

基于结构化知识蒸馏的轻量级伪装目标检测模型

宋建华, 刘淳, 张龔

引用本文

宋建华, 刘淳, 张龔. [基于结构化知识蒸馏的轻量级伪装目标检测模型](#)[J]. 计算机科学, 2026, 53(4): 299-307.

SONG Jianhua, LIU Chun, ZHANG Yan. [Lightweight Camouflaged Object Detection Model Based on Structured Knowledge Distillation](#) [J]. Computer Science, 2026, 53(4): 299-307.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于循环一致性约束的LLM增强型语言模型训练框架](#)

LLM-augmented Training Framework with Cycle-Consistency Constraints
计算机科学, 2026, 53(4): 377-383. <https://doi.org/10.11896/jsjcx.250600032>

[基于多任务强化学习的优先级加权软模块化方法:SM-PHT](#)

SM-PHT:Robust,Scalable,and Efficient Method for Multi-task Reinforcement Learning
计算机科学, 2026, 53(4): 366-376. <https://doi.org/10.11896/jsjcx.250700198>

[基于对比学习的双通道源代码漏洞检测模型](#)

Dual-channel Source Code Vulnerability Detection Model Based on Contrastive Learning
计算机科学, 2026, 53(3): 424-432. <https://doi.org/10.11896/jsjcx.250200124>

[融合ByteTrack的EAP-YOLOv8无人机Marker点检测与追踪](#)

Integrate ByteTrack's EAP-YOLOv8 UAV Marker Point Detection and Tracking
计算机科学, 2026, 53(3): 266-276. <https://doi.org/10.11896/jsjcx.241100115>

[计算机视觉在轨道交通中的应用](#)

Computer Vision Applications in Rail Transit Systems
计算机科学, 2026, 53(3): 214-224. <https://doi.org/10.11896/jsjcx.250400009>

基于结构化知识蒸馏的轻量级伪装目标检测模型

宋建华^{1,3,4} 刘淳² 张龔^{2,3}

1 湖北大学网络空间安全学院 武汉 430062

2 湖北大学计算机与信息工程学院 武汉 430062

3 智能感知系统与安全教育部重点实验室 武汉 430062

4 智能网联汽车网络安全湖北省工程研究中心 武汉 430062

(sjhhubu@126.com)

摘要 伪装目标检测在自然场景分析与安全监控中发挥着重要作用,但伪装目标的复杂性和多样性对检测模型的性能提出了严峻挑战。现有知识蒸馏方法多用于模型压缩,通过对教师网络与学生网络的输出层特征对齐,实现轻量化。然而,现有知识蒸馏方法通常忽略了教师网络中间特征的丰富语义信息。此外,固定学习率策略难以适应教师和学生模型规模差距过大的情况,导致蒸馏过程不稳定。为此,设计了一种基于结构化知识蒸馏的轻量级伪装目标检测模型,利用结构化知识改进传统的软硬标签损失计算,从而显著提升蒸馏效果。同时,将学习率优化问题建模为一个最优化任务,以稳定蒸馏过程中的性能波动。实验结果表明,该方法在 COD10K-V3 和 CAMO 伪装目标检测数据集上, S_m 分别达到 82.9% 和 81.0%,且训练时间减少至 6.5h。

关键词: 知识蒸馏; 伪装目标检测; 目标检测; 结构化知识; 学习率

中图分类号 TP391

Lightweight Camouflaged Object Detection Model Based on Structured Knowledge Distillation

SONG Jianhua^{1,3,4}, LIU Chun² and ZHANG Yan^{2,3}

1 School of Cyber Science and Technology, Hubei University, Wuhan 430062, China

2 School of Computer Science and Information Engineering, Hubei University, Wuhan 430062, China

3 Key Laboratory of Intelligent Sensing System and Security, Ministry of Education, Wuhan 430062, China

4 Hubei Engineering Research Center of Cyber Security for Intelligent Connected Vehicles, Wuhan 430062, China

Abstract Camouflaged object detection(COD) plays a crucial role in natural scene analysis and security monitoring. However, the complexity and diversity of camouflaged objects pose significant challenges to the performance of detection models. Existing knowledge distillation methods are primarily used for model compression by aligning the output features of teacher and student networks to achieve lightweight models. Nonetheless, these methods often overlook the rich semantic information contained in the intermediate features of teacher networks. Additionally, fixed learning rate strategies struggle to adapt to the significant scale differences between teacher and student models, leading to instability during the distillation process. To address these issues, this paper proposes a lightweight camouflaged object detection model based on structured knowledge distillation. The method leverages structured knowledge to improve the traditional soft and hard label loss calculation, significantly enhancing the distillation performance. Furthermore, the learning rate optimization problem is modeled as an optimization task to stabilize performance fluctuations during the distillation process. Experimental results demonstrate that the proposed method achieves an S_m of 82.9% and 81.0% on the COD10K-V3 and CAMO camouflaged object detection datasets, respectively, while reducing training time to 6.5 hours.

Keywords Knowledge distillation, Camouflaged object detection, Object detection, Structured knowledge, Learning rate

到稿日期:2025-01-16 返修日期:2025-03-28

基金项目: 国家自然科学基金(62377009); 湖北省重大攻关项目(JD)(2023BAA018); 湖北省重点研发计划重点项目(2021BAA184, 2021BAA188); 湖北省高等学校人文社会科学重点研究基地绩效评价信息管理研究中心课题(2020JX01); 湖北省科技计划重大科技专项(2024BAA008)

This work was supported by the National Natural Science Foundation of China(62377009), Major Project of Hubei Province(JD)(2023BAA018), Key Project of Hubei Provincial Key R & D Program(2021BAA184, 2021BAA188), Research Center for Performance Evaluation and Information Management of Key Research Bases for Humanities and Social Sciences in Hubei Provincial Colleges and Universities(2020JX01) and Major Science and Technology Special Project of Hubei Science and Technology Plan(2024BAA008).

