基于金字塔知识的自蒸馏HRNet目标分割方法

郑云飞[1,2,3],王晓兵[1,2],张雄伟[1],曹铁勇[1],孙 蒙[1]
(1. 陆军工程大学指挥控制工程学院,江苏南京 210007; 2. 陆军炮兵防空兵学院,安徽合肥 230031;
3. 安徽省偏振成像与探测重点实验室,安徽合肥 230031)

摘 要: 知识蒸馏能有效地将教师网络的表征能力迁移到学生网络,无须改变网络结构即可提升网络的性能.因此,在性能优异的目标分割主干网HRNet(High-Resolution Net)中构建自蒸馏学习模型具有重要意义.针对HRNet并行结构中深层与浅层信息充分融合导致直接蒸馏难以实现的挑战,本文提出一种基于多尺度池化金字塔的结构化自蒸馏学习模型:在HRNet分支结构中引入多尺度池化金字塔表示模块,提升网络的知识表示和学习能力;构造"自上而下"和"一致性"两种蒸馏模式;融合交叉熵损失、KL(Kullback-Leibler)散度损失和结构化相似性损失进行自蒸馏学习.在四个包含显著性目标和伪装目标的分割数据集上的实验表明:本文模型在不增加资源开销的前提下,有效提升了网络的目标分割性能.

关键词: 自蒸馏学习;并行结构网络;多尺度池化金字塔;结构化相似性;目标分割
基金项目: 国家自然科学基金(No.61801512, No.62071484);江苏省自然科学基金(No.BK20180080)
中图分类号: TP391;TP183 **文献标识码:** A **文章编号:** 0372-2112(2023)03-0746-11
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20210169

The Self-Distillation HRNet Object Segmentation Based on the Pyramid Knowledge

ZHENG Yun-fei[1,2,3], WANG Xiao-bing[1,2], ZHANG Xiong-wei[1], CAO Tie-yong[1], SUN Meng[1]
(1. *Institude of Command and Control Engineering, Army Engineering University of PLA, Nanjing, Jiangsu* 210007, *China*;
2. *The Army Artillery and Defense Academy of PLA, Hefei, Anhui* 210031, *China*;
3. *The Key Laboratory of Polarization Imaging Detection Technology, Hefei, Anhui* 210031, *China*)

Abstract: The knowledge distillation can effectively transfer the representation ability of a teacher network to a student network, and improve the performance of the network without changing the network structure. Therefore, it is of great significance to construct a self-distillation learning model in the backbone network of the HRNet (High-Resolution Net)with an excellent performance in the object segmentation tasks. Aiming to the challenge that parallel integration architecture of deep and shallow information in HRNet makes direct distillation difficult to achieve, a structured self-distillation learning framework based on multi-scale pooling pyramid is proposed in this paper. Firstly, the multiscale pooling pyramid feature modules are introduced into the branch structure in the HRNet to improve knowledge representation and learning ability. Secondly, the top-down and consistency distillation modes are constructed. Meanwhile the cross entropy loss, KL (Kullback-Leibler)divergence loss and structural similarity loss are combined for the self-distillation learning framework. The experiments on four segmentation datasets including saliency and camouflaged objects demonstrate that the proposed model improves the performance of the object segmentation of the network without increasing resource costs.

Key words: knowledge distillation; parallel network; multi-scale pooling pyramid; structural similarity; object segmentation
Foundation Item(s): National Natural Science Foundation of China (No.61801512, No.62071484); Natural Science Foundation of Jiangsu Province (No.BK20180080)

收稿日期:2021-01-26;修回日期:2022-05-15;责任编辑:梅志强

1 引言

近年来,深度神经网络在图像识别、目标检测、图像分割等计算机视觉领域[1~8]取得了巨大成功. 然而,以往的研究表明:要取得更好性能,往往要构建规模更大、结构更复杂的深度神经网络模型. 例如,ResNet(Residual Net)[1]、DenseNet(Dense Net)[6]需要增加网络的深度提升其分类能力. WRNet(Wide Residual Net)[9]需要增加残差模块的通道数(即网络的宽度),提升其性能.

网络深度或宽度的增加意味着网络运行需要消耗更多的计算和存储资源,然而许多应用场景要求模型既要有高准确率又有低复杂度,如手持终端、嵌入式设备、边缘计算设备等. 因此,如何获得性能优良的轻量级模型成为人工智能研究的一个热点问题. 2015年,深度学习领域的先驱Hinton教授提出知识蒸馏方法[10],通过知识蒸馏过程赋予轻量级模型更强的特征表示能力. 所谓知识蒸馏是指将“教师模型”(已完成学习的大规模参数模型)输出的“软目标”(模型输出的类别概率向量)作为监督信息训练“学生模型”(待学习的轻量级模型),通过“软目标”中“暗知识”(Dark knowledge)的传递,实现模型表征能力的迁移.

Hinton教授提出的知识蒸馏方法引起了学术界的广泛关注,许多研究者拓展了这一知识迁移方法,并将其应用到多种计算机视觉任务中. 如Adriana、Zagoruyko等人[11,12]将网络中间层特征的匹配程度加入到知识蒸馏的学习目标中,提升知识蒸馏的效果. Zhang、Chen等人[13,14]为提升知识蒸馏的效率,提出了将教师模型与学生模型同步学习的在线蒸馏方法. 随着研究的深入,许多研究者探索出了不需要教师网络的自蒸馏方法[15,16],即在网络自身结构中构建适当的蒸馏结构,将网络不同层或训练不同阶段的输出信息作为监督信号实现蒸馏学习,能在不增加存储和计算资源开销的前提下提升网络的性能.

2019年,Zhang[15]等人抽取残差网络的4组中间层特征,并通过瓶颈层、全连接层、Softmax函数级联输出4组类别概率. 将网络末端的概率输出作为教师端,其他组输出作为学生端,构建自上而下蒸馏结构,结合KL(Kullback-Leibler)散度、L2距离和交叉熵损失实现自蒸馏学习,提升残差网络的图像识别性能. 2019年,Yang等人[16]通过网络训练不同轮次中产生的模型快照构建自蒸馏结构. 将前一轮训练产生的网络作为教师端、下一轮待训练的网络作为学生端,以KL散度作为蒸馏损失,利用循环学习率策略实现蒸馏学习,该方法有效提升了残差网络、密集网络的图像识别性能. 2020年,Li等人[17]提取残差网络、密集网络的中间层和输出层特征,用类似文献[15]的结构将其输出为类别概率,将几组概率输出互相作为蒸馏学生端和教师端,并利用KL散度和交叉熵损失实现一致性蒸馏,促使网络中间层特征相互模仿、融合,从而实现网络目标识别性能的提升.

以上自蒸馏方法研究均面向图像识别任务,而对于目标分割这一像素级密集预测问题,尚未见有成熟的研究. 为此,本文重点研究目标分割中的自蒸馏学习方法,针对目前在目标分割任务上性能优异的主干网HRNet(High-Resolution Net)[18],探索构建有效的自蒸馏学习结构,在不增加资源开销的前提下提升其目标分割性能. 更进一步地,本文通过分析特征学习的过程,揭示自蒸馏学习的作用机制.

