.3　超参数分析

所提方法使用交叉熵损失、知识蒸馏损失以及多级特征知识蒸馏损失共同指导网络学习。借鉴方法FRSKD（Ji等，2021）和CAT-KD（Guo等，2023）的经验，将超参数α设置为1。对于β分别取$\beta\in\{0,10,50,100,150,200,250,300,350,400\}$，并保持除超参数$\beta$变化外，其他所有设置与3.2节相同，对应实验结果见图6。

图6展示了β取不同值时在3个数据集上的Top-1分类精度，当β取200时模型在CUB上取得最高Top-1分类精度为76.04%，当β取100时模型在CAR上取得最高Top-1分类精度为91.11%，当β取50时模型在AIR上取得最高Top-1分类精度为87.64%。从图中可以发现，ECASKD对超参数取值具有鲁棒性，但不同的β取值仍有一定表现差异。

对产生这种结果的原因进行分析，当β取值过小时，主干网主要根据真实标签和自教师子网输出的类预测结果进行训练，不能有效地学习到自教师子网提取图像判别性区域特征的能力；当β取值过大时，主干网主要根据自教师子网输出的跨尺度特征进行训练，而自教师子网的输出特征可能包含了错误提取图像判别性区域特征的信息，骨干网更易受到自教师子网“过度自信”的负面影响。所以需要正确地使用超参数来平衡3种损失，以保证模型获得最佳分类精度。

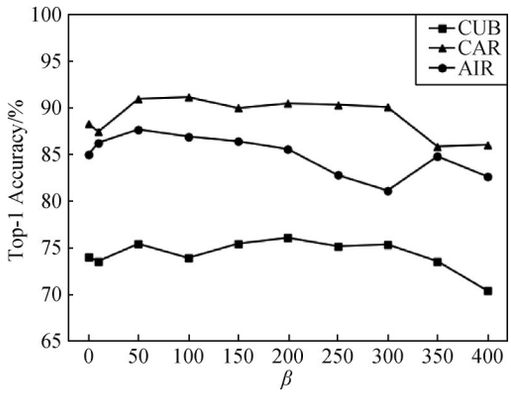


图6　超参数β分析

Fig. 6　Analysis of the hyperparameter β

3.3.4　模型复杂度分析

为了准确评价所提方法在模型规模大小与精度性能上的平衡性，将所提方法与4种当前主流模型方法在参数量、浮点运算量和Top-1分类精度上进行了对比，如表3所示。其中浮点运算量采用大小为224×224像素的图像计算得出。

表3　不同大小模型的分类精度

Table 3　Classification accuracy for models of varying sizes

方法	骨干网	参数量/M	浮点运算量/G	CUB/%	CAR/%	AIR/%
ResNet18（He等，2016）	ResNet18	11.5	1.8	57.5	83.5	77.0
ECASKD	ECA-ResNet18	11.9	2.0	76.0	91.1	87.6
ResNet50（He等，2016）	ResNet50	24.7	4.1	85.5	91.7	90.1
Vision Transformer（Dosovitskiy等，2020）	ViT-Base	86.6	17.5	**90.8**	92.5	90.0
Swin Transformer（Liu等，2021）	Swin-Transformer-B	88.0	15.4	89.7	**94.8**	**92.9**

