基于查询引导和语义增强的小样本目标检测方法

谢斌红，石宇飞，张睿，张英俊
（太原科技大学计算机科学与技术学院，山西 太原 030024）

摘　要： 针对元学习范式中原型关键信息欠缺、对查询图像适应性不足以及检测器对新类方差敏感导致误分类问题，提出一种基于查询引导策略和语义增强机制的小样本目标检测（FSOD）方法。查询引导模块（QGM）通过学习查询与支持特征之间的相关性，将查询感知信息有条件地耦合到支持特征中，旨在为每个查询图像生成特定且具有代表性的原型。而视觉语义增强模块（VSEM）从文本语义信息中蒸馏与新类视觉特征相匹配的知识，并自适应地对这些特征增强，提高其可判别性，缓解方差敏感，以更好地分类。此外，将分类和回归任务解耦，在分类分支上执行语义增强，以促进模型对目标语义的理解。实验结果表明，相较于目前已知最新的 SMPCCNet 方法，所提出方法在 PASCAL VOC 数据集上的新类平均精度（nAP）提升了 2.2 百分点，在 MS COCO 数据集上的平均精度（AP）提升了 1.0 百分点，证明了其有效性。

关键词： 目标检测；小样本学习；元学习；查询引导原型；语义增强
源代码链接： https://github.com/yufeishi123/QGSE.git
中图分类号： TP391　　**文献标志码：** A　　**DOI：** 10.19678/j.issn.1000-3428.0070093

Few-shot Object Detection Method Based on Query Guidance and Semantic Enhancement

XIE Binhong, SHI Yufei, ZHANG Rui, ZHANG Yingjun
(School of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan 030024, Shanxi, China)

【Abstract】 This study proposes a Few-Shot Object Detection (FSOD) method based on a query-guided strategy and semantic enhancement mechanism to address the following concerns: the lack of prototypical key information, insufficient adaptation to query images in the meta-learning paradigm, and the detector′s sensitivity to the variance of the novel class leading to misclassification. The Query Guidance Module (QGM) conditionally couples query-aware information into support features by learning the correlation between the query and support features, aiming to generate specific and representative prototypes for each query image. The Visual Semantic Enhancement Module (VSEM) distils the knowledge from textual semantic information that matches the novel class of visual features and adaptively enhances these features to improve their discriminability and mitigate variance sensitivity for better classification. In addition, the classification and regression tasks are decoupled, and semantic enhancement is performed on the classification branch to facilitate the model′s understanding of the target semantics. The experimental results demonstrate that, compared to the currently known state-of-the-art SMPCCNet method, the proposed approach achieves an average improvement of 2.2 percentage points in novel Average Precision (nAP) on the PASCAL VOC dataset and an average improvement of 1.0 percentage points in Average Precision (AP) on the MS COCO dataset, validating its effectiveness.

【Key words】 object detection; few-shot learning; meta learning; query-guided prototype; semantic enhancement

0　引言

目标检测是计算机视觉中的一项基本任务，近年来，基于深度学习的目标检测器取得了显著进展，并被广泛应用于现实场景。然而，这些检测器往往依赖大量的标注数据，训练成本高昂。当样本数量有限时，性能会明显下降，相比之下，人类仅需接触少量带注释的样本即可准确识别物体。因此，为了模拟人类的学习模式并减轻对大规模数据的依赖，研究人员开始关注小样本目标检测（FSOD）[1-2]任务，即在训练期间仅给出少量新类的注释样本进行学习，期望检测器在测试时能正确预测这些物体的

基金项目： 山西省基础研究计划面上项目（20210302123216）；吕梁市引进高层次科技人才重点研发项目（2022RC08）；山西省产教融合研究生联合培养示范基地项目（2022JD11）。

作者简介： 谢斌红（CCF 会员），男，教授、硕士，主研方向为智能化软件工程、机器学习；石宇飞（通信作者），硕士研究生；张睿，副教授、博士；张英俊，教授级高级工程师、硕士。

收稿日期： 2024-07-09　　**修回日期：** 2024-08-29　**E-mail：** s202220210972@stu.tyust.edu.cn

位置和类别。

目前,FSOD 方法大致可分为两类:基于微调范式和基于元学习范式。基于微调的 FSOD 方法[3-5]采用两阶段训练框架,首先在数据充足的基类上预训练检测器,然后使用少量基类和新类样本微调检测器。该类方法训练方式简单,但其对类间差异敏感,当差异明显时可能会产生负迁移。

最近,基于元学习的 FSOD 方法表现出了良好的性能,其构建由支持集和查询集组成的任务来模拟小样本场景,通过若干个任务的训练使模型学会学习,进而快速泛化至新任务。在每个任务中,由支持集生成类级原型,然后利用这些原型对查询图像中的目标进行检测。因此,原型的质量会直接影响查询图像的检测性能。然而,大多数方法生成的类级原型固定且缺乏细节信息,无法适应不同的查询图像,主要表现在以下 2 个方面:1)支持图像与查询图像的相关性挖掘不充分,一些方法[6-8]在生成原型时,未考虑查询图像在视角、形态和光照等方面的差异,导致其无法自适应不同的查询图像,另外一些方法[9-10]虽然考虑到查询图像和支持图像间的关系,但仍存在一定的局限性,例如,SMPCCNet 方法[10]仅学习图像级的相关性,这会引入冗余信息;2)支持图像显著信息淹没,其将支持特征直接通过平均池化的方式提取类原型,这种简单的操作很难捕获显著区域。

为了解决上述问题,本文提出一种独特的查询引导模块(QGM),利用查询表示引导支持特征的增强,即深入挖掘查询与支持特征的相关性,生成具有查询感知信息引导的耦合支持特征,从而为每个查询图像获得定制的原型。此外,还引入了一个背景抑制模块(BGSM),利用自适应阈值有效抑制查询图像背景噪声的干扰,以辅助提升查询感知的准确性和可靠性。另一方面,同一目标类别图像因为语义差异和透视失真而存在着类内方差,在小样本的情况下,新类的类内方差更加显著。因此,有限的视觉信息难以准确地表征目标类别,导致学习到的新类表示判别性不足,在最终分类时会产生误分类问题。借鉴人类在少样本下学习新概念时通常利用辅助性信息(如语言或音频)促进认知过程[11]的思想,提出了视觉语义增强模块(VSEM),通过知识蒸馏方法融合丰富的语义信息来提升视觉特征的表征能力。特别地,还引入了语义一致性对齐(SCA)损失以促进两种不同模态的特征,实现更有效的融合。

本文的主要贡献如下:

1)提出了查询引导模块,它利用查询目标级表示引导支持集图像特征的增强,从而为查询图像获取与之相适应的类级原型。同时引入背景抑制模块,通过动态阈值降低查询图像背景的干扰,确保模型更加专注于前景目标的检测。

2)提出了视觉语义增强模块,该模块融合语义嵌入向量丰富的先验知识,旨在增强视觉特征的表征能力。

3)对 FSOD 任务的两个基准(即 PASCAL VOC 和 MS COCO)进行了广泛的实验,结果表明所提出的模型优于现有大多数方法。

1 相关工作

1.1 小样本学习

小样本学习旨在使用少量标记样本进行学习,现有方法主要包括:

1)迁移学习[12]:利用从其他任务上学到的知识迁移泛化到新的任务中。

2)元学习:通过模拟学习任务的过程,使模型更好地适应新任务。其又分为两类,其中,基于度量学习的方法[13]通过设计度量标准来比较样本对以区分不同类别,基于优化的方法[14]旨在学习良好的初始化参数,经过少量迭代快速收敛至新任务。

