

基于自引导原型增强的小样本语义分割方法

——一种改进的基学习与元学习分割方法

陈 涵

合肥工业大学计算机与信息学院, 安徽 合肥

收稿日期: 2024年4月26日; 录用日期: 2024年5月24日; 发布日期: 2024年5月31日

摘 要

近期, 小样本语义分割研究备受关注并取得了显著进展。先前的方法主要依赖于分类任务的元学习框架以实现泛化能力, 然而这种训练方式往往导致模型对已见类别存在偏见, 未能达到理想的类别无关性。最新的研究中, 一种名为基学习与元学习的方法被提出, 以识别基础类别目标并有效地区分背景部分。然而, 该方法在强调对背景特征的识别时, 忽略了前景特征的增强。因此, 我们对该方法进行了进一步改进。我们引入了一种自引导原型学习的方法, 通过生成辅助原型并用它生成激活特征图, 从而增强原型特征, 有效促进模型对前景特征的识别。在PASCAL-5i数据集上的实验结果表明, 我们提出的方法在1-shot和5-shot情况下的mIoU分别达到了68.01和71.12, 证明了该方法能够有效提升小样本语义分割的精确度。

关键词

小样本语义分割, 原型学习, 特征增强

A Few-Shot Semantic Segmentation Method Based on Self Guided Prototype Enhancement

—An Improved Segmentation Method of Base Learning and Meta Learning

Han Chen

School of Computer Science and Information Engineering, Hefei University of Technology, Hefei Anhui

Received: Apr. 26th, 2024; accepted: May 24th, 2024; published: May 31st, 2024

Abstract

Recently, the research on few-shot semantic segmentation frameworks has gained significant attention and made notable progress. Previous methods mainly relied on meta-learning frameworks of classification tasks to achieve generalization, but this training approach often resulted in biases towards seen classes and failed to achieve ideal class-agnosticism. In the latest research, a method based on meta-learner was proposed to identify base class objects and effectively differentiate background regions. However, this method overlooked the enhancement of foreground features while emphasizing the recognition of background features. Therefore, we further improved this method. We introduced a self-guided approach that enhanced prototype features by generating auxiliary prototypes and using them to generate activation feature maps, thus effectively promoting the recognition of foreground features. Experimental results on the PASCAL-5i dataset showed that our proposed method achieved mIoU of 68.01 and 71.12 in 1-shot and 5-shot scenarios, respectively, demonstrating the effectiveness of the method in improving the accuracy of few-shot semantic segmentation.

Keywords

Few-Shot Semantic Segmentation, Prototypes, Feature Enhancement

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

语义分割作为计算机视觉的基本任务之一，广泛应用于医学图像识别、自动驾驶、地质勘探等领域 [1] [2]。基于卷积神经网络的语义分割模型需要大量具有像素级别标注的数据，然而获取大量的、高质量的像素级别标注数据费时费力 [3] [4] [5] [6] [7]，限制了语义分割在很多领域的应用。为了缓解语义分割模型需要大量标注数据的压力，使用少量的带标注样本完成语义分割任务的小样本语义分割问题被提出并受到广泛关注 [8]。小样本语义分割，也称小样本分割 (Few-shot semantic segmentation, FSS)。小样本分割主要采用基于度量的元学习方法进行 [9] [10]。利用度量的思想，使用度量工具 (例如，余弦相似度、卷积等) 度量少量带标签样本特征和待分割图片特征在高维空间的距离，对待分割图片中的每个像素点做出预测。

受小样本分类任务的推动，大多数小样本语义分割任务采取了元学习来实现模型的泛化。这些方法在基础数据集中采样一系列学习任务，模拟包含新类别的小样本场景来进行训练。然而这样训练的结果往往并不够充分的。当在基础数据集上进行元训练时，使用丰富的注释样本不可避免地引入了对已知类别的偏见从而妨碍了模型对新类别的学习。所以一些研究人员尝试调整包含足够训练样本的基础数据集的使用 [11]。他们除了使用小样本分割常见的元学习器外还引入了一个额外分支即基学习器，他们利用基学习器去预测基础类别的目标物体，通过自适应集成这两个学习器并行输出的粗略结果去生成准确的预测结果。这种操作可以帮助模型在传统范式的基础上更好地分割查询集中的易混淆区域。这种方法被称作基学习与元学习结合的小样本分割方法 (Base and Meta) [12]，后文简称 BAM。此外在目前的小样本语义分割方法中，文献 [13] [14] [15] 等方法都采用原型学习范式进行元推理，通过平均池化操作生成单个或

多个原型，为查询对象推理提供了必要的前景线索。然而，当面对具有模糊对象边界和无关对象干扰等因素的困难小样本任务时，它们的分割能力可能会大打折扣。我们认为这些方法失败的一个可能原因在于未充分利用支持图像-掩膜对中的信息，仅依靠压缩的与对象相关的特征来引导分割是远远不够的。甚至在支持集引导的过程中即使查询集图像与支持集图像一样也无法很好地引导分割[16]。

为此，我们在 BAM 方法中引入一种原型特征增强的方法，使模型对目标类别可以有更强的分辨能力。该方法利用基学习器完成对查询集背景中基类区域有效识别，同时原型特征增强策略进一步加强了对原有原型中学习不充分的部分，这使得模型同时对待分割图像的前景和背景完成了特征增强。具体的，我们给出一种自引导的方式，利用支持集生成辅助原型，再通过辅助原型引导查询集生成激活特征图，最终使用基学习器和元学习器得到的结果进行有效集成得到更好的分割结果。该方法在 Pascal 数据集上取得了较为明显的效果。

