

# 基于教师模型正确预测的图像分类知识蒸馏算法

邵春阳<sup>1</sup>, 刘宁钟<sup>1,2</sup>

(1. 南京航空航天大学 计算机科学与技术学院, 江苏 南京 211106;

2. 南京航空航天大学 模式分析与机器智能工业和信息化部重点实验室, 江苏 南京 211106)

**摘要:** 知识蒸馏旨在简化大型模型的复杂性。Logit 蒸馏是一种普遍采用的策略, 它利用教师模型的输出 Logit 作为软标签来辅助学生模型的训练。尽管这种算法在很多情况下都非常有效, 但教师模型在预测时的错误同样会对知识转移产生不利影响。换句话说, 教师可能会将错误的知识传递给学生, 这可能会导致学生模型在模仿教师时产生偏差, 从而影响其最终的判断力。为了解决这一问题, 提出了一种基于教师正确预测的知识蒸馏算法。该算法能够过滤教师预测错误的实例, 使学生模型专注于正确预测的实例, 提高教师的教学质量。此外, 该算法不仅能够传递不同实例中蕴含的信息, 还能传递类别级别的信息, 这为学生模型提供了更全面的语义信息。实验表明, 该算法在 CIFAR-100 和 ImageNet 等多个数据集上均显著提升了学生模型分类准确率, 验证了该算法的有效性。

**关键词:** 深度学习; 模型压缩; 知识蒸馏; 图像分类; 卷积神经网络

中图分类号: TP391.4

文献标识码: A

文章编号: 1673-629X(2025)06-0094-06

doi: 10.20165/j.cnki.ISSN1673-629X.2025.0021

## Knowledge Distillation Algorithm for Image Classification Based on Correct Predictions of Teacher Model

SHAO Chun-yang<sup>1</sup>, LIU Ning-zhong<sup>1,2</sup>

(1. School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics,

Nanjing 211106, China;

2. MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, Nanjing 211106, China)

**Abstract:** Knowledge distillation aims to simplify the complexity of large models. Logit distillation is a commonly adopted strategy, which utilizes the output Logit of the teacher model as a soft label to assist in the training of the student model. Although the algorithm is highly effective in many cases, the errors of the teacher model in prediction can also have an adverse influence on knowledge transfer. In other words, teachers may pass on incorrect knowledge to students, which may cause the student model to deviate when imitating the teacher, thereby affecting its final judgment. To solve the problem, we introduce a simple yet effective algorithm based on correct instance predictions (CIPKD), which can filter the instances where teachers make incorrect predictions, enabling the student model to focus on the instances with correct predictions and improving the teaching quality of teachers. Moreover, CIPKD can not only convey the information contained in different instances, but also convey the information at the category level, which provides more comprehensive semantic information for the student model. Experiments show that the proposed algorithm significantly improves the classification accuracy of the student model on multiple datasets such as CIFAR-100 and ImageNet, verifying its effectiveness.

**Key words:** deep learning; model compression; knowledge distillation; image classification; convolutional neural network

## 0 引言

近些年来,深度学习技术广泛应用于各个领域<sup>[1-5]</sup>。然而,随着模型尺寸的不断扩大,在计算资源有限的环境中部署它们变得越来越困难。为了应对这一挑战,知识蒸馏技术应运而生<sup>[6]</sup>。通过知识传递,小模型可以继承大模型的“智慧”,即使在资源受限的情况下,

它们也可以表现出接近大模型的性能。

知识蒸馏根据是否利用中间层信息,通常分为 Logit 蒸馏和特征蒸馏。Logit 蒸馏专注于模型输出层的知识提取,而特征蒸馏则通过匹配教师模型与学生模型之间的中间层特征分布来进行知识传递,它需要额外的计算和存储资源来提取深层特征,并且对中间

收稿日期: 2024-10-07

修回日期: 2025-02-11

基金项目: 国家自然科学基金(62276129)