在HRNet中构建有效自蒸馏学习模型的挑战在于:(1)HRNet独特的并行分支结构实现了浅层特征与深层特征的有效融合,现有的蒸馏结构[15]在这种并行架构网络中难以奏效. (2)图像分割是像素级预测,在知识蒸馏过程中用于传递“暗知识”的分类概率层的特征维度远大于图像分类任务,因此更难以实现有效地蒸馏过程.

本文提出一种基于多尺度池化金字塔的结构化自蒸馏学习模型,主要创新在于:(1)在HRNet各分支的特征输出端构建特征金字塔模块,提升网络的特征表示能力. (2)分别将每个子分支与主分支视为教师-学生蒸馏对象,构建基于一致性和自上而下模式的自蒸馏结构. (3)在蒸馏学习常用的KL散度和交叉熵损失基础上引入结构化相似性损失,有效地捕捉分割结果的结构化差异,在分割任务中实现更有效的自蒸馏学习. 在两个伪装目标分割数据集COD(Camouflaged Object Detection)[19]、CPD(Camouflaged People Detection)[20,21]和两个显著性目标分割数据集DUT-OMRON(Dalian University of Technology-OMRON)[22]、PASCAL-S(Pattern Analysis, Statistical Modelling and Computational Learning- Subset)[23]的实验表明,本文提出的自蒸馏学习模型有效提升了HRNet基准模型的目标分割性能,并且没有增加计算和存储资源开销.

2 HRNet结构分析

目标分割常用的主干网络如VGG(Visual Geometry Group)[24]、ResNet[1]、DenseNet[6]均为串行结构,其优势在于通过卷积层、激活层、池化层等计算单元的堆叠在空间上对图像进行层次化表示,能更有效地对图像的语义信息进行建模. 串行结构网络的这一特性决定了:越深层的特征包含越多的目标语义信息,但由于网络中带步长卷积、池化运算带来的下采样效应,越深层的特征图尺寸越小[25]. 目标分割任务需要包含完整语义信息且密集的特征图,才能获得高精确度的像素级预测结果.

因此,串行网络在语义信息与特征图尺寸之间的矛盾是制约其在分割任务上性能提升的主要因素.

许多目标分割模型将新的连接结构嵌入到串行主干网中,试图解决以上矛盾,提升目标分割性能. 如DeconvNet(Deconvolution Net)[26]、U-Net(U-Shape Net)[27]、SegNet(Segmentation Net)[28]等.

HRNet直接从架构入手,构建具有并行结构的主干网,获得高分辨率语义特征. 图1所示为HRNet的网络结构:在经过两次带步长卷积运算后,特征图分辨率降为原图的1/4. 在此之后随着深度加深,网络分出四个分支. 这四个分支抽取的特征类似于串行结构网络不同阶段的中间层特征,分辨率分别为原图的1/4、1/8、1/16、1/32. 与经典串行结构网络不同的是,HRNet设计了四个分支的交互结构,使得目标语义特征与底层特征充分融合. 最后,将四个分支输出特征统一到1/4分辨率后级联,获得目标的高分辨率特征表示.

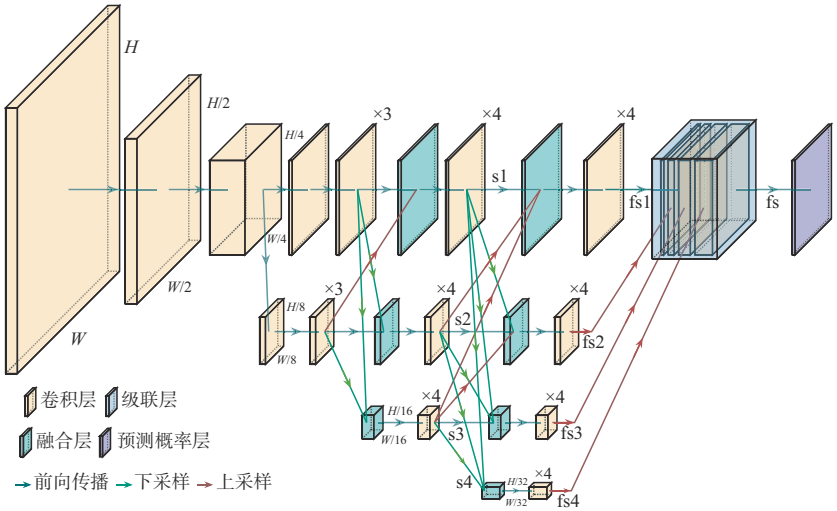


图1 HRNet结构示意图

为深入分析HRNet交互融合结构的影响,本文将HRNet中间层特征在通道维度聚合,聚合后的特征图为单通道特征图,能直观反映网络中间层特征的优劣. 对于一组大小为$W\times H\times N$(宽W、高H、通道N)的中间层特征,其聚合的公式为:

$$\mathrm{AF}(j)=\frac{1}{N}\sum_{i=1}^{N}F_i(j) \tag{1}$$

式中$F_i(j)$为第j个像素的第i个特征值,$\mathrm{AF}(j)$为第j个像素对应的聚合特征值. 图2所示为HRNet几组典型的中间层聚合特征图,图2(d)~(g)为图1所示HRNet四个分支s1、s2、s3、s4的特征热图,四组特征图表示了不同空间尺度的特征. 图2(h)~(k)为四个分支交互融合后的特征fs1、fs2、fs3、fs4. 图中可见,HRNet独特的交互融合结构有效地激活了目标区域,同时抑制了背景区域. 图2(c)为HRNet最终输出特征,网络通过四个分支特征的融合,获得了精确的高分辨率特征.

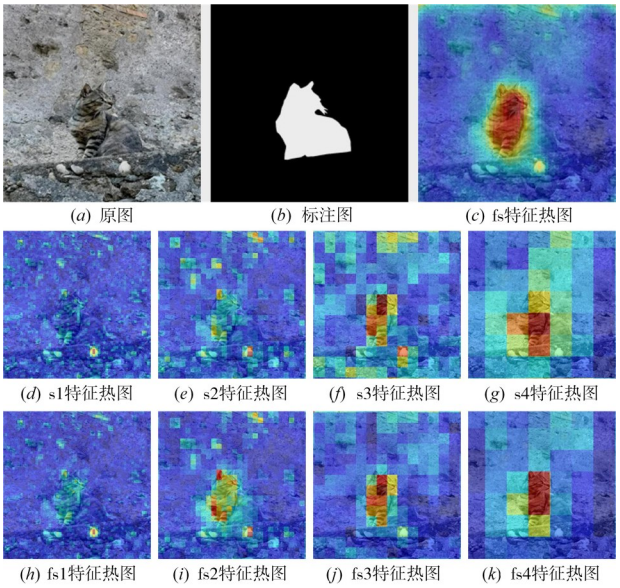


图2 HRNet输出特征图

近年来,研究者主要采取直接提取网络中间层特征来构建自蒸馏学习模型的方法. 在此类模型中[15,17,29,30],自蒸馏学习的主要作用在于促进网络浅层特征与深层语义特征之间的融合,从而提升网络的性能. 对于HRNet而言,其底层特征与语义特征已充分融合,直接在分支结构上构建蒸馏学习结构难以实现性能提升. 表1为基于文献[15]构建的HRNet自蒸馏模型(HRNet18-YOT、HRNet48-YOT)在四个目标分割任务上的效果(数据集和评价指标见4.1节和4.2节). 从表中可见,直接提取中间层特征构建自蒸馏学习模型

对HRNet的目标分割性能提升十分有限.