2. 相关工作

2.1. 小样本学习

计算机视觉领域多年来一直致力于使网络具备对新类别的泛化能力。目前在小样本学习领域中，大多数方法都遵循[17]中提出的元学习框架，其中从基础数据集中采样一组学习任务以模拟小样本场景。基于此，小样本学习方法可以进一步分为以下三个类别：第一是基于度量的方法[17][18]该方法使用特定的距离度量(如欧氏距离和余弦距离)确定支持原型与查询特征之间的关联性。第二是基于优化的方法[19][20]，第三是基于特征增强的方法[21][22]。

2.2. 小样本语义分割

为了解决稀疏注释数据集上的泛化问题，将小样本学习方法应用于语义分割领域。度量学习是小样本学习方法之一，在小样本分割语义任务中得到了广泛应用，例如使用原型向量学习知识。一些小样本语义分割方法[23][24]开始探索建立像素级别的特征匹配与引导。他们有选择地从支持图像中提取有用的信息，通过聚合查询和支持特征之间的关系并在这种关系的引导下进行语义分割。然而这种方法往往需要耗费巨大的计算量。另一些小样本分割方法主要基于原型向量[13][14]。它们对支持特征的前景区域进行全局平均池化，得到全局特征向量，并将查询特征的空间位置与这些特征向量进行比较。一些工作[16][24]尝试添加多个包含目标类别的部分感知特征的原型向量，以丰富支持特征表示。

3. 问题设定

小样本语义分割是一个新兴的研究方向，它结合了语义分割和小样本学习的思想。其主要目标是通过使用少量带标签的未知类别来训练网络，使其能够对该类别的图像进行准确的分割。在这个任务仍然需要训练集和测试集对模型进行训练和评估，但是这个任务的训练集与测试集设定与别的深度学习任务中的设定有一定的区别。

具体而言，我们将训练集称为 D_{train} ，测试集称为 D_{test} 分别用于训练和测试模型。但是训练集中作为分割目标的类别 C_{train} 和测试集中作为分割目标的类别 C_{test} 是完全不同的，并且没有交集，即 $C_{train} \cap C_{test} = \emptyset$ 。除此以外，在小样本语义分割最初的工作 OSLSM [8]中引入一个“episode”范式，也叫片段范式。每一个片段的构成是同一目标类别的支持集图片和查询集图片。每个片段由同一类别的支持集和查询集图片组成。将支持集称作 S ， S 中包含了通过在包含同一类别的 N 张图片中随机采样得到的 K 张图片，将其写作 $S = \{I_s^i, M_s^i\}_{i=1}^K$ ，其中 I_s^i 代表第 i 张支持集图片。 M_s^i 代表对应支持集的二进制掩码。同样的，将查询集称作 Q ， Q 包含带掩码 M_q 的图片 I_q ，即 $Q = \{I_q, M_q\}$ ，而查询集真实的标签掩码

M_q 仅在训练阶段使用。

此外，我们将训练集和测试集的支持集与查询集分别记为 S_{train} 和 Q_{train} 以及 S_{test} 和 Q_{test} 。在训练阶段，我们使用 (I_q, M_q, I_s, M_s) 来训练模型。一旦模型训练完成，在测试阶段，我们不需要对网络进行微调，而是从测试集中随机采样 (I_q, I_s, M_s) ，输入模型并获得查询集的预测结果。

4. 方法

在本节中，我们将详细介绍如何在 BAM 中实现原型特征增强。简而言之，我们利用自引导的方式对支持集进行分割，在对分割正确和分割错误的部分进行特征聚合得到表达信息更完整的辅助原型。最后利用辅助原型结合基学习与元学习对查询集进行分割，得到更加精细的分割结果。

4.1. BAM 简述

当前小样本分割的泛化性能严重依赖元学习，即使进行了微调过程，其能力也不达标。具体来说，训练的小样本语义模型由于数据分布不平衡会从而对基类产生偏见。在这个领域中，很少有研究明确研究泛化性能下降的问题，而是专注于设计两个分支之间的高容量交互模块。然而 Lang [12] 为了减轻当前小样本分割方法中的偏见问题，提出建立一个利用额外学习器的方法，明确预测查询图像中基础类别的区域，从而促进新对象的分割。所提出的方法主要由三个组件构成，包括一对互补学习器(即基学习器和元学习器)和一个集成模块。使用共享主干的两个学习器分别用于识别基础类别和不可见类。然后，集成模块接收它们的粗略预测结果和利用浅层特征计算得到的一个调整因子 ψ ，以抑制基础类别错误激活的区域，从而进一步产生准确的分割结果，具体如图 1 所示。我们将该方法分为第一阶段和第二阶段，第一阶段结合基学习器与元学习生成粗糙结果，第二阶段利用浅层特征得到的调整因子对基学习预测结果和元学习预测结果进行集成调整最后得到最终的结果，我们的改进在第一阶段上进行。