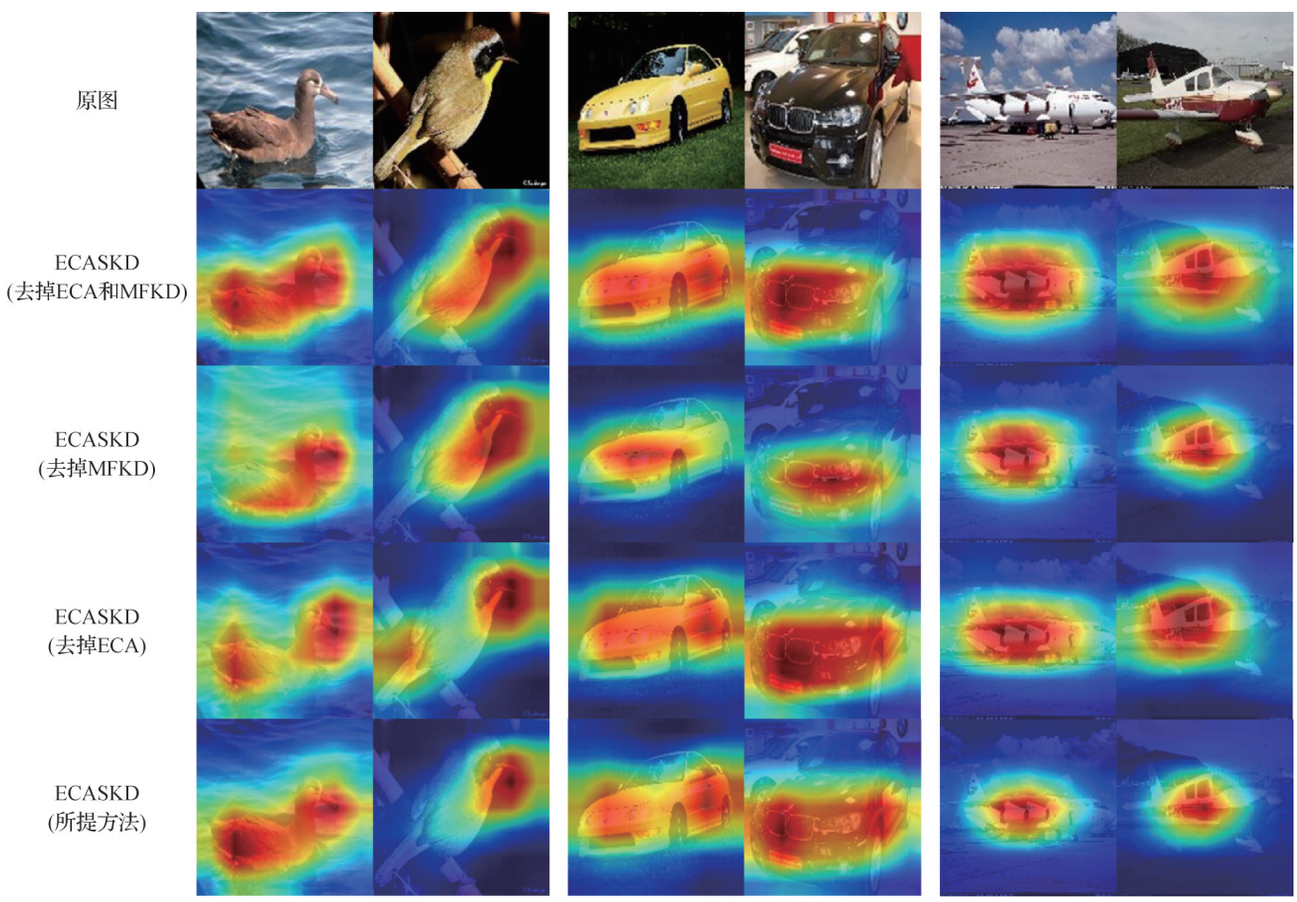
注：加粗字体表示各列最优结果。

从表3中可以观察到，与ResNet18相比，所提方法采用的ECA-ResNet18骨干网能够在少量增加参数量和计算量的基础上大幅提升分类精度。与较大规模的ResNet50相比，所提方法在参数量和计算量方面不足ResNet50的1/2，但在CAR数据集上的分类精度与ResNet50相差仅为0.6%。与更大规模的ViT-Base（vision Transformer）和Swin-Transformer-B相比，所提方法在参数量和计算量方面约为两者的1/8，在CAR和AIR数据集上的分类精度比最优的Swin-Transformer-B低3.7%和5.3%，而在CUB数据集上还存在较大差距，比最优的ViT-Base低约14.8%。综合各模型的复杂度和分类精度表现来

看，所提方法模型的分类精度虽然较最佳方法模型的分类精度有一定差距，但其参数量和浮点运算量较最佳方法的参数量和浮点运算量低得多，且能够在少量增加参数量和浮点运算量的基础上大幅提高分类精度。因此，所提方法能够更好地平衡计算复杂度和分类精度需求，更适合落地资源受限设备上。

