计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
63(10): 2682−2696, 2026

基于跨标签分配知识蒸馏的端到端目标检测方法

胡志良[1,2] 陈 思[1,2] 王大寒[1,2] 严 严[3]

[1](厦门理工学院计算机与信息工程学院 福建厦门 361024)

[2](福建省模式识别与图像理解重点实验室(厦门理工学院) 福建厦门 361024)

[3](厦门大学信息学院 福建厦门 361100)

(hzliang@whut.edu.cn)

Cross-Label Assignment Knowledge Distillation Based End-to-End Object Detection

Hu Zhiliang[1,2], Chen Si[1,2], Wang Dahan[1,2], and Yan Yan[3]

[1](*School of Computer and Information Engineering*, *Xiamen University of Technology*, *Xiamen*, *Fujian* 361024)

[2](*Fujian Key Laboratory of Pattern Recognition and Image Understanding* (*Xiamen University of Technology*), *Xiamen*, *Fujian* 361024)

[3](*School of Informatics*, *Xiamen University*, *Xiamen*, *Fujian* 361100)

Abstract Recently, the one-to-one label assignment rule has played a great role in removing non-maximum suppression (NMS) in the post-processing step and building an NMS-free end-to-end detection paradigm. However, this strict sample matching rule leads to a significant reduction in the number of positive samples during the training process, making it difficult for the model to fully mine the latent semantic information in the data during the feature representation learning stage, resulting in low learning efficiency. To this end, we propose a cross-label assignment knowledge distillation (CAKD) method for end-to-end object detection. This method combines the transfer learning mechanism of knowledge distillation with the end-to-end object detection algorithm to make up for the defects of the existing one-to-one label assignment rule and establishes an effective knowledge transfer path for the knowledge distillation framework to train end-to-end detection models. Specifically, the multi-scale features of the student model are first transferred to the detection head of the teacher model across models. Then, the student-teacher cross-label assignment predictions are compared with the teacher's predictions to calculate the distillation loss. This method avoids directly forcing the student to imitate the teacher's features, which leads to feature confusion and semantic misalignment. In addition, we also design an effective task-aware matching metric that comprehensively considers the quality of classification and regression, thus avoiding the problem of irrelevance between classification and localization tasks. We conduct a large number of experiments on COCO dataset to demonstrate the effectiveness of this method. By applying the cross-label assignment knowledge distillation method proposed in this paper to the non-end-to-end FCOS baseline model, it achieves a 38.8% non-NMS detection performance in terms of mean average precision (*mAP*), which is a 2.2 percentage point accuracy gain compared with the original detection performance with NMS.

Key words end-to-end object detection; NMS-free detection; knowledge distillation; cross-label assignment; computer vision

收稿日期: 2025-10-14; **修回日期**: 2026-04-28

基金项目: 国家自然科学基金项目(62571466); 厦门市自然科学基金项目(3502Z202573073)

This work was supported by the National Natural Science Foundation of China (62571466) and the Natural Science Foundation of Xiamen (3502Z202573073).

通信作者: 陈思(chensi@xmut.edu.cn)

摘　要　1对1标签分配规则在去除后处理步骤中的非极大值抑制（non-maximum suppression, NMS）和构建无NMS的端到端检测范式中发挥了巨大作用。然而，这种严格的样本匹配规则导致训练过程中正样本数量显著减少，使得模型在特征表示学习阶段难以充分挖掘数据中的潜在语义信息，学习效率低下。为此，提出了一种跨标签分配知识蒸馏（cross-label assignment knowledge distillation, CAKD）方法，用于实现端到端的目标检测。该方法将知识蒸馏的迁移学习机制与端到端目标检测算法相结合以弥补现有1对1标签分配规则的缺陷，为知识蒸馏框架用在训练端到端检测模型建立了一条有效的知识迁移路径。具体地，该方法首先将学生模型的多尺度特征传递到教师模型的检测头中。然后，将学生—教师跨标签分配预测与教师的预测用于计算蒸馏损失。这种方式避免了直接强制学生模仿教师的特征导致的特征混乱、语义错位的问题。此外，设计了一种有效的任务感知匹配度量方式，综合考虑了分类和回归质量，避免了分类和定位任务之间没有相关性的问题。在COCO数据集上进行了大量实验证明了提出方法的有效性。提出的跨标签分配知识蒸馏方法在非端到端FCOS基线模型上的均值平均精度（mean average precision, *mAP*）达到了38.8%的无NMS检测性能，其相较于原始有NMS的检测性能实现了2.2个百分点的精度增益。

关键词　端到端目标检测；无NMS检测；知识蒸馏；跨标签分配；计算机视觉

中图法分类号　TP311.13；TP309

DOI: 10.7544/issn1000-1239.202550737　　**CSTR:** 32373.14.issn1000-1239.202550737

目标检测是计算机视觉领域的一项长期任务，旨在检测图像中预定目标的类别和位置[1]。目前，目标检测已被广泛应用于视觉相关任务中，例如实例分割[2]、机器人导航[3]、目标跟踪[4]和自动驾驶[5]。

当前，主流目标检测框架[6-11]普遍采用非端到端检测范式。如图1上半部分所示，这类检测范式通常采用1对多标签分配规则[12-13]，每个真实实例通常会对应多个候选区域或锚框。使用这种分配策略可以大幅增加正样本的数量，从而为模型提供更加丰富的监督信息。然而，这种策略会带来新的问题，即在推理中单个目标对象出现多个近似预测框，不可避免带来冗余检测，从而需要依赖后处理步骤中的非极大值抑制（non-maximum suppression, NMS）去除冗余检测框。虽然，NMS及其变体[14-16]在过去持续推动了目标检测方法的进步，但其局限性在复杂场景中越发明显。首先，检测器的性能优劣依赖人工设定的NMS阈值参数，同时参数还需要针对不同场景手动调整，缺乏泛化性。其次，NMS在训练与推理时并不统一，模型训练时以独立预测框的损失函数为优化目标，而推理时却通过启发式规则对预测框之间的关系进行后处理，导致训练与推理的逻辑割裂，形成“局部最优但全局次优”的矛盾。更关键的是，NMS无法处理高度重叠或遮挡目标的检测需求，其重叠目标可能被NMS误判为冗余框而剔除，导致目标丢失问题。

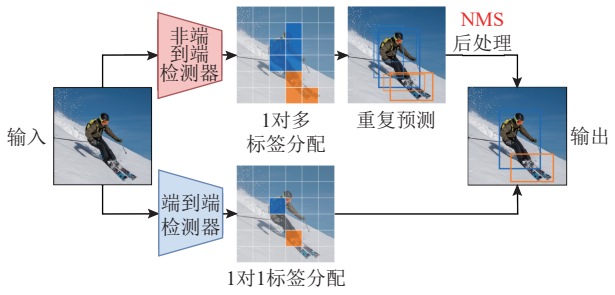


Fig. 1　The main difference between non-end-to-end and end-to-end detectors

图1　非端到端检测器和端到端检测器的主要区别

为了解决以上问题，学术界开始探索端到端目标检测范式[17-19]。如图1下半部分所示，与传统非端到端目标检测方法不同，端到端检测器通过使用1对1标签分配规则，在训练阶段强制每个真实目标仅与1个候选区域或锚框匹配，从而避免冗余预测，无需NMS直接输出无冗余的检测结果。然而，这种严格的样本匹配机制导致训练过程中前景样本数量显著减少，使得模型在特征表示学习阶段难以充分挖掘特征中的潜在语义信息，导致模型收敛效率低下和性能较低等问题。现有研究主要通过改进标签分配策略（如引入动态匹配或软权重机制），但这类方法需要更多的超参数调整，这可能会增加训练的不稳定性。其次，通过增强解码器结构复杂度来提升模型性能。然而，这些方法往往需要增加网络复杂度或依赖特定场景的先验知识，难以在效率与泛化性之间取得平衡。

值得注意的是，知识蒸馏（knowledge distillation,

KD)通过迁移大模型的知识到小模型中，已被证明是提高小模型性能和泛化能力的有效方式。自 Hinton 等人[20]提出软标签蒸馏范式以来，知识蒸馏已被广泛应用于图像分类任务中，其优势在于无需额外标注数据即可提升学生模型的泛化能力。在目标检测领域，Chen 等人[21]通过迁移教师模型的多尺度特征图与预测框置信度，增强学生模型对小目标与遮挡场景的检测鲁棒性。一个自然的问题是：能否将知识蒸馏的迁移学习机制与端到端目标检测算法相结合，以弥补现有 1 对 1 标签分配规则的缺陷。具体而言，使用 1 对多标签分配规则的教师模型在训练过程中学习到的表征信息可能蕴含丰富的判别性知识，若将其提炼至学生模型，或可在不增加推理开销的前提下，显著增强使用 1 对 1 标签分配规则的学生模型对复杂场景的适应能力。然而，目前基于知识蒸馏的目标检测方法在设计时主要面向 2 阶段或基于密集预测的检测器，蒸馏策略往往未能充分考虑端到端检测器[17-19]中 1 对 1 标签分配方法的特殊性。这种分配方法虽然简化了推理流程，但也限制了学生模型在复杂场景中的鲁棒性与表现力。如何设计既与端到端检测器兼容的蒸馏范式，又能充分挖掘教师模型 1 对多匹配过程中蕴含的判别性知识，仍是当前研究的难点。

因此，本文提出了一种用于端到端目标检测器的跨标签分配知识蒸馏(cross-label assignment knowledge distillation, CAKD)方法。该方法为知识蒸馏框架用在训练端到端检测模型建立了一条有效的知识迁移路径，有效弥补了现有 1 对 1 标签分配规则的缺陷。本文的主要贡献如下：

1)采用了有效的跨标签分配知识蒸馏路径将学生模型的多尺度特征传递给教师模型的检测头，避免了直接强制学生模仿教师的特征，接收到与其端到端优化目标矛盾的监督信号，解决了由于教师和学生使用不同标签分配策略导致的特征冲突和目标冲突的问题。