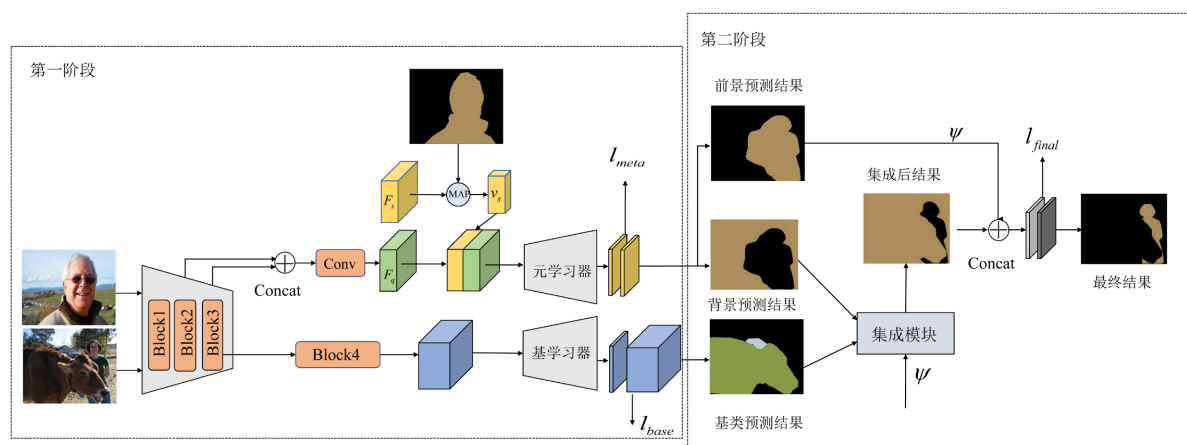


Figure 1. Illustration of the combination of base learning and meta-learning for few-shot semantic segmentation

图 1. 基学习与元学习结合的小样本语义分割示意图

4.2. 原型特征增强方法概述

如图 2 所示，该方法结合基学习与元学习方法，在此之上利用自引导机制生成辅助原型，从而感知原本模型中可能会忽略掉的区域，利用辅助原型激活查询集的不同区域生成激活特征图，最后将激活特征图与拓展维度的原型和原本查询集特征聚合，得到增强后特征，利用增强后的特征引导查询集分割。如此一来该方法不仅增强了前景特征，还通过基类学习器增强了模型对背景的感知。

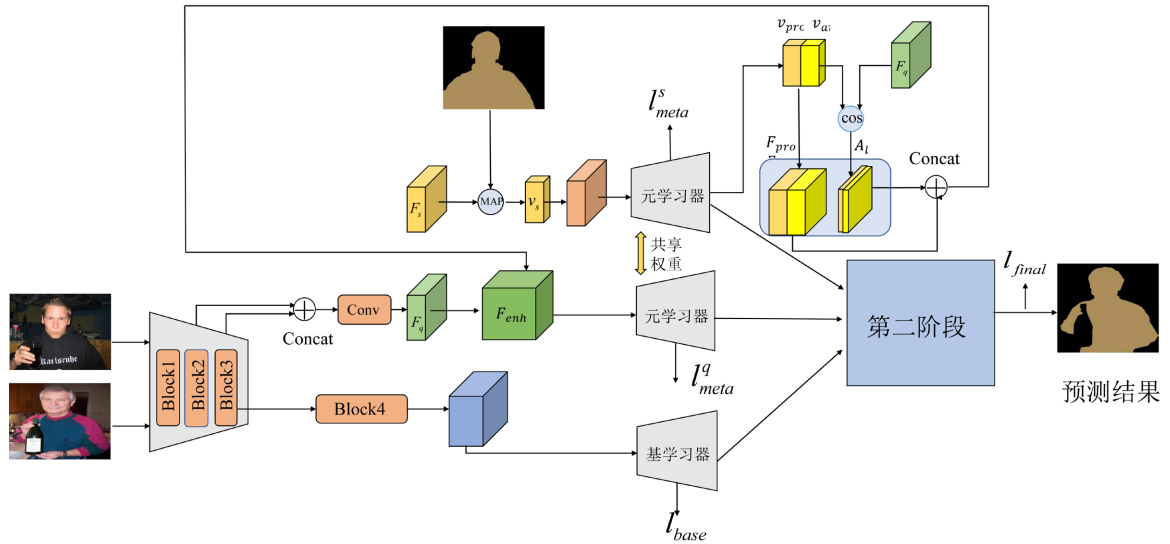


Figure 2. Diagram of the method we proposed
图 2. 提出的方法

4.3. 自引导辅助原型生成

原型学习是小样本分割一种重要的方法，这类方法利用原型网络的思想，首先将支持集图像输入骨干网络中得到支持集图像的高维特征，再结合支持集图像的掩膜进行全局平均池化得到原型。该原型表示通过支持集图像计算得到的该目标类在高维空间中的特征中心。同样的，将查询集图像也送入同样权重的骨干网络得到查询集在高维空间的特征表达。最后利用无参度量工具如余弦函数或欧式距离公式衡量高维特征与原型之间的距离完成对图像像素点的分类，得到最终的预测图。然而这样得到的原型往往会损失不少有用的信息。为此，引入自引导模块可以更加有效利用支持集信息，提供更加全面的信息引导查询集进行分割。