通信作者: 张龔(zhangyan@hubu.edu.cn)

1 引言

伪装目标检测(Camouflaged Object Detection, COD)是一项识别隐藏于复杂背景中的目标(也称伪装目标)的计算机视觉任务,广泛存在于军事侦察、生物多样性保护和生态监测等领域^[1]。伪装目标通过改变颜色、纹理或形状,与背景融为一体,从而达到伪装自身的目的。例如,变色龙可根据环境变色,章鱼和竹节虫通过形态伪装与背景融为一体。同样,在军事任务中,人类通过迷彩服或遮蔽物等伪装手段规避侦测,以有效提升生存能力和隐蔽性。

随着科技的发展,伪装目标检测模型的潜在应用场景日益丰富,被广泛应用在无人机等资源受限设备中。然而,这些设备因算力和存储容量的限制,对模型的轻量化提出了更高的要求。目前,大多数 COD 方法依赖于复杂的网络结构实现精细的特征提取和精准预测,但复杂网络带来的高计算需求和存储成本限制了其在实际场景中的部署能力。现有的轻量级伪装目标检测模型(如 SINet 系列)采用高效的编解码器结构来提升模型的推理能力,但是这些模型的特征提取能力不足,背景建模能力有限,并且对小目标不敏感,难以满足任务对细粒度特征和复杂背景信息的要求。因此,设计轻量级、高效能的 COD 模型,成为该领域亟待解决的核心问题。

为此,研究者探索了多种轻量化技术,如模型剪枝、参数量化、神经架构搜索和知识蒸馏^[2]等。模型剪枝虽然能减少计算量,但会削弱模型捕捉细微特征的能力;参数量化和神经架构搜索计算资源消耗巨大。相比之下,知识蒸馏(Knowledge Distillation)将教师网络知识传递给学生网络,不仅能有效捕捉细微特征,还可以显著降低计算开销,使蒸馏后的学生网络更适合应用在资源受限场景。因此,知识蒸馏成为实现轻量级伪装目标检测模型的优选方案。

传统的知识蒸馏方法主要匹配教师模型和学生模型的输出概率分布,或使用注意力分布等显式特征,存在一些局限性:首先,传统的知识蒸馏只关注网络的输出层信息,而忽略了中间层中丰富的层次化特征;其次,伪装目标与背景之间存在显著的多尺度特征差异和复杂的上下文关联,使得传统方法难以有效捕捉关键细节;此外,当教师网络和学生网络之间存在显著的规模差距时,蒸馏效果会受到明显影响。

针对上述问题,本文通过提取结构化知识^[3]来提高学生网络的性能。结构化知识是网络中间特征层的层次化表达,包括空间位置、纹理分布和语义关联等信息。与传统的知识蒸馏方法相比,结构化知识能为模型提供更多的细粒度特征及背景信息,从而弥补学生模型在 COD 任务中推理能力不足的缺陷。为进一步提升蒸馏的稳定性,本文通过学习率监督优化策略,将学习率优化问题建模为动态最优化任务,指导学生网络更高效地学习关键知识,从而实现模型性能和蒸馏稳定性的同步提升。

综合上述策略,本文设计了一种结合结构化知识蒸馏与动态学习率优化策略的轻量级伪装目标检测模型 LCKD(Lightweight Camouflaged Object Detection Model Based on Structured Knowledge Distillation)。本文的主要贡献如下:

(1)设计了结构化知识损失计算模块。首先,提取教师网

络的中间特征层,并与学生网络的输出特征进行匹配,计算中间层损失。同时,针对教师网络和学生网络的最终输出特征,分别计算损失。最后,将中间层损失与最终输出损失按照一定权重进行融合,从而得到整体的蒸馏损失。

(2)将学习率优化问题建模为动态最优化任务。改进学习率调度方法,使模型在训练过程中自动搜寻最优学习率并应用,从而显著降低蒸馏效果的不稳定性,有效提升模型的性能。

(3)实验验证。在伪装目标检测数据集上对 LCKD 进行了实验。结果显示,LCKD 在伪装目标检测任务中取得了显著的性能提升, S_m 指标达到 82.9%,训练时间减少至 6.5h,推理时间比已有方法缩短了 40%。

2 相关工作

2.1 伪装目标检测

基于深度学习的伪装目标检测方法可以分为 4 类,包括生物启发式、多任务式、多源信息式和 Transformer 式。

(1)生物启发式 COD:其灵感来源于自然界伪装生物的捕猎行为及其独特的视觉特征。经典方法包括 SINet^[1]及其改进版 SINetV2^[4]模型,分别引入纹理增强与联级反转注意力模块优化检测表现。D²CNet^[5]通过自优化注意力机制进一步提升了检测精度。然而,这类方法仅模仿生物视觉的浅层特征,难以达到真实生物的高效信息处理能力,因而在复杂环境中性能有限。

(2)多任务式 COD:通过任务间信息协同,深度挖掘伪装目标特征,以提高检测准确度。代表性方法包括 RankNet^[6],TANet^[7]和 DGNet^[8],它们分别通过排名机制、亲和力损失和分支设计提升检测精度。Sun 等^[9]提出的 BGNet 利用目标边界信息优化检测结果。然而,多任务模型在处理小目标区域和复杂场景时仍存在局限,未能充分解决纹理细节和跨场景适应性等问题。

(3)多源信息式 COD:通过融合深度、频域与 RGB 等多种信息通道,实现伪装目标的高精度检测。DCNet^[10]通过深度估计结合多模态损失提升性能,但未充分考虑模态间的差异性。Zhong 等^[11]设计的频域感知模型,通过离散余弦变换优化特征对齐,但不同模态的互补性与差异性限制了模型性能的进一步提升。

(4)Transformer 式 COD:Transformer 具备强大的全局特征捕获能力,能够在大范围内提取上下文信息,因此在复杂背景下的伪装目标检测任务中比 CNN 表现更优。多头注意力机制能够从不同特征维度提取伪装目标的前景和背景信息,从而提高检测精度。此外,Transformer 在多模态特征融合方面表现出显著优势,能够有效整合 RGB、深度信息和频域特征,为检测任务提供更全面的支持。Yang 等^[12]提出的 UGTR 模型,通过贝叶斯学习关注不确定区域,但对全局范围的目标特征挖掘不足,可能忽略背景中的潜在目标;此外,不确定性分析增加了计算开销和模型复杂性,进一步加大了对硬件资源的需求。Liu 等^[13]提出的双任务交互式 Transformer(DTINet)中的双任务交互机制显著增加了模型的参数量,使其难以适应轻量化场景;且任务间的交互机制存在依赖