作者简介: 邵春阳(2000-),男,硕士,通信作者,研究方向为计算机视觉;刘宁钟(1975-),男,博士,教授,研究方向为计算机视觉和模式识别。

特征图的访问可能带来安全风险。

Logit 蒸馏避免了这些问题,因为它不需要访问中间层特征。然而,现有的 Logit 蒸馏算法仍存在一些不足。具体而言,当前的知识蒸馏算法没有充分考虑教师模型错误预测对学生模型的负面影响,这可能导致不准确的知识传递。此外,教师模型通常只传递每个实例在不同类别下的概率分布,而忽略了同一类别在不同实例间的概率分布关系。笔者认为,同一类别在不同实例中的概率分布也蕴含着重要信息,值得学生模型进一步学习。

针对现有问题,该文提出了一种基于教师模型正确预测的知识蒸馏算法,对传统 Logit 蒸馏进行了两方面的改进。首先,从教师模型的 Logit 输出中筛选出正确预测的实例,从而消除错误信息对学生模型的不利影响。其次,不仅传递实例级知识,还引入了类别级知识的转移,将同一类别在不同实例中的概率分布传递给学生模型,使其能够获取更广泛、更丰富的信息。实验结果表明,该算法能显著提高学生模型的分

### 1 相关工作

作为最重要的模型压缩技术之一,知识蒸馏最早由 Hinton 等人引入<sup>[6]</sup>。它使用教师模型生成的软标签来指导学生模型的训练,标签软化的程度由一个称为温度的超参数控制。这项基础研究表明,利用教师的产出分布可以有效地传递知识,优于传统的硬标签算法。根据是否利用了中间层的信息,知识蒸馏可分

为 Logit 蒸馏<sup>[6-9]</sup>和特征蒸馏<sup>[10-13]</sup>。

Logit 蒸馏仍然是最广泛采用的算法之一。Xu 等人<sup>[14]</sup>提出了一种蒸馏算法框架,利用自我监督任务从教师模型中转移知识,从而增强学生模型的泛化能力。Zhao 等人<sup>[15]</sup>将传统知识蒸馏损失解耦为目标类知识和非目标类知识,将这两种知识分别转移到学生模型中。Yang 等人<sup>[16]</sup>进一步将非目标 Logit 进行归一化处理,使其分布与教师模型对齐。这种算法有效解决了学生与教师模型在非目标概率总和上的差异问题。然而,这些算法忽略了教师模型在预测过程中可能产生的错误知识对学生模型的负面影响。当教师模型提供错误的预测结果时,学生模型可能会学习到不准确或误导性的知识,从而影响模型的整体性能和泛化能力。

该文通过对教师模型的预测结果进行筛选,只选择那些教师模型在给定任务上正确预测的实例,确保学生模型只从高质量的知识中进行学习。这样一来,学生模型能够更专注于有效信息,从而进一步提高模型的准确性和性能表现,同时避免了错误知识的负迁移。

### 2 算法研究

在本节中,将详细介绍基于教师正确预测的知识蒸馏算法(CIPKD)。该算法由两个主要步骤组成:首先,从教师模型的输出中筛选出正确预测的实例;其次,针对这些正确实例分别计算实例蒸馏损失和类蒸馏损失。图 1 展示了算法的整体框架。