3　基于池化金字塔的自蒸馏学习模型

Li等人近期的研究[31]表明:知识蒸馏可看成一种特殊的标签平滑正则化方法,并且发现知识蒸馏与监督学习联用通常能获得更好的性能. 因此,本文将知识蒸馏看作一个带有正则化效果的有监督学习过程.

基于上述思想,本文构建自蒸馏模型主要考虑两个要素:蒸馏模型的结构和学习度量. 模型结构涉及到如何在HRNet中构建蒸馏网络,获得有效的监督信息. 模型的学习度量涉及到如何构建损失函数,促进各子分支的知识迁移,总体上提升模型的性能.

表1　自蒸馏学习 F_β 值效果对比　　单位:%

模 型	COD	CPD	DUT-OMRON	PASCAL-S	性能提升
HRNet18-BL	63.93	63.52	81.36	82.33	
HRNet18-YOT	64.02(+0.09)	63.58(+0.06)	81.41(+0.05)	82.37(+0.04)	+0.060
HRNet48-BL	70.29	71.91	84.46	85.79	
HRNet48-YOT	70.33(+0.04)	71.94(+0.03)	84.49(+0.03)	85.80(+0.01)	+0.027

3.1　自蒸馏模型结构

图3所示为本文的自蒸馏模型结构. 结构的设计基于如下考虑:

(1) 蒸馏结构对蒸馏效果有重要的影响,其在网络中的嵌入位置关系到能否从网络中提取有效的知识表达,以及蒸馏梯度信息能否在主干网中高效传播[26]. 以往的研究经验表明,网络中的浅层特征缺乏目标的语义知识,直接利用其分类将对预测结果产生负面影响[31,32]. 如图2所示,HRNet分支的输出端包含了丰富且具有差异性的目标语义知识. 因此,本文将蒸馏结构嵌入在HRNet四个子分支fs1、fs2、fs3、fs4和主分支fs的输出端.

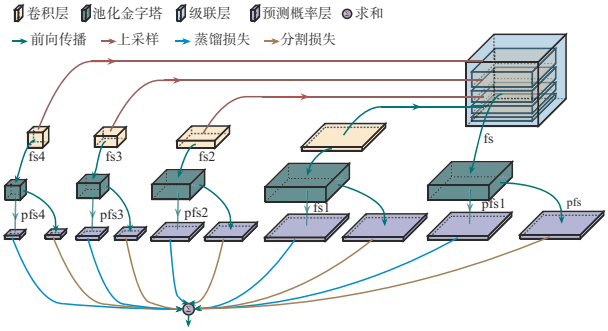


图3　自蒸馏模型结构

(2) 如前所述,HRNet分支的输出特征已融合语义知识与底层知识,难以直接蒸馏. 然而,从有监督学习的角度看,自蒸馏学习是一个将更精准的表示信息迁移到目标蒸馏模型的过程. 越准确的监督信息将越能提升模型表示学习的效果. 由此,本文将多尺度池化金字塔模块[33]引入到自蒸馏结构中,提升网络的特征表示能力.

图4所示为多尺度池化金字塔的具体结构,其通过多尺度池化、卷积运算有效地捕获图像的全局上下文信息,对原有特征有提升、精炼的作用. 在本文蒸馏学习结构中:五组金字塔特征一方面作为教师端,为蒸馏结构提供了更准确的监督信息;另一方面其作为学生端,为知识蒸馏过程提供了更好的特征表示,更易于实现蒸馏学习过程.

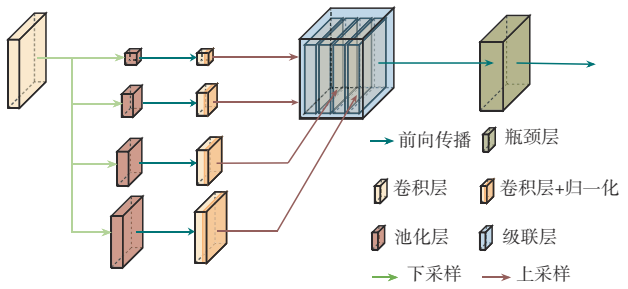


图4　多尺度池化金字塔模块

本文自蒸馏结构具体如图3所示,在HRNet原有输出结构的基础上分别添加蒸馏分支:将池化金字塔模块串联在fs1、fs2、fs3、fs4、fs之后,对HRNet原输出特征进一步提炼,获得更准确的特征表示pfs1、pfs2、pfs3、pfs4、pfs. 在此基础上构建蒸馏路径,通过蒸馏学习提升HRNet原网络分支的特征表示能力.

(3) 如图5所示,本文自蒸馏模型融合两种蒸馏模式:

一是四个子分支的一致性蒸馏模式. 四个子分支的输出特征表示了不同空间尺度的知识,在学习过程中容易产生优化目标不一致的情况,导致模型产生次优结果[29]. 一致性蒸馏促使四个子分支之间相互模仿,约束子分支向共同的目标优化学习,从而提升模型的性能. 具体而言,依次将四个子分支中的一个子分支作为教师端,其他三个子分支分别作为学生端,构成三组蒸馏对. 共计生成十二组蒸馏对,实现子分支的一致性蒸馏.

二是从主分支到子分支的自上而下蒸馏模式. 主分支特征是HRNet最深层特征,也是四个子分支特征融合的结果,且在推断阶段直接生成图像分割结果,实际指示了模型总体的优化方向. 因此,将其作为教师

端,四个子分支分别作为学生端,构成4组蒸馏对,实现自上而下的蒸馏模式.

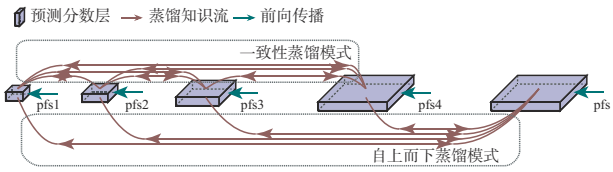


图5 自蒸馏知识迁移模式

3.2 自蒸馏模型的学习

现有蒸馏方法通过类别概率的KL散度传递"暗知识",同时结合预测损失(如交叉熵损失)共同完成学习.对于目标分割任务而言,单个像素的类别属性与其空间邻域的上下文信息有着强依赖关系.然而,现有蒸馏学习框架中的KL散度损失和交叉熵损失实质上都将每个像素作为独立个体看待,无法从空间上度量预测图与标注图之间的结构化差异.因此,本文将结构化相似性引入自蒸馏学习框架,量化空间结构化信息,通过融合结构化相似性损失[34]、KL散度损失、交叉熵损失构建更准确的自蒸馏度量学习空间.