性,若一个任务性能下降,可能拖累整体检测效果;此外,在处理复杂背景时,该方法对目标与背景之间的差异性信息捕捉有限,可能会影响检测效果。Yin 等^[14]设计的 CamoFormer 中,多头注意力虽然提升了前景与背景区域的检测精度,但显著增加了计算成本,导致训练和推理时间太长;此外,多层次的分割策略对小目标区域的检测不够敏感,在复杂场景中容易出现漏检。Transformer 式 COD 模型尽管具备全局特征捕获能力,但仍面临两大问题:一是模型复杂,训练时间长且资源消耗大;二是当场景较复杂时,难以捕获目标与背景之间的差异信息。

为解决上述问题,本文采用知识蒸馏来轻量化 Transformer 式 COD 模型,设计结构化知识损失计算模块,用以提取 Transformer 式 COD 模型中间层的结构化知识,从而弥补在复杂场景中损失的特征。

2.2 知识蒸馏

知识蒸馏由 Hinton 等提出,起初被应用于图像分类任务。其核心理念是利用一个复杂的大型网络(即教师网络)指导较小的简单网络(即学生网络)完成训练,从而将教师网络的知识迁移到学生网络中。知识蒸馏的研究可以分为 4 类,即基于网络输出的改进,探索知识的表示形式,探索知识的学习方法,以及小规模网络蒸馏训练。

(1)基于网络输出的改进:主要聚焦于改进传统知识蒸馏的损失函数。Meng 等^[15]提出,学生模型在训练中应评估教师模型输出的可靠性。当教师模型的分类型结果不准确时,学生应直接依赖真实标签学习,而非基于错误的教师输出进行学习。Li 等^[16]通过加权机制提升蒸馏效果,利用教师和学生网络输出差异对训练样本进行自适应更换。然而,神经网络的输出信息量有限,未充分挖掘网络内部的知识。

(2)探索知识的表示形式:除了网络输出中包含的隐藏信息,神经网络中还有更多信息可供挖掘。Romero 等^[17]提出了提示损失(Hint Loss),通过对教师网络的特征图进行维度对齐来匹配学生网络和教师网络的对应层。Gao 等^[18]提出渐进式的方法,使学生网络逐层学习教师网络的特征图。Zhou 等^[19]让教师网络和学生网络的浅层进行共享,并在模型的输出部分使用蒸馏。这些方法对网络结构的要求过高,要尽量保证教师网络和学生网络特征图大小相同,并且直接拟合需要大量的计算。

(3)探索知识的学习方法:这是一种区别于传统教师-学生网络结构的在线蒸馏方法。Zhu 等^[20]提出了将学生网络的输出整合生成在线蒸馏信号的方法。Si 等^[21]提出了一个多助教的知识蒸馏框架。然而,在线蒸馏仍然需要一个合成的教师网络来提升性能;此外,同时训练多个网络可能会导致训练难度显著增加。

(4)小规模网络蒸馏训练:一些研究者发现,当教师网络和学生网络的模型规模差距过大时,蒸馏的效果会受到明显影响。学生网络的容量有限,难以接受教师网络传递的庞大知识,导致学生网络学习到的知识不足以有效检测伪装目标。由于使用的教师网络是 Transformer 式的 COD 模型,教师网络的知识较为复杂,难以完全传递给容量有限的学生网络,从而影响了蒸馏效果。

因此,本文针对小规模网络蒸馏训练设计了基于结构化知识蒸馏的轻量级伪装目标检测模型,利用结构化知识来改进知识蒸馏损失的计算过程,以降低 Transformer 式 COD 模型的训练成本,解决教师网络和学生网络规模差距过大时引起的蒸馏效果的波动,提高训练效率。

3 基于结构化知识蒸馏的轻量级伪装目标检测模型

3.1 模型概览

本文提出的基于知识蒸馏的模型框架如图 1 所示,采用 BiRefNet^[22] 作为教师网络,ResNet101 作为学生网络。使用轻量级的学生网络替代复杂的教师网络,可以实现高效的伪装目标检测。与现有方法相比,本文模型注重从教师网络中提取并传递结构化知识,从而增强学生网络的特征表达能力。

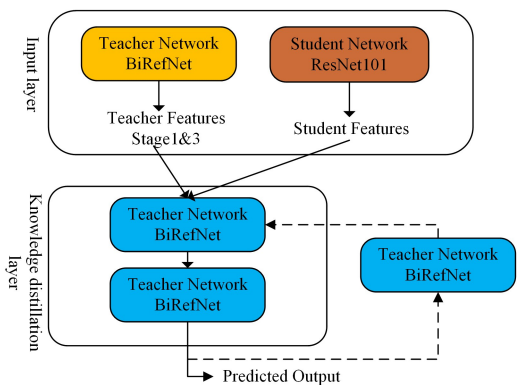


图 1 知识蒸馏模型

Fig. 1 Knowledge distillation model

结构化知识是深度学习模型各层中学习的层次化特征信息,涵盖从低级到高级特征的层次关系及其局部的互相作用。这些信息不仅包含局部特征的细粒度表示,还反映特征图之间的层次化组织。传统的知识蒸馏往往只关注输出层的知识传递,而忽略了中间层的丰富语义信息,这可能导致学生网络无法充分理解复杂的目标特征。因此,本文针对教师网络,从多个中间层提取结构化知识,包括低级特征(如边缘和纹理)、语义抽象信息以及高级特征表达。通过特征对齐,将教师网络中间层特征与学生网络的输出对齐,有效传递知识并提升检测性能。

针对蒸馏损失计算,本文设计了两种蒸馏策略。其一,采用均方误差(MSE)度量教师网络中间层与学生网络间的特征差异,确保学生网络能捕捉到关键的局部和全局信息;其二,结合 MSE 和绝对值损失函数(L1Loss),量化学生网络和教师网络最终输出的不同之处,进一步强化学生网络的全局预测能力。这两种策略的结合,在保障特征提取一致性的同时,能够提高模型的泛化能力。

此外,为解决教师与学生网络规模差距过大导致蒸馏效果不稳定的问题,本文将学习率优化建模为最优化任务。通过动态调整学习率,适应学生网络的学习节奏,模型能够充分吸收伪装目标检测中的关键知识,避免训练过程中的不稳定性。