3)数据增强[15]:通过数据或特征增强的方式增加训练样本的多样性来提高模型性能。但这些方法仅关注分类,无法扩展到目标检测任务。

1.2 小样本目标检测

基于元学习范式的方法通过“学会学习”提取可以泛化到不同任务的元知识,其更适应于新类检测,目前主要分为基于特征聚合、基于度量学习、基于数据增强以及基于原型生成 4 类方法。基于特征聚合的方法旨在将查询和支持特征聚合来增强查询特征表示,如 FAN 等[16]、徐守坤等[17]和 LEE 等[18]提出的方法,但其需要解决空间对齐和多尺度变化问题。基于度量学习方法[19-20]通过学习相似性度量以促进类内紧凑和类间分离,从而有效区分新类。该类方法需要设计合理的度量方式和学习良好的特征分布表示。FSOD 任务的挑战是样本量少,通过数据或特征增强[21-22]来提升其分布的多样性是最朴素直接的做法,但使用过多的数据增强策略可能会引入无关噪声信息。

在元学习范式下,原型通常用于指导模型在查询图像中检测属于支持集类别的目标,因此,其是衡量查询图像预测结果的关键因素,本文聚焦于此类方法。QUAN 等[8]通过计算支持集内实例间的相关性并重加权生成原型,但其并未考虑对不同查询

图像的适应性。ZHANG 等[9]通过学习查询和支持图像间的相似性权重，以聚合不同的支持图像生成原型，但它忽略了查询特征中的指导作用。XU 等[10]利用查询特征执行类注意力操作增强支持特征并生成原型，但该过程会引入噪声和非目标类别信息。针对上述方法存在的问题，本文提出一种为每个查询图像生成特定支持特征的方法，其通过相关性学习将查询图像中的信息有条件地耦合到支持特征中，以生成查询引导的支持特征，从而提高原型的质量。

此外，考虑到小样本设置下新类的视觉信息具有显著的类内方差，致使模型难以学习到具有区分性的类别表示，导致误分类问题，探索一种利用文本语义知识增强新类不全面的视觉表达的机制，并使用 SCA 损失对齐不同模态特征，最终提高模型分类的性能。

2 本文方法

2.1 问题定义

在小样本目标检测任务中有两个类集：基类 C_{base} 和新类 C_{novel}，$C_{\text{base}} \cap C_{\text{novel}} = \varnothing$。对于基类，有大量标注的训练图像，对于新类，只有每个类别的 K（$K=1$，5，10）个标记实例。小样本目标检测的目的是期望检测器能同时检测 C_{base} 和 C_{novel} 中的目标。与之前基于元学习范式的方法[7,16]相同，遵循以任务为单元的训练策略训练检测器，每个任务按照 N-way、K-shot（N 个类别且每个类别有 K 个样本）的方式从训练集中随机抽取图像来构建支持集，同时随机选取少量图像作为查询集，通过借助支持集图像来检测出查询图像中的目标。

2.2 网络总体框架

本文在 Meta R-CNN[7]的基础上进行改进，提出一种基于查询引导策略和语义增强机制的 FSOD 方法（QGSE），其主要由主干特征提取器、背景抑制模块（BGSM）、查询引导模块（QGM）、特征聚合模块（FAM）、区域提议网络（RPN）、视觉语义增强模块（VSEM）和检测头 7 个部分组成。其中，背景抑制模块旨在突出查询图像的前景目标区域，强化感知信息。查询引导模块通过学习查询特征与支持特征之间的相关性，为支持特征引入精细化的查询感知信息以生成高质量原型。而视觉语义增强模块将语义信息进行蒸馏并自适应地增强到视觉特征中，从而促进模型学习更具辨别性的类别表示。

如图 1 所示，QGSE 方法具体检测流程如下：首先将支持图像和查询图像输入到共享参数的主干特征提取器；然后针对查询图像中可能存在的背景噪声，BGSM 对查询特征使用动态阈值筛选来缓解背景噪声的负面影响；接着将支持特征和查询特征输入到 QGM 进行查询感知引导，并通过类别平均池化（CAP）生成特定于不同查询图像的类级原型。为了提高查询图像中目标的召回率，特征聚合模块利用类级原型进行深度互相关操作，将其作为卷积核在查询特征图上滑动以匹配和突出相似的感兴趣区域。增强后的查询特征被传递给 RPN 生成候选框并进行感兴趣区域（RoI）对齐获得 RoI 特征。

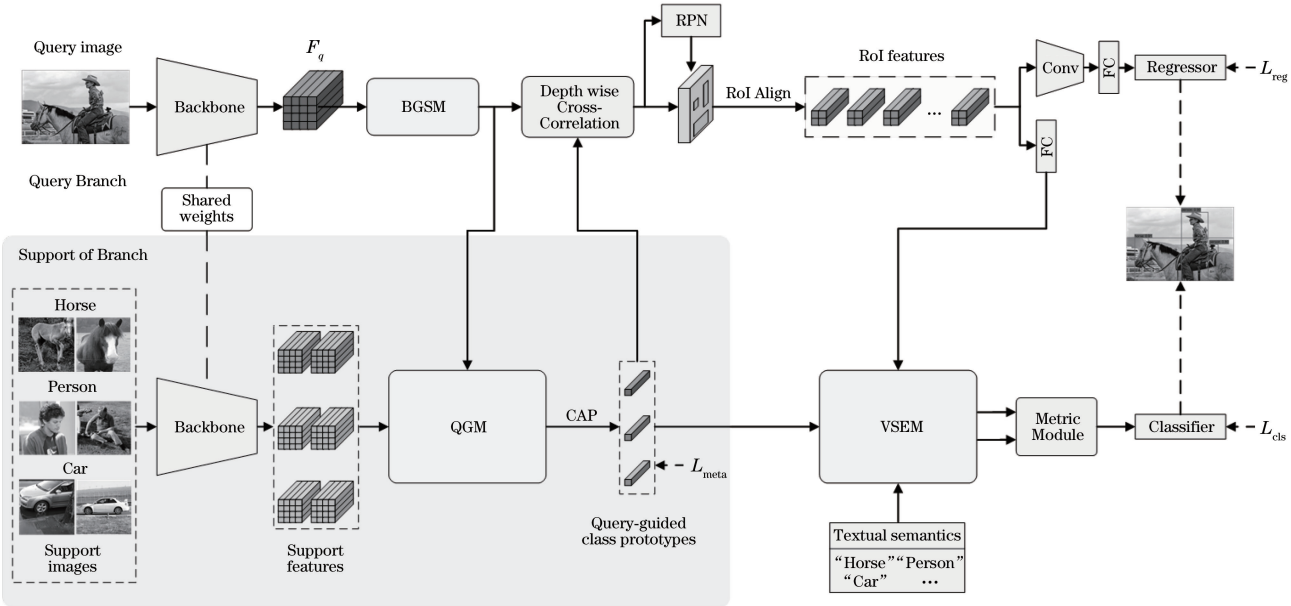


图 1 QGSE 方法总体框架

Fig. 1 The overall framework of QGSE method

最后，将检测头中的分类回归任务解耦，分别学习和优化两个特征空间，提高检测器的性能和鲁棒性。参照之前的方法[19]，通过一个残差卷积块来代替全连接层以进行更好地回归。而对于分类头，先将原型特征和 RoI 特征发送到 VSEM 中去增强其视觉特征表示，在此基础上通过度量模块进行相似度匹配分类，并且使用皮尔逊距离来获取每个 RoI 特征和原型之间的相似度。