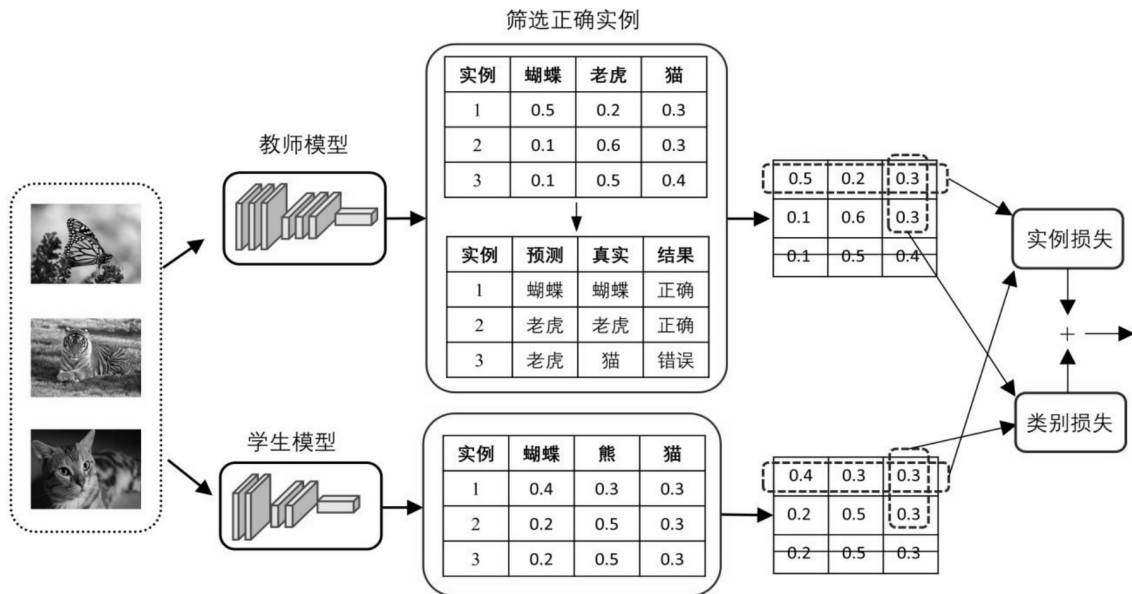


图 1 基于教师模型正确预测的图像分类知识蒸馏算法整体框架

#### 2.1 基于教师模型正确预测的实例筛选算法

将模型的输出视为一个整体批次进行处理。首

先,输入数据通过教师模型正向传播,获得批处理中所有实例的预测 Logit,这些 Logit 表示模型中每个类别

的非标准化预测分数。

对于批次中的每个实例,选择 Logit 值最高的类别作为该实例在教师模型中的预测类别。接着,将教师模型的预测类别与实际标签(目标)进行对比。如果预测类别与目标类别一致,则该实例被标记为“正确实例”;否则,标记为“不正确实例”。

为了确保学生模型在训练过程中专注于教师模型正确预测的实例,从教师和学生模型的输出中移除不正确的实例,仅保留正确预测的实例用于训练。通过过滤教师模型中预测错误的实例,学生模型能够集中精力学习具有代表性和准确性的实例,从而提升训练效率和模型的泛化性能。最后,将模型的 Logit 输出表示为  $Z \in R^{B \times C}$ ,其中  $B$  表示教师预测正确实例的数量, $C$  表示类别的数量。

## 2.2 实例蒸馏算法

首先,需要将教师模型和学生模型输出的 Logit 转换为概率分布。Logit 是模型输出的未归一化分数,表示每个类别的相对可能性。然而,由于 Logit 并未标准化,直接比较它们并不合适。因此,使用 softmax 函数对这些 Logit 进行转换,将其转换为一个范围在 0 到 1 之间的概率分布。

$$P_i = \frac{\exp(z_i/T)}{\sum_{j=1}^N \exp(z_j/T)} \quad (1)$$

其中, $P_i$  表示一维向量中第  $i$  个元素的预测概率, $N$  表示该一维向量的长度。 $T$  是超参数温度,用于对概率分布进行“软化”。

为了衡量教师模型和学生模型之间的差异,遵循原始的知识蒸馏算法,使用 KL 散度作为损失函数。

$$L_{\text{ins}} = \frac{T^2}{B} \sum_{i=1}^B \text{KL}(P_i^{\text{Tea}} \parallel P_i^{\text{Stu}}) = \frac{T^2}{B} \sum_{i=1}^B \sum_{j=1}^C P_{i,j}^{\text{Tea}} \log\left(\frac{P_{i,j}^{\text{Tea}}}{P_{i,j}^{\text{Stu}}}\right) \quad (2)$$