2)设计了任务感知匹配度量，其将分类和回归质量同时纳入到选择度量中，避免了分类和定位任务之间没有相关性的问题。

3)联合了 1 对多与 1 对 1 标签分配规则，构成双标签分配规则，进一步弥补特征表示学习的不足。

4)探索了将知识蒸馏框架用在端到端目标检测方法上，通过提出的方法，能够将传统非端到端目标检测器训练为端到端检测器，并实现具有竞争力的端到端检测性能。

1 相关工作

1.1 知识蒸馏方法

知识蒸馏是一种旨在将大规模、高性能教师模型中蕴含的知识传递给轻量级学生模型的模型压缩技术，其基本操作流程是先利用大量数据训练出性能卓越但计算复杂度较高的教师模型，然后在学生模型的训练过程中，通过设计适当的蒸馏损失函数，尽可能地模仿教师模型的输出分布和中间特征表示。当前，用于知识蒸馏的知识主要分为基于响应的知识和基于特征的知识 2 类。

基于响应的知识蒸馏方法的主要思想是让学生模型直接模仿教师模型最终预测 logits(未归一化的类别概率)，从而实现知识的迁移。该方法最早用于图像分类任务并由 Hinton 等人[20]提出，通过引入温度参数对 logits 进行软化处理，使得学生模型不仅能学习教师模型的类别预测结果，更能捕捉类别间相对关系所隐含的暗知识。目标检测任务中的响应还包含有关边界框偏移量的 logits。Zheng 等人[22]提出的定位蒸馏(location distillation, LD)方法，将传统边界框的回归问题转化为概率分布建模，通过软化教师模型的定位知识，指导学生模型学习更精细的定位信息。

基于特征的知识蒸馏方法通过迁移教师模型中间层特征所蕴含的语义信息和空间关系，引导学生模型优化特征表示能力，已成为提升轻量化模型性能的重要范式。与上述基于响应的方法不同，该方法的核心在于构建教师模型与学生模型特征空间的对齐机制。早期研究中，Romero 等人[23]提出的 FitNets 方法利用教师网络中间层的特征响应作为提示(hint)，通过最小化学生网络适配层输出的 L2 距离实现特征知识迁移，在图像分类任务中验证了深层特征指导的有效性。此后，Zagoruyko 等人[24]引入注意力迁移(attention transfer)，将教师模型特征图的通道注意力图作为监督信号，使学生模型能够聚焦于关键区域的特征学习。

在目标检测领域，基于特征的蒸馏面临特征尺度多样性、空间敏感性和任务耦合性等挑战，Wang 等人[25]提出利用特征响应跨位置差异的细粒度特征模仿方法，通过物体锚点位置上的特征响应差异指导学生模型在这些位置上模仿老师，以获得增强的性能。Liu 等人[26]提出了实例关系图(IRG)用于知识

蒸馏，通过建模实例特征、实例关系和特征空间变换，有效迁移了整个网络的知识。Guo 等人[27]提出的 DeFeat 方法通过解耦特征蒸馏策略，将目标检测中的前景与背景区域特征分离，并差异化分配蒸馏权重。Yang 等人[28]提出的焦点和全局蒸馏（FGD）方法，通过将前背景分离，迫使学生专注于老师的关键像素和通道。

尽管上述知识蒸馏方法在一定程度上提升了学生模型在目标检测任务中的性能，但当前用于端到端目标检测的知识蒸馏方法仍存在一些关键不足。首先，上述方法对教师与学生模型的输出或中间特征进行强对齐，没有考虑到端到端检测器和非端到端检测器在特征响应上的根本不同，故存在特征冲突。其次，在优化目标上端到端的稀疏检测与非端到端的密集检测存在冲突，这会导致训练优化时学生网络和教师网络之间的知识迁移时可能造成信息混乱或学习不稳定。在本文中，我们提出了一种用于端到端目标检测器的跨标签分配知识蒸馏方法，该方法通过构建学生—教师知识蒸馏路径，避免了直接强制学生模仿教师的特征，解决了由于教师和学生使用不同标签分配策略导致的特征冲突和目标冲突的问题，为知识蒸馏框架用在训练端到端检测模型建立了一条有效的知识迁移路径。

1.2 端到端目标检测方法

在目标检测的发展中，一个重要趋势是移除 NMS 以实现端到端检测。早期的研究[29-30]，一些循环神经网络（recurrent neural network, RNN）框架使用长短期记忆（long short-term memory, LSTM）单元将图像内容解码为一组可变长度的预测框。尽管它们实现了端到端的检测，但它们仅在有限的小数据集上得到了证明的有效性。由 Hu 等人[31]提出的关系网络（relational network）是一种将注意力机制引入目标检测的创新方法。它通过建模目标之间的成对关系，将重复检测问题转化为二元分类问题，从而减少冗余检测。这种方法不仅提升了模型的预测能力，还为端到端检测提供了一种新的思路。在此基础上，Carion 等人[18]提出的 DETR 方法成为了端到端目标检测方法的代表性工作，该方法利用 Transformer 架构，通过 1 组固定的可学习查询对象与图像特征进行交互，并使用 1 对 1 匹配取代传统的候选框分配机制，在训练时全局优化学习直接输出稀疏的检测结果，无需依赖 NMS 等后处理。此后，各种 DETR 变体[32-35]被提出以提升性能和效率。RT-DETR[36]进一步设计了高效的混合编码器，并提出了基于匈牙利匹配算法的不确定性最小查询选择。Wang 等人[37]提出的 RT-DETRv3 方法通过引入一个新的包含自注意力扰动的学习策略使得模型训练时能够跨多个查询组对前景样本进行多样化的标签分配从而丰富正监督。然而，Transformer 模块包含大量冗余参数，导致模型复杂度较高，难以部署在资源受限的设备上。

最近，一些基于卷积神经网（convolutional neural network, CNN）的端到端目标检测方法在训练过程中采用了 1 对 1 标签分配规则，该分配规则在训练期间为每个真实实例只匹配 1 个前景样本，为端到端目标检测提供了一种更有效的解决方法。与 DETR 相比，这些基于 CNN 的方法收敛速度更快。其中代表性的方法是由 Sun 等人[17]提出的 OneNet 方法，通过最小代价标签匹配与全卷积架构设计，在单阶段检测框架中实现了无 NMS 的纯端到端检测范式。由 Wang 等人[19]提出的 DeFCN 方法通过预测感知 1 对 1（prediction-aware one-to-one, POTO）标签分配策略在训练期间为每个真实实例只分配 1 个前景样本，并通过动态计算预测框与真实框的联合概率分布，在单阶段框架内实现端到端匹配。尽管这些方法通过使用 1 对 1 标签分配策略进行训练，在推理时不需要 NMS 即可直接输出检测结果，但是存在鲁棒性下降和收敛困难的问题。在相同的设置下，其性能不如最初使用 1 对多分配的检测器。文献 [38] 提出端到端检测器性能降低是因为 1 对 1 标签分配为监督提供了较少的前景样本，导致检测模型在训练时无法充分进行特征表示学习。同时，前景样本太少还加剧了前景样本与背景样本不均的问题。因此，本文首次探索了将知识蒸馏框架用在端到端目标检测方法上，将知识蒸馏的迁移学习机制与端到端目标检测算法相结合，充分利用 1 对 1 与 1 对多标签分配方式的优势，以弥补现有端到端检测器特征表示学习能力不足的问题。

2 本文方法

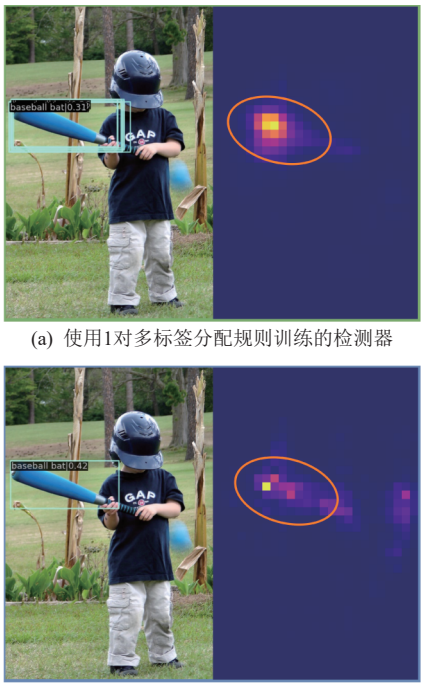
2.1 动机和框架

通常，检测模型依赖于人工设定的标签分配规则进行训练，即确定前景样本和背景样本[38]。在传统蒸馏模型中，教师检测器与学生检测器中的标签分配规则通常是相同的，以保证在训练中检测器可以准确地重现分配器的标签，避免在知识蒸馏中教师与学生目标之间发生冲突[39]。本文将利用基于知识蒸馏的迁移学习机制以弥补现有端到端检测器特

征表示能力不足的问题。然而教师模型是使用传统1对多标签分配训练的密集检测器，学生模型是使用1对1标签分配训练的端到端检测器。因此，学生模型使用了1对1分配这种与教师模型1对多标签分配截然不同的分配规则，从而导致两者优化目标冲突。如图1所示，与1对多标签分配规则不同，1对1分配规则在训练时只将1个前景样本与1个真实实例进行匹配。通过这种分配方式，给学生模型提供去除重复的监督信号，促使检测器从密集检测转向稀疏检测，以保证学生模型可以不需要NMS就直接输出检测结果。