3.2 结构化知识损失计算模块

传统知识蒸馏方法在计算损失时通常只关注教师网络和学生网络最终输出的概率分布,而忽略了教师网络中间层所

蕴含的丰富特征信息。过于依赖网络最终输出的概率分布,可能限制学生网络对目标特征的充分学习。此外,输出分布过于平滑时可能缩小感受野的覆盖范围,从而降低模型在复杂背景中准确定位伪装目标的能力。

为了解决上述问题,本文设计了一种结构化知识损失计

算模块。该模块采用 BiRefNet 作为教师网络,并按照特征抽象程度划分为 4 个阶段(Stage 层),如图 2 所示。较低层的特征主要包含细节信息,而较高层的特征则表现出更强的语义抽象性。因此,中间特征层的选择应遵循能最大限度地捕获更多尺度特征的原则。

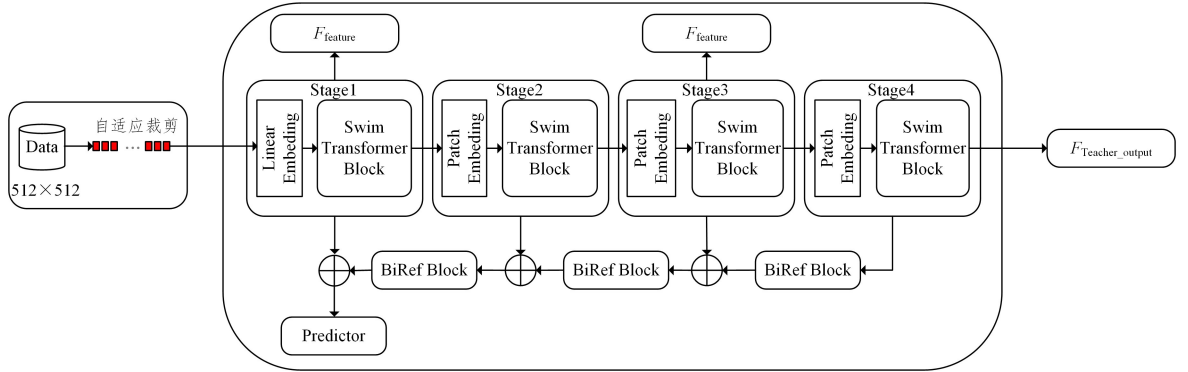


图 2 教师网络中间特征图的提取

Fig. 2 Extraction of intermediate feature maps in the teacher network

在 BiRefNet 中:

- 1) Stage1 层(早期卷积层)由初始卷积层和一个残差块构成,主要用于捕获低级特征;
- 2) Stage2 和 Stage3 层分别负责语义信息的提取与融合,其中 Stage3 特别集成了中间特征提取模块;
- 3) Stage4 生成最终的输出 $F_{teacher_output}$ 作为教师网络的全局特征表示。

基于上述分析,本文选择提取 Stage1 和 Stage3 层的特征图(Feature Map)。在结构化知识损失计算模块中,从教师网络提取的 Stage1 和 Stage3 特征图首先与 $F_{student_output}$ 在维度上进行匹配处理。设 f_l^s 为教师网络第 l 层的中间特征表示, L 表示教师网络总层数,将教师网络的中间层特征与学生网络的最终输出进行均方误差计算,得到中间层的结构化知识损失:

$$L_{feature_loss} = \sum_{l=1}^L MSE(f^s, f_l^t) \quad (1)$$

3.3 知识蒸馏损失计算

实验发现,当教师网络的输出趋于平滑分布(例如接近均匀分布或值趋于一致)时,蒸馏效果会显著下降。这种现象是由于软标签无法在平滑分布的条件下充分传递深层次的特征信息而产生的。此外,硬标签的误差梯度特征也存在局限性。硬标签的惩罚多以固定的离散值(如分类任务中的 0 或 1)表示,这种特征对于较小或者较复杂的错误难以提供有效的误差梯度,导致学习过程不够平滑。特别是当学生网络较小时,这一局限性会显著放大,进一步加剧蒸馏效果的不稳定性。

为此,本文对传统软硬标签的损失计算方法进行了优化。如图 3 所示,整体损失计算包含两部分:第一部分为结构化知识损失 $L_{feature_loss}$,前文已经计算完成;第二部分为对教师网络输出 $F_{teacher_output}$ 和学生网络输出 $F_{student_output}$ 进行缩放后的损失计算。

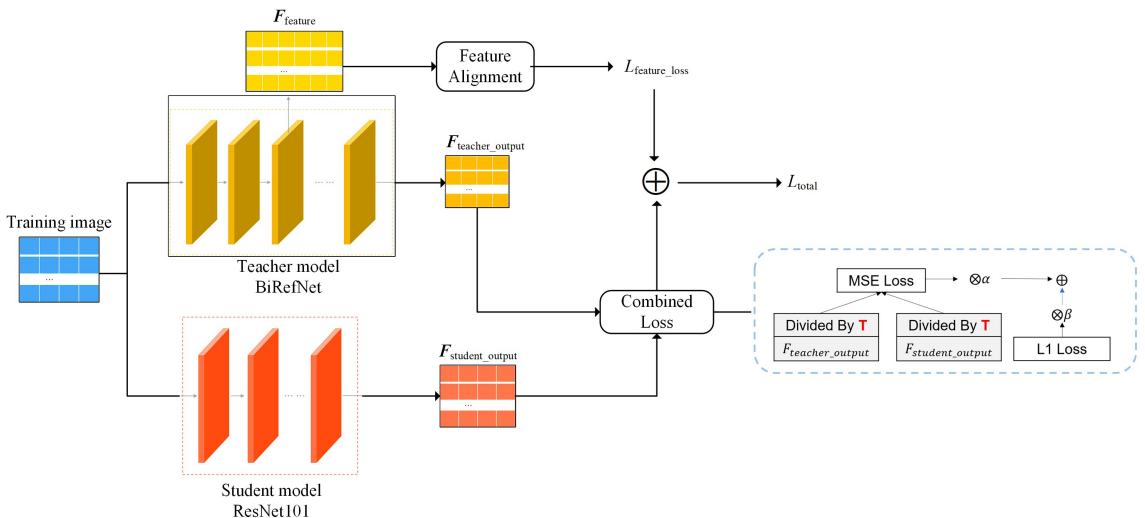


图 3 知识蒸馏损失计算

Fig. 3 Knowledge distillation loss calculation

均方误差损失(MSELoss)因对较大误差具有高敏感性,在学生输出和教师输出差异显著时能够提供较强的惩罚

信号,从而促进学生网络快速优化。L1Loss 则在处理较小误差时表现出更高的稳定性,并对异常值的影响更具鲁棒性。

为此,本文引入 L1Loss,并通过权重系数调控 MSELoss 的作用,以进一步增强系统鲁棒性。最终,对二者的缩放输出进行 MSELoss 计算损失,得到:

$$L_{\text{combined_loss}} = \alpha \cdot \text{MSE}\left(\frac{F_s}{T}, \frac{F_t}{T}\right) + \beta \cdot L1\left(\frac{F_s}{T}, \frac{F_t}{T}\right) \quad (2)$$