其中, $P_{i,j}^{\text{Tea}}$  和  $P_{i,j}^{\text{Stu}}$  分别表示教师和学生模型对第  $i$  个实例的第  $j$  个类别的预测概率。 $L_{\text{ins}}$  表示从单个实例中获得的蒸馏损失。它衡量了学生模型对特定实例的预测与教师模型对相同实例的预测之间的差异。通过 KL 散度衡量教师和学生模型的输出分布差异,帮助学生模型从教师模型的知识中学习。温度系数  $T$  使得蒸馏过程更关注教师模型的全局知识,而非单纯的最优类。

## 2.3 类别蒸馏算法

笔者认为,同一类别在不同实例上的概率分布包含丰富的信息。不同实例可能展现出与该类别不同程度的相似性。以“猫”这一类别为例,老虎的图像相比鸟类更具较高的概率分值,因为老虎与猫同属猫科动物,二者更加相似。笔者希望学生模型能够捕捉到这

些相似性的细微差异。为此,提出了类蒸馏的概念,通过计算每个类别在不同实例上的概率分布,利用 KL 散度缩小教师模型与学生模型在这些分布上的差异,从而提升学生模型的性能。类蒸馏损失可以形式化为以下公式:

$$L_{\text{class}} = \frac{T^2}{C} \sum_{j=1}^C \text{KL}(P_j^{\text{Tea}} \parallel P_j^{\text{Stu}}) = \frac{T^2}{C} \sum_{j=1}^C \sum_{i=1}^B P_{j,i}^{\text{Tea}} \log\left(\frac{P_{j,i}^{\text{Tea}}}{P_{j,i}^{\text{Stu}}}\right) \quad (3)$$

该公式计算了每个类别在不同实例上的概率分布差异。对于每一个类别  $j$ ,通过 KL 散度来度量教师模型和学生模型在该类别上的预测差异。最终的类蒸馏损失  $L_{\text{class}}$  是对所有类别的损失求平均,旨在缩小教师和学生模型在类别层面上的预测差距,使学生模型更好地捕捉到类别间的细微相似性和差异性。

## 2.4 损失函数

在本节中,提出的知识蒸馏算法结合了实例蒸馏损失  $L_{\text{ins}}$  和类蒸馏损失  $L_{\text{class}}$ ,以确保学生模型能够从教师模型中有效学习到丰富的知识。最终的知识蒸馏损失函数  $L_{\text{CIPKD}}$  表达式如下:

$$L_{\text{CIPKD}} = \alpha L_{\text{ins}} + \beta L_{\text{class}} \quad (4)$$

其中, $\alpha$  和  $\beta$  分别表示两个超参数,用于平衡实例蒸馏损失和类别蒸馏损失的贡献。

最终损失函数  $L$  将分类损失与蒸馏损失结合在一起,对学生模型进行训练。最终损失函数的表达式如下:

$$L = \lambda L_{\text{CE}} + (1 - \lambda) L_{\text{CIPKD}} \quad (5)$$

其中, $L_{\text{CE}}$  表示针对当前分类任务的标签监督损失, $\lambda$  是一个平衡因子,用于在总损失函数中平衡不同损失项的影响力。

## 3 实验结果与分析

### 3.1 数据集描述

CIFAR-100 是图像分类领域中广泛使用的经典数据集,以其细粒度的类别结构和低分辨率的图像成为模型评估的重要基准之一。该数据集包含 60 000 张  $32 \times 32$  像素的彩色图像,其中训练集 50 000 张,测试集 10 000 张。拥有 100 个不同的类别,每个类别有 600 张图像,涵盖了哺乳动物、鱼类、昆虫、车辆和植物等各种物体。

ImageNet 是一个大型的图像数据库,包含超过 120 万张高分辨率的标注图像,涵盖 1 000 个物体类别,涉及现实世界中常见的各种物体,例如动物、植物、日常用品、交通工具等。

### 3.2 实验设置

将师生模型分为同构架构和异构架构两大类,以

便更细致地研究不同网络架构下知识蒸馏算法的表现。

在同构架构中,表示教师模型和学生模型属于相同类型的网络架构。具体而言,它们的网络基本结构相似,只是网络的深度或宽度不同。例如 ResNet56 和 ResNet20 都是残差网络(ResNet)架构,但前者比后者更深。