给定训练数据集 $D=\{(\boldsymbol{x}_i,\boldsymbol{y}_i)|i=1,2,\cdots,N\}$,其中 $\boldsymbol{x}_i$ 表示数据集中第 i 个图像数据,N 为数据集包含的图像数量,$\boldsymbol{y}_i$ 表示其对应的像素级标注图.H、W 分别为图像的高和宽,N 为数据集包含的图像数量,K 为预测类别数量(本文中 K 等于2).对于一个 L 层的深度神经网络,$\boldsymbol{W}_m$ 为网络主体(不涉及自蒸馏网络)的权重矩阵,$\boldsymbol{W}_s=\{\boldsymbol{W}_s^l|l=1,2,\cdots,M\}$ 为自蒸馏结构中的辅助分类网络权重矩阵.如图3所示,本文的自蒸馏辅助网络由5个(M=5)自蒸馏分支组成,将第 l 个自蒸馏分支的权重矩阵记为 $\boldsymbol{W}_s^l$.将自蒸馏分支与网络主体的连接位置记为 $A=\{m_i|i=1,2,\cdots,M\}$.本文的自蒸馏学习是通过最小化目标函数学习到一个参数映射函数 $f(\boldsymbol{W}_m;\boldsymbol{x}):\boldsymbol{X}\rightarrow\boldsymbol{Y}$,其中目标函数可表示为:

$$\arg\min_{\boldsymbol{W}_m,\boldsymbol{W}_s} L_m(\boldsymbol{W}_m;D)+L_s(\boldsymbol{W}_m,\boldsymbol{W}_s;D)+L_k(\boldsymbol{W}_m,\boldsymbol{W}_s;D) \tag{2}$$

式(2)中 L_m 为网络主体的学习目标函数,在本文中为交叉熵损失函数:

$$L_m(\boldsymbol{W}_m;D)=\frac{1}{N\times I}\sum_{j=1}^{N}\sum_{i=1}^{I}H(\boldsymbol{y}_{ij},f(\boldsymbol{W}_m;\boldsymbol{x}_{ij})) \tag{3}$$

其中 $\boldsymbol{x}_{ij}$ 表示训练集中第 i 个图像中第 j 个像素值,$\boldsymbol{y}_{ij}$ 为其对应的标签值,I 为第 i 个图像包含的像素数量(I=W×H).$f(\boldsymbol{W}_m;\boldsymbol{x}_{ij})$ 表示网络主体对像素 $\boldsymbol{x}_{ij}$ 的预测概率向量.$H(\boldsymbol{y}_{ij},f(\boldsymbol{W}_m;\boldsymbol{x}_{ij}))$ 具体表示为:

$$H(\boldsymbol{y}_{ij},f(\boldsymbol{W}_m;\boldsymbol{x}_{ij}))=-\sum_{k=1}^{K}\boldsymbol{Y}_{ij}^k\log f^k(\boldsymbol{W}_m;\boldsymbol{x}_{ij}) \tag{4}$$

其中 $f^k(\boldsymbol{W}_m;\boldsymbol{x}_{ij})$ 表示网络主体对像素 $\boldsymbol{x}_{ij}$ 在第 k 个类别上的预测概率,$\boldsymbol{Y}_{ij}$ 为标注值 $\boldsymbol{y}_{ij}$ 的one-hot编码向量,$\boldsymbol{Y}_{ij}^k$ 为向量中第 k 个值.

式(2)中 L_s 为自蒸馏网络的预测结果相对于标注图产生的分割损失.在本文中为交叉熵损失与结构化相似性损失的加权和,可表示为:

$$\begin{aligned}L_s(\boldsymbol{W}_m,\boldsymbol{W}_s;D)&=\frac{1}{N\times I\times M}\sum_{j=1}^{N}\sum_{i=1}^{I}\sum_{l\in A}\Big(H(\boldsymbol{y}_{ij},f(\boldsymbol{W}_m,\boldsymbol{w}^l;\boldsymbol{x}_i))\Big)\\&+\frac{1}{N\times P\times M}\sum_{j=1}^{N}\sum_{p=1}^{P}S(\boldsymbol{y}_{pj},f(\boldsymbol{W}_m,\boldsymbol{w}^l;\boldsymbol{x}_{pj}))\end{aligned} \tag{5}$$

其中 N 为数据集中图像的数量,I 为图像中的像素数量,M=5为自蒸馏分支的数量.式(5)中第一和第二项分别为交叉熵损失和结构化相似性损失,具体如式(6)、式(7)所示:

$$H(\boldsymbol{y}_{ij},f(\boldsymbol{W}_m,\boldsymbol{w}_s^l;\boldsymbol{x}_i))=-\sum_{k=1}^{K}\boldsymbol{Y}_{ij}^k\log f^k(\boldsymbol{W}_m,\boldsymbol{w}_s^l;\boldsymbol{x}_{ij}) \tag{6}$$

$$S(\boldsymbol{y}_{pj},f(\boldsymbol{W}_m,\boldsymbol{w}^l;\boldsymbol{x}_{pj}))=\sum_{k=1}^{K}\sum_{p=1}^{P}\Big(1-\mathrm{SSIM}\Big(\boldsymbol{Y}_{pj}^k,f^k(\boldsymbol{W}_m,\boldsymbol{w}^l;\boldsymbol{x}_{pj})\Big)\Big) \tag{7}$$

式(6)中,$f^k(\boldsymbol{W}_m,\boldsymbol{w}_s^l;\boldsymbol{x}_{ij})$ 表示网络主体联合第 l 个自蒸馏分支对像素 $\boldsymbol{x}_{ij}$ 在第 k 个类别上的预测概率;式(7)中SSIM(·,·)为结构化相似相度量,其用图像的亮度、对比度、结构相关度量两幅图像的结构化差异.在具体计算中,先用矩形窗将图像分为 P 个图像块(本文中 P=24×24),再计算标注图图像块与预测图像块之间的结构化相似性度量.最后,用 P 个图像块结构化相似性的均值表示两个图像之间的结构化相似性.$\boldsymbol{Y}_{pj}^k$ 为数据集中第 j 个图像的第 p 个图像块在 k 类别上的标注one-hot编码值(其值为该图像块包含所有像素的one-hot编码的平均值).$f^k(\boldsymbol{W}_m,\boldsymbol{w}^l;\boldsymbol{x}_{pj})$ 为网络主体联合第 l 个自蒸馏分支对第 p 个图像块 $\boldsymbol{x}_{pj}$ 在第 k 个类别上的预测概率,将其简写为 $f_l^k(\boldsymbol{x}_{pj})$,式(7)中结构化相似性具体为:

$$\mathrm{SSIM}\Big(\boldsymbol{y}_{pj}^k,f_l^k(\boldsymbol{x}_{pj})\Big)=\frac{(2\mu_{\boldsymbol{y}_{pj}^k}\mu_{f_l^k(\boldsymbol{x}_{pj})}+C_1)(2\sigma_{\boldsymbol{y}_{pj}^k f_l^k(\boldsymbol{x}_{pj})}+C_2)}{(\mu_{\boldsymbol{y}_{pj}^k}^2+\mu_{f_l^k(\boldsymbol{x}_{pj})}^2+C_1)(\sigma_{\boldsymbol{y}_{pj}^k}^2+\sigma_{f_l^k(\boldsymbol{x}_{pj})}^2+C_2)} \tag{8}$$