2.3 背景抑制模块

ZHA 等[23]指出，特征图的每个通道都可以被认为是一个模式检测器，如果特征图中的某个空间位置对于大多数通道具有高激活值，则它很可能对应于信息丰富的区域。受此启发，引入了背景抑制模块(BGSM)，如图 2 所示。给定输入查询图像 x_q 对应的特征图 $F_q=\theta(x_q)\in\mathbb{R}^{H\times W\times C}$，先使用深度可分离卷积(DW-Conv)将 F_q 聚合为 $A_F\in\mathbb{R}^{H\times W\times 1}$，然后为 A_F 引入动态阈值 φ_{A_F} 来确定哪个位置是目标区域的一部分，如式(1)所示：

$$\varphi_{A_F}=\frac{\sum_{j=1}^{w}\sum_{i=1}^{h}A_F(i,j)}{h\times w} \tag{1}$$

接着将 A_F 的每个元素来与阈值 φ_{A_F} 进行比较来生成前景掩码M_{A_F}。对于特定位置(i,j)，如果 $A_F(i,j)$的激活值大于阈值 φ_{A_F}，则对应的 $M_{A_F}(i,j)$ 设置为 1，否则设置为 0.5。最后将查询特征 F_q 和掩码 M_{A_F} 对应元素相乘得到 $\hat{F}_q$，如式(2)、式(3)所示：

$$M_{A_F}=\begin{cases}1, & A_F(i,j)>\varphi_{A_F}\\ 0.5, & \text{others}\end{cases} \tag{2}$$

$$\hat{F}_q=F_q\odot M_{A_F} \tag{3}$$

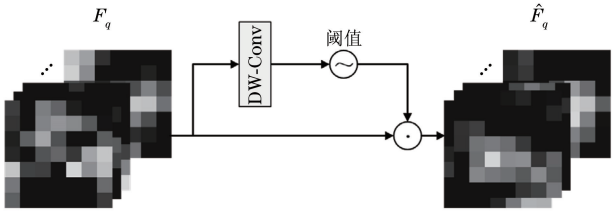


图 2 背景抑制模块

Fig. 2 Background suppression module

2.4 查询引导模块

针对现有 FSOD 方法[8-10]生成类级原型存在未能充分利用查询特征的引导性信息而导致对查询图像适应性不足，以及在利用查询引导时引入了不相关的噪声和类别信息而导致关键细节信息不突出等问题，本文提出一个查询引导模块(QGM)，为每个查询图像生成特定的查询引导支持特征，如图 3 所示。该模块包括 3 个部分：重构查询特征，学习耦合信息掩码和生成查询引导特征。

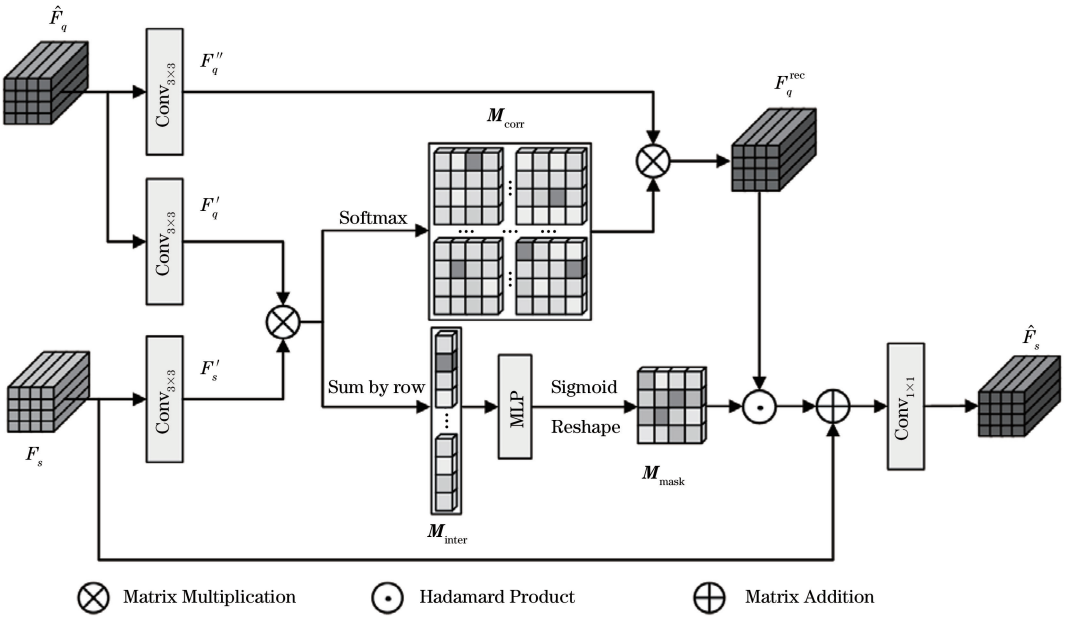


图 3 查询引导模块

Fig. 3 Query guidance module

首先，由于查询图像中存在多个目标或类别，信息较为复杂，要想实现对支持特征的精确引导，需要获得与支持特征相符的查询感知引导信息，而不是直接地进行特征匹配。因此，本文通过学习两个特征之间的相关性，重构查询特征以得到支持特征特定的感知信息。其次，为了避免引入过多冗余信息干扰支持特征原有的结构，学习一个耦合信息掩码。最后，将掩码作为条件，指导查询感知信息更好地与支持特征耦合以生成查询引导的支持特征。

具体地，为了重构查询特征来获取与支持特征相关的感知信息，首先，用 3 个并行的 3×3 卷积块对查询特征 $\hat{F}_q \in \mathbb{R}^{H_1\times W_1\times C}$ 和支持特征 $F_s \in \mathbb{R}^{H_2\times W_2\times C}$ 编码得到 F'_q、F''_q、F'_q，并将 F'_s 和 F'_q 进行矩阵相乘以及归一化得到相关性矩阵 $\boldsymbol{M}_{\text{corr}} \in \mathbb{R}^{H_2W_2\times H_1W_1}$：

$$\boldsymbol{M}_{\text{corr}}=\text{Soft}\left(\left(\frac{F'_s}{\|F'_s\|_2}\right)\left(\frac{F'_q}{\|F'_q\|_2}\right)^{\text{T}}\right) \tag{4}$$

式中：Soft(·)指的是 Softmax 归一化函数，该相关矩阵 $\boldsymbol{M}_{\text{corr}}$ 表示 F'_q 和 F'_s 的每个位置之间的相关权重。然后，将相关矩阵与 F''_q 加权求和得到重构后的查询特征 $F_q^{\text{rec}} \in \mathbb{R}^{H_2\times W_2\times C}$，其不仅突出表征了与支持特征最具相似性细节的特征，还确保了与支持特征的空间对齐。

$$F_q^{\text{rec}}=\boldsymbol{M}_{\text{corr}}\otimes F''_q \tag{5}$$

接着，若直接将重构查询特征 F_q^{rec} 和支持特征 F_s 耦合可能会影响原有支持特征的结构和表示，因此，本文利用条件耦合掩码来引导其有效地流入支持特征中。将 F'_q 和 F'_s 相乘后的矩阵逐行元素相加得到式(6)，该矩阵表示查询特征对支持特征每个区域的关注度。

$$\boldsymbol{M}_{\text{inter}}=\text{Sum by row}\left(\left(\frac{F'_s}{\|F'_s\|_2}\right)\left(\frac{F'_q}{\|F'_q\|_2}\right)^{\text{T}}\right) \tag{6}$$

之后，经过可学习的多层感知机(MLP)以及激活函数归一化生成条件耦合掩码 $\boldsymbol{M}_{\text{mask}} \in \mathbb{R}^{H_2\times W_2\times 1}$：

$$\boldsymbol{M}_{\text{mask}}=\text{Res}(\partial(\text{MLP}(M_{\text{inter}}))) \tag{7}$$

式中：Res(·)是重塑操作；∂(·)是 Sigmoid 函数；$\boldsymbol{M}_{\text{mask}}$ 用于指示查询特征与支持特征高度相关的部分，元素值越大表示该位置上的信息应当被保留或强调，反之应当抑制。