按照在问题设定中的定义，支持集图像被设定为 I_s ，我们选择基于 ResNet50 [26] 拓展得到的 PSPNet [2] 作为模型的骨干网络，支持集图片和查询集图片分别输入到骨干网络得到的特征为 $\{F_s^m, F_q^m\}$ 。其中 s 和 q 分别代表支持集和查询集， m 表骨干网络的层数。对于支持集而言，我们聚合骨干网络中得到的第二层和第三层特征作为骨干网络的输出进行下一步处理。具体的，将第三层和第四层特征在通道维度进行拼接得到待处理的特征 F_s 。这里的 $Concat(\bullet)$ 表示拼接操作

$$F_s = Concat\left(\left[F_s^2, F_s^3\right]\right) \quad (1)$$

之后，为了利用支持集的掩膜对所得特征进行处理生成待分割类别的原型，我们需要将掩膜下采样到和特征 F_s 一样的大小。有：

$$M_s = DownSample(M) \quad (2)$$

其中 M 代表支持集原始的掩膜。 M_s 代表对原始掩膜下采样后得到的与特征 F_s 高度和宽度大小一致的掩膜。完成上述处理，我们遵循之前工作的操作利用掩膜全局平均池化生成初始原型[3] [4] [5]。

$$v_s = \frac{\sum_{i=1}^{hw} F_s(i) \cdot [M_s(i)=1]}{\sum_{i=1}^{hw} [M_s(i)=1]} \quad (3)$$

这里的 i 表示特征图中空间位置的索引。 h 和 w 分别是特征图的高度和宽度。 $[\bullet]$ 是艾弗森括号。它代表中括号中的内置条件如果成立那么这个表达式的值取 1, 如果不成立这个表达式的值取 0。在这里 M_s 表示一个二进制掩膜, $M_s(i)=1$ 表示在第 i 个位置上的这个像素点是属于待分割的目标类别。

然而, 通过这种方式所求得到的原型往往是对待分割类别信息的粗略表达, 在掩膜全局平均池化这一过程中, 一些可以有效指导查询集图像分割的信息会被忽略掉, 没有在接下来引导分割的任务中保存下来。这样即使使用与查询集图片完全一样的支持集图片引导分割也会损失一定的分割细节。为此我们设置了一种自引导方式, 使用之前计算得到的初始原型 v_s 首先对支持集进行引导分割。首先参考[3][6]的做法, 将初始原型拓展到和支持集特征相同的尺寸, 新得到的原型特征称作 V_s 。然后将新得到的原型特征 V_s 与经过骨干网络得到的支持集特征沿着通道维度进行拼接得到一个新的特征图 F_v 。

$$F_v = \text{Concat}([F_s, V_s]) \quad (4)$$

通过这样一个融合了初始原型的特征, 我们可以顺利对支持集图像进行分割, 这样一个代表支持集图像的概率图可以通过解码器被求出来, 对此有:

$$P_s = \text{soft max}(\text{Decoder}(F_v)) \quad (5)$$

这里的 P_s 是对支持集图片预测的概率图, 其中 $P_s \in R^{2 \times h \times w}$, $\text{Decoder}(\bullet)$ 表示解码器, 这个解码器与后续对查询集分割的解码器共享权重。共享权重的解码器解码器的具体细节内容会在 4.4 节介绍。 soft max 是一个归一化函数。接下来利用支持集的概率特征图利用一个求最大值的函数完成对支持集预测掩膜的生成。

对此有:

$$\hat{M}_s = \text{argmax}(P_s) \quad (6)$$

这里的 \hat{M}_s 依然是一个二进制的掩膜, 在这个掩膜当中如果该位置值为 0 则代表这里被预测为支持集图像的背景, 如果值为 1 则代表这个像素点被预测为目标类。通过对支持集预测的掩膜和支持集掩膜的真实值进一步处理, 我们可以生成用来完成原型特征增强的两个辅助原型。

$$v_{pro} = \frac{\sum_{i=1}^{hw} F_s(i) \cdot [M_s(i)=1] [\hat{M}_s(i)=1]}{\sum_{i=1}^{hw} [M_s(i)=1] [\hat{M}_s(i)=1]} \quad (7)$$

$$v_{aux} = \frac{\sum_{i=1}^{hw} F_s(i) \cdot [M_s(i)=1] [\hat{M}_s(i) \neq 1]}{\sum_{i=1}^{hw} [M_s(i)=1] [\hat{M}_s(i) \neq 1]} \quad (8)$$

通过公式 7 和公式 8 可以计算得到用来实现原型特征增强的辅助原型, 在公式 7 中 $[M_s(i)=1][\hat{M}_s(i)=1]$ 表示使用初始原型 v_s 作为引导分割支持集时正确预测的前景掩膜, 它代表在这些位置支持集待分割的目标被精确地分割出来。而公式 8 中 $[M_s(i)=1][\hat{M}_s(i) \neq 1]$ 表示初始原型 v_s 引导没有被分割出来的部分, 即缺失的前景掩膜。从公式 7 可以看到 v_{pro} 可以有效聚合模型预测正确的信息, 而 v_{aux} 可以聚合之前分割中被遗忘的细节信息。通过这样的操作, 之前初始原型中利用支持集图像得到泛化性信息被保留, 而在初始原型中被忽略的细节公式 8 重新搜集。因此得到的辅助原型可以更加充分表达目标类的主体与细节信息, 从而在接下来的步骤中更加有效地引导查询集进行分割。