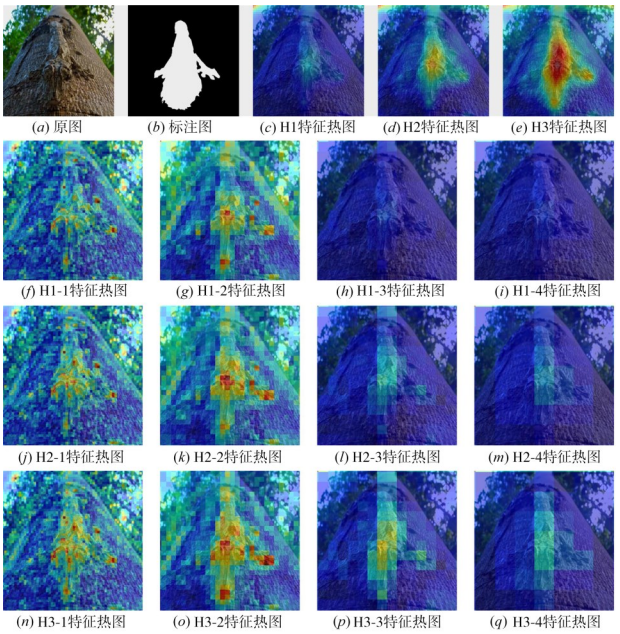
其中 $\mu_{\boldsymbol{y}_{pj}^k}$、$\mu_{f_l^k(\boldsymbol{x}_{pj})}$、$\sigma_{\boldsymbol{y}_{pj}^k}^2$、$\sigma_{f_l^k(\boldsymbol{x}_{pj})}^2$ 分别为 $\boldsymbol{y}_{pj}^k$ 和 $f_l^k(\boldsymbol{x}_{pj})$ 的均值与标准差,$\sigma_{\boldsymbol{y}_{pj}^k f^k(\boldsymbol{x}_{pj})}$ 为 $\boldsymbol{y}_{pj}^k$ 和 $f_l^k(\boldsymbol{x}_{pj})$ 之间的协方差.$C_1=0.01^2$、$C_2=0.03^2$,为两个常数.

式(2)中 L_k 为本文自蒸馏结构的蒸馏损失,具体如式(9)所示:

$$L_k(\boldsymbol{W}_m,\boldsymbol{W}_s;\ D)$$
$$=\frac{1}{N\times I\times M_p}\sum_{j=1}^{N}\sum_{i=1}^{I}\sum_{t=m_1}^{m_5}\sum_{\substack{s=m_2\\ s\neq t}}^{m_5}\left(\mathrm{KL}(f_t(\boldsymbol{x}_{ij}),f_s(\boldsymbol{x}_{ij}))\right)$$
$$+\frac{1}{N\times P\times M_p}\sum_{j=1}^{N}\sum_{p=1}^{P}\sum_{t=m_1}^{m_5}\sum_{\substack{s=m_2\\ s\neq t}}^{m_5}S(f_t(\boldsymbol{x}_{pj}),f_s(\boldsymbol{x}_{pj})) \quad (9)$$

其中,N为数据集中图像的数量,I为图像中的像素数量,M_p为自蒸馏对的数量(本文中M_p=12). 式中第一项为两个蒸馏辅助分类器输出的KL损失,KL(·, ·)为KL散度. 第二项为两者的结构化相似性损失. $f_t(\boldsymbol{x}_{ij})$、$f_s(\boldsymbol{x}_{ij})$、$f_t(\boldsymbol{x}_{pj})$、$f_s(\boldsymbol{x}_{pj})$分别为第t个和第s个自蒸馏分支对像素$\boldsymbol{x}_{ij}$和图像块$\boldsymbol{x}_{pj}$的预测概率,其中第t个和第s个自蒸馏分支分别作为蒸馏学习的教师端和学生端,m_1为主分支位置标记,m_2~m_5为子分支位置标记.

图6所示为原始HRNet和两种自蒸馏学习HRNet的输出特征对比图. 由于目标区域与背景的颜色、纹理等外观特征极为相似,原始HRNet(图6(*c*)、图6(*f*)~图6(*i*))难以有效激活目标区域. 图6(*d*)、图6(*j*)~图6(*m*)为通过交叉熵损失合并KL散度损失训练的自蒸馏HRNet输出特征,可见本文的自蒸馏学习结构有效提升了HRNet四个分支的特征表示能力,并最终提升了主分支的特征表示能力,更有效地激活了目标区域. 图6(*e*)、图6(*n*)~图6(*q*)为增加了结构化相似性损失训练的自蒸馏HRNet输出特征,相比前一自蒸馏学习模型,结构化相似性损失能更准确地度量目标与背景的差异,赋予模型更强的特征表示能力,在极低对比度的伪装条件下也能显著地激活目标区域.



(H1: 原始HRNet,H2:交叉熵+KL散度自蒸馏学习HRNet,H3: 交叉熵+KL散度+结构化相似性自蒸馏学习HRNet,H1/2/3-1: HRNet分支1,H1/2/3-2: HRNet分支2,H1/2/3-3: HRNet分支3,H1/2/3-4: HRNet分支4)

图6　原始HRNet与两种自蒸馏学习HRNet输出特征对比图

4　实验与分析

由于目前未见有专门用于目标分割任务的自蒸馏模型,本文在四个难度较大的目标分割数据集上对原始HRNet与本文的自蒸馏学习HRNet进行定性与定量试验比较,验证本文自蒸馏学习模型对HRNet分割任务的性能提升效果.

4.1　评价数据

本文在伪装目标分割和显著性目标分割两个任务的四个公开数据集上试验比较.

伪装目标数据集包括:(1)COD[19]是一个自然伪装目标数据集,包含了10个超类和78个子类的10 000幅自然伪装图像,其中5 066幅伪装图像,3 000幅背景图像,1 934幅无伪装的图像. 总体而言,由于自然目标的伪装属性,该数据集是目前公开的难度最大的目标分割数据集. (2)CPD[20,21]是一个迷彩伪装单兵数据集,内容为穿着与背景内容相应的迷彩服装士兵的图片. 场景包括了丛林、雪地、荒漠三种,迷彩类型26种,共2 600副图像. 由于迷彩的伪装属性和场景的复杂性,该数据集也是分割难度较大的数据集. 显著性目标数据集包括:(1)DUT-OMRON[22]是一个包含5 168副图像显著性目标数据集,每副图像中有1~2个显著性目标,该数据集的目标与背景的对比度较大,但目标本身较为复杂,是有一定难度的目标分割数据集. (2)PASCAL-S[23]包含850副图像,是一个图像的背景与目标都较为复杂的目标分割数据集.

4.2　评价指标

本文使用目标分割任务中常用的F_β值(F-measure)[35]评价并比较基准模型和自蒸馏模型的性能.