其中, F_s 和 F_t 分别是学生网络和教师网络的输出, T 为温度。最后,将得到的 $L_{\text{combined_loss}}$ 和 $L_{\text{feature_loss}}$ 进行加权,得到最终的知识蒸馏损失:

$$L_{\text{total_loss}} = \alpha \cdot \text{MSE}\left(\frac{F_s}{T}, \frac{F_t}{T}\right) + \beta \cdot L1\left(\frac{F_s}{T}, \frac{F_t}{T}\right) + \sum_{i=1}^k \text{MSE}(f^s, f^t) \quad (3)$$

3.4 动态最优化学习率

Zhao 等^[23] 和 Mirzadeh 等^[24] 的研究表明,当教师网络与学生网络在模型规模上存在显著差距时,蒸馏效果会被显著削弱。这是由于学生网络的容量有限,从教师网络学习的“知识”波动较大,从而减弱了蒸馏的效果。此外,学生网络的简单结构使其梯度信号微弱,进一步限制了学习的效率。

为了解决这一问题,本文参考 Loshchilov 等^[25] 提出的余弦退火学习率调度方法,并对其进行优化设计,将学习率优化问题建模为动态最优化学习率任务,通过余弦函数的周期性特性实现学习率的平滑调整。其余弦函数以及学习率退火的性质可以表示为:

$$y = A \cos(Cx(x + P_s)) + V_s \quad (4)$$

其中, A 是振幅, C 是根据训练周期而定的常数, x 是训练过程中的迭代次数, P_s 是周期参数, V_s 是学习率的最低值。

动态最优化学习率的计算如式(5)所示,学习率从较高值逐渐平滑至较低值。

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min}) \left(1 + \cos\left(\frac{T_{\text{cur}}}{T_{\text{max}}} \pi\right)\right) \quad (5)$$

其中, η_t 表示 t 时的学习率, η_{\min} 是学习率的最小值, η_{\max} 为学习率的最大值, T_{cur} 为当前的迭代次数, T_{max} 为总迭代次数。

与传统的阶梯式学习率衰减(StepLR)相比,动态最优化学习率策略能够避免固定迭代点导致的突降现象,保证学习率从高值到低值的平滑过渡,提升模型训练的稳定性。

进而,对该策略进一步优化。当学习率衰减至接近 0 时,采用反弹机制,使学习率逐步回升以跳出局部最优解。这一策略分阶段调整学习率:初期以较高学习率进行探索性训练,以快速定位最优区域;中期逐渐减小学习率,降低参数更新幅度,以减少过拟合风险;后期平滑调整学习率,避免剧烈变化带来的不稳定性,同时增强模型搜索全局最优解的能力。

当学生网络从教师网络中学习对伪装目标检测无用的“知识”时,动态调整学习率能够引导其关注更关键的信息。通过这一策略,学生网络在有限容量下更专注于最有价值的知识,从而提升蒸馏效果的稳定性和性能。

LCKD 方法结合了中间层特征提取与损失计算的改进,不仅能捕获更多特征,还通过动态学习率指导学生网络关注伪装目标检测任务中最重要的特征,有效缓解教师网络和学生网络差距过大的负面影响。基于 LCKD 的策略如算法 1 所示。

算法 1 结构化知识和动态监督学习率的知识蒸馏框架

输入:网络对应尺寸的图像

输出:训练好的学生网络 Student

1. /* 蒸馏训练方法设置过程 */
2. 初始化:设定 method=KD,EPOCHS=Epochs,k=0,K=10(预设参数)/* 设定训练方法、总训练轮数,并定义计数器 k 用于判断是否切换到非蒸馏模式 */
3. for i ← 1 to n do
4. tea_features=get_teacher_features(teacher,i);/* 教师网络提取第 i 个样本特征 */
5. stu_features=get_student_features(student,i);/* 学生网络提取第 i 个样本特征 */
6. loss=distillation_loss(tea_features,stu_features);/* 最终损失 */
7. /* 动态最优化学习率 */
8. 得到当前的最佳学习率 learning_rate = LR.adjust_learning_rate();
9. 优化学生网络,使用 loss 和 learning_rate 反向传播
10. /* 判断学生网络性能 */
11. if student.accuracy ≤ teacher.accuracy then
12. k=k+1;
13. if k ≥ K then 切换训练方法 method=non_KD;
14. else 重置计数器 k=0;
15. end for
16. 输出训练好的学生网络 Student

4 实验与结果分析

4.1 实验环境及参数设置

实验环境:实验基于 Linux 平台进行,硬件配置为 NVIDIA GeForce RTX 4080(16GB 显存),内存 64GB。开发环境包括 CUDA 12.2,Torch 2.0.1 和 Python 3.10。

实验设置:教师网络采用 BiRefNet,学生网络为预训练的 ResNet101。输入图像尺寸统一调整为 512×512 像素,训练阶段的批量大小(Batch_size)为 2,优化器为 Adam,初始学习率为 1×10^{-3} ,训练总轮数为 50 轮。

训练耗时:50 轮训练时间为 6.5 h。

4.2 数据集描述与训练设置

训练和测试:为了验证本文模型的泛化性,使用两个常用的伪装目标检测数据集 CAMO 和 COD10K-V3 进行训练,如表 1 所列。

表 1 数据集

Table 1 Datasets

Dataset	Train	Test
CAMO	1000	250
COD10K-V3	6000	4000

CAMO 数据集包含 1000 张图像作为训练集和 250 张图像作为测试集;COD10K-V3 数据集总共有 10000 张图像,其中 6000 张用于训练,4000 张用于测试。为了保证准确性和公平性,使用与文献[22]相同的训练集和测试集,在整个 CAMO 和 COD10K-V3 测试集上评估本文模型。