最后，使用 $\boldsymbol{M}_{\text{mask}}$ 作为条件，激活 F_q^{rec} 中的区域以进行信息耦合生成查询引导的支持特征，如式(8)所示：

$$\hat{F}_s=\text{Conv}_{1\times 1}(\boldsymbol{M}_{\text{mask}}\odot F_q^{\text{rec}}+F_s) \tag{8}$$

式中：$\text{Conv}_{1\times 1}$(·)是 1×1 卷积；⊙是哈达玛积(Hadamard Product)。在获得 $\hat{F}_s$ 后，通过类别平均池化(CAP)得到查询引导的类级原型。与原始支持特征以及之前的方法相比，$\hat{F}_s$ 通过查询引导模块融合了精细的查询感知信息，这些信息能够为每个查询图像获取独特且定制的原型。同时，通过重构查询特征和条件耦合掩码的方式可缓解噪声和不相关信息的引入，实现高效的信息耦合。

2.5 视觉语义增强模块

在小样本条件下，由于新类存在显著的类内方差，模型难以学习到完整的类别表示，容易出现错误分类的问题。而当视觉信息不足时，人类通常会将语义信息作为先验知识，增强视觉模态特征中类别的广义信息。受此启发，本文提出了视觉语义增强模块(VSEM)，通过融合语义信息来提升模型对新类的理解和表征能力，如图 4 所示。该模块由语义编码器、视觉特征映射器以及语义知识蒸馏器 3 个部分组成。其中，语义编码器利用预训练的语言模型(CLIP)编码获得目标类别的标签嵌入，视觉特征映射器旨在将视觉特征投影至与标签嵌入相同的特征空间中，而语义知识蒸馏器是将文本模态的标签信息作为先验知识蒸馏到视觉模态中，以增强视觉特征的表示。

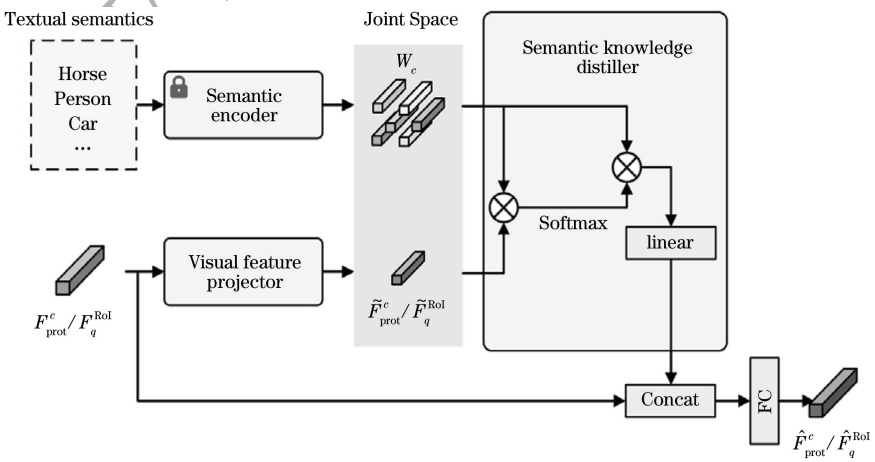


图 4 视觉语义增强模块

Fig. 4 Visual semantic enhancement module

具体来说，以增强 RoI 特征为例，为了实现语义信息和视觉特征之间的交互，应该学习标签嵌入和视觉特征在同一表示空间的联合表示。首先通过语义编码器获得不同类别的语义表示(也称文本原型)$W_c \in \mathbb{R}^{N\times d}$，其中，$N$ 是类别数量，d 是词向量的维度。然后采用视觉特征映射器将视觉特征投影至与标签

嵌入相同的表示空间 S_{joint} 中，如式(9)所示：

$$\widetilde{F}_q^{\text{RoI}}=\delta_{\text{visual}}(F_q^{\text{RoI}}) \tag{9}$$

式中：$\delta_{\text{visual}}(\cdot)$表示 MLP，用于将视觉特征投影到 d 维的空间上。

为了确保视觉特征与文本原型在语义上的兼容性，本文引入语义一致性对齐(SCA)损失来对齐这两种模态，如式(10)、式(11)所示：

$$m_c=\text{Softmax}(\widetilde{F}_q^{\text{RoI}}\cdot\boldsymbol{W}_c) \tag{10}$$

$$L_{\text{sca}}=\frac{1}{N_{\text{RoI}}}\sum_{N_{\text{RoI}}}\text{CrossEntropy}(m_c,m_g) \tag{11}$$

式中：m_c 是视觉特征和文本原型之间的匹配概率；m_g 是真值标签，表示视觉特征与其对应的文本原型之间的真实匹配；N_{RoI} 表示用于计算损失的 RoI 特征数量。

采用交叉注意力机制对语义知识进行蒸馏，具体公式如式(12)、式(13)所示：

$$\boldsymbol{Q}=\boldsymbol{W}_q\widetilde{F}_q^{\text{RoI}},\ \boldsymbol{K}=\boldsymbol{W}_k\boldsymbol{W}_c,\ \boldsymbol{V}=\boldsymbol{W}_v\boldsymbol{W}_c \tag{12}$$

$$F_{\text{fusion}}=\text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\text{T}}}{\sqrt{d_K}}\right)\boldsymbol{V} \tag{13}$$

式中：$\boldsymbol{W}_q$、$\boldsymbol{W}_k$ 和 $\boldsymbol{W}_v$ 是可学习的权重矩阵，以更好地满足不同输入的需求。

为了将“蒸馏”后的语义知识和视觉特征有效融合，先将其映射至与视觉特征同维度空间。由于语义知识和视觉特征侧重于不同层级的信息，因此采用拼接操作进行融合而非直接相加，计算公式如式(14)所示：

$$\hat{F}_q^{\text{RoI}}=\text{FC}(\text{Concat}(\delta(F_{\text{fision}}),F_q^{\text{RoI}})) \tag{14}$$

式中：$\delta(\cdot)$是线性变换层，确保特征在同维度下融合；Concat(·)是拼接操作；FC 是全连接层；$\hat{F}_q^{\text{RoI}}$ 为最终增强的视觉特征表示，通过这种方式可以自适应地增强新类的视觉特征，并有助于学习到一定的类别相关关系，提高后续分类的性能。

此外，在得到语义增强的视觉特征后，本文将其送入度量模块进行相似性匹配分类，并采用皮尔逊距离进行度量，计算公式如式(15)所示：

$$P_{\text{PR}}(\hat{F}_q^{\text{RoI}},\hat{F}_{\text{prot}}^c)=\frac{\sum_{i=1}^{d}(\hat{F}_q^{\text{RoI}}(i)-\overline{\hat{F}_q^{\text{RoI}}})(\hat{F}_{\text{prot}}^c(i)-\overline{\hat{F}_{\text{prot}}^c})}{\|\hat{F}_q^{\text{RoI}}(i)-\overline{\hat{F}_q^{\text{RoI}}}\|\cdot\|\hat{F}_{\text{prot}}^c(i)-\overline{\hat{F}_{\text{prot}}^c}\|} \tag{15}$$

式中：$\hat{F}_q^{\text{RoI}}$ 是查询图像的 RoI 特征；$\hat{F}_{\text{prot}}^c$ 是类级原型特征。

2.6 损失函数

与 Meta R-CNN[7]类似，除了由交叉熵(CE)损失实现的分类损失和 L1 损失实现的回归损失组成的通用损失之外，还有由 CE 损失实现的元损失 L_{meta} 用于对类级原型进行优化，确保模型能够生成清晰、可区分的类别中心，从而减少预测时的歧义。因此，本文模型的总体损失为：

$$L=L_{\text{rpn}}+L_{\text{cls}}+L_{\text{reg}}+L_{\text{meta}}+\lambda L_{\text{sca}} \tag{16}$$