正确检出的目标区域面积占标注图中目标区域面积的比例为准确率(precision),准确率侧重于衡量算法检测目标区域的准确程度. 正确检出的目标区域面积占算法检出所有目标区域的比例为召回率(recall),召回率侧重于衡量算法检测目标区域的完整程度. F_β值是融合检测准确率和召回率的综合评价指标,F_β值的计算公式如式(10)所示. 按照以往的经验[30],公式中β设置为0.3.

$$F_\beta=\frac{(1+\beta)\times\mathrm{precision}\times\mathrm{recall}}{\beta\times\mathrm{precision}+\mathrm{recall}} \quad (10)$$

4.3　实验模型与训练参数

实验将HRNet18与HRNet48两个网络[18]作为基准模型分别比较自蒸馏学习效果. HRNet48与HRNet18

有相同的网络结构，区别在于主干网卷积核通道数不同，具有不同的参数规模．在训练阶段，HRNet基准模型与自蒸馏学习模型在相同的数据集上使用相同的参数进行训练．在测试阶段，自蒸馏学习模型移除蒸馏分支，使用与基准模型相同的结构进行前向推断，获得目标分割结果．

基准模型与自蒸馏模型均使用随机梯度下降法进行训练，训练图像统一缩放到尺寸288×288．训练与测试数据的数量分别为数据集总数量的60%和40%．训练参数统一设置为：学习率0.01，批处理数量8，权重衰减系数0.05，训练迭代次数20轮．

4.4 模型效果与分析

4.4.1 定性比较

图7所示为HRNet18、HRNet48基准模型（HRNet18-BL、HRNet48-BL）及相应的自蒸馏学习模型（HRNet18-SDL、HRNet48-SDL）在四个数据集上具有代表性的分割效果示例．如图中所示，对于自然伪装目标（COD）和迷彩伪装目标（CPD），HRNet18自蒸馏模型分割出了更完整的目标（图7HRNet18-SDL分割图中第1、2、3、4列），而HRNet48自蒸馏模型比HRNet48基准模型显著提升了目标分割的准确性（图7HRNet48-SDL分割图中第1、2、3、4列）．对于通用场景下的显著性目标（DUTS-OMRON、PASCAL-S），HRNet模型已能较完整地分割出目标．经过自蒸馏学习后，HRNet18、HRNet48模型更有效地抑制了背景干扰，分割出了更准确的目标区域．

4.4.2 定量比较

表2所示为HRNet18、HRNet48基准模型及其相应的自蒸馏学习模型在四个数据集上的F_β值效果对比．如表中所示，在四个难度较大的目标分割数据集上，本文方法对两种参数规模模型的性能均有较大提升．因此，本文构建的自蒸馏学习模型能有效增强HRNet的特征表示能力，提升其在目标分割任务上的性能．

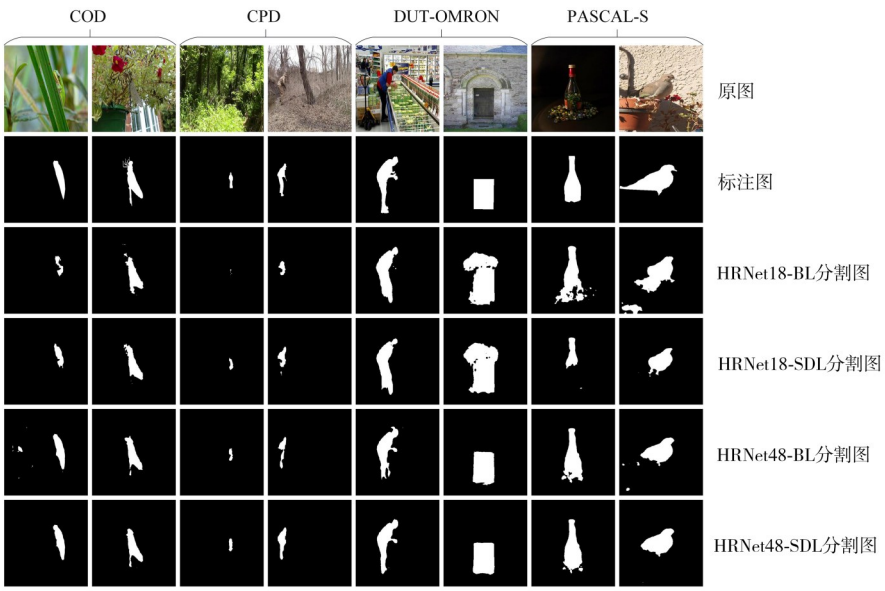


图7 自蒸馏学习模型分割图对比

表2 自蒸馏学习模型F_β值效果对比 单位：%

模 型	COD	CPD	DUT-OMRON	PASCAL-S	性能提升
HRNet18-BL	63.93	63.52	81.36	82.33	
HRNet18-SDL	65.67(+1.73)	66.43(+2.91)	83.25(+1.64)	85.29(+2.44)	+2.178
HRNet48-BL	70.29	71.91	84.46	85.79	
HRNet48-SDL	71.66(+1.40)	73.61(+1.50)	85.75(+1.35)	87.08(+1.80)	+1.513

4.4.3 消融实验

本文模型在结构上涉及一致性蒸馏、自上而下蒸馏两种蒸馏模式，在学习度量上涉及交叉熵、KL散度、结构化相似性三种损失函数．以下设计了两组消融实验，研究这些要素在自蒸馏学习模型中的作用．

蒸馏模式消融实验：分别设置一致性蒸馏、自上而下蒸馏、一致性+自上而下三种蒸馏模式，使用相同的损失函数（交叉熵+KL散度+结构化相似性）完成模型的训练．

表3为蒸馏模式消融实验获得的F_β值，从表中可看出：（1）自上而下蒸馏模式对模型性能的提升略优于一致性蒸馏模式．（2）两种蒸馏模式融合对模型的提升有

加性效果. 如在COD数据集上,自上而下蒸馏和一致性蒸馏对HRNet18基准模型的性能提升分别为1.21和0.62,两种蒸馏模式融合后提升了1.73,大于单独的两种蒸馏模式,略小于两种蒸馏模式之和1.83.

学习度量消融实验:蒸馏模式统一为一致性+自上而下蒸馏,分别设置损失函数为交叉熵、交叉熵+KL散度、交叉熵+KL散度+结构化相似性三种学习度量进行训练.

表4为学习度量消融实验获得的F_β值. 如表所示:(1)交叉熵损失对HRNet18和HRNet48的性能提升的贡献分别为32.9%和32.4%. 与基准模型相比,交叉熵损失训练模式通过更准确的特征表示、更有效的监督方式实现了性能提升,其实质是基于池化金字塔的多尺度有监督学习. (2)在交叉熵基础上加上蒸馏损失后,模型性能进一步提升了36.9%、37%. 这表明蒸馏学习增强了特征信息、监督信息从蒸馏辅助分支到主干网的迁移,实现了主干网性能的进一步提升. (3)融合结构化相似性损失后,模型的性能进一步提升了30.2%、30.6%,这表明在目标分割的自蒸馏学习中,结构化相似性损失能从不同于交叉熵、KL散度的角度,即图像结构化信息的角度度量特征图之间的差异,有效提升知识迁移效率、增强主干网的目标分割性能.