4.3 评估指标与定义

为了进行全面评估,采用被广泛使用的指标 S-meas-

sure^[26] (S_m)、最大/平均/加权 F-measure^[27] ($F_\beta^s/F_\beta^m/F_\beta^w$)、最大/平均 E-measure^[28] (E_ξ^s/E_ξ^m) 和平均绝对误差 (MAE) 来评估性能。

结构度量 (S-measure, S_o) 是一种衡量显著图与其对应 GT 图之间结构相似度的指标:

$$S_\alpha = \alpha \cdot S_o + (1-\alpha) \cdot S_r \quad (6)$$

其中, S_o 和 S_r 分别表示对象感知和区域感知的结构相似度, 默认设置为 0.5, 参考了文献[26]。 S_o 的评估速度很快, 无须二值化。

F-measure^[27] (F_β) 用于评估精度和召回率的加权调和平均值:

$$F_\beta = \frac{(1+\beta^2) \cdot Precision \cdot Recall}{\beta^2 \cdot Precision + Recall} \quad (7)$$

显著性图的输出以不同的阈值进行二值化, 以获得一组二值显著性预测。将预测的显著性图和 GT 图进行比较, 以获得精度和召回率值。参考文献[29], 本文将 β^2 设置为 0.3, 以强调精度相较召回率的重要性。使用整个数据集的最佳阈

值获得的最大 F-measure 得分记为 F_β^s 。

增强对齐度量 (E-measure, E_ξ)^[28] 作为一种感知度量, 用于评估预测图和 GT 图在局部和全局上的相似性:

$$E_\xi = \frac{1}{WH} \sum_{x=1}^W \sum_{y=1}^H \mathcal{O}_\xi(x, y) \quad (8)$$

其中, \mathcal{O}_ξ 表示增强对齐矩阵。类似于 F-measure, 同样采用最大 E-measure (E_ξ^s) 和平均 E-measure (E_ξ^m) 作为评估指标。

平均绝对误差 (MAE, ϵ) 是一个简单的像素级评估指标, 用于测量非二值化预测结果 M 和 GT 图像 G 之间的绝对差异:

$$\epsilon = \frac{1}{WH} \sum_{x=1}^W \sum_{y=1}^H |\hat{Y}(x, y) - G(x, y)| \quad (9)$$

4.4 实验结果分析

本文模型 LCKD 在 CAMO 和 COD10K-V3 数据集上与一些 SOTA 模型的定量比较结果如表 2 所列, 其中向上箭头表示数值越大效果越好, 向下箭头表示数值越小效果越好。

表 2 伪装目标检测与本文方法的对比实验结果

Table 2 Experiment results of camouflaged object detection and the proposed method

方法	CAMO				COD10K-V3			
	$S_m \uparrow$	$E_\xi^m \uparrow$	$F_\beta^m \uparrow$	MAE \downarrow	$S_m \uparrow$	$E_\xi^m \uparrow$	$F_\beta^m \uparrow$	MAE \downarrow
SINet ₂₀ ^[1]	0.751	0.771	0.606	0.100	0.771	0.806	0.551	0.051
SINetV2 ₂₂ ^[4]	0.820	0.882	0.743	0.070	0.815	0.887	0.680	0.037
UGTR ₂₁ ^[12]	0.787	0.823	0.686	0.086	0.818	0.853	0.667	0.035
FEDER ₂₃ ^[33]	0.802	0.867	0.738	0.071	0.822	0.900	0.716	0.032
E-UGTR ₂₄ ^[34]	0.717	0.801	0.581	0.114	0.754	0.823	0.541	0.055
ZoomNet ₂₂ ^[30]	0.820	0.878	0.752	0.066	0.838	0.888	0.729	0.029
FSPNet ₂₃ ^[31]	0.856	0.899	0.799	0.050	0.851	0.895	0.735	0.026
SegMaR ₂₂ ^[32]	0.815	0.874	0.753	0.071	0.833	0.899	0.724	0.034
LCKD(Ours)	0.810	0.823	0.801	0.176	0.829	0.829	0.720	0.155

从表 2 结果可以看出, LCKD 在多个关键指标上取得了较好的性能。

(1) 与轻量级模型 (SINet 和 SINetV2) 对比

在两个数据集上, LCKD 的 S_m 指标达到了 0.810 和 0.829, 超越了 SINet 的 0.751 和 0.771; 但在图像数量过少的 CAMO 数据集上, 未完全超越 SINetV2。在 F_β^m 指标下, LCKD 达到了 0.823 和 0.829, 相比于 SINet 的 0.771 和 0.806 有一定的提升, 但略低于 SINetV2。

(2) 与 Transformer-based 方法 (UGTR, E-UGTR 和 FEDER) 相比

在两个数据集上, LCKD 的 S_m 指标达到了 0.810 和 0.829, 优于 UGTR 的 0.787 和 0.818、E-UGTR 的 0.717 和 0.754, 以及 FEDER 的 0.802 和 0.822。 F_β^m 指标下, LCKD 达到了 0.823 和 0.829, 高于 E-UGTR 的 0.801 和 0.823, 但低于 FEDER。

(3) 与 SOTA 方法 (ZoomNet, FSPNet 和 SegMaR) 相比

LCKD 尽管在一些指标上略低于 SOTA 方法, 但是作为一种轻量级的伪装目标检测模型, 其主要目标是部署到边缘设备上, 实现高效的实时推理。相比之下, 许多 SOTA 模型虽然在性能上表现优异, 但计算开销大, 训练和推理时间过长, 难以满足要求。因此, LCKD 作为一种轻量级方法, 在实际应用中更具优势。

实验结果表明, 在 COD10K 和 CAMO 验证集上, 本文方法的性能可以达到或超过一些现有的伪装目标检测模型, 同时蒸馏训练 50 轮的时间为 6.5 h, 在 COD10K 数据集上的推理时间缩短至 90 s。

与现有的 SOTA 方法相比, LCKD 不仅在推理速度方面显著领先, 而且推理时间在 COD10K-V3 和 CAMO 数据集上均大幅优于其他模型。如表 3 所列, LCKD 的推理时间仅为 90 s (COD10K-V3) 和 40 s (CAMO), 远短于教师网络 BiRefNet 以及轻量化模型 UGTR。与此同时, LCKD 的模型参数数量和浮点计算量也比其他方法少。如表 4 所列, LCKD 的参数数量为 4.2×10^7 , 浮点计算量为 2.7×10^9 , 较 UGTR 和 BiRefNet 大幅降低, 进一步验证了 LCKD 在模型压缩和计算效率上的优势。