在异构架构中,教师模型和学生模型在网络架构设计上存在显著差异。例如,ResNet32x4 是一种较深且宽的残差网络,而 ShuffleNet-V1 则是一种轻量级的网络,采用不同的模块设计和计算方式。

此外,实验中还包括其他常见的神经网络架构,如 Wide ResNet(WRN)、VGG、ShuffleNet 和 MobileNet。对比算法包括基于特征和基于 Logit 的算法,例如 FitNet<sup>[10]</sup>、RKD<sup>[13]</sup>、CRD<sup>[17]</sup>、OFD<sup>[18]</sup>、ReviewKD<sup>[11]</sup>、DKD<sup>[15]</sup> 以及基准算法 KD<sup>[6]</sup>。

### 3.3 实验细节

实验均在 Ubuntu 18.04 操作系统下,使用 Nvidia RTX3090 进行训练,Pytorch 版本为 1.10.1。

对于 CIFAR-100 数据集,输入图像通过随机裁剪和翻转进行数据增强。教师模型和学生模型训练 240 个周期。批量大小(batch\_size)为 64,ShuffleNet 和 MobileNet-V2 的学习率为 0.01,而其他系列(如 VGG、ResNet 和 WRN)的学习率为 0.05。在第 150、

180 和 210 个周期时,将学习率降低 90%。使用 SGD (Stochastic Gradient Descent) 作为优化器。

对于 ImageNet 数据集,对输入图像通过随机裁剪和翻转来进行数据增强。训练 100 个周期。批量大小为 256,学习率初始化为 0.2,并在每 30 个周期后降低 90%。权重衰减设置为 1e-4。

### 3.4 评价指标

采用准确率(Accuracy)作为模型性能的主要评价指标。准确率定义为模型预测正确的样本数量与总样本数量的比值,用于衡量模型在给定测试集上的整体分类表现。准确率越高,分类效果越好。在后续的实验分析中,所有模型的性能比较均基于准确率这一指标展开,对于准确率最高的算法,将其加粗以示突出。

### 3.5 结果分析

表 1 和表 2 分别展示了 CIPKD 算法在 CIFAR-100 数据集上同构架构和异构架构师生模型的实验结果。在同构架构师生对中,与基准 KD 算法相比,该算法的准确率提升了 1.07 百分点至 3.81 百分点;在异构架构师生对中,提升了 2.19 百分点至 2.85 百分点。总体来看,在大多数师生对中,该算法均优于其他对比算法。表 3 展示了在大规模数据集 ImageNet 上的实验结果,该算法相较基准 KD 算法实现了 1.68 百分点至 3.6 百分点的提升,同样优于其他对比算法。

表 1 CIFAR-100 同构架构师生对模型准确率 %

Teacher	Student			Feature					Logit		
				FitNet	RKD	CRD	OFD	ReviewKD	KD	DKD	Ours
ResNet56	72.34	ResNet20	69.06	69.21	69.61	71.16	70.98	71.89	70.66	<b>71.97</b>	71.77
ResNet110	74.31	ResNet32	71.14	71.06	71.82	73.48	73.23	73.89	73.08	74.11	<b>74.15</b>
ResNet32x4	79.42	ResNet8x4	72.50	73.50	71.90	75.51	74.95	75.63	73.33	76.32	<b>77.14</b>
Wm-40-2	75.61	Wm-16-2	73.26	73.58	73.35	75.48	75.24	76.12	74.92	76.24	<b>76.30</b>
Wm-40-2	75.61	Wm-40-1	71.98	72.24	72.22	74.14	74.33	75.09	73.54	74.81	<b>75.61</b>
VGG13	74.64	VGG8	70.36	71.02	71.48	73.94	73.95	74.84	72.98	74.68	<b>74.78</b>