4.4. 原型增强

4.4.1. 伪掩膜生成

遵循大多数小样本分割任务[12] [14], 我们也使用了伪掩膜作为最后解码器的输入。利用伪掩膜可以在没有可学习参数的情况下大致定位前景对象。具体而言, 它通过测量编码器输出的查询集高层特征和支持集前景特征的余弦相似度实现。在这之后, 得到的相似度被归一化, 并作为查询集特征成为前景即待分割目标的概率值。我们利用编码器最后一层特征 F_s^4 和 F_q^4 以及变化到对应尺寸的掩膜进行上述操作。

$$x_s = F_s^4 \odot M_{stmp} \quad (9)$$

其中 x_s 代表支持集前景特征, M_{stmp} 代表变化后尺寸的支持集掩膜。之后对所得前景特征与查询集特征进行余弦相似度计算得到:

$$\cos(F_{imp}, F_q^4) = \frac{F_{imp}^\top F_q^4}{\|F_{imp}^\top\| \|F_q^4\|} \quad (10)$$

$$c_q = \max(\cos(F_{imp}, F_q^4)) \quad (11)$$

$$C_Q = [c_1, c_2, \dots, c_{hw}] \quad (12)$$

之后将 C_Q 形状变为三维得到 Y_Q , 其中 ε 被设定为 10^{-7} 。

$$Y_Q = \frac{Y_Q - \min(Y_Q)}{\max(Y_Q) - \min(Y_Q) + \varepsilon} \quad (13)$$

4.4.2. 激活特征图

在激活注意力图的第一步首先进行特征匹配操作, 密集特征匹配是小样本语义分割中的常见方法, 它通常将查询集特征图 F_q 与全局特征向量在所有空间位置上进行比较实现特征匹配。具体而言我们分别在每个空间位置评估它们与查询特征的余弦相似性。由于在之前的步骤中得到了两个全新的原型, 因此要将该操作拓展为多原型设置。值得注意的是, 这里的查询集特征与求我们将两个原型按照通道维度进行拼接。具体的有:

$$F_q = \text{Concat}([v_{pro}, v_{aux}]) \quad (14)$$

$$F_q = \text{Concat}([F_q^2, F_q^3]) \quad (15)$$

$$A_l = \frac{F_{pro} F_q}{\|F_{pro}\| \|F_q\|} \quad (16)$$

A_l 是一个双通道的激活特征图, 每一个通道与之前求得的辅助原型相对应。如此一来, 两个辅助原型感兴趣的区域都可以在查询集中被激活。

4.4.3. 增强特征生成

最终增强的特征包含生成的辅助原型、激活特征图、查询集特征和伪掩膜。为了使这些特征融合在一起, 我们首先将得到的辅助原型拓展到和查询集特征图相同的尺寸, 然后用一种简单的通道拼接的方式将不同特征连接起来, 最后通过一层卷积层完成对通道维度的压缩, 一方面压缩通道可以降低模型的参数量提高模型的效率和推理速度, 另一方面可以使特征更好的聚合在一起, 形成一个有效的增强特征。这里的 $\text{Expand}(\cdot)$ 表示对特征图拓展尺寸大小。

$$F_{v_{pro}} = \text{Expand}(v_{pro}) \quad (17)$$

$$F_{v_{aux}} = \text{Expand}(v_{aux}) \quad (18)$$

$$F_{enh} = \text{Concat}\left(\left[F_{v_{pro}}, F_{v_{aux}}, A_l, F_q, Y_Q\right]\right) \quad (19)$$

4.5. 解码器

该方法的解码器与 BAM [12]方法中的解码器保持一致，具体结构如图 3：编码器得到的特征先经过空洞空间卷积池化金字塔[27]，随后经过几个由卷积和激活函数堆叠而成的模块，第一个模块由卷积核大小为 1 的卷积核和 ReLU [28]激活函数组成，第二个模块由一个卷积核大小为 3 的卷积和 ReLU 激活函数，第三个模块由 2 个连续的卷积核大小为 3 的卷积和 ReLU 激活函数构成。值得注意的是第二个模块和第三个模块之间有一个残差连接。最后一个模块依次由卷积核大小为 3 的卷积、ReLU 激活函数、Dropout [29]，卷积核大小为 3 的卷积连续堆叠而成。

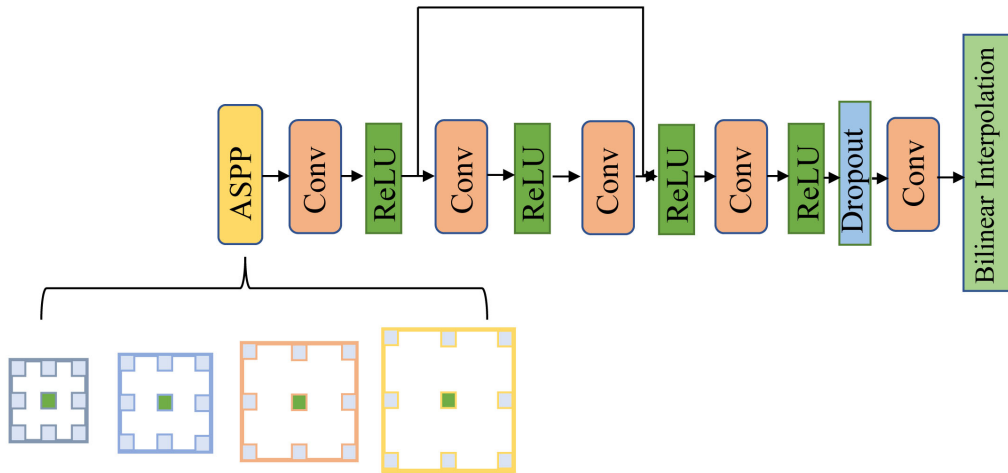


Figure 3. Diagram of the decoder architecture
图 3. 解码器结构示意图

4.6. 损失函数

为了保证 v_{pro} 可以从支持特征图中学习到大部分有用的信息，我们使用交叉熵损失函数对预测结果进行约束。在公式 20 到公式 23 中 P_s, P_q, P_{result} 与 M_s, M_q 分别代表模型输出的掩码概率图和对应的真实值。