3.3.5　注意力可视化

为了观察所提模型对目标判别性区域的关注情况，分别对CUB、CAR和AIR 3个数据集中的部分测试图像的类激活图进行了展示，如图7所示。第1行是6幅原始图像，第2行展示了同时删除ECA模块和未采用MFKD损失的类激活图效果，第3行展示了仅引入ECA模块但未采用MFKD所得方法的类激活图效果，第4行表示仅采用MFKD而未引入ECA模块所得方法的类激活图效果，最后一行则为同时引入ECA模块和采用MFKD的本文方法的类激活图。



(a) 鸟类激活图　(b) 汽车类激活图　(c) 飞机类激活图

图7　不同模型类激活图对比

Fig. 7　Comparison of class activation maps for different models

((a) class activation maps of birds; (b) class activation maps of cars; (c) class activation maps of aircrafts)

对于细粒度图像分类任务，提取图像中目标对象具有类判别性关键局部区域的特征至关重要(Wei等，2019)。因此，相比于感兴趣区域范围大小，使最感兴趣范围准确定位至具有类判别性关键局部区域对细粒度图像分类模型更为重要。从图7可以观察到，相较于第2行的类激活图效果，所提方法的类激活图虽然整体感兴趣区域范围更小，但最为关注的范围能够更加精确到目标的判别性区域位置，如鸟的喙、汽车的车灯、飞机的机翼等。相较于第3、4行的类激活图效果，所提方法的类激活图不会丢失需要关注的区域，也不会过多关注非重点的区域。因此，本文所提方法具有能够更加精确地关注到图像类判别性区域的能力。

4 结 论

本文提出了一种融合高效通道注意力的细粒度图像分类自知识蒸馏学习方法。该方法将高效通道注意力机制嵌入自知识蒸馏框架中，设计了ECA残差模块并构建ECA-ResNet18骨干网和ECA-BiFPN模块，实现不同尺度特征的有效提取和融合，并提出基于多级特征知识蒸馏的联合损失函数增强模型对图像判别性区域的关注。实验表明，所提方法在3个细粒度图像分类数据集上较已有自知识蒸馏方法表现出了更优的分类性能。

所提自知识蒸馏细粒度图像分类方法以较低的计算耗费取得了不错的性能，下一步将面向低资源设备继续优化模型方法落地应用；同时，与非自知识蒸馏的相关方法相比，所提方法还存在一定的性能差距，有待进一步探索提升。

参考文献（References）

Chen P G, Liu S, Zhao H S and Jia J Y. 2021. Distilling knowledge via knowledge review//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 5008-5017 [DOI: 10.1109/CVPR46437.2021.00497]

Cho Y, Ham G, Lee J H and Kim D. 2023. Ambiguity-aware robust teacher (ART): enhanced self-knowledge distillation framework with pruned teacher network. Pattern Recognition, 140: #109541 [DOI: 10.1016/j.patcog.2023.109541]

Chollet F. 2017. Xception: deep learning with depthwise separable convolutions//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 1800-1807 [DOI: 10.1109/CVPR.2017.195]

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J and Houlsby N. 2021. An image is worth 16 × 16 words: Transformers for image recognition at scale [EB/OL]. [2023-12-12]. https://arxiv.org/pdf/2010.11929.pdf

Gou J P, Yu B S, Maybank S J and Tao D C. 2021. Knowledge distillation: a survey. International Journal of Computer Vision, 129(6): 1789-1819 [DOI: 10.1007/s11263-021-01453-z]

Guo Z Y, Yan H N, Li H and Lin X D. 2023. Class attention transfer based knowledge distillation//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 11868-11877 [DOI: 10.1109/CVPR52729.2023.01142]

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Hinton G, Vinyals O and Dean J. 2015. Distilling the knowledge in a neural network [EB/OL]. [2023-12-12]. https://arxiv.org/pdf/1503.02531.pdf

Hou Y N, Ma Z, Liu C X and Loy C C. 2019. Learning lightweight lane detection CNNs by self attention distillation//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 1013-1021 [DOI: 10.1109/ICCV.2019.00110]

Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]