表3　蒸馏模式消融实验F_β值效果对比(T2D:自上而下蒸馏模式,CST:一致性蒸馏模式)　　单位:%

模 型	COD	CPD	DUT-OMRON	PASCAL-S	性能提升
HRNet18-BL	63.93	63.52	81.36	82.33	
HRNet18-T2D	65.15(+1.21)	65.78(+2.26)	82.70(+1.09)	83.19(+0.86)	+1.163
HRNet18-CST	64.56(+0.62)	65.35(+1.83)	82.27(+0.66)	83.61(+1.28)	+1.048
HRNet18-T2D+CST	71.66(+1.40)	73.61(+1.50)	85.75(+1.35)	87.08(+1.80)	+1.513
HRNet48-BL	70.26	72.11	84.40	85.28	
HRNet48-T2D	71.02(+0.76)	73.01(+1.10)	85.32(+0.92)	87.00(+1.21)	+0.938
HRNet48-CST	71.24(+0.98)	72.68(+0.77)	85.18(+0.78)	86.35(+1.07)	+0.755
HRNet48-T2D+CST	71.66(+1.40)	73.61(+1.50)	85.75(+1.35)	87.08(+1.80)	+1.513

表4　学习度量消融实验F_β值效果对比(CE:交叉熵损失,KL:KL散度损失,SSIM:结构化相似性损失)　　单位:%

模 型	COD	CPD	DUT-OMRON	PASCAL-S	性能提升
HRNet18-BL	63.94	63.52	81.61	82.85	
HRNet18-CE	64.49(+0.55)	64.37(+0.85)	82.44(+0.63)	83.69(+0.84)	+0.718/32.9
HRNet18-CE+KL	65.12(+1.18)	66.08(+1.76)	82.77(+1.16)	84.83(+1.98)	+1.520/69.8
HRNet18-CE+KL+SSIM	65.67(+1.73)	66.43(+2.91)	83.25(+1.64)	85.29(+2.44)	+2.178
HRNet48-BL	70.26	72.11	84.40	85.28	
HRNet48-CE	70.87(+0.61)	72.60(+0.49)	84.82(+0.42)	85.72(+0.44)	+0.490/32.4
HRNet48-CE+KL	71.21(+0.95)	73.01(+1.10)	85.26(+0.86)	86.57(+1.29)	+1.050/69.4
HRNet48-CE+KL+SSIM	71.66(+1.40)	73.61(+1.50)	85.75(+1.35)	87.08(+1.80)	+1.513

4.5　自蒸馏学习机制分析

近年来的知识蒸馏研究侧重于构建更有效的蒸馏学习模型,而探究知识蒸馏学习机制的研究较少. 本文尝试从模型输出端特征学习的角度探索自蒸馏学习模型的作用机制,揭示自蒸馏模型如何在学习过程中提升模型的性能.

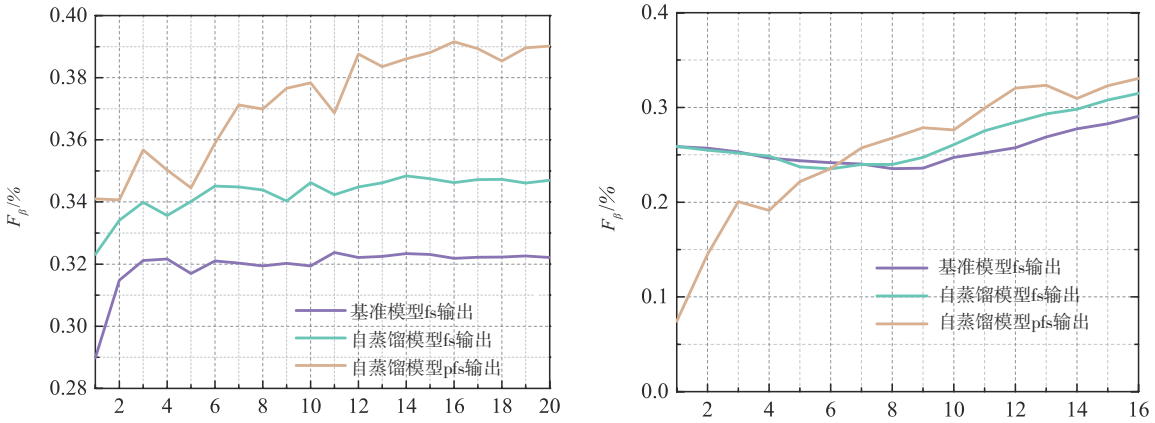
就本文HRNet自蒸馏模型而言,自蒸馏学习过程与基准模型学习过程的主要区别在于:自蒸馏学习过程中有多尺度特征金字塔辅助分支向主干网传播梯度信息. 从4.4节的实验结果看,这一过程无疑增强了主干网的特征表示能力.

为验证以上设想,本文设计以下实验:在COD数据集的训练过程中,保存每轮基准模型与自蒸馏模型,输出基准模型的fs端特征图、自蒸馏模型的fs端和pfs端的特征图,再计算以上输出端聚合特征图的F_β值. 图8(a)所示为每轮模型输出特征图的F_β值曲线图,可知在训练过程中,pfs端特征始终优于fs端特征,且其一直向网络主干传播梯度信息. 这一过程相当于pfs端向模型提供了额外的监督信息,使得自蒸馏学习模型的fs端输出始终优于基准模型的fs端输出,即增强了模型的特征表示能力.

然而,pfs端与fs端有相同的初始条件,其何时获得更强的特征表示能力在上图中并无体现. 为此,本文将从训练初始状态到第一轮训练完成过程中每批次的训练模型保存. 与以上实验相似,计算相应的fs端、pfs端聚合特征图的F_β值,得到的曲线如图8(b)所示. 由图8(b)可知,在训练起始阶段基准模型与自蒸馏模型的fs端输出并无明显性能差异. 然而,由于pfs端特征是在fs

端特征基础上的增强，其在学习时有更快的收敛速率．在第6~8批次训练后，pfs端性能迅速超过了fs端．此时，pfs端已可以通过蒸馏学习向fs端传递额外的知识信息辅助训练，使得自蒸馏模型fs端比基准模型fs端有更快的性能提升，即增强了HRNet主干网的特征表示能力．

由以上分析可知，本文自蒸馏模型起作用的关键在于：在结构上引入了特征表示能力更强的多尺度池化金字塔辅助分支，其通过蒸馏学习过程为网络主干提供了额外的知识和监督信息，从而增强了模型表示能力．



(*a*) 20轮训练泛化曲线/轮　　(*b*) 第1轮中批次数据训练的训练泛化曲线/批次

图8 自蒸馏学习曲线对比

5 结论

本文针对最新主干网HRNet的结构特点，提出了基于金字塔知识迁移的自蒸馏学习模型．构建基于多尺度池化金字塔的自蒸馏特征表示结构，为蒸馏过程提供更准确且丰富的知识表示信息，增强分支结构作为蒸馏学生端的学习能力．根据HRNet多分支结构的特点，构建了分支一致性和自上而下两种蒸馏模式，约束四个分支在学习过程中保持一致且准确的优化方向．通过实验验证了本文的自蒸馏模型能在不改变网络结构的前提下提升HRNet目标分割性能，且从特征学习的角度揭示了自蒸馏学习的作用机制．

参考文献

[1] 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.