$$l_{meta}^s = \frac{1}{n_e} \sum_{i=1}^{n_e} BCE(P_s, M_s) \quad (20)$$

类似的有对查询集预测结果进行训练

$$l_{meta}^q = \frac{1}{n_e} \sum_{i=1}^{n_e} BCE(P_q, M_q) \quad (21)$$

上面的公式表示的是利用交叉熵损失函数对查询集分割结果与真实值的差异做评估。其中 n_e 表示训练时的批次数量。最后对模型在第二阶段输出的结果在训练阶段利用 l_{final} 进行约束。最终我们将几个损失函数相加得到在元学习阶段的损失函数：

$$l_{final} = \frac{1}{n_e} \sum_{i=1}^{n_e} BCE(P_{result}, M_q) \quad (22)$$

$$l = l_{final} + l_{meta}^s + l_{meta}^q \quad (23)$$

同样的，有在基学习阶段的损失函数，其中 n 表示训练时的批次数。

$$l_{base} = \frac{1}{n} \sum_{i=1}^n BCE(P_{base}, M_{base}) \quad (24)$$

5. 实验

5.1. 实施细节

标准小样本分割是在 PASCAL-5i [8]数据集上进行的。该数据集是基于 PASCAL VOC 2012 [30]中创建的，其中包含来自 SDS [31]的额外掩码注释，由 20 个语义类别组成，均匀分为 4 份。模型其中 3 份上进行训练，并以交叉验证方式在剩余的一份上进行测试。我们从网络未见的那份数据集中随机采样 1,000 个片段进行评估。

遵循之前的小样本语义分割方法[13]，我们采用平均交并比(mIoU)作为评估指标，其中 mIoU 指标主要用于反映不同类别的整体性能。

为了与现有的方法进行公平比较，我们采用了相同的骨干网络(ResNet50 [26])。根据[13]的方法，我们对这些骨干网络进行了预训练，并在小样本分割任务训练期间冻结它们以提高泛化能力。预处理技术的设置与[13]相同。使用 SGD 优化器，在 PASCAL-5i 上使用学习率为 0.005 进行 200 个 epoch 的训练，所有实验都在 NVIDIA RTX 3090GPU 上使用 PyTorch 框架进行。我们报告了使用 4 个不同 fold 进行训练测试的平均结果，以减少所选的支持 - 查询图像对性能的影响

5.2. 方法比较

我们将提出的模型与最近的几个小样本分割方法进行比较。在表 1 中显示了在 ResNet50 骨干网络下 PASCAL-5i 数据集上的结果。从表格中可以看出，在 1-shot 和 5-shot 场景相比基线网络 BAM [25]，我们的方法的性能都有一定程度的提升。尤其是在 1-shot 下，每一次交叉验证中都取得了更好的效果。

Table 1. Performance on PASCAL-5i in mIoU
表 1. PASCAL-5i 数据集上 mIoU 指标性能对比

方法	1-shot					5-shot				
	5 ⁰	5 ¹	5 ²	5 ³	mean	5 ⁰	5 ¹	5 ²	5 ³	mean
CANet [13]	52.50	65.90	51.30	51.90	55.40	55.50	67.80	51.90	53.20	57.10
PGNet [24]	56.00	66.90	50.60	50.40	56.00	57.70	68.70	52.90	54.60	58.50
CRNet [32]	-	-	-	-	55.70	-	-	-	-	58.80
骨干网络										
PPNet [33]	48.58	60.58	55.71	46.47	52.84	58.85	68.28	66.77	57.98	62.97
ResNet50										
PFENet [14]	61.70	69.50	55.40	56.30	60.80	63.10	70.70	55.80	57.90	61.90
SCL_1 [16]	56.80	67.30	53.50	52.50	57.50	59.50	68.50	54.90	53.70	59.20
SCL_2 [16]	63.00	70.00	56.50	57.70	61.80	64.50	70.90	57.30	58.70	62.90
HSNet [23]	64.30	70.70	60.30	60.50	63.90	70.30	73.20	67.40	67.10	69.50
DCP [25]	63.81	70.54	61.16	55.69	62.80	67.19	73.15	66.39	64.48	67.80
BAM [12]	68.97	73.59	67.55	61.13	67.81	70.59	75.05	70.79	67.20	70.91
Ours	69.22	73.77	67.88	61.79	68.02	71.05	74.90	71.12	67.41	71.12

5.3. 消融实验

Table 2. Comparison results of ablation study on PASCAL-5i

表 2. PASCAL-5i 数据集上消融实验性能对比

v_s	v_{pro}	v_{aux}	Activation Map	$mIoU$
√				67.81
	√	√		67.84
	√	√	√	68.02

我们在 PASCAL-5i 上进行了消融实验，消融实验结果如表 2 所示。对最后强化特征中仅保留了原始原型、辅助原型，辅助原型和激活特征图(Activation Map)的三种情况进行了消融实验，可以发现仅仅依靠辅助原型对模型性能提升并不大，但是和激活特征图配合使用会让模型性能有一定的提升。