Huang Z H, Yang S Z, Lin W, Ni J, Sun S L, Chen Y W and Tang Y. 2022. Knowledge distillation: a survey. Chinese Journal of Computers, 45(3): 624-653 (黄震华, 杨顺志, 林威, 倪娟, 孙圣力, 陈运文, 汤庸. 2022. 知识蒸馏研究综述. 计算机学报, 45(3): 624-653) [DOI: 10.11897/SP.J.1016.2022.00624]

Ji M, Shin S, Hwang S, Park G and Moon I C. 2021. Refine myself by teaching myself: feature refinement via self-knowledge distillation//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 10664-10673 [DOI: 10.1109/CVPR46437.2021.01052]

Jiang L Y, Zheng Y F, Chen C, Li G H and Zhang W J. 2023. Review of optimization methods for supervised deep learning. Journal of Image and Graphics, 28(4): 963-983 (江铃燚, 郑艺峰, 陈澈, 李国和, 张文杰. 2023. 有监督深度学习的优化方法研究综述. 中国图象图形学报, 28(4): 963-983) [DOI: 10.11834/jig.211139]

Krause J, Stark M, Deng J and Li F F. 2013. 3D object representations for fine-grained categorization//Proceedings of 2013 IEEE International Conference on Computer Vision Workshops. Sydney, Australia: IEEE: 554-561 [DOI: 10.1109/ICCVW.2013.77]

Krizhevsky A, Sutskever I and Hinton G. 2012. Imagenet classification with deep convolutional neural networks//Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, USA: Curran Associates Inc.: 1097-1105

Lee H, Hwang S J and Shin J. 2020. Self-supervised label augmentation via input transformations//Proceedings of the 37th International Conference on Machine Learning. Virtual Event, JMLR.org: 5714-5724

Liu Z, Lin Y T, Cao Y, Hu H, Wei Y X, Zhang Z, Lin S and Guo B N. 2021. Swin Transformer: hierarchical vision Transformer using shifted windows//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 10012-

10022 [DOI: 10.1109/ICCV48922.2021.00986]

Maji S, Rahtu E, Kannala J, Blaschko M and Vedaldi A. 2013. Fine-grained visual classification of aircraft [EB/OL]. [2023-12-12]. https://arxiv.org/pdf/1306.5151.pdf

Mirzadeh S I, Farajtabar M, Li A, Levine N, Matsukawa A and Ghasemzadeh H. 2020. Improved knowledge distillation via teacher assistant//Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA: AAAI Press: 5191-5198 [DOI: 10.1609/aaai.v34i04.5963]

Romero A, Ballas N, Kahou S E, Chassang a, Gatta C and Bengio Y. 2015. Fitnets: hints for thin deep nets [EB/OL]. [2023-12-12]. https://arxiv.org/pdf/1412.6550.pdf

Si Z F and Qi H G. 2023. Survey on knowledge distillation and its application. Journal of Image and Graphics, 28(9): 2817-2832 (司兆峰, 齐洪钢. 2023. 知识蒸馏方法研究与应用综述. 中国图象图形学报, 28(9): 2817-2832) [DOI: 10.11834/jig.220273]

Sun D W, Yao A B, Zhou A J and Zhao H. 2019. Deeply-supervised knowledge synergy//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 6997-7006 [DOI: 10.1109/CVPR.2019.00716]

Tan M X, Pang R M and Le Q V. 2020. EfficientDet: scalable and efficient object detection//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 10781-10790 [DOI: 10.1109/CVPR42600.2020.01079]

Tian Y L, Krishnan D and Isola P. 2020. Contrastive representation distillation [EB/OL]. [2023-12-12]. https://arxiv.org/pdf/1910.10699.pdf

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, N. Gomez A, Kaiser Ł and Polosukhin I. 2017. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 5998-6008

Wah C, Branson S, Welinder P, Perona P and Belongie S. 2011. The Caltech-UCSD Birds-200-2011 Dataset. Pasadena: California Institute of Technology

Wang Q L, Wu B G, Zhu P F, Li P H, Zuo W M and Hu Q H. 2020. ECA-Net: efficient channel attention for deep convolutional neural networks//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 11534-11542 [DOI: 10.1109/CVPR42600.2020.01155]