式中：λ 是平衡检测损失、元损失和语义一致性对齐损失之间的损失权重，在实验设置中 $\lambda=0.5$。

3 实验结果与分析

数据集及评估指标：在实验中使用 PASCAL VOC[24]和 MS COCO[25]两个流行的小样本目标检测基准数据集以评估所提出方法的有效性。本文遵循最近的工作[7,16]来构建训练、测试设置，具体做法是利用 VOC2007 和 VOC2012 的训练及验证集进行训练，并选用 VOC2007 的测试集进行测试。为了公平比较，使用与之相同的 3 个基类、新类划分，从 VOC 数据集的 20 个类别中选择 15 个类别作为基类，其余 5 个类别作为新类。3 个新类划分分别是{“bird”，“bus”，“cow”，“motorbike”，“sofa”}、{“aeroplane”，“bottle”，“cow”，“horse”，“sofa”}、{“boat”，“cat”，“motorbike”，“sheep”，“sofa”}，且只为每个新类采样 $K\in\{1,2,3,5,10\}$个带注释的样本进行学习。本文使用阈值为 0.5(AP@0.50)交并比(IoU)的平均精度来评估检测器性能。

对于 COCO 数据集，在由 80 000 张训练图像和 35 000 张验证图像组成的训练集上训练，剩余的 5 000 张验证图像用于测试。VOC 的 20 个类别作为新类，其他 60 个类别作为基类。采样的样本数量 K 设置为 10 和 30，并采用标准的 COCO 度量指标来评估本文的方法。

实验环境及参数设置：本文实验环境所使用的操作系统为 Ubuntu 20.04，CPU 型号为 Intel® Xeon® Platinum 8163 CPU@2.50 GHz，在两张型号为 Nvidia RTX 3080Ti 的 GPU 上进行训练和测试。所采用的深度学习框架 PyTorch 版本为 1.12.0，CUDA 版本为 11.3，Python 版本为 3.9.0，MMDetection 版本为 2.25.0，并将在 ImageNet 上预训练的 ResNet-101 作为模型的主干网络。模型训练时使用随机梯度下降(SGD)作为优化器，动量为 0.9，权重衰减为 0.000 1，批量大小为 4。元训练阶段在 VOC 和 COCO 数据集上分别进行 36 000 次和 120 000 次迭代，初始学习率为 0.005。元微调阶段分别进行 4 000 次和 10 000 次迭代，学习率设置为 0.001。此外，使用 CLIP 作为语义编码器，遵循之前的方法[11]在训练过程中冻结其参数。

3.1 对比实验

为了验证本文方法的有效性，在 PASCAL VOC 和 MS COCO 数据集上与最近主流的 FSOD 方法进行了对比实验。

PASCAL VOC：由表 1 可知，本文的方法在 3 个新类划分的不同 shot 设置下优于其他基于微调和元学习的方法。其中，加粗数字为最优值，加下划线数字为次优值。以 3-shot 设置为例，与 3 个划分的次佳方法比较，获得了 4.2 百分点(58.4% vs 54.2%)、1.7 百分点(45.4% vs 43.7%)和 1.3 百分点(49.6% vs 48.3%)的提升。并且，随着标注实例数量的增加，模型的性能逐渐提高，尤其是在极少量样本的设置下，随着样本数量的增加，性能显著提升。例如，当仅增加一个实例从 1-shot 到 2-shot 以及从 2-shot 到 3-shot 时，所有新类划分的平均性能分别提高了 7.5 和 4.9 百分点。这是因为当样本数量较少时，添加一个样本可以为原型提供相对丰富的信息。

表 1 不同模型在 PASCAL VOC2007 数据集上的性能(nAP@0.5)对比

Table 1 Comparison of the performance of different models on the PASCAL VOC2007 dataset (nAP@0.5) %

Type	Model	Novel Set 1					Novel Set 2					Novel Set 3				
		1-shot	2-shot	3-shot	5-shot	10-shot	1-shot	2-shot	3-shot	5-shot	10-shot	1-shot	2-shot	3-shot	5-shot	10-shot
Fine-tuning	TFA w/fc[3]	36.8	29.1	43.6	55.7	57.0	18.2	29.0	33.4	35.5	39.0	27.7	33.6	42.5	48.7	50.2
	TFA w/cos[3]	39.8	36.1	44.7	55.7	56.0	23.5	26.9	34.1	35.1	39.1	30.8	34.8	42.8	49.5	49.8
	NP-RepMet[4]	37.8	40.3	41.7	47.3	49.4	**41.6**	**43.0**	43.4	47.4	49.1	33.3	38.0	39.8	41.5	44.8
	Retentive RCNN[5]	42.4	45.8	45.9	53.7	56.1	21.7	27.8	35.2	37.0	40.3	30.2	37.6	43.0	49.7	50.1
	DMNet[26]	39.0	48.9	50.7	58.6	**62.5**	31.2	32.4	40.3	47.6	52.0	**41.7**	41.8	42.7	50.3	52.1
	TeSNet[27]	44.6	46.6	48.2	58.2	61.1	26.1	31.2	41.6	42.8	43.6	40.1	41.8	48.3	54.2	55.6
Meta-learning	FSRW[6]	14.8	15.5	26.7	33.9	47.2	15.7	15.3	22.7	30.1	40.5	21.3	25.6	28.4	42.8	45.9
	Meta R-CNN[7]	19.9	25.5	35.0	45.7	51.5	10.4	19.4	29.6	34.8	45.4	14.3	18.2	27.5	41.2	48.1
	FSOD[16]	37.8	43.6	51.6	56.5	58.6	22.5	30.6	40.7	43.1	47.6	31.0	37.9	43.7	51.3	49.8
	DCNet[22]	33.9	37.4	43.7	51.1	59.6	23.2	24.8	30.6	36.7	46.6	32.3	34.9	39.7	42.6	50.7
	CME[20]	41.5	47.5	50.4	58.2	60.9	27.2	30.2	41.4	42.5	46.8	34.3	39.6	45.1	48.3	51.5
	ISAM-QSAM[18]	31.1	35.7	39.2	50.7	59.4	22.9	28.4	32.1	35.4	42.7	24.3	29.1	35.0	50.0	53.6
	CAReD[8]	36.5	45.2	47.1	50.8	58.8	26.4	31.0	37.9	43.5	51.1	20.2	33.8	41.6	48.3	55.3
	FM-FSOD[19]	**47.5**	50.6	53.4	57.2	59.9	37.9	39.8	43.6	46.7	49.8	33.4	36.3	40.5	46.8	48.9
	SMPCCNet[10]	45.5	50.3	54.2	56.1	—	27.9	35.6	43.7	46.2	—	37.7	44.5	48.2	55.6	—
QGSE(Ours)		44.8	**54.1**	**58.4**	**59.8**	61.7	31.5	38.7	**45.4**	**48.5**	**52.1**	39.8	**45.9**	**49.6**	**55.8**	**58.4**

表 2～表 4 提供了 VOC2007 数据集在 3-shot 设置下每个新类、基类的详细检测结果。观察 3 个不同的基类、新类划分可知，与基线及其他方法相比，所提出的 QGSE 方法在提升新类性能的同时，也保持了基类的性能，这表明其具有更好的泛化能力。

表 2 不同模型在 VOC2007 数据集 3-shot 设置下的新类和基类(AP@0.5)对比 1

Table 2 Comparison of novel and base classes (AP@0.5) of different models under VOC2007 dataset 3-shot setting 1 %