其次，不同标签分配策略带来的特征冲突问题导致两者的特征空间难以对齐。图2详细展示了两者在特征空间分布的不同。1对多标签分配训练的模型需同时优化多个候选框的分类和回归，特征中隐含冗余预测的竞争关系，同一位置的特征需响应多个候选框，特征图更“分散”，更偏向于密集检测。而使用1对1标签分配训练的模型，特征需高度聚焦于唯一匹配的预测框，仅在与目标中心强相关的稀疏区域产生高响应(稀疏激活)，更偏向于单一目标的精确检测，导致特征表示过度局部化。故这2种不同的标签分配策略存在密集-稀疏特征分布的结构性差异。由于教师模型使用1对多标签分配，而学生模型使用1对1标签分配，若直接强制学生模仿教师的输出特征，学生会被迫学习分散的响应模式，接收到与其端到端优化目标相矛盾的监督信号，导致特征混乱而发生语义错位。

因此，本文提出了一种跨标签分配知识蒸馏(cross-label assignment knowledge distillation, CAKD)方法。CAKD方法的框架如图3所示。该方法将生成



(a) 使用1对多标签分配规则训练的检测器

(b) 使用1对1标签分配规则训练的检测器

Fig. 2 Visualization of output feature maps trained by detectors with different label assignment rules

图2 使用不同标签分配规则训练的检测器输出特征图可视化

的学生—教师跨标签分配预测与教师的预测计算跨标签分配知识蒸馏损失L_{CAKD}，避免了直接强制学生模仿教师的特征。同时，还为CAKD方法引入了双标签分配规则，在学生—教师前向路径中使用1对多标签分配策略进行训练来计算1对多分配损失L_{o2m}，以补充表征信息。学生模型前向路径中使用1对1标签分配策略进行训练来计算1对1分配损失L_{o2o}，以保证模型能够获得抑制冗余检测的能力。并且为了解决当前预测分类和回归质量不一致的问题，即

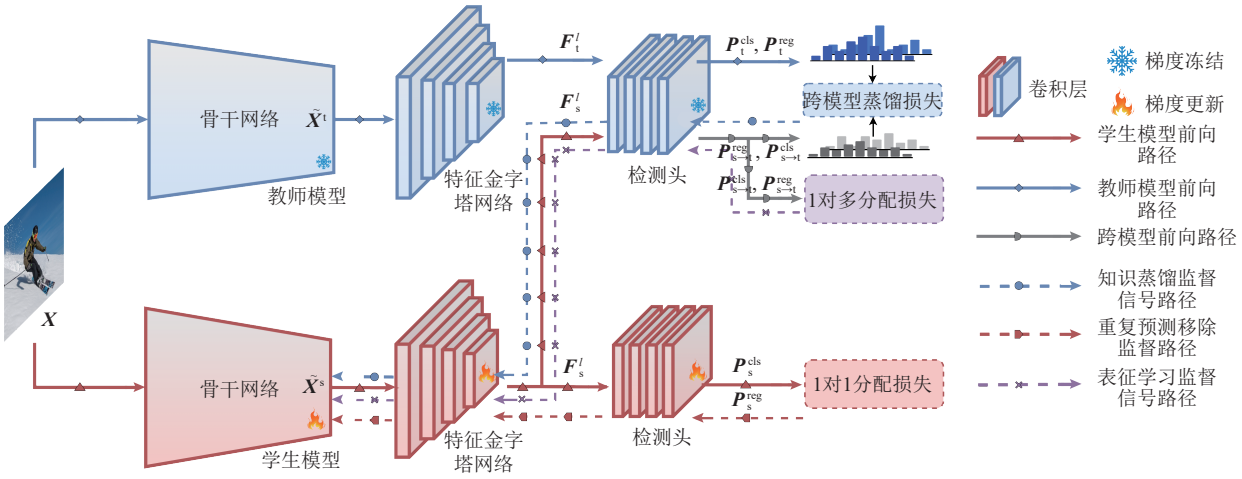


Fig. 3 Network structure diagram of CAKD

图3 CAKD网络结构图

分类得分最高的位置通常不是回归实例边界的最佳位置，还设计了任务感知匹配度量。

2.2 跨标签分配知识蒸馏

跨标签分配知识蒸馏方法可以使用多种基准模型，例如 RetinaNet[9]，FCOS[10]，GFL[40]。我们以 FCOS 作为基准框架。如图 3 所示，以输入图片 $\boldsymbol{X}$ 为例，我们提出的 CAKD 方法的详细过程如下。蒸馏框架分为 2 块，上半部分为教师模型，下半部分为学生模型。2 种模型都拥有相同的结构，包括骨干网络（backbone）、特征金字塔网络（FPN）和检测头（head）。教师模型和学生模型的特征提取过程如式（1）所示。

$$\begin{aligned}\tilde{\boldsymbol{X}}^{\mathrm{t}} &= Backbone_{\mathrm{t}}(\boldsymbol{X}),\\ \tilde{\boldsymbol{X}}^{\mathrm{s}} &= Backbone_{\mathrm{s}}(\boldsymbol{X}),\end{aligned} \tag{1}$$

式中 $Backbone_{\mathrm{t}}$ 和 $Backbone_{\mathrm{s}}$ 分别表示教师与学生的骨干网络。图片 $\boldsymbol{X}$ 经过教师和学生模型中的骨干网络进行特征提取分别得到 $\tilde{\boldsymbol{X}}^{\mathrm{t}}$ 和 $\tilde{\boldsymbol{X}}^{\mathrm{s}}$。

随后，$\tilde{\boldsymbol{X}}^{\mathrm{t}}$ 和 $\tilde{\boldsymbol{X}}^{\mathrm{s}}$ 输入进 FPN 中通过特征金字塔网络提取多尺度特征。多尺度特征提取过程如式（2）所示。

$$\begin{aligned}\boldsymbol{F}_{\mathrm{t}}^{l} &= FPN_{\mathrm{t}}(\tilde{\boldsymbol{X}}_{\mathrm{t}}),\\ \boldsymbol{F}_{\mathrm{s}}^{l} &= FPN_{\mathrm{s}}(\tilde{\boldsymbol{X}}_{\mathrm{s}}),\end{aligned}\quad \text{s.t. } l\in\{1,2,\cdots,n\}, \tag{2}$$

式中 FPN_{t} 和 FPN_{s} 分别表示教师和学生模型的特征金字塔网络。假设教师和学生模型中FPN的输出有 n 层，则输出多尺度特征 $\boldsymbol{F}$ 分别为 $\{\boldsymbol{F}_{\mathrm{t}}^{1},\boldsymbol{F}_{\mathrm{t}}^{2},\cdots,\boldsymbol{F}_{\mathrm{t}}^{n}\}$ 和 $\{\boldsymbol{F}_{\mathrm{s}}^{1},\boldsymbol{F}_{\mathrm{s}}^{2},\cdots,\boldsymbol{F}_{\mathrm{s}}^{n}\}$。后续步骤以取中间特征 $\boldsymbol{F}_{\mathrm{t}}^{l}$ 和 $\boldsymbol{F}_{\mathrm{s}}^{l}$ 为例，其中 $l\in\{1,2,\cdots,n\}$ 表示多尺度层级。

之后，2 组特征将被输入检测头。检测头是目标检测网络的最后一部分，其将从特征提取网络中获得的高层特征转换为最终的检测结果，包括目标的分类概率 $\boldsymbol{P}^{\mathrm{cls}}$ 和回归位置 $\boldsymbol{P}^{\mathrm{reg}}$。其中，教师特征 $\boldsymbol{F}_{\mathrm{t}}^{l}$ 输入教师检测头 $Head_{\mathrm{t}}$，输出分类概率 $\boldsymbol{P}_{\mathrm{t}}^{\mathrm{cls}}$ 和回归位置 $\boldsymbol{P}_{\mathrm{t}}^{\mathrm{reg}}$。需要注意的是，在本文的方法中，跨标签分配特征模仿路径将学生模型的多尺度特征传递给教师模型的检测头，避免了直接强制学生模仿教师的特征，接收到与其端到端优化目标矛盾的监督信号，缓解由于教师和学生使用不同标签分配策略导致的特征冲突和目标冲突的问题。因此，学生特征 $\boldsymbol{F}_{\mathrm{s}}^{l}$ 的输入将有 2 条：一条是跨标签分配前向路径，将学生特征 $\boldsymbol{F}_{\mathrm{s}}^{l}$ 输入到教师模型的检测头 $Head_{\mathrm{t}}$ 中。另一条是将学生特征 $\boldsymbol{F}_{\mathrm{s}}^{l}$ 输入到自身的学生模型的检测头 $Head_{\mathrm{s}}$ 中。

因此，本方法包含 3 组分类和回归输出，分别为：1）教师特征 $\boldsymbol{F}_{\mathrm{t}}^{l}$ 到教师检测头 $Head_{\mathrm{t}}$ 的输出分类概率 $\boldsymbol{P}_{\mathrm{t}}^{\mathrm{cls}}$ 和回归位置 $\boldsymbol{P}_{\mathrm{t}}^{\mathrm{reg}}$；2）学生特征 $\boldsymbol{F}_{\mathrm{s}}^{l}$ 到教师检测头 $Head_{\mathrm{t}}$ 的输出分类概率 $\boldsymbol{P}_{\mathrm{s}\to\mathrm{t}}^{\mathrm{cls}}$ 和回归位置 $\boldsymbol{P}_{\mathrm{s}\to\mathrm{t}}^{\mathrm{reg}}$；3）学生特征 $\boldsymbol{F}_{\mathrm{s}}^{l}$ 到学生检测头 $Head_{\mathrm{s}}$ 的输出分类概率 $\boldsymbol{P}_{\mathrm{s}}^{\mathrm{cls}}$ 和回归位置 $\boldsymbol{P}_{\mathrm{s}}^{\mathrm{reg}}$。最终的输出如式（3）所示。

$$\begin{aligned}\boldsymbol{P}_{\mathrm{t}}^{\mathrm{cls}},\boldsymbol{P}_{\mathrm{t}}^{\mathrm{reg}} &= Concat\left(Head_{\mathrm{t}}\left(\boldsymbol{F}_{\mathrm{t}}^{l}\right)\right),\\ \boldsymbol{P}_{\mathrm{s}\to\mathrm{t}}^{\mathrm{cls}},\boldsymbol{P}_{\mathrm{s}\to\mathrm{t}}^{\mathrm{reg}} &= Concat\left(Head_{\mathrm{t}}\left(\boldsymbol{F}_{\mathrm{s}}^{l}\right)\right),\\ \boldsymbol{P}_{\mathrm{s}}^{\mathrm{cls}},\boldsymbol{P}_{\mathrm{s}}^{\mathrm{reg}} &= Concat\left(Head_{\mathrm{s}}\left(\boldsymbol{F}_{\mathrm{s}}^{l}\right)\right),\end{aligned} \tag{3}$$