5.4. 可视化

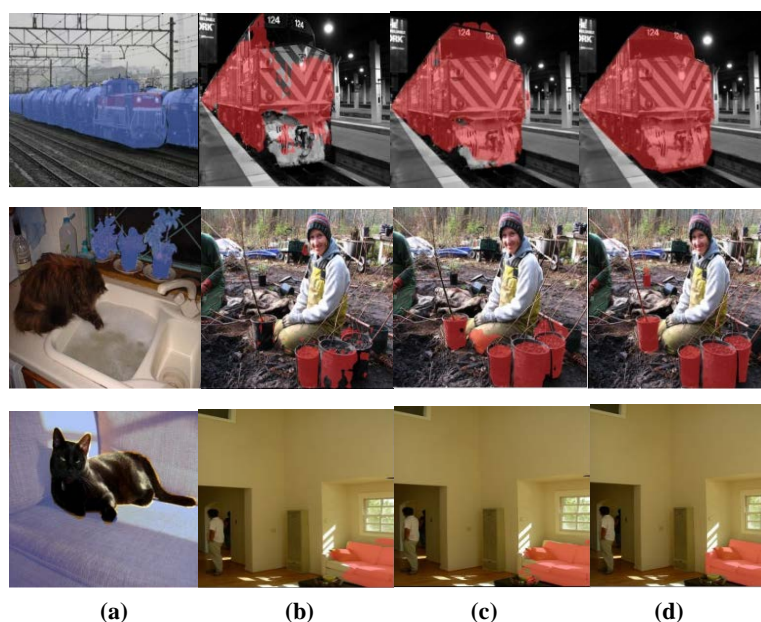


Figure 4. Diagram of Visualization results

图 4. 可视化结果展示

本小节将对模型在 BAM 上运行的结果和我们提出的方法上运行的结果进行比较，具体情况如图 4 所示，第(a)列是支持集图像，第(b)列是 BAM 方法的分割结果，第(c)列是本文提出的方法分割结果，第(d)列是查询集图像的标签，其中支持集目标对象以蓝色蒙版呈现，其余图像中的目标对象以品红色蒙版展现。可以看出我们的方法在细节上的分割比 BAM 做的更加出色。

6. 结论

在本文中，我们在 BAM 方法的基础上提出一个原型特征增强的方法。我们的方法充分利用了支持集的信息，通过自引导的方式构建了辅助原型和激活特征图，完成了原型特征增强。这种方法弥补了 BAM 对前景特征细节关注不够的问题，我们在 PASCAL-5i 数据集上进行实验，验证了方法的有效性。在未来，可以尝试更细致的方式完成对前景特征的充分提取，以此提高小样本语义分割的精确性。

参考文献

- [1] Dong, G., Yan, Y., Shen, C., *et al.* (2020) Real-Time High-Performance Semantic Image Segmentation of Urban Street Scenes. *IEEE Transactions on Intelligent Transportation Systems*, **22**, 3258-3274. <https://doi.org/10.1109/TITS.2020.2980426>
- [2] Zhang, J., Xie, Y., Xia, Y., *et al.* (2021) Dodnet: Learning to Segment Multi-Organ and Tumors from Multiple Partially Labeled Datasets. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, Nashville, 20-25 June 2021, 1195-1204. <https://doi.org/10.1109/CVPR46437.2021.00125>
- [3] Badrinarayanan, V., Kendall, A. and Cipolla, R. (2017) Segnet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 2481-2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
- [4] Lin, D., Dai, J., Jia, J., *et al.* (2016) Scribblesup: Scribble-Supervised Convolutional Networks for Semantic Segmentation. 2016 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 3159-3167. <https://doi.org/10.1109/CVPR.2016.344>
- [5] Sun, W. and Wang, R. (2018) Fully Convolutional Networks for Semantic Segmentation of Very High Resolution Remotely Sensed Images Combined with DSM. *IEEE Geoscience and Remote Sensing Letters*, **15**, 474-478. <https://doi.org/10.1109/LGRS.2018.2795531>
- [6] Peng, C., Zhang, X., Yu, G., *et al.* (2017) Large Kernel Matters—Improve Semantic Segmentation by Global Convolutional Network. 2017 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 1743-1751. <https://doi.org/10.1109/CVPR.2017.189>
- [7] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015*, Munich, 5-9 October 2015, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28
- [8] Shaban, A., Bansal, S., Liu, Z., *et al.* (2017) One-Shot Learning for Semantic Segmentation. arXiv:1709.03410. <https://doi.org/10.48550/arXiv.1709.03410>
- [9] Vinyals, O., Blundell, C., Lillicrap, T., *et al.* (2016) Matching Networks for One Shot Learning. *Proceedings of the 30th International Conference on Neural Information Processing Systems*, Barcelona, 5-10 December 2016, 3637-3645.
- [10] Huisman, M., Van Rijn, J.N. and Plaat, A. (2021) A Survey of Deep Meta-Learning. *Artificial Intelligence Review*, **54**, 4483-4541. <https://doi.org/10.1007/s10462-021-10004-4>
- [11] Fan, Z., Ma, Y., Li, Z., *et al.* (2021) Generalized Few-Shot Object Detection without Forgetting. 2021 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 4525-4534. <https://doi.org/10.1109/CVPR46437.2021.00450>
- [12] Lang, C., Cheng, G., Tu, B., *et al.* (2022) Learning What not to Segment: A New Perspective on Few-Shot Segmentation. 2022 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, 18-24 June 2022, 8047-8057. <https://doi.org/10.1109/CVPR52688.2022.00789>
- [13] Zhang, C., Lin, G., Liu, F., *et al.* (2019) CANet: Class-Agnostic Segmentation Networks with Iterative Refinement and Attentive Few-Shot Learning. 2019 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 5212-5221. <https://doi.org/10.1109/CVPR.2019.00536>
- [14] Tian, Z., Zhao, H., Shu, M., *et al.* (2020) Prior Guided Feature Enrichment Network for Few-Shot Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44**, 1050-1065. <https://doi.org/10.1109/TPAMI.2020.3013717>
- [15] Zhang, X., Wei, Y., Yang, Y., *et al.* (2020) SG-One: Similarity Guidance Network for One-Shot Semantic Segmentation. *IEEE Transactions on Cybernetics*, **50**, 3855-3865. <https://doi.org/10.1109/TCYB.2020.2992433>
- [16] Zhang, B., Xiao, J. and Qin, T. (2021) Self-Guided and Cross-Guided Learning for Few-Shot Segmentation. 2021 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 8308-8317. <https://doi.org/10.1109/CVPR46437.2021.00821>
- [17] Li, H., Eigen, D., Dodge, S., *et al.* (2019) Finding Task-Relevant Features for Few-Shot Learning by Category Traversal. 2019 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 1-10. <https://doi.org/10.1109/CVPR.2019.00009>
- [18] Sung, F., Yang, Y., Zhang, L., *et al.* (2018) Learning to Compare: Relation Network for Few-Shot Learning. 2018 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 1199-1208. <https://doi.org/10.1109/CVPR.2018.00131>
- [19] Finn, C., Abbeel, P. and Levine, S. (2017) Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, Sydney, 6-11 August 2017, 1126-1135.