表 2 CIFAR-100 异构架构师生对模型准确率 %

Teacher	Student			Feature				Logit		
				FitNet	RKD	CRD	OFD	KD	DKD	Ours
ResNet32x4	79.42	ShuffleNet-V1	70.50	73.59	72.28	75.11	75.98	74.07	76.45	<b>76.78</b>
Wm-40-2	75.61	ShuffleNet-V1	70.50	73.73	72.21	76.05	75.85	74.83	76.70	<b>77.02</b>
VGG13	74.64	MobileNet-V2	64.60	64.14	64.52	69.73	69.48	67.37	69.71	<b>70.15</b>
ResNet50	79.34	MobileNet-V2	64.60	63.16	64.43	69.11	69.04	67.35	<b>70.35</b>	69.64
ResNet32x4	79.42	ShuffleNet-V2	71.82	73.54	73.21	75.65	76.82	74.45	77.07	<b>77.30</b>

表 3 ImageNet 模型准确率 %

	Top-1	Top-5	Top-1	Top-5
teacher	ResNet34		ResNet50	
	73.31	91.42	76.16	92.86
student	ResNet18		MobileNet-V2	
	69.75	89.07	68.87	88.76
AT	70.69	90.01	69.56	89.33
OFD	70.81	89.98	71.25	90.34
CRD	71.17	90.13	71.37	90.41
KD	70.66	89.88	68.58	88.98
DKD	71.70	90.41	72.05	<b>91.05</b>
Ours	<b>72.34</b>	<b>90.77</b>	<b>72.18</b>	90.52

表 4 展示了消融实验的结果,其中,使用 ResNet32×4 作为教师网络,ResNet8×4 作为学生网络。当仅进行实例蒸馏时,算法退化为基准 KD 算法。

表 4 CIFAR-100 数据集消融实验结果

实例蒸馏	正确实例筛选	类别蒸馏	准确率/%
√	×	×	73.33
√	√	×	74.97
√	×	√	75.28
√	√	√	77.14

单独筛选教师正确预测的样本,准确率提升了 1.64 个百分点。单独引入类别蒸馏,准确率提升了 1.95 个百分点。最终融合所有算法后,整体提升达 3.81 个百分点。

表 5 展示了实例蒸馏损失和类别蒸馏损失在选用不同的权重时的结果,使用 ResNet32×4 作为教师网络,ResNet8×4 作为学生网络。当  $\alpha$  和  $\beta$  权重比例为 [1.0,2.0] 时,学生模型达到了最佳性能,Top-1 准确率为 77.14%。

表 5 不同超参数设置时模型准确率 %

$[\alpha, \beta]$	Top-1
[0.5,1.0]	76.72
[1.0,0.5]	76.31
[1.0,1.0]	76.59
[1.0,2.0]	77.14
[1.0,4.0]	76.69

图 2 展示了 CIPKD 算法和基准蒸馏算法 KD 在 CIFAR-100 数据集上的训练精度对比。从结果来看,无论是在早期训练阶段的上升速度,还是在后期收敛的准确度,CIPKD 算法都优于 KD 算法。这表明学生模型能够获得更好的分类能力。

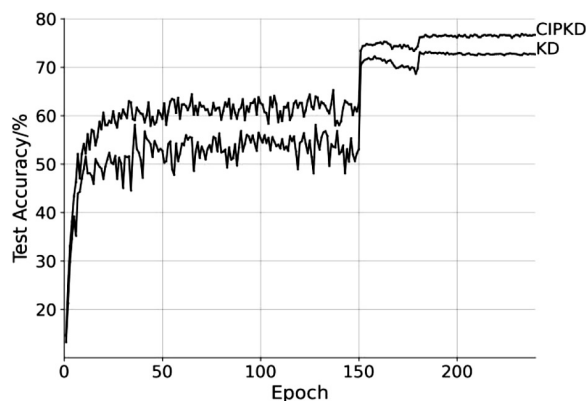


图 2 KD 和 CIPKD 的训练精度变化对比

#### 4 结束语

该文提出了一种基于教师正确预测的知识蒸馏算法。该算法能有效地利用教师模型的输出 Logit 信息。首先过滤教师模型中错误预测的实例,确保学生模型仅从正确预测的实例中学习。这一过滤步骤显著提高了传递给学生模型知识的质量。此外,CIPKD 引入了一种结合实例级和类别级知识蒸馏策略。在实例级,学生模型学习每个实例的输出分布;在类别级,学生模型学习同一类别中多个实例之间的概率分布关系,从而更好地理解一个类别在不同实例中的表现。这种双层次的蒸馏丰富了学生的知识,并增强了其泛化能力。在 CIFAR-100 和 ImageNet 数据集上的实验结果表明,该算法显著提升了学生模型的分类能力。