式中 $Concat$ 表示拼接操作。$l\in\{1,2,\cdots,n\}$ 表示多尺度层级。

最后，将生成的学生—教师跨标签分配预测 $\boldsymbol{P}_{\mathrm{s}\to\mathrm{t}}$ 与教师模型的预测计算蒸馏损失。L_{CAKD} 的损失计算公式如式（4）所示。

$$L_{\mathrm{CAKD}}=L_{\mathrm{CAKD}}^{\mathrm{cls}}(\boldsymbol{P}_{\mathrm{t}}^{\mathrm{cls}},\boldsymbol{P}_{\mathrm{s}\to\mathrm{t}}^{\mathrm{cls}})+L_{\mathrm{CAKD}}^{\mathrm{reg}}(\boldsymbol{P}_{\mathrm{t}}^{\mathrm{reg}},\boldsymbol{P}_{\mathrm{s}\to\mathrm{t}}^{\mathrm{reg}}), \tag{4}$$

式中 $\boldsymbol{P}_{\mathrm{t}}^{\mathrm{cls}}$ 和 $\boldsymbol{P}_{\mathrm{s}\to\mathrm{t}}^{\mathrm{cls}}$ 分别表示教师和学生—教师跨标签分配预测的分类向量。$\boldsymbol{P}_{\mathrm{t}}^{\mathrm{reg}}$ 和 $\boldsymbol{P}_{\mathrm{s}\to\mathrm{t}}^{\mathrm{reg}}$ 分别表示教师和学生—教师跨标签分配预测的回归向量。L_{CAKD} 表示总的蒸馏损失，根据目标检测任务分为分类损失和回归蒸馏损失。$L_{\mathrm{CAKD}}^{\mathrm{cls}}$ 是用于分类的蒸馏损失，通常使用 KL 散度进行计算。$L_{\mathrm{CAKD}}^{\mathrm{reg}}$ 是用于回归的蒸馏损失，通常使用 L1 损失和定位蒸馏（localization distillation, LD）损失。

2.3 任务感知匹配度量

优秀的目标检测器应该能够预测同时具有高分类分数和精确定位的边界框。现有的传统检测器，如 RetinaNet[9]，FCOS[10]，SSD[41] 都按照数据分布中固有的中心先验（即定位质量）知识分配样本。这种策略导致分类任务和定位任务之间没有相关性，其中分类得分最高的位置通常不是回归实例边界的最佳位置。文献 [17, 19, 38, 42] 指出，最佳匹配应该考虑每个样本的预测质量，包括定位和分类质量。据此将分类和回归得分纳入选择度量中，匹配质量公式如式（5）所示。

$$\boldsymbol{Q}_{\mathrm{s,s}\to\mathrm{t}}=\mathbb{I}[\boldsymbol{P}\in\Pi_{\boldsymbol{G}}]\times L_{\mathrm{cls}}(\boldsymbol{P}_{\mathrm{s,s}\to\mathrm{t}}^{\mathrm{cls}},\boldsymbol{G}^{\mathrm{cls}})^{\alpha}\times L_{\mathrm{reg}}(\boldsymbol{P}_{\mathrm{s,s}\to\mathrm{t}}^{\mathrm{reg}},\boldsymbol{G}^{\mathrm{reg}})^{\beta}, \tag{5}$$

其中下标 $\mathrm{s,s}\to\mathrm{t}$ 用于表示是属于学生分支还是跨标签分配分支；$\boldsymbol{Q}_{\mathrm{s,s}\to\mathrm{t}}\in[0,1]$ 表示真实实例 $\boldsymbol{G}$ 和模型预测 $\boldsymbol{P}$ 之间的匹配质量；$\mathbb{I}[\boldsymbol{P}\in\Pi_{\boldsymbol{G}}]$ 表示空间先验；$\Pi_{\boldsymbol{G}}$ 表示真实实例的边界框的内部空间。当预测 $\boldsymbol{P}_{\mathrm{s,s}\to\mathrm{t}}^{\mathrm{cls}}$ 属于边界框 $\Pi_{\boldsymbol{G}}$ 内，则为 1，反之为 0。$\boldsymbol{G}^{\mathrm{cls}}$ 和 $\boldsymbol{G}^{\mathrm{reg}}$ 表示真实实例的类别和边界框。$\boldsymbol{P}_{\mathrm{s,s}\to\mathrm{t}}^{\mathrm{cls}}$ 和 $\boldsymbol{P}_{\mathrm{s,s}\to\mathrm{t}}^{\mathrm{reg}}$ 分别表示学生前向路径和学生—教师前向路径预测的分类和回归分数。L_{cls} 和 L_{reg} 分别表示分类损失和回归损失函数。通常使用 Focal Loss[9] 和 GIoU[43]/Smooth L1[7] 作为这

2 种损失。α和β表示平衡系数。通过使用该匹配度量,本文从联合优化的角度动态关注同时拥有高质量定位和分类分数的锚点,避免删除 NMS 造成的性能损失。

2.4 双标签分配规则

虽然,一些目标检测方法[17-19]通过引入 1 对 1 标签分配规则,将密集检测转为稀疏检测实现了无 NMS 的端到端检测。然而,这种严格的样本匹配机制导致训练过程中正样本数量显著减少,使得模型在特征表示学习阶段难以充分挖掘数据中的潜在语义信息,收敛效率低下。因此,为了弥补 1 对 1 标签分配特征表示学习的不足,我们引入 1 对多标签分配规则,以组成双标签分配规则。具体而言,在学生模型前向路径需要使用 1 对 1 标签分配规则计算损失函数L_{o2o},在学生—教师前向路径需要使用 1 对多标签分配规则计算损失函数L_{o2m}。其中学生模型前向路径的检测损失L_{o2o}表示如式(6)所示。

$$L_{\text{o2o}}=BCE\left(\boldsymbol{P}^{\text{cls}}_{\arg\max(\boldsymbol{Q}_{\text{s}})},1\right)+BCE\left(\boldsymbol{P}^{\text{cls}}_{\mathcal{B}\notin\arg\max(\boldsymbol{Q}_{\text{s}})},0\right)+GIoU\left(\boldsymbol{P}^{\text{reg}}_{\arg\max(\boldsymbol{Q}_{\text{s}})},\boldsymbol{G}^{\text{reg}}\right),\tag{6}$$

式中,$\arg\max(\boldsymbol{Q}_{\text{s}})$表示从学生模型的匹配质量$\boldsymbol{Q}_{\text{s}}$中选择匹配质量最优样本(即前景样本)的坐标。$\mathcal{B}\notin\arg\max(\boldsymbol{Q}_{\text{s}})$表示除了前景样本以外的背景样本集合。$\boldsymbol{P}^{\text{cls}}_{\arg\max(\boldsymbol{Q}_{\text{s}})}$和$\boldsymbol{P}^{\text{cls}}_{\mathcal{B}\notin\arg\max(\boldsymbol{Q}_{\text{s}})}$分别表示学生模型的前景和背景样本的分类分数。$\boldsymbol{P}^{\text{reg}}_{\arg\max(\boldsymbol{Q}_{\text{s}})}$表示学生模型的前景样本的定位分数。$\boldsymbol{G}^{\text{reg}}$表示真实实例的边界框。$GIoU$表示回归损失函数。$BCE$表示二值交叉熵损失函数。其次,学生—教师前向路径损失L_{o2m}如式(7)所示。

$$L_{\text{o2m}}=FL\left(\boldsymbol{P}^{\text{cls}}_{TopK(\boldsymbol{Q}_{\text{s}\to\text{t}},k)},1\right)+BCE\left(\boldsymbol{P}^{\text{cls}}_{\mathcal{B}\notin TopK(\boldsymbol{Q}_{\text{s}\to\text{t}},k)},0\right)+GIoU\left(\boldsymbol{P}^{\text{reg}}_{TopK(\boldsymbol{Q}_{\text{s}\to\text{t}},k)},\boldsymbol{G}^{\text{reg}}\right),\tag{7}$$

式中,$TopK(\boldsymbol{Q}_{\text{s}\to\text{t}},k)$表示从学生—教师前向路径的匹配质量$\boldsymbol{Q}_{\text{s}\to\text{t}}$中选择匹配质量排名前$k$的样本作为前景样本。$FL$表示焦点损失函数[9]。

因此,整个网络的双标签分配损失函数如式(8)所示。

$$L_{\text{det}}=L_{\text{CAKD}}+\lambda_1 L_{\text{o2o}}+\lambda_2 L_{\text{o2m}},\tag{8}$$

其中λ_1和λ_2分别表示L_{o2o}和L_{o2m}的权重。

3 实验对比与结果分析

3.1 训练细节

本文提出的 CAKD 方法是基于 Python 编程语言和 Pytorch 深度学习框架实现,并使用 MMDetection 工具箱[44]快速搭建。所有实验均在 8 张 GeForce RTX 2080Ti GPU 上训练,批次大小为 16(每张 GPU 训练 2 张图像)。在 ImageNet[45]上预训练的 ResNet-50[46]和 FPN[47]用作所有消融实验的骨干网络。使用 SGD[48]作为优化器来训练本方法,初始学习率设置为 0.01,权重衰减设置为 0.000 1,动量设置为 0.9。在 1 倍学习计划(训练 12 轮)中,输入图像的大小调整为1 333×800,学习率在训练至第 8 轮和第 11 轮时分别衰减 10 倍。在 3 倍学习计划(训练 36 轮)中,输入图像采用多个尺度,学习率衰减分别发生在训练至第 24 轮和第 33 轮。首先,在 COCO 数据集[49]上,我们将本文提出的 CAKD 方法与基于 MMDetection 工具箱的近年最先进方法进行模型的性能对比。然后,本文将 CAKD 方法与主流知识蒸馏方法进行模型性能对比。最后,我们对 CAKD 在 COCO 验证集上进行消融实验。