Type	Model	Novel Classes						Base Classes														
		bird	bus	cow	mbike	sofa	mean	aero	bike	boat	bottle	car	cat	chair	table	dog	horse	person	plant	sheep	train	tv	mean
Novel Set 1	FSRW[6]	26.1	19.1	40.7	20.4	27.1	26.7	73.6	73.1	56.7	41.6	76.1	78.7	42.6	66.8	72.0	77.7	68.5	42.0	57.1	74.7	70.7	64.8
	Meta R-CNN[7]	30.1	44.6	50.8	38.8	10.7	35.0	67.6	70.5	59.8	50.0	75.7	81.4	44.9	57.7	76.3	74.9	76.9	34.7	58.7	74.7	67.8	64.8
	NP-RepMet[4]	12.9	60.5	39.9	43.1	52.2	41.7	79.8	82.5	66.9	73.8	71.6	57.6	52.9	64.1	49.6	70.7	71.8	58.7	74.2	55.0	69.5	66.6
	FSOD[16]	35.8	61.2	57.6	60.2	44.7	51.9	68.0	73.3	58.6	54.1	79.5	81.7	48.4	62.9	79.1	83.6	76.3	36.6	65.2	75.4	62.3	67.0
	KFSOD[28]	39.8	61.9	59.6	58.3	45.7	53.1	78.1	73.4	60.3	58.2	79.4	81.0	52.2	61.1	83.3	74.9	80.8	41.3	76.6	71.7	70.5	69.5
	Baseline	41.3	63.5	56.6	60.2	48.1	54.0	76.1	70.0	63.4	54.1	75.2	78.3	49.8	64.9	77.2	77.6	79.1	42.4	69.4	71.1	69.7	67.9
	QGSE(Ours)	43.8	69.6	62.5	64.9	51.3	**58.4**	79.2	72.7	65.1	57.7	77.4	80.5	53.4	67.6	79.2	78.4	81.4	46.5	70.2	73.0	70.6	**70.2**

表 3 不同模型在 VOC2007 数据集 3-shot 设置下的新类和基类(AP@0.5)对比 2

Table 3 Comparison of novel and base classes (AP@0.5) of different models under VOC2007 dataset 3-shot setting 2 %

Type	Model	Novel Classes						Base Classes														
		aero	bottle	cow	horse	sofa	mean	bird	bike	boat	bus	car	cat	chair	table	dog	mbike	person	plant	sheep	train	tv	mean
Novel Set 2	Baseline	42.6	15.1	53.9	52.2	46.7	42.1	73.2	74.3	49.9	76.6	75.4	77.1	47.2	57.4	75.3	68.7	79.5	39.5	67.8	69.2	70.1	66.8
	QGSE(Ours)	45.8	18.6	56.7	56.3	49.6	**45.4**	75.1	76.7	51.4	78.5	77.0	79.6	49.2	59.1	77.2	71.4	81.4	41.8	68.2	70.8	71.6	**68.6**

表 4 不同模型在 VOC2007 数据集 3-shot 设置下的新类和基类(AP@0.5)对比 3

Table 4 Comparison of novel and base classes (AP@0.5) of different models under VOC2007 dataset 3-shot setting 3 %

Type	Model	Novel Classes						Base Classes														
		boat	cat	mbike	sheep	sofa	mean	aero	bird	bike	bottle	bus	cow	car	chair	table	dog	horse	person	plant	train	tv	mean
Novel Set 3	Baseline	38.2	59.1	60.4	28.6	45.1	46.3	74.3	74.6	72.2	58.6	78.7	68.6	80.9	44.1	54.3	71.8	79.6	75.8	38.5	71.5	70.9	67.6
	QGSE(Ours)	41.4	62.6	63.9	32.1	47.8	**49.6**	75.6	76.2	74.1	60.3	81.6	71.7	81.5	47.3	57.8	74.4	81.7	77.1	40.6	73.1	72.4	**69.7**

MS COCO:尽管 MS COCO 包含更具挑战性的图像,检测难度明显增加,但通过观察表 5 中的数据可知,与 SMPCCNet 方法相比,AP 值分别在 10-shot 和 30-shot 下有 0.5 和 1.6 百分点的显著提升,在更严格的 AP@0.75 指标下,也有 2.2 和 0.3 百分点的提升,表明本文的方法实现了良好的泛化性和鲁棒性。

表 5 不同模型在 MS COCO 数据集上的性能对比

Table 5 Comparison of the performance of different models on the MS COCO dataset %

Type	Model	10-shot			30-shot		
		AP	AP@0.5	AP@0.75	AP	AP@0.5	AP@0.75
Fine-tuning	TFA w/fc[3]	10.0	19.2	9.2	13.4	24.7	13.2
	TFA w/cos[3]	10.0	19.1	9.3	13.7	24.9	13.4
	Retentive RCNN[5]	10.5	—	—	13.8	—	—
	DMNet[26]	10.0	17.4	10.4	17.1	29.7	17.7
	TeSNet[27]	14.1	25.1	14.0	16.5	27.7	17.4
Meta-learning	FSRW[6]	5.6	12.3	4.6	9.1	19.0	7.6
	Meta R-CNN[7]	8.7	19.1	6.6	12.4	25.3	10.8
	FSOD[16]	11.1	20.4	10.6	—	—	—
	DCNet[22]	12.8	23.4	11.2	18.6	32.6	17.5
	CME[20]	15.1	24.6	16.4	16.9	28.0	17.8
	ISAM-QSAM[18]	13.4	**30.6**	9.1	17.1	35.2	14.7
	CAReD[8]	15.5	25.1	14.9	18.4	30.1	17.7
	FM-FSOD[19]	13.1	26.9	8.4	—	—	—
	SMPCCNet[10]	16.8	—	14.5	18.5	—	18.9
QGSE(Ours)		**17.3**	28.1	**16.7**	**20.1**	**35.6**	**19.2**

3.2 消融实验

为了验证所提出模块的有效性,在 PASCAL VOC 数据集上进行了消融实验,并复现基线以验证各个模块的优越性,实验结果如表 6 所示,其中,√表示加入该组件模块。从表 6 的第 8 行可以看出,通过叠加所有的模块达到了最好的性能,而对比第 5～第 7 行发现,当缺少任一模块时,模型性能有所下降,证明了不同模块对模型性能都有着积极贡献。同时,观察第 1～第 4 行可知,当使用单一模块时,模型性能也有不同程度的提升,例如,背景抑制模块(BGSM)的引入带来了约 0.6 百分点的平均性能提升,表明其可以弱化查询图像中的一些背景噪声信息;当只加入查询引导模块(QGM)时,性能提升最为显著,平均性能增益提升 2.3 百分点,说明其在生成针对查询图像定制的高质量原型时,有效增强了查询感知和关键细节信息;此外,视觉语义增强模块(VSEM)也带来了 1.2 百分点的平均性能提升,证明了语义信息可以增强新类视觉特征的表示。

表 6 不同模块的消融实验

Table 6 Ablation experiments with different modules %

序号	Component			Novel Set 1					Novel Set 2					Novel Set 3				
	BGSM	QGM	VSEM	1-shot	2-shot	3-shot	5-shot	10-shot	1-shot	2-shot	3-shot	5-shot	10-shot	1-shot	2-shot	3-shot	5-shot	10-shot
1				40.2	50.1	54.0	55.2	56.6	27.8	34.8	42.1	44.9	47.9	35.2	41.6	46.3	52.4	53.9
2	√			40.8	50.7	54.5	55.9	57.3	28.3	35.4	42.7	45.6	46.7	35.9	42.3	46.9	53.0	54.5
3		√		42.5	52.3	56.3	57.7	59.7	29.9	36.8	44.1	47.0	50.3	37.8	44.1	48.3	54.5	56.6
4			√	41.6	51.3	55.3	56.7	58.3	29.0	35.9	43.0	45.8	49.0	36.3	42.8	47.2	53.4	55.2
5	√	√		43.2	52.8	56.9	58.2	60.3	30.2	37.5	44.6	47.3	50.6	38.6	44.6	48.5	54.9	57.0
6		√	√	43.9	53.5	57.6	59.0	60.8	30.8	37.9	44.8	48.0	51.4	39.3	45.2	49.1	55.0	57.6
7	√		√	42.4	52.0	56.1	57.4	58.9	29.5	36.6	43.4	46.4	49.9	37.1	43.6	47.7	53.9	55.9
8	√	√	√	**44.8**	**54.1**	**58.4**	**59.8**	**61.7**	**31.5**	**38.7**	**45.4**	**48.5**	**52.1**	**39.8**	**45.9**	**49.6**	**55.8**	**58.4**