综上所述,该算法提升了传统知识蒸馏算法的性能,使学生模型能够更好地学习教师模型传授的知识,从而更便于在实际应用中进行落地部署。

#### 参考文献:

[1] ZHANG X,ZHOU X,LIN M,et al. Shufflenet: an extremely efficient convolutional neural network for mobile devices [C]//Proceedings of the IEEE conference on computer vi-

- sion and pattern recognition. Salt Lake City: IEEE, 2018: 6848–6856.
- [2] 赵 风, 耿苗苗, 刘汉强, 等. 卷积神经网络与视觉 Transformer 联合驱动的跨层多尺度融合网络高光谱图像分类方法[J]. 电子与信息学报, 2024, 46(5): 2237–2248.
- [3] 陈立玮, 周新志. 基于特征自注意力的图像分类知识蒸馏算法[J]. 现代计算机, 2023, 29(4): 49–53.
- [4] 程凯伦, 胡晓兵, 陈海军, 等. 基于改进 YOLOv5s 的遥感图像目标检测方法[J]. 激光与光电子学进展, 2024, 61(18): 285–291.
- [5] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Boston: IEEE, 2015: 3431–3440.
- [6] HINTON G. Distilling the knowledge in a neural network [J]. arXiv:1503.02531, 2015.
- [7] LI X C, FAN W S, SONG S, et al. Asymmetric temperature scaling makes larger networks teach well again[J]. Advances in Neural Information Processing Systems, 2022, 35: 3830–3842.
- [8] MIRZADEH S I, FARAJTABAR M, LI A, et al. Improved knowledge distillation via teacher assistant[C]//Proceedings of the AAAI conference on artificial intelligence. New York: AAAI, 2020: 5191–5198.
- [9] 常晓琴, 李雅梦, 李子成, 等. 基于多教师多学生知识蒸馏的情感分类方法[J]. 中文信息学报, 2024, 38(10): 127–134.
- [10] ROMERO A, BALLAS N, KAHOU S E, et al. Fitnets: hints for thin deep nets[J]. arXiv:1412.6550, 2014.
- [11] CHEN P, LIU S, ZHAO H, et al. Distilling knowledge via knowledge review[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Nashville: IEEE, 2021: 5008–5017.
- [12] 杨传广, 陈路明, 赵二虎, 等. 基于图表征知识蒸馏的图像分类方法[J]. 电子学报, 2024, 52(10): 3435–3447.
- [13] PARK W, KIM D, LU Y, et al. Relational knowledge distillation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Long Beach: IEEE, 2019: 3967–3976.
- [14] XU G, LIU Z, LI X, et al. Knowledge distillation meets self-supervision[C]//European conference on computer vision. Glasgow: Springer, 2020: 588–604.
- [15] ZHAO B, CUI Q, SONG R, et al. Decoupled knowledge distillation[C]//Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition. New Orleans: IEEE, 2022: 11953–11962.
- [16] YANG Z, ZENG A, LI Z, et al. From knowledge distillation to self-knowledge distillation: a unified approach with normalized loss and customized soft labels[C]//Proceedings of the IEEE/CVF international conference on computer vision. Paris: IEEE, 2023: 17185–17194.
- [17] TIAN Y, KRISHNAN D, ISOLA P. Contrastive representation distillation[J]. arXiv:1910.10699, 2019.
- [18] HEO B, KIM J, YUN S, et al. A comprehensive overhaul of feature distillation[C]//Proceedings of the IEEE/CVF international conference on computer vision. Seoul: IEEE, 2019: 1921–1930.