3.2 数据集

本文主要在 COCO 数据集[49]上进行实验,该数据集由微软研究院主导推出,最早发布于 2014 年,后续更新到 2017 版本,是计算机视觉领域中一个非常重要且广泛使用的基准数据集,其主要目标是为图像识别、物体检测、实例分割、关键点检测以及图像描述等任务提供一个具有挑战性的、多样化且真实反映现实场景的资源。它包含约 328 000 张图像,注释了超过 200 万个实例,涵盖 80 个物体类别、每张图像 5 个描述性文字等信息。其中 COCO 2017 版本包括 118 287 张训练图像、5 000 张验证图像和 40 670 张测试图像。

3.3 评价指标

本文主要采用平均精度(average precision, AP)和均值平均精度(mean average precision, mAP)进行模型的性能评估。

AP 包含多个指标,在 COCO 数据集中,AP 默认采用 0.5~0.95(步长 0.05)的 10 个阈值下的交并比(intersection over union, IoU)阈值平均值,记为 AP@[0.5:0.95],用于全面衡量模型在不同定位精度下的表现。例如,AP_{50}为 IoU 阈值为 0.5 时计算的 AP 值,常用于评估宽松条件下目标的检出能力。AP_{75}为当 IoU 阈值为 0.75 时计算的 AP 值,反映模型在精确边界框回归上的性能。此外,还有根据目标面积将检测结果 AP 分为 3 类:AP_{S}针对面积小于32^2像素的小目标计算的 AP 值,AP_{M}针对面积介于32^2至96^2像素之间的中等目标计算的 AP 值,AP_{L}针对面积大于96^2像素的大目标计算的 AP 值。

1）AP的计算是对 PR(precision-recall)曲线的一个数值化描述，通常是计算不同召回率($Recall$)下精准率($Precision$)的平均值，反映了单一类别检测的整体性能，其计算公式为

$$AP=\int_0^1 Precision(Recall)\,\mathrm{d}(Recall)\ 。\tag{9}$$

2）mAP是对所有类的AP取平均值计算均值平均精度，用于多类目标检测任务，常被用作综合性能评价指标，其计算公式为

$$mAP=\frac{1}{C\times T}\sum_{i=1}^{C}\sum_{t=1}^{T}AP_i^t\,,\tag{10}$$

式中C表示当前目标检测任务的类别数量。T表示不同 IoU 阈值个数，例如本文所使用的 COCO 数据集，从 0.5~0.95，步长为 0.05 共 10 个 IoU 阈值的数量。AP_i^t表示在第i个类别的第t个 IoU 阈值下的AP值。

3.4 与其他算法的性能比较

为了证明 CAKD 的有效性，本文首先将提出的 CAKD 方法与具有代表性的端到端和非端到端目标检测方法进行模型的性能对比，然后再与主流蒸馏方法进行性能对比。

如表 1 所示，与近几年最先进的方法相比，CAKD 方法取得了最好的性能。需要注意的是，CAKD 使用 ResNet-50 作为学生网络，ResNet-101 作为教师网络。在训练时，教师网络是加载已训练好的权重文件，并且冻住梯度不参与网络训练，只有学生网络更新网络参数。在推理时，CAKD 仅加载学生网络进

Table 1 Performance Comparison of CAKD Method and Representative Object Detection Methods on COCO Dataset

表 1 CAKD 方法与代表性目标检测方法在 COCO 数据集上的性能对比

方法	来源	推理网络	训练轮数	性能指标/%						NMS
				mAP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	
RetinaNet[9]	CVPR'17	ResNet-50	12	36.9	56.7	39.0	20.7	40.2	49.8	√
FCOS[10]	ICCV'19	ResNet-50	12	36.6	56.0	38.8	21.0	40.6	47.0	√
FCOS[10]	ICCV'19	ResNet-50	12	17.8	23.8	19.7	13.9	22.8	24.0	×
FreeAnchor[50]	NIPS'19	ResNet-50	12	38.7	**57.3**	41.5	21.5	**42.3**	50.5	√
FSAF[51]	CVPR'19	ResNet-50	12	37.2	57.2	39.4	21.0	41.2	49.7	√
OneNet[17]	ICML'21	ResNet-50	12	35.7	54.3	38.4	17.9	39.3	48.6	×
DeFCN[19]	CVPR'21	ResNet-50	12	37.0	53.8	40.6	**23.4**	40.6	46.8	×
DeFCN+Aux[19]	CVPR'21	ResNet-50	12	37.6	55.6	41.8	21.1	41.3	48.7	×
SAM-DETR[52]	CVPR'22	ResNet-50	12	33.1	54.2	33.7	13.9	36.5	**51.7**	×
DAB-DETR[53]	ICLR'22	ResNet-50	12	34.9	55.3	36.5	16.4	38.1	51.3	×
CAKD（本文）		ResNet-50†	12	**38.8**	56.2	**42.9**	22.5	42.1	50.1	×
RetinaNet[9]	CVPR'17	ResNet-50	36	39.4	58.8	42.2	23.8	43.2	50.3	√
FCOS[10]	ICCV'19	ResNet-50	36	41.4	60.0	44.9	25.5	44.8	53.0	√
FCOS[10]	ICCV'19	ResNet-50	36	19.4	25.4	21.4	14.5	25.0	26.6	×
DETR[18]	ECCV'20	ResNet-50	500	42.0	**62.4**	44.2	21.2	44.8	**61.1**	×
OneNet[17]	ICML'21	ResNet-50	36	38.9	57.5	42.2				×
DeFCN[19]	CVPR'21	ResNet-50	36	40.6	57.9	44.7	26.1	44.2	50.8	×
DeFCN+Aux[19]	CVPR'21	ResNet-50	36	41.4	59.5	35.6	26.0	44.7	52.0	×
Deformable-DETR[54]	ICLR'21	ResNet-50	50	39.4	59.6	42.3				×
SAM-DETR[52]	CVPR'22	ResNet-50	50	39.8	61.8	41.6				×
PSS[55]	TMM'23	ResNet-50	36	42.0						×
DHLA[38]	TCSVT'24	ResNet-50	36	42.2	59.5	**46.5**	26.3	44.9	53.6	×
YOLOv8-MS-N[56]	TPAMI'25	CSPNet	300	40.2	56.5	43.3	20.9	44.1	55.5	×
YOLOv9-MS-T[56]	TPAMI'25	CSPNet	300	38.5	53.7	41.9	19.3	42.8	52.9	×
YOLOv10-N[57]	NIPS'24	CSPNet	500	38.5	53.8	41.7	19.0	42.3	54.7	×
CAKD（本文）		ResNet-50†	36	**42.5**	61.3	45.9	**27.3**	**46.3**	54.6	×

注："†"表示蒸馏架构，模型训练时使用 ResNet-101 作为教师网络，ResNet-50 作为学生网络，推理时仅使用 ResNet-50 学生网络上进行测试；"√"表示使用，"×"表示不使用；黑体数值表示在同类算法中表现最好。

行推理。所以,CAKD 方法与传统使用相同主干的检测模型在训练和推理的耗时上没有差别。从表 1 中可以看出,CAKD 方法在 COCO 数据集上的检测性能相比近年多种主流检测框架均有较为明显的提升。具体而言,在使用 ResNet-50 作为学生主干网络、ResNet-101 作为教师主干网络训练 12 轮时,CAKD 方法使用 ResNet-50 进行推理、不使用 NMS 时在 *mAP* 指标上可获得 38.8% 的端到端检测性能,远远优于使用当前主流的非端到端检测器,例如 RetinaNet[9] 和 FCOS[10]。同时,CAKD 方法也优于其他基于相同推理主干的端到端检测器,包括 OneNet[17](*mAP* 值上表现为 35.7%)、DeFCN[19](*mAP* 值上表现为 37.0%)、DeFCN+Aux[19](*mAP* 值上表现为 37.6%)、SAM-DETR[52](*mAP* 值上表现为 33.1%)和 DAB-DETR[53](*mAP* 值上表现为 34.9%)。在检测小目标(AP_S)与大目标(AP_L)指标上也分别达到了 22.5% 和 50.1% 的精度,展现出对不同尺度目标的良好适应性。这说明通过跨标签分配之间的特征蒸馏,CAKD 方法在多尺度目标上都能够获得更为丰富和精细的表征能力。进一步地,当训练轮次增加到 36 轮后,CAKD 方法在 *mAP* 指标上提升至 42.5%,不仅超过了 DeFCN[19](*mAP* 值上表现为 40.6%)、Deformable-DETR[54](*mAP* 值上表现为 39.4%)、PSS[55](*mAP* 值上表现为 42.0%)和 DHLA[38](*mAP* 值上表现为 42.2%)等近期较为先进的方法,还超过了需要更长训练时间的 DETR[18](训练 500 轮)和 YOLOv8-MS-N[56](训练 300 轮)、YOLOv9-MS-T[56](训练 300 轮)和 YOLOv10-N[57](训练 500 轮)方法。这说明通过跨标签分配蒸馏策略,CAKD 方法有效地提升了检测器的特征表达能力。