-
- [20] Ravi, S. and Larochelle, H. (2016) Optimization as a Model for Few-Shot Learning. *ICLR 2017 Conference Track 5th International Conference on Learning Representations*, Toulon, 24-26 April 2017.
- [21] Chen, Z., Fu, Y., Chen, K., *et al.* (2019) Image Block Augmentation for One-Shot Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, **33**, 3379-3386. <https://doi.org/10.1609/aaai.v33i01.33013379>
- [22] Chen, Z., Fu, Y., Wang, Y.X., *et al.* (2019) Image Deformation Meta-Networks for One-Shot Learning. 2019 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 8672-8681. <https://doi.org/10.1109/CVPR.2019.00888>
- [23] Min, J., Kang, D. and Cho, M. (2021) Hypercorrelation Squeeze for Few-Shot Segmentation. 2021 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, 10-17 October 2021, 6921-6932. <https://doi.org/10.1109/ICCV48922.2021.00686>
- [24] Zhang, C., Lin, G., Liu, F., *et al.* (2019) Pyramid Graph Networks with Connection Attentions for Region-Based One-Shot Semantic Segmentation. 2019 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, 27 October-02 November 2019, 9586-9594. <https://doi.org/10.1109/ICCV.2019.00968>
- [25] Lang, C., Tu, B., Cheng, G., *et al.* (2022) Beyond the Prototype: Divide-and-Conquer Proxies for Few-Shot Segmentation. *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, Messe Wien, 23-29 July 2022, 1024-1030. <https://doi.org/10.24963/ijcai.2022/143>
- [26] He, K., Zhang, X., Ren, S., *et al.* (2016) Deep Residual Learning for Image Recognition. 2016 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [27] Chen, L.C., Papandreou, G., Kokkinos, I., *et al.* (2017) DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **40**, 834-848. <https://doi.org/10.1109/TPAMI.2017.2699184>
- [28] Nair, V. and Hinton, G.E. (2010) Rectified Linear Units Improve Restricted Boltzmann Machines. *Proceedings of the 27th International Conference on International Conference on Machine Learning*, Haifa, 21-24 June 2010, 807-814.
- [29] Hinton, G.E., Srivastava, N., Krizhevsky, A., *et al.* (2012) Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors. arXiv:1207.0580. <https://doi.org/10.48550/arXiv.1207.0580>
- [30] Everingham, M., Van Gool, L., Williams, C.K.I., *et al.* (2010) The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, **88**, 303-338. <https://doi.org/10.1007/s11263-009-0275-4>
- [31] Hariharan, B., Arbeláez, P., Bourdev, L., *et al.* (2011) Semantic Contours from Inverse Detectors. 2011 *International Conference on Computer Vision*, Barcelona, 6-13 November 2011, 991-998. <https://doi.org/10.1109/ICCV.2011.6126343>
- [32] Liu, W., Zhang, C., Lin, G., *et al.* (2020) CRNet: Cross-Reference Networks for Few-Shot Segmentation. 2020 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 4164-4172. <https://doi.org/10.1109/CVPR42600.2020.00422>
- [33] Liu, Y., Zhang, X., Zhang, S., *et al.* (2020) Part-Aware Prototype Network for Few-Shot Semantic Segmentation. *Computer Vision-ECCV 2020*, Glasgow, 23-28 August 2020, 142-158. https://doi.org/10.1007/978-3-030-58545-7_9