[2] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[3] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 3431-3440.

[4] 郑云飞, 张雄伟, 曹铁勇, 等. 基于全卷积网络的语义显著性区域检测方法研究[J]. 电子学报, 2017, 45(11): 2593-2601.
ZHENG Y F, ZHANG X W, CAO T Y, et al. The semantic salient region detection algorithm based on the fully convolutional networks[J]. Acta Electronica Sinica, 2017, 45(11): 2593-2601. (in Chinese)

[5] 李雅倩, 盖成远, 肖存军, 等. 基于细化多尺度深度特征的目标检测网络[J]. 电子学报, 2020, 48(12): 2360-2366.
LI Y Q, GAI C Y, XIAO C J, et al. Object detection networks based on refined multi-scale depth feature[J]. Acta Electronica Sinica, 2020, 48(12): 2360-2366. (in Chinese)

[6] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 2261-2269.

[7] 张锦, 李阳, 任传伦, 等. 基于帧间高级特征差分的跨场景视频前景分割算法[J]. 电子学报, 2021, 49(10): 2032-2040.
ZHANG J, LI Y, REN C L, et al. Cross-scene foreground segmentation algorithm based on high-level feature differencing between frames[J]. Acta Electronica Sinica, 2021, 49(10): 2032-2040. (in Chinese)

[8] 权宇, 李志欣, 张灿龙, 等. 融合深度扩张网络和轻量化网络的目标检测模型[J]. 电子学报, 2020, 48(2): 390-397.
QUAN Y, LI Z X, ZHANG C L, et al. Fusing deep dilated convolutions network and light-weight network for object

detection[J]. Acta Electronica Sinica, 2020, 48(2): 390-397. (in Chinese)

[9] ZAGORUYKO S, KOMODAKIS N. Wide residual networks[C]//Proceedings of the British Machine Vision Conference 2016. York: BMVA Press, 2016: 81-87.

[10] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[EB/OL]. (2015-03-05). https://arxiv.org/abs/1503.02531.

[11] ADRIANA R, NICOLAS B, EBRAHIMI K S, et al. Fitnets: Hints for thin deep nets[C]//Proceedings of the European International Conference on Learning Representations. Piscataway: IEEE, 2015: 1-13.

[12] ZAGORUYKO S, KOMODAKIS N. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer[EB/OL]. (2016-12-12). https://arxiv.org/abs/1612.03928.

[13] ZHANG Y, XIANG T, HOSPEDALES T M, et al. Deep mutual learning[C]//2018 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 4320-4328.

[14] CHEN D F, MEI J P, WANG C, et al. Online knowledge distillation with diverse peers[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(4): 3430-3437.

[15] ZHANG L F, SONG J B, GAO A N, et al. Be your own teacher: Improve the performance of convolutional neural networks *via* self distillation[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2019: 3712-3721.

[16] YANG C L, XIE L X, SU C, et al. Snapshot distillation: teacher-student optimization in one generation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 2859-2868.

[17] LI D, CHEN Q F. Dynamic hierarchical mimicking towards consistent optimization objectives[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 7642-7651.

[18] WANG J D, SUN K, CHENG T H, et al. Deep high-resolution representation learning for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(10): 3349-3364.

[19] FAN D P, JI G P, SUN G L, et al. Camouflaged object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 2774-2784.

[20] ZHENG Y F, ZHANG X W, WANG F, et al. Detection of people with camouflage pattern via dense deconvolution network[J]. IEEE Signal Processing Letters, 2019, 26(1): 29-33.

[21] FANG Z, ZHANG X W, DENG X T, et al. Camouflage people detection via strong semantic dilation network[C]//Proceedings of the ACM Turing Celebration Conference. New York: ACM, 2019: 1-7.

[22] YANG C, ZHANG L H, LU H C, et al. Saliency detection via graph-based manifold ranking[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2013: 3166-3173.

[23] LI Y, HOU X D, KOCH C, et al. The secrets of salient object segmentation[C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2014: 280-287.

[24] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04). https://arXiv.org/abs/1409.1556.

[25] ZHOU B L, BAU D, OLIVA A, et al. Interpreting deep visual representations via network dissection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(9): 2131-2145.

[26] NOH H, HONG S, HAN B. Learning deconvolution network for semantic segmentation[C]//2015 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2015: 1520-1528.

[27] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015: 234-241.

[28] BADRINARAYANAN V, KENDALL A, CIPOLLA R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.

[29] LAN X, ZHU X T, GONG S G. Knowledge distillation by on-the-fly native ensemble[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: ACM, 2018: 7528-7538.

[30] SUN D W, YAO A B, ZHOU A J, et al. Deeply-supervised knowledge synergy[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 6997-7006.

[31] YUAN L, TAY F E, LI G L, et al. Revisiting knowledge

distillation via label smoothing regularization[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 3902-3910.

[32] HUANG G, CHEN D L,et al. Multi-scale dense networks for resource efficient image classification[EB/OL]. (2017-03-29). https://arXiv.org/abs/1703.09844.

[33] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 6230-6239.

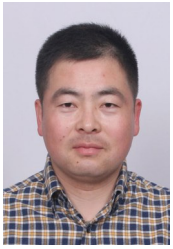
[34] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: From error visibility to structural similarity [J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

[35] ACHANTA R, HEMAMI S, ESTRADA F, et al. Frequency-tuned salient region detection[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2009: 1597-1604.

作者简介

郑云飞 男,1983 年 9 月出生于安徽省滁州市. 解放军理工大学信息与通信工程专业博士,现为陆军炮兵防空兵学院南京校区讲师、陆军工程大学指挥控制工程学院计算机与科学博士后流动站在站博士后,主要研究方向为目标分割、深度学习.

E-mail: 597184353@qq.com



王晓兵 男,1981 年出生于安徽省滁州市. 解放军炮兵防空兵学院作战指挥学博士. 陆军炮兵防空兵学院南京校区副教授、陆军工程大学指挥控制工程学院计算机与科学博士后流动站在站博士后,从事智能任务规划、人工智能方向的研究工作.

E-mail: 23813083@qq.com

张雄伟 男,1965 年出生于浙江嘉兴. 现为陆军工程大学指挥控制工程学院教授、博导. 主要研究方向为人工智能、多媒体信息处理.

E-mail: xwzhang9898@163.com

曹铁勇(通讯作者) 男,1971 年出生于江苏省. 现为陆军工程大学指挥控制工程学院教授、博导. 主要研究方向为人工智能、图像处理.

E-mail: cty_ice@sina.com

孙 蒙 男,1984 年出生于山东齐河,比利时鲁汶大学信息与通信工程博士. 现为陆军工程大学指挥控制工程学院副教授. 主要研究方向为机器学习、语音信号处理.

E-mail: sunmengccjs@163.com