然后,我们将 CAKD 与一些主流蒸馏方法进行了模型性能对比。如表 2 所示,CAKD 方法在 COCO 数据集上的检测性能相较于这些主流蒸馏方法,在 *mAP* 主要指标上取得了最好的性能。具体而言,使用 FCOS 作为基线模型,ResNet-101 作为教师网络,ResNet-50 作为学生网络训练 12 轮时,CAKD 方法(无 NMS)在 *mAP* 指标上取得了 38.8% 的最好性能,超过了 FGD[28](*mAP* 值上表现为 38.0%)、MGD[58](*mAP* 值上表现为 38.2%)、FreeKD[59](*mAP* 值上表现为 38.3%)和 DMKD[60](*mAP* 值上表现为 38.5%)等主流蒸馏方法(有 NMS)。至关重要的是,CAKD 方法还消除了对 NMS 的需求,实现了端到端的检测,在没有牺牲性能的同时还大幅度提高了无 NMS 时的性能。例如,对于 $FCOS_{ResNet\text{-}50}$ 学生模型不使用 NMS 时 *mAP* 仅有 17.8%,而使用 CAKD 方法后,无 NMS 时的性能大幅度提高了 21.0 个百分点。这一现象,在训练 36 轮时也有所体现,无 NMS 时的性能提高了 23.1 个百分点,这进一步说明了 CAKD 方法的有效性。

3.5 消融实验

表 3 测试了不同度量结合方式和不同平衡超参数对训练结果的影响。如式(5)所示,超参数α和

Table 2 Performance Comparison of CAKD Method and Mainstream Distillation Methods on COCO Dataset

表 2 CAKD 方法与主流蒸馏方法在 COCO 数据集上的的性能对比

教师网络	学生网络	训练轮数	性能指标/%						NMS
			mAP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	
$FCOS_{ResNet\text{-}101}$[10]	$FCOS_{ResNet\text{-}50}$	12	36.6	56.0	38.8	21.0	40.6	47.0	√
	+FGD[28]	12	38.0 (+1.4)			21.4	31.5	49.5	√
	+MGD[58]	12	38.2 (+1.6)			21.5	31.8	49.7	√
	+FreeKD[59]	12	38.3 (+1.7)			21.7	41.9	49.7	√
	+DMKD[60]	12	38.5 (+1.9)			22.0	42.1	49.8	√
	+SAMKD[61]	12	38.8 (+2.2)			21.9	**42.3**	**50.1**	√
	$FCOS_{ResNet\text{-}50}$	12	17.8	23.8	19.7	13.9	22.8	24.0	×
	+CAKD (本文)	12	**38.8 (+21.0)**	**56.2**	**42.9**	**22.5**	42.1	**50.1**	×
	$FCOS_{ResNet\text{-}50}$	36	41.4	60.0	44.9	25.5	44.8	53.0	√
	+FGD[28]	36	42.3 (+0.9)	60.8	45.8	26.1	**46.7**	53.3	√
	+FRS[62]	36	42.2 (+0.8)	60.6	45.6	27.1	46.5	53.0	√
	$FCOS_{ResNet\text{-}50}$	36	19.4	25.4	21.4	14.5	25.0	26.6	×
	+CAKD (本文)	36	**42.5 (+23.1)**	**61.3**	**45.9**	**27.3**	46.3	**54.6**	×

注:"√"表示使用,"×"表示不使用;黑体数值表示在同类算法中表现最好。网络名称"+"后的表示增加的网络。括号内的值表示相比 $FCOS_{ResNet\text{-}50}$ 网络 *mAP* 的提升值。

β用于控制分类质量和回归质量之间的比例。当设置$\beta=0$时，意味着仅考虑分类质量，这会降低无NMS时的端到端检测性能。然而，只考虑回归质量（即当$\alpha=0$时）可能会导致重复预测并增加去除NMS的成本。同时，此实验中还评估了不同度量的结合方式。其中，“Mul”表示式(5)形式，“Add”表示$L_{\text{cls}}(\boldsymbol{P}^{\text{cls}}_{\text{s,s}\to\text{t}},\boldsymbol{G}^{\text{cls}})^{\alpha}+L_{\text{reg}}(\boldsymbol{P}^{\text{reg}}_{\text{s,s}\to\text{t}},\boldsymbol{G}^{\text{reg}})^{\beta}$。实验对比发现“Mul”度量方式更适合端到端检测，与最佳的“Add”度量方式相比，在*mAP*指标上提高了0.9个百分点。因此，后续实验将α和β分别设置为0.2和0.8。

Table 3 Impact of Hyper-Parameters α and β on Performance in Ablation Study

表3 消融实验中超参数α、β对性能的影响

匹配方式	α	β	性能指标/%			NMS
			mAP	AP_{50}	AP_{75}	
Add	0.8	0.2	36.5	52.9	41.9	×
	0.6	0.4	37.2	53.8	42.2	×
	0.4	0.6	37.3	53.8	42.2	×
	0.2	0.8	37.9	54.7	42.6	×
Mul	1.0	0.0	34.7	52.3	40.2	×
	0.8	0.2	36.4	54.2	42.2	×
	0.6	0.4	37.8	54.6	41.8	×
	0.4	0.6	38.3	55.9	42.8	×
	0.2	0.8	**38.8**	**56.2**	**42.9**	×
	0.0	1.0	9.3	13.6	10.3	×

注：“√”表示使用NMS，“×”表示不使用NMS；黑体数值表示最优值。

表4进一步探讨CAKD方法在式(4)上执行分类$L^{\text{cls}}_{\text{CAKD}}$和回归$L^{\text{reg}}_{\text{CAKD}}$计算对模型性能的影响。如$L^{\text{cls}}_{\text{CAKD}}$所示，我们可以看出在2个分支上计算蒸馏损失取得了最高的*mAP*(38.8%)，比单独使用分类蒸馏损失高1.3个百分点，也比单独使用回归蒸馏损失高1.6个百分点。

Table 4 Distillation Loss in Ablation Study

表4 消融实验中的蒸馏损失

$L^{\text{cls}}_{\text{CAKD}}$	$L^{\text{reg}}_{\text{CAKD}}$	性能指标/%			NMS
		mAP	AP_{50}	AP_{75}	
√		37.5	55.4	40.1	×
	√	37.2	54.1	39.8	×
√	√	**38.8**	**56.6**	**42.6**	×

注：“√”表示使用，“×”表示不使用；黑体数值表示最优值。

表5测试了不同权重下的双标签分配损失对模型性能的影响。如式(8)所示，我们调节双标签分配损失函数中的权重值λ_1和λ_2。为了保证端到端检测性能，我们首先将权重λ_1固定为1.0，并逐步增加权重λ_2。随着1对多损失权重λ_2从0.3逐步增加到1.0，模型的*mAP*，AP_{50}，AP_{75}指标逐渐上升，并在1对多损失权重λ_2设置为1.0时*mAP*指标取得最佳值(38.8%)。类似地，我们也将权重λ_2固定为1.0，并逐步减少权重λ_1，此时模型的*mAP*，AP_{50}，AP_{75}指标均呈现明显下降趋势。因此，在本方法中，我们设置$\lambda_1=1.0,\lambda_2=1.0$以获得最佳的检测性能。

Table 5 Dual-Label Assignment Loss Weights in Ablation Study

表5 消融实验中的双标签分配损失权重

λ_1	λ_2	性能指标/%			NMS
		mAP	AP_{50}	AP_{75}	
1.0	0.3	37.4	52.9	41.2	×
1.0	0.6	37.8	53.4	41.6	×
1.0	1.0	**38.8**	**56.6**	**42.6**	×
0.6	1.0	37.0	53.3	40.8	×
0.3	1.0	36.6	52.8	40.4	×

注：“√”表示使用NMS，“×”表示不使用NMS；黑体数值表示最优值。

表6将CAKD方法用在了3种主流检测器上，即RetinaNet[9]，FCOS[10]，ATSS[63]，来进一步证明CAKD方法在端到端目标检测上的有效性。实验首先对比了以上3种检测器在不使用NMS时造成的性能衰减。经过对比发现传统的检测器在不使用NMS时，性能衰减极其严重，最严重的性能衰减超过一半。例如，使用NMS的RetinaNet-50在*mAP*指标上达到了36.9%，但在不使用NMS时，*mAP*性能仅有14.4%，衰减了

Table 6 Comparison of Ablation Study of CAKD Method on Different Detection Models

表6 CAKD方法在不同检测模型上的消融实验对比

教师网络	学生网络	性能指标/%						NMS
		mAP	AP_{50}	AP_{75}	AP_{S}	AP_{M}	AP_{L}	
$\text{RetinaNet}_{\text{ResNet-101}}$[9]	$\text{RetinaNet}_{\text{ResNet-50}}$	36.9	**56.7**	39.0	20.7	40.2	**49.8**	√
		14.4	21.7	18.4	12.5	20.7	19.8	×
	+CAKD（本文）	**38.1**	55.1	**41.9**	**25.5**	**42.1**	46.9	×
$\text{FCOS}_{\text{ResNet-101}}$[10]	$\text{FCOS}_{\text{ResNet-50}}$	36.6	56.0	38.8	21.0	40.6	47.0	√
		17.8	23.8	19.7	13.9	22.8	24.0	×
	+CAKD（本文）	**38.8**	**56.6**	**42.6**	**22.2**	**42.1**	**49.6**	×
$\text{ATSS}_{\text{ResNet-101}}$[63]	$\text{ATSS}_{\text{ResNet-50}}$	**39.4**	**57.6**	42.8	**23.6**	**42.9**	**50.3**	√
		19.8	24.2	23.5	16.7	25.2	26.2	×
	+CAKD（本文）	39.3	57.2	**42.9**	22.5	42.2	50.1	×

注：“√”表示使用，“×”表示不使用；黑体数值表示最优值。网络名称“+”后表示增加的网络。

22.5个百分点。在以上3种非端到端检测器中，ATSS检测器由于使用自适应的标签分配策略，取得了最高的*mAP*性能(19.8%)，但缺失NMS造成的性能衰减依然十分明显，这表明了传统检测器对于NMS的严重依赖。然后，将CAKD方法用在了这3种检测器上。需要注意的是，CAKD使用了ResNet-101骨干网络训练的3种检测器作为教师模型。实验结果表明，使用CAKD的3种检测器都达到了具有竞争力的端到端性能表现。例如，FCOS在不使用NMS时的*mAP*性能仅有17.8%，加上CAKD后，不使用NMS的性能提高至38.8%，性能增益达到21.0个百分点。在其他2个检测器上，性能增益也达到23.7个百分点和19.5个百分点。这证明了CAKD方法在端到端检测上的有效性。