3.3 可视化分析

为了更加直观地展示查询引导模块(QGM)的效果，本文通过梯度加权类激活映射(Grad-CAM++)对支持特征图进行可视化，来验证其在引导支持特征生成原型时的作用，如图 5 所示(彩色效果见《计算机工程》官网 HTML 版，下同)。其中，第 1 列是支持图像，第 2 列描绘了基线方法的效果，第 3、第 4 列为查询引导模块的效果。可以看出，针对于不同的查询图像，基线所提取的支持特征是固定的，而本文的查询引导模块引导查询感知信息准确耦合到支持图像的相关区域中，有效减少了冗余信息，从而生成了适应于查询图像的原型。另外，观察图中第 4、第 5 行可以发现，该模块对其他类别的类级原型生成并不会造成较大影响。

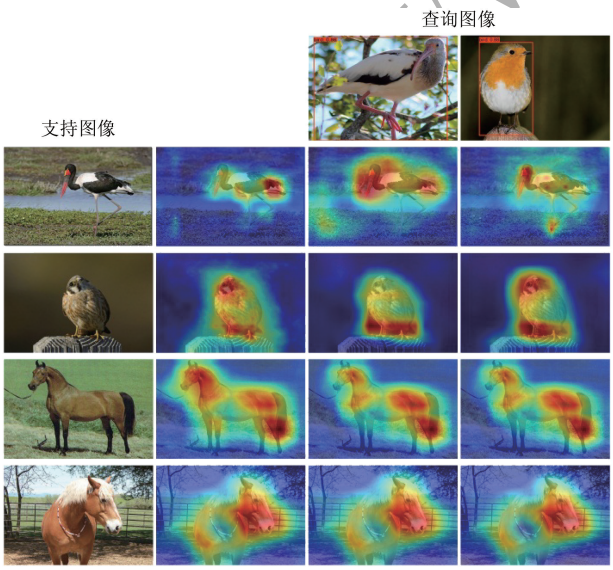


图 5 查询引导模块可视化

Fig. 5 Visualization of the query guidance module

在得到类级原型后，通常会将其与查询特征聚合，本文展示了特征聚合后的对比效果可视化，如图 6 所示。在图 6 中，第 1 列是新类查询图片，第 2 列是基线方法聚合后的特征图可视化，第 3 列是其检测结果，第 4 列是使用查询引导的类级原型聚合后的特征图可视化，第 5 列是本文的检测结果。通过对比可视化和检测结果可以看出，本文的方法更加突出了查询图像中的前景区域，同时在更复杂的 COCO 数据集图片中可以很好地发现潜在的新类目标，这也进一步证明了本文生成的原型能为查询图像的检测提供强有力的支持。

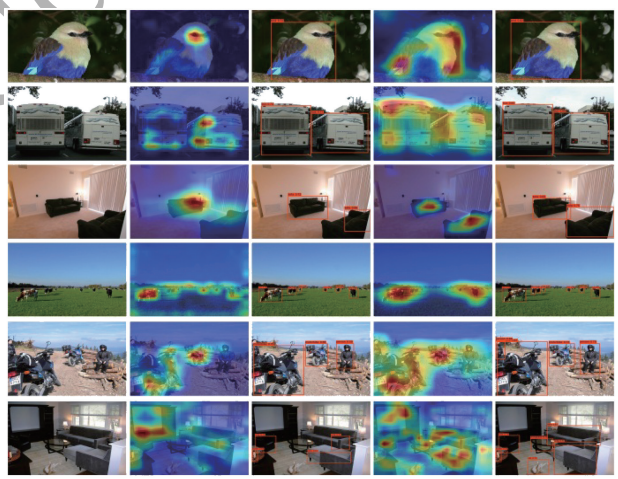
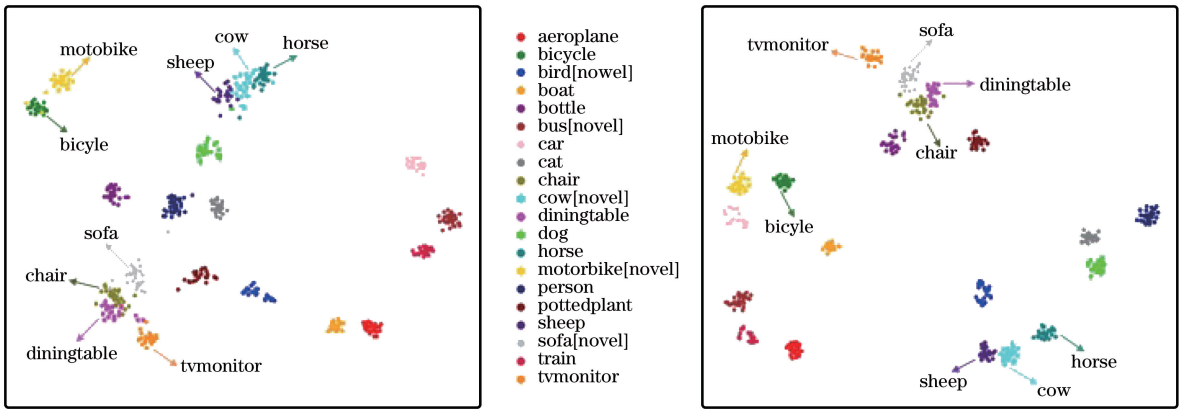


图 6 特征聚合可视化对比

Fig. 6 Visualization comparison of feature aggregation

为了直观地展示视觉语义增强模块(VSEM)的效果，从 PASCAL VOC 的测试集中随机选取了 200 张图像进行测试，并通过 t-SNE 可视化定性地说明该模块在提升模型对视觉语义理解上的作用，如图 7 所示。如果没有使用该模块，在嵌入空间中，图 7(a)会出现不同类特征重叠和误分类问题，例如“horse”、“cow”和“sheep”的特征交叉重叠，“motorbike”被误分类为“bicycle”等。通过语义信息的增强，可以看到图 7(b)中新类和基类被很好地分开，缓解了分类混淆的问题，证明该模块在减少类内方差和形成更明确的决策边界方面的有效性。



(a)基线 (b)基线+VSEM

图 7 t-SNE 可视化

Fig.7 Visualization of t-SNE

同时，λ 用于平衡语义对齐损失和其他损失，因此，本文展示了式(16)中不同的平衡系数 λ 对模型性能的影响，如图 8 所示。可以观察到，在 3-shot、5-shot、10-shot 设置下，当 $\lambda=0.50$ 时，nAP 达到最大值，而 λ 为其他值时，nAP 值反而下降，这表明合适的 λ 值能够促进语义知识和视觉特征更有效地融合，而过大或过小的 λ 值都会降低模型的整体性能。并且在 1-shot、2-shot 设置下，当 $\lambda=0.50$ 时也达到了次优的性能。综合考虑，选择 $\lambda=0.50$ 作为本文的平衡系数值。