表 3 推理时间对比

Table 3 Comparison of inference time

Networks	Times/s
DTINet	220
UGTR	121
BiRefNet	334
LCKD	90
DTINet	120
UGTR	67
BiRefNet	68
LCKD	40

表4 模型参数数量和浮点计算量 d 对比

Table 5 Comparison of model parameters and floating point operations(FLOPs)

Networks	Params	FLOPs
BiRefNet	2.200×10^8	1.0×10^{10}
UGTR	4.887×10^7	3.1×10^9
LCKD	4.200×10^7	2.7×10^9

LCKD与UGTR模型的推理对比如图4所示,在识别并分割伪装目标边缘时,LCKD相比于UGTR对边缘细节和物体定位更加准确。

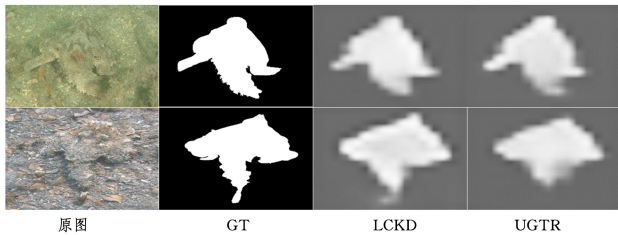


图4 LCKD与UGTR模型推理结果的对比

Fig. 4 Comparison of inference results between LCKD and UGTR models

综上所述,LCKD在伪装目标检测任务中明显优于传统工具和其他基线模型。

4.5 消融实验

本节对结构化知识损失计算模块和动态最优化学学习率进行详细的消融实验分析,以验证它们对LCKD性能的贡献。通过移除或替换LCKD中的关键模块,分析其对最终结果的影响。实验在COD10K-V3数据集上进行,因为该数据集是当前伪装目标检测任务中最具挑战性和规模最大的公开数据集,并且本文的主要实验部分同样也是基于该数据集进行性能评测,从而提高实验的可信度,以验证所提方法在伪装目标检测任务中的有效性。

为了验证结构化知识损失计算模块代替传统知识蒸馏损失计算的合理性,对照使用标准的Softmax和传统软硬标签损失计算的方法,将其命名为Baseline;将中间层结构化知识提取部分移除,保留输出层计算损失的方法(仅保留网络输出层MSELoss和L1Loss),将其命名为LCKD w/o Mid-Feature。实验结果如表5所列。

表5 结构化损失计算模块的消融实验结果

Table 5 Ablation experimental results of the structural loss calculation module

方法	$S_m/\%$	Params
Baseline	70.3	4.2×10^7
LCKD w/o Mid-Feature	77.4	4.2×10^7
LCKD w/o Final-Output	78.6	4.2×10^7
LCKD	82.9	4.2×10^7

由表5可知,与Baseline相比,LCKD在伪装目标检测任务中 S_m 值从70.3%提升到了82.9%,表明结构化知识损失计算模块的引入显著提升了学生网络的推理能力。这是因为LCKD能够通过提取中间层的丰富特征和语义信息,增强网

络的特征表达能力。同时,引入结构化知识损失计算并未增加额外的模型参数。此外,相较于LCKD w/o Mid-Feature,完整的LCKD通过对中间层特征进行结构化处理,使得损失计算更精确,提升了模型性能。相较于LCKD w/o Final-Output,LCKD由于包括了教师和学生模型最终层的输出,在特征上比只提取中间层要多,因此展现出了更好的模型性能。同时,结构化知识的引入主要体现在损失计算上,对网络结构的影响有限,在训练时间的对比上使用均方误差(MSE)和绝对误差(L1)损失的计算复杂度与使用KL散度基本一致甚至略低,所以也不会显著增加训练时间。

为了验证动态最优化学学习率策略的有效性,将以固定学习率进行训练的方法(命名为LCKD w/o LR Supervision,学习率为 1×10^{-3})、使用传统阶梯式学习率衰减的方法(命名为StepLR-LCKD),以及使用本文自动监督学习率策略的方法进行对比实验。将每个蒸馏实验进行5次,并以均值的方式分别求取LCKD w/o LR Supervision,StepLR-LCKD和LCKD在数据集上的准确度和训练时间,实验结果如表6所列。

表6 动态最优化学学习率的消融实验结果

Table 6 Ablation experimental results of dynamic learning rate

optimization		
方法	$S_m/\%$	Train Time/h
LCKD w/o LR Supervision	70.3	6.5
StepLR-LCKD	77.7	7.0
LCKD	82.9	6.5

由表6中的对比结果可知,相比LCKD w/o LR Supervision,LCKD在训练过程中可以根据学生网络的学习情况来动态调整学习率,从而更好地指导学生网络学习教师网络传递的有效知识,因此性能得到大幅度的提升。同时,学习率策略没有引入额外的预测结构,参数量不变。尽管前期花费大量时间搜寻最佳学习率,但由于后期加快了模型的收敛速度,因此总时间基本持平。相比于StepLR-LCKD,LCKD不会因为学习率衰减至0而产生较大的误差,并且通过回弹机制帮助模型跳出局部最优解,因此获得了更稳定的模型效果和更好的性能。

上述消融实验,验证了结构化知识损失计算模块和动态最优化学学习率的有效性。结构化知识损失计算模块通过提取网络中间层特征图并将其转换为结构化知识,提高了蒸馏损失计算的准确性,从而提升了模型精度。动态最优化学学习率通过动态调整学习率,提高了模型的学习效率和收敛速度。两者协同作用,使得LCKD模型在伪装目标检测任务中表现出色。

结束语 本文设计的用于伪装目标检测任务的轻量级知识蒸馏网络,在改进现有知识蒸馏框架的基础上引入了结构化知识损失计算模块,并通过动态最优化学学习率策略提升了蒸馏过程的稳定性,从而解决了学生网络难以有效学习教师网络高阶特征的问题。实验结果表明,该方法在COD多个数据集上均取得了显著的性能提升。然而,蒸馏过程对教师网络特征表达的依赖性较强,在某些情况下可能引入不必要的