表7对比了常规蒸馏方法[26](即直接将学生的输出与教师模型的输出进行拟合)和CAKD方法在使用不同基准模型的消融实验结果。在本实验中，我们选择FCOS[10]和ATSS[63]检测器作为对比的基准模型。其中，ResNet-50作为骨干训练的学生网络，ResNet-101作为骨干训练的教师网络。当使用常规知识蒸馏方式进行训练，在FCOS基准模型中发生了网络无法收敛的现象。这说明由于常规蒸馏方法使用1对1训练的学生网络的输出和使用1对多训练的教师网络的输出直接蒸馏，模型训练遇到了严重的目标冲突和特征冲突。常规蒸馏方法直接简单地强制对齐学生和教师模型的输出，使学生模型在端到端和非端到端优化目标上产生矛盾，导致网络无法收敛。而CAKD方法避免了直接强制学生模仿教师的特征，而是通过间接特征模仿的方式学习到教师网络中丰富的表征信息。使用FCOS作为基准网络，CAKD方法在*mAP*指标上实现38.8%的无NMS检测性能。在使用ATSS作为基准网络，CAKD方法在*mAP*指标上实现39.3%的无NMS检测性能。

Table 7 Experimental Comparison of CAKD and Conventional Knowledge Distillation Methods on Different Detection Models

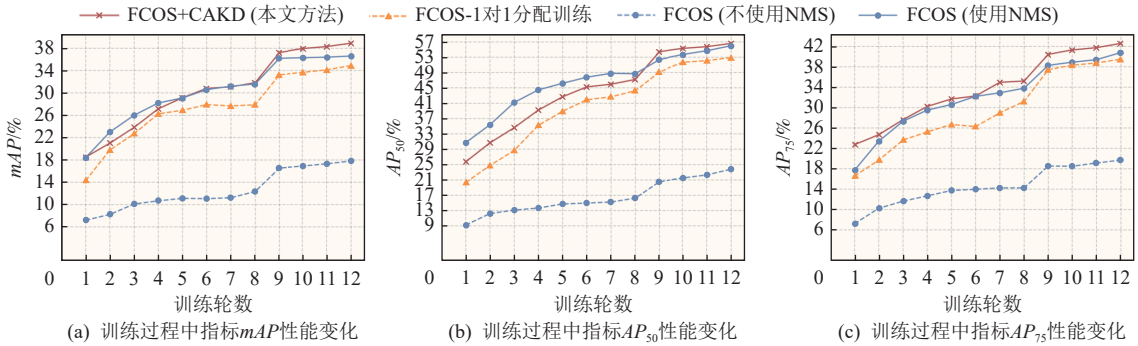
表7 CAKD和常规知识蒸馏方法在不同检测模型上的实验对比

方法	性能指标/%						NMS
	mAP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	
$FCOS_{ResNet-50}$	17.8	23.8	19.7	13.9	22.8	24.0	×
$FCOS_{ResNet-101}$	40.8	59.8	44.2	24.2	44.3	52.4	√
常规蒸馏[26]	-	-	-	-	-	-	×
CAKD（本文）	38.8	56.6	42.6	22.5	42.1	50.1	×
$ATSS_{ResNet-50}$	19.8	24.2	23.5	16.7	25.2	26.2	×
$ATSS_{ResNet-101}$	41.5	59.9	45.2	24.2	45.9	53.3	√
常规蒸馏[26]	9.7	13.2	7.2	2.3	8.2	9.1	×
CAKD（本文）	39.3	56.3	43.0	22.6	42.2	50.1	×

注："√"表示使用，"×"表示不使用；"-"表示模型训练无法收敛。

3.6 可视化分析

图4展示了在COCO数据集上基准模型和CAKD方法在训练过程中的性能。图中叉形实线表示使用了CAKD方法的FCOS基准模型性能变化；三角虚线表示了仅使用1对1标签分配训练的FCOS模型性能变化；圆点虚线表示基准模型使用默认1对多标签分配方法训练但在推理阶段不使用NMS的性能变化；圆点实线表示基准模型使用默认1对多标签分配方法训练并且使用NMS的性能变化图。根据观察，当传统密集检测器不使用NMS进行预测框的去重后，性能发生了巨大的下降。当使用1对1标签分配训练时，不使用NMS的性能出现了提升，但性能相较于默认1对多标签分配方法训练的检测器有下降。这可能是由于1对1分配所能提供的表征学习有限，而将CAKD方法用在基准检测器后，展现出了优异的无NMS检测性能，超过了原先基准检测器的使用NMS检测性能。这证明了CAKD能将使用



(a) 训练过程中指标*mAP*性能变化 (b) 训练过程中指标AP_{50}性能变化 (c) 训练过程中指标AP_{75}性能变化

Fig. 4 Performance changes of different methods on *mAP*, AP_{50}, and AP_{75} metrics during the training process

图4 训练过程中不同方法在*mAP*，AP_{75}，AP_{75}指标上的性能变化

教师模型在训练过程中学习到的表征信息知识迁移至学生模型，弥补了学生模型由于使用 1 对 1 分配引起的表征学习不足的问题。

为了展现 CAKD 方法在实际检测效果上抑制重复预测的有效性，图 5 展示了在 COCO 数据集上对基准模型 FCOS(不使用 NMS)、FCOS+本文 CAKD(不使用 NMS)方法进行目标检测的效果可视化。如图 5(b)所示，传统的非端到端不使用 NMS 的 FCOS 基准检测器对单个目标输出了多个预测框。相比之下，如图 5(c)所示 CAKD 方法在不使用 NMS 的情况下重复预测的现象被显著抑制。虽然 CAKD 提高了端到端检测的性能，也显著抑制了重复预测的发生，但我们也发现 CAKD 在前景与背景目标存在严重遮挡时无法完全消除重复预测，如图 5(c)中对 Person 类的检测。这是由于当前景与背景目标存在严重遮挡时，学生模型输出的预测框质量大幅降低，学生—教师间的跨标签分配出现匹配偏差，基于该类无效匹配对计算的蒸馏损失会产生负优化。针对严重遮挡场景性能退化问题，在未来工作中，我们将在蒸馏损失中设置匹配置信度阈值，过滤低质量的学生—教师预测匹配对，规避无效匹配对模型优化的干扰。



(a) 真实实例 (b) FCOS (不使用NMS) (c) FCOS+CAKD (不使用NMS)

Fig. 5 Visualization of the detection effect of CAKD method and the baseline models on COCO dataset

图 5 CAKD 方法和基线模型在 COCO 数据集上的检测效果可视化

4 结 论

本文提出 CAKD 的目标检测方法。该方法将学生模型的多尺度特征传递到教师模型的检测头中。然后，将生成的学生—教师跨标签分配预测与教师模型的预测用来计算蒸馏损失。这种方式避免了直接强制学生模仿教师特征而产生的训练目标矛盾和

特征结构不同导致的特征冲突问题。该方法为知识蒸馏框架用在训练端到端检测模型建立了一条有效的知识迁移路径。此外,我们设计了一种有效的任务感知匹配度量方式,综合考虑了分类质量和回归质量,避免了分类和定位任务之间没有相关性的问题。在COCO数据集上,我们进行了详细的对比实验、消融实验以及可视化实验。大量的实验表明,本文所提出的CAKD方法在端到端目标检测任务上获得了具有竞争力的表现。

作者贡献声明:胡志良、陈思负责主体设计工作并撰写论文;王大寒、严严负责设计思路指导和提出改进建议。

参 考 文 献

[1] Li Chengye, Zhang Zhen, Liang Zheheng, et al. Survey on object detection models[J]. Journal of Computer Research and Development, 2025, 62(12): 2895−2928 (in Chinese)
(李承烨, 张震, 梁哲恒, 等. 目标检测模型综述[J]. 计算机研究与发展, 2025, 62(12): 2895−2928)

[2] Zhang Xiaoliang, Li Hongliang, Meng Fanman, et al. Segmenting beyond the bounding box for instance segmentation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32(2): 704−714

[3] Luo R, Lai C. Multisensor fusion-based concurrent environment mapping and moving object detection for intelligent service robotics[J]. IEEE Transactions on Industrial Electronics, 2013, 61(8): 4043−4051

[4] Xue Wanli, Zhang Zhibin, Pei Shenglei, et al. Mixing tokens from target and search regions for visual object tracking[J]. Journal of Computer Research and Development, 2024, 61(2): 460−469 (in Chinese)
(薛万利, 张智彬, 裴生雷, 等. 混合目标与搜索区域令牌的视觉目标跟踪[J]. 计算机研究与发展, 2024, 61(2): 460−469)

[5] Zheng Junhao, Lin Chenhao, Zhao Zhengyu, et al. Towards transferable and stealthy attacks against object detection in autonomous driving systems[J]. Journal of Computer Research and Development, 2025, 62(8): 2111−2124 (in Chinese)
(郑君豪, 蔺琛皓, 赵正宇, 等. 针对自动驾驶系统目标检测器的迁移隐蔽攻击方法[J]. 计算机研究与发展, 2025, 62(8): 2111−2124)

[6] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2014: 580−587

[7] Girshick R. Fast R-CNN[C]//Proc of the IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2015: 1440−1448

[8] Ren Shaoqing, He Kaiming, 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, 2016, 39(6): 1137−1149

[9] Lin T, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proc of the IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2017: 2980−2988

[10] Tian Zhi, Shen Chunhua, Chen Hao, et al. FCOS: Fully convolutional one-stage object detection[C]//Proc of the IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 9627−9636