Wang L and Yoon K J. 2022. Knowledge distillation and student-teacher learning for visual intelligence: a review and new outlooks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(6): 3048-3068 [DOI: 10.1109/TPAMI.2021.3055564]

Wang X C, Han P C and Guo L. 2023. Lightweight self-knowledge distillation with multi-source information fusion [EB/OL]. [2023-12-12]. https://arxiv.org/pdf/2305.09183.pdf

Wei X S, Song Y Z, Aodha O M, Wu J X, Peng Y X, Tang J H, Yang J and Belongie S. 2022. Fine-grained image analysis with deep learning: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(12): 8927-8948 [DOI: 10.1109/TPAMI.2021.3126648]

Wei X S, Wu J and Cui Q. 2019. Deep learning for fine-grained image analysis: a survey [EB/OL]. [2023-12-12]. https://arxiv.org/pdf/1907.03069.pdf

Wei X S, Xu Y Y and Yang J. 2022. Review of webly-supervised fine-grained image recognition. Journal of Image and Graphics, 27(7): 2057-2077 (魏秀参, 许玉燕, 杨健. 2022. 网络监督数据下的细粒度图像识别综述. 中国图象图形学报, 27(7): 2057-2077) [DOI: 10.11834/jig.210188]

Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]

Xu T B and Liu C L. 2019. Data-distortion guided self-distillation for deep neural networks//Proceedings of the 33rd AAAI Conference on Artificial Intelligence. Honolulu, USA: AAAI Press: 5565-5572 [DOI: 10.1609/aaai.v33i01.33015565]

Yang C G, An Z L, Zhou H L, Cai L H, Zhi X, Wu J W, Xu Y and Zhang Q. 2022. MixSKD: self-knowledge distillation from mixup for image recognition//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 534-551 [DOI: 10.1007/978-3-031-20053-3_31]

Yuan L, Tay F E, Li G L, Wang T and Feng J S. 2020. Revisiting knowledge distillation via label smoothing regularization//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 3903-3911 [DOI: 10.1109/CVPR42600.2020.00396]

Yun S, Park J, Lee K and Shin J. 2020. Regularizing class-wise predictions via self-knowledge distillation//Proceedings of 2010 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 13876-13885 [DOI: 10.1109/CVPR42600.2020.01389]

Zagoruyko S and Komodakis N. 2017. Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer [EB/OL]. [2023-12-12]. https://arxiv.org/pdf/1612.03928.pdf

Zhang H Y, Cisse M, Dauphin Y N and Lopez P D. 2018. Mixup: beyond empirical risk minimization [EB/OL]. [2023-12-12]. https://arxiv.org/pdf/1710.09412.pdf

Zhang L F, Bao C L and Ma K S. 2021. Self-distillation: towards efficient and compact neural networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(8): 4388-4403 [DOI: 10.1109/TPAMI.2021.3067100]

Zhang L F, Song J B, Gao A N, Chen J W, Bao C L and Ma K S. 2019. Be your own teacher: improve the performance of convolutional

neural networks via self-distillation//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 3713-3722 [DOI: 10.1109/ICCV.2019.00381]

Zhao B R, Cui Q, Song R J, Qiu Y Y and Liang J J. 2022. Decoupled knowledge distillation//Proceedings of 2022 IEEE/CVF International Conference on Computer Vision. New Orleans, USA: IEEE: 11953-11962 [DOI: 10.1109/CVPR52688.2022.01165]

Zhou B L, Khosla A, Lapedriza A, Oliva A and Torralba A. 2016. Learning deep features for discriminative localization//Proceedings of 2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 2921-2929 [DOI: 10.1109/CVPR.2016.319]

作者简介

张睿，男，教授，主要研究方向为数据工程和信息融合。
E-mail：3959966@qq.com
王家宝，通信作者，男，副教授，主要研究方向为计算机视觉与机器学习。E-mail：jiabao_1108@163.com
陈瑶，女，硕士研究生，主要研究方向为模型轻量化方法。
E-mail：1916664304@qq.com
李阳，男，副教授，主要研究方向为计算机视觉与图像检索。
E-mail：solarleeon@outlook.com
张旭，男，副教授，主要研究方向为模型可解释性。
E-mail：494848647@qq.com