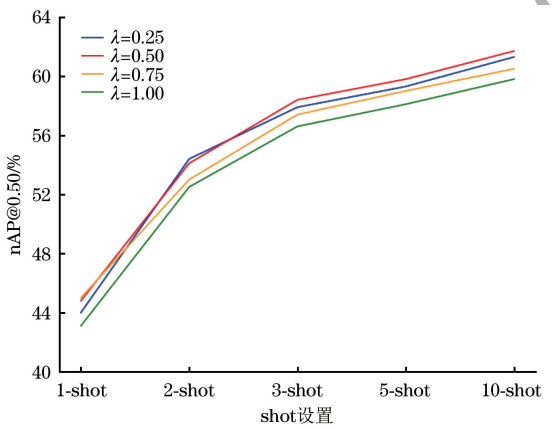
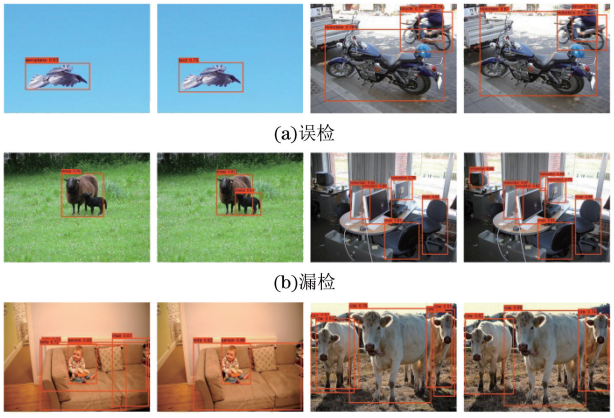


图 8 平衡系数 λ 不同值的影响

Fig.8 Impact of different values of the balance coefficient λ

最后，本文还选取了一些代表性的图像检测结果，如图 9 所示。图 9(a)中基线存在将新类错误识别为与其相似度较高的基类，例如将鸟误判为飞机，摩托车误判为自行车等情况，QGSE 能正确检测。在图 9(b)中，相比基线，QGSE 对位置信息更敏感，能准确定位，减少漏检。在图 9(c)中，同一个目标出现了多个不完整的框，而 QGSE 可以完整准确地框出目标。通过一系列图例，可以发现本文的方法有效地缓解了误检、漏检及不确定检测等问题，进一步证实了其在 FSOD 任务中的有效性。



(a)误检 (b)漏检 (c)不确定检测

图 9 部分测试结果对比

Fig.9 Comparison of some test results

4 结束语

基于元学习的范式，本文提出一种基于查询引导策略和视觉语义增强机制的小样本目标检测方法。通过查询引导模块增强支持特征中的查询感知信息，从而生成针对每个查询图像特定的高质量原型。为了提高查询感知信息的精确性，额外引入了一个辅助性的背景抑制模块。对于误分类问题，借助视觉语义增强模块利用文本语义信息对视觉特征进行自适应增强，以补充不完整的类别表示。最后采用度量模块学习更具辨别力的嵌入空间，同一类别特征聚集，不同类特征彼此远离。在 PASCAL VOC 和 MS COCO 数据集上的实验结果表明，相比现有主流的方法，本文方法实现了最佳的检测性能。未来将探索更准确的类概念表示，并考虑改进视觉和语义信息的交互，进一步优化小样本目标检测器的性能。

参考文献

[1] 史燕燕, 史殿习, 乔子腾, 等. 小样本目标检测研究综述[J]. 计算机学报, 2023, 46(8): 1753-1780.
SHI Y Y, SHI D X, QIAO Z T, et al. A survey on recent advances in few-shot object detection[J]. Chinese Journal of Computers, 2023, 46(8): 1753-1780. (in Chinese)

[2] KOHLER M, EISENBACH M, GROSS H M. Few-shot object detection: a comprehensive survey [J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(9): 11958-11978.

[3] WANG X, HUANG T E, DARRELL T, et al. Frustratingly simple few-shot object detection [C] // Proceedings of the 37th International Conference on Machine Learning. Washington D. C., USA: IEEE Press, 2020: 9919-9928.

[4] YANG Y, WEI F, SHI M, et al. Restoring negative information in few-shot object detection[C]//Proceedings of the Advances in Neural Information Processing Systems. Cambridge, USA: MIT Press, 2020: 3521-3532.

[5] FAN Z, MA Y, LI Z, et al. Generalized few-shot object detection without forgetting[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2021: 4527-4536.

[6] KANG B Y, LIU Z, WANG X, et al. Few-shot object detection via feature reweighting [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2019: 8420-8429.

[7] YAN X P, CHEN Z L, XU A N, et al. Meta R-CNN: towards general solver for instance-level low-shot learning [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2019: 9577-9586.

[8] QUAN J N, GE B Z, CHEN L. Cross attention redistribution with contrastive learning for few shot object detection[J]. Displays, 2022, 72: 102162.

[9] ZHANG L, ZHOU S G, GUAN J H, et al. Accurate few-shot object detection with support-query mutual guidance and hybrid loss[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE Press, 2021: 14424-14432.

[10] XU J B, WANG Y, HE X Y, et al. Support-query mutual promotion and classification correction network for few-shot object detection[J]. IEEE Signal Processing Letters, 2024, 31: 201-205.

[11] 姚涵涛, 余璐, 徐常胜. 视觉语言模型引导的文本知识嵌入的小样本增量学习[J]. 软件学报, 2024, 35(5): 2101-2119.
YAO H T, YU L, XU C S. Few-shot incremental learning with textual-knowledge embedding by visual-language model[J]. Journal of Software, 2024, 35(5): 2101-2119. (in Chinese)

[12] WERTHEIMER D, HARIHARAN B. Few-shot learning with localization in realistic settings[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE Press, 2019: 6558-6567.

[13] SUNG F, YANG Y X, ZHANG L, et al. Learning to compare: relation network for few-shot learning [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 1199-1208.

[14] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks [C] // Proceedings of the IEEE International Conference on Machine Learning. Washington D. C., USA: IEEE Press, 2017: 1126-1135.

[15] XU J Y, LE H, HUANG M Z, et al. Variational feature disentangling for fine-grained few-shot classification [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2021: 8812-8821.

[16] FAN Q, ZHUO W, TANG C K, et al. Few-shot object detection with attention-RPN and multi-relation detector [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE Press, 2020: 4013-4022.

[17] 徐宇坤, 张路军, 石林, 等. 意图注意力引导的小样本 3D 点云目标检测[J]. 计算机工程, 2024, 50(12): 288-295.
XU S K, ZHANG L J, S L et al. Few-shot 3D point cloud object detection guided by intention-attention[J]. Computer Engineering, 2024, 50(12): 288-295. (in Chinese)

[18] LEE H, LEE M, KWAK N. Few-shot object detection by attending to per-sample-prototype [C] // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Washington D. C., USA: IEEE Press, 2022: 2445-2454.

[19] LI Y W, FENG W Q, LYU S C, et al. Feature reconstruction and metric based network for few-shot object detection[J]. Computer Vision and Image Understanding, 2023, 227: 103600.

[20] LI B H, YANG B Y, LIU C, et al. Beyond max-margin: class margin equilibrium for few-shot object detection[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2021: 7363-7372.

[21] LI A X, LI Z G. Transformation invariant few-shot object detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2021: 3094-3102.

[22] HU H Z, BAI S, LI A X, et al. Dense relation distillation with context-aware aggregation for few-shot object detection [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2021: 10185-10194.

[23] ZHA Z C, TANG H, SUN Y L, et al. Boosting few-shot fine-grained recognition with background suppression and foreground alignment[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(8): 3947-3961.

[24] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes challenge [J]. International Journal of Computer Vision, 2010, 88(2): 303-338.

[25] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[M]. Berlin, Germany: Springer, 2014.

[26] LU Y, CHEN X, WU Z, et al. Decoupled metric network for single-stage few-shot object detection [J]. IEEE Transactions on Cybernetics, 2023, 53(1): 514-525.

[27] ZHAO X W, LIU X L, MA Y Q, et al. Temporal speciation network for few-shot object detection[J]. IEEE Transactions on Multimedia, 2023, 25: 8267-8278.

[28] ZHANG S, WANG L, MURRAY N, et al. Kernelized few-shot object detection with efficient integral aggregation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2022: 19207-19216.

文字编辑 索书志
栏目编辑 宋 圆