噪声;并且动态最优化学学习率策略的适应性仍有提升空间,尤其在目标遮挡严重的情况下,仍可能存在不稳定的情况。未来可进一步探索更强鲁棒性的特性,以进一步提升模型的检测精度。

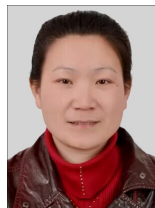
参 考 文 献

- [1] FAN D P,JI G P,SUN G L,et al. Camouflaged object detection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020;2777-2787.
- [2] HINTON G,VINYALS O,DEAN J. Distilling the knowledge in a neural network [J]. arXiv:1503.02531,2015.
- [3] CHEN H,WANG Y,XU C,et al. Learning student networks via feature embedding [J]. IEEE Transactions on Neural Networks and Learning Systems,2020,32(1):25-35.
- [4] FAN D P,JI G P,CHENG M M,et al. Concealed object detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2021,44(10):6024-6042.
- [5] WANG K,BI H,ZHANG Y,et al. D²C-Net: A Dual-Branch, Dual-Guidance and Cross-Refine Network for Camouflaged Object Detection [J]. IEEE Transactions on Industrial Electronics, 2021,69(5):5364-5374.
- [6] LYU Y,ZHANG J,DAI Y,et al. Simultaneously localize, segment and rank the camouflaged objects [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021;11591-11601.
- [7] REN J,HU X,ZHU L,et al. Deep texture-aware features for camouflaged object detection [J]. IEEE Transactions on Circuits and Systems for Video Technology,2021,33(3):1157-1167.
- [8] JI G P,FAN D P,CHOU Y C,et al. Deep gradient learning for efficient camouflaged object detection [J]. Machine Intelligence Research,2023,20(1):92-108.
- [9] SUN Y,WANG S,CHEN C,et al. Boundary-guided camouflaged object detection [J]. arXiv:2207.00794,2022.
- [10] ZHANG J,LYU Y, XIANG M,et al. Depth confidence-aware camouflaged object detection [J]. arXiv:2106.05608,2021.
- [11] ZHONG Y,LI B,TANG L,et al. Detecting camouflaged object in frequency domain [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 4504-4513.
- [12] YANG F,ZHAI Q,LI X,et al. Uncertainty-guided transformer reasoning for camouflaged object detection [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021;4146-4155.
- [13] LIU Z,ZHANG Z,TAN Y,et al. Boosting camouflaged object detection with dual-task interactive transformer [C]//2022 26th International Conference on Pattern Recognition(ICPR). IEEE, 2022;140-146.
- [14] YIN B,ZHANG X,FAN D P,et al. Camoformer: Masked separable attention for camouflaged object detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024,46(12):10362-10374.
- [15] MENG Z,LI J,ZHAO Y,et al. Conditional teacher-student learning [C]//2019 IEEE International Conference on Acoustics,Speech and Signal Processing(ICASSP 2019). IEEE,2019: 6445-6449.
- [16] LI T,LI J,LIU Z,et al. Few sample knowledge distillation for efficient network compression [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020;14639-14647.
- [17] ROMERO A,BALLAS N,KAHOUS E,et al. Fitnets: Hints for thin deep nets [J]. arXiv:1412.6550,2014.
- [18] GAO M,SHEN Y,LI Q,et al. An embarrassingly simple approach for knowledge distillation [J]. arXiv:1812.01819,2018.
- [19] ZHOU G,FAN Y,CUI R,et al. Rocket launching: A universal and efficient framework for training well-performing light net [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2018.
- [20] ZHU X,GONG S. Knowledge distillation by on-the-fly native ensemble[C]//Advances in Neural Information Processing Systems. 2018.
- [21] SI Y H,CHENG Q,HUANG J C. A dynamic multi-teacher setting method for knowledge distillation [J]. Computer Science, 2025,52(5):241-247.
- [22] ZHENG P,GAO D,FAN D P,et al. Bilateral reference for high-resolution dichotomous image segmentation [J]. arXiv:2401.03407,2024.
- [23] ZHAO H W,WU H,MA K,et al. Image classification framework based on knowledge distillation[J]. Journal of Jilin University (Engineering and Technology Edition), 2024, 54(8): 2307-2312.
- [24] MIRZADEH S I,FARAJTABAR M,LI A,et al. Improved knowledge distillation via teacher assistant [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020: 5191-5198.
- [25] LOSHCHILOV I,HUTTER F. Sgdr: Stochastic gradient descent with warm restarts [J]. arXiv:1608.03983,2016.
- [26] FAN D P,CHENG M M,LIU Y,et al. Structure-measure: A new way to evaluate foreground maps [C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 4548-4557.
- [27] ACHANTA R,HEMAMI S,ESTRADA F,et al. Frequency-tuned salient region detection [C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009: 1597-1604.
- [28] FAN D P,GONG C,CAO Y,et al. Enhanced-alignment measure for binary foreground map evaluation [J]. arXiv:1805.10421, 2018.
- [29] BORJI A,CHENG M M,JIANG H,et al. Salient object detection: A benchmark [J]. IEEE Transactions on Image Processing,2015,24(12):5706-5722.
- [30] PANG Y,ZHAO X, XIANG T Z,et al. Zoom in and out;

A mixed-scale triplet network for camouflaged object detection [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022;2160-2170.

- [31] HUANG Z, DAI H, XIANG T Z, et al. Feature shrinkage pyramid for camouflaged object detection with transformers [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023;5557-5566.
- [32] JIA Q, YAO S, LIU Y, et al. Segment, magnify and reiterate: Detecting camouflaged objects the hard way [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022;4713-4722.
- [33] HE C, LI K, ZHANG Y, et al. Camouflaged object detection with feature decomposition and edge reconstruction [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023;22046-22055.
- [34] SONG J Y, LUO X Z. A lightweight camouflaged object detection model based on attention mechanism and knowledge distil-

lation [J]. Journal of Hubei University (Natural Science Edition), 2024, 46(5):611-620.



SONG Jianhua, born in 1973, Ph.D, professor, postgraduate supervisor, is a member of CCF (No. 27785M). Her main research interests include network and information security.



ZHANG Yan, born in 1974, Ph.D, professor, doctoral supervisor. His main research interests include code security and so on.

(责任编辑:柯颖)