[11] Zhang Xiuzai, Qiu Ye, Shen Tao. Small pedestrian target detection in subway scene based on improved SSD algorithm[J]. Journal of Computer Research and Development, 2025, 62(2): 397−407 (in Chinese)
(张秀再, 邱野, 沈涛. 基于改进SSD算法的地铁场景小行人目标检测[J]. 计算机研究与发展, 2025, 62(2): 397−407)

[12] Zhu Benjin, Wang Jianfeng, Jiang Zhengkai, et al. Autoassign: Differentiable label assignment for dense object detection[J]. arXiv preprint, arXiv: 2007.03496, 2020

[13] Li Shuai, He Chenhang, Li Ruihuang, et al. A dual weighting label assignment scheme for object detection[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 9387−9396

[14] Bodla N, Singh B, Chellappa R, et al. Soft-NMS-Improving object detection with one line of code[C]//Proc of the IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2017: 5561−5569

[15] He Yihui, Zhang Xiangyu, Savvides M, et al. Softer-NMS: Rethinking bounding box regression for accurate object detection[J]. arXiv preprint, arXiv: 1809.08545, 2018

[16] Liu Songtao, Huang Di, Wang Yunhong. Adaptive NMS: Refining pedestrian detection in a crowd[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 6459−6468

[17] Sun Peize, Jiang Yi, Xie Enze, et al. What makes for end-to-end object detection?[C]//Proc of the Int Conf on Machine Learning. New York: PMLR, 2021: 9934−9944

[18] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//Proc of the European Conf on Computer Vision. Berlin: Springer, 2020: 213−229

[19] Wang Jianfeng, Song Lin, Li Zeming, et al. End-to-end object detection with fully convolutional network[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 15849−15858

[20] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network[J]. arXiv preprint, arXiv: 1503.02531, 2015

[21] Chen Guobin, Choi Wongun, Yu Xiang, et al. Learning efficient object detection models with knowledge distillation[J/OL]. Advances in Neural Information Processing Systems, 2017[2026-3-21]. https://proceedings.neurips.cc/paper/2017/hash/e1e32e235eee1f970470a3a6658dfdd5-Abstract.html

[22] Zheng Zhaohui, Ye Rongguang, Wang Ping, et al. Localization distillation for dense object detection[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 9407−9416

[23] Romero A, Ballas N, Kahou S E, et al. FitNets: Hints for thin deep

nets[J]. arXiv preprint, arXiv: 1412.6550, 2014

[24] Zagoruyko S, Komodakis N. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer[J]. arXiv preprint, arXiv: 1612.03928, 2016

[25] Wang Tao, Yuan Li, Zhang Xiaopeng, et al. Distilling object detectors with fine-grained feature imitation[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 4933−4942

[26] Liu Yufan, Cao Jiajiong, Li Bing, et al. Knowledge distillation via instance relationship graph[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 7096−7104

[27] Guo Jianyuan, Han Kai, Wang Yunhe, et al. Distilling object detectors via decoupled features[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 2154−2164

[28] Yang Zhendong, Li Zhe, Jiang Xiaohu, et al. Focal and global knowledge distillation for detectors[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 4643−4652

[29] Stewart R, Andriluka M, Ng A Y. End-to-end people detection in crowded scenes[C]//Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 2325−2333

[30] Ren Mengye, Zemel R S. End-to-end instance segmentation with recurrent attention[C]//Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 6656−6664

[31] Hu Han, Gu Jiayuan, Zhang Zheng, et al. Relation networks for object detection[C]//Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 3588−3597

[32] Chen Qiang, Chen Xiaokang, Wang Jian, et al. Group DETR: Fast detr training with group-wise one-to-many assignment[C]//Proc of the IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2023: 6633−6642

[33] Zong Zhuofan, Song Guanglu, Liu Yu. DETRs with collaborative hybrid assignments training[C]//Proc of the IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2023: 6748−6758

[34] Jia Ding, Yuan Yuhui, He Haodi, et al. DETRs with hybrid matching[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 19702−19712

[35] Zhao Chuyang, Sun Yifan, Wang Wenhao, et al. MS-DETR: Efficient detr training with mixed supervision[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2024: 17027−17036

[36] Zhao Yian, Lv Wenyu, Xu Shangliang, et al. DETRs beat yolos on real-time object detection[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2024: 16965−16974

[37] Wang Shuo, Xia Chunlong, Lv Feng, et al. RT-DETRv3: Real-time end-to-end object detection with hierarchical dense positive supervision[C]//Proc of the Winter Conf on Applications of Computer Vision. Hawaii: WACV, 2025: 1628−1636

[38] Hu Zhiliang, Chen Si, Hua Yang, et al. DHLA: Dynamic hybrid label assignment for end-to-end object detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 35(2): 1055−1069

[39] Li Zhihui, Xu Pengfei, Chang Xiaojun, et al. When object detection meets knowledge distillation: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(8): 10555−10579

[40] Li Xiang, Wang Wenhai, Wu Lijun, et al. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection[J]. Advances in Neural Information Processing Systems, 2020, 33: 21002−21012

[41] Liu Wei, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]//Proc of the European Conf on Computer Vision. Berlin: Springer, 2016: 21−37

[42] Li Shuai, Li Minghan, Li Ruihuang, et al. One-to-few label assignment for end-to-end dense detection[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 7350−7359

[43] Rezatofighi H, Tsoi N, Gwak J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 658−666

[44] Chen Kai, Wang Jiaqi, Pang Jiangmiao, et al. MMDetection: Open mmlab detection toolbox and benchmark[J]. arXiv preprint, arXiv: 1906.07155, 2019

[45] Deng Jia, Dong Wei, Socher R, et al. ImageNet: A large-scale hierarchical image database[C]//Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2009: 248−255

[46] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition[C]//Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 770−778

[47] Lin T, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 2117−2125

[48] Bottou L. Large-scale machine learning with stochastic gradient descent[C]//Proc of the 19th Int Conf on Computational Statistics-Paris France (COMPSTAT'2010). Heidelberg: Physica-Verlag HD, 2010: 177−186

[49] Lin T, Maire M, Belongie S, et al. Microsoft COCO: Common objects in context[C]//Proc of the European Conf on Computer Vision. Berlin: Springer, 2014: 740−755

[50] Zhang Xiaosong, Wan Fang, Liu Chang, et al. FreeAnchor: Learning to match anchors for visual object detection[J/OL]. Advances in Neural Information Processing Systems, 2019[2026-03-21]. https://proceedings.neurips.cc/paper/2019/hash/43ec517d68b6edd3015b3edc9a11367b-Abstract.html

[51] Zhu Chenchen, He Yihui, Savvides M. Feature selective anchor-free module for single-shot object detection[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 840−849

[52] Zhang Gongjie, Luo Zhipeng, Yu Yingchen, et al. Accelerating DETR convergence via semantic-aligned matching[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 949−958

[53] Liu Shilong, Li Feng, Zhang Hao, et al. DAB-DETR: Dynamic anchor boxes are better queries for detr[J]. arXiv preprint, arXiv: 2201.12329, 2022

[54] Zhu Xizhou, Su Weijie, Lu Lewei, et al. Deformable DETR: Deformable Transformers for end-to-end object detection[J]. arXiv preprint, arXiv: 2010.04159, 2020

[55] Zhou Qiang, Yu Chaohui. Object detection made simpler by eliminating heuristic NMS[J]. IEEE Transactions on Multimedia, 2023, 25: 9254−9262

[56] Chen Yuming, Yuan Xinbin, Wang Jiabao, et al. YOLO-MS: Rethinking multi-scale representation learning for real-time object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(6): 4240−4252

[57] Wang Ao, Chen Hui, Liu Lihao, et al. YOLOv10: Real-time end-to-end object detection[J]. Advances in Neural Information Processing Systems, 2024, 37: 107984−108011

[58] Yang Zhendong, Li Zhe, Shao Mingqi, et al. Masked generative distillation[C]//Proc of the European Conf on Computer Vision. Berlin: Springer, 2022: 53−69

[59] Zhang Yuan, Huang Tao, Liu Jiaming, et al. FreeKD: Knowledge distillation via semantic frequency prompt[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2024: 15931−15940

[60] Yang Guang, Tang Yin, Wu Zhijiang, et al. DMKD: Improving feature-based knowledge distillation for object detection via dual masking augmentation[C]//Proc of the 2024 IEEE Int Conf on Acoustics, Speech and Signal Processing (ICASSP). Piscataway, NJ: IEEE, 2024: 3330−3334

[61] Zhang Zhourui, Li Jun, Li Jiayan, et al. SAMKD: Spatial-aware Adaptive masking knowledge distillation for object detection[J]. arXiv preprint, arXiv: 2501.07101, 2025

[62] Du Zhixing, Zhang Rui, Chang Ming, et al. Distilling object detectors with feature richness[J]. Advances in Neural Information Processing Systems, 2021, 34: 5213−5224

[63] Zhang Shifeng, Chi Cheng, Yao Yongqiang, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//Proc of the IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 9759−9768



Hu Zhiliang, born in 1999. Master candidate. His main research interests include computer vision and object detection.

胡志良, 1999 年生。硕士研究生。主要研究方向为计算机视觉、目标检测。



Chen Si, born in 1987. PhD, professor, master supervisor. Her main research interests include computer vision, machine learning, and pattern recognition.

陈　思, 1987 年生。博士，教授，硕士生导师。主要研究方向为计算机视觉、机器学习、模式识别。



Wang Dahan, born in 1983. PhD, professor, master supervisor. His main research interests include pattern recognition, text detection and recognition, and semantic segmentation.

王大寒, 1983 年生。博士，教授，硕士生导师。主要研究方向为模式识别、文本检测与识别、语义分割。



Yan Yan, born in 1984. PhD, professor, PhD supervisor. His main research interests include computer vision and pattern recognition.

严　严, 1984 年生。博士，教授，博士生导师。主要研究方向为计算机视觉、模式识别。