

基于多尺度先验融合的小样本语义分割方法

谭文博

同济大学电子与信息工程学院, 上海

收稿日期: 2024年1月23日; 录用日期: 2024年2月22日; 发布日期: 2024年2月29日

摘要

小样本语义分割侧重于模型泛化能力的提升, 用有限的样本为未见类提供像素级分割。现有方法在分割性能上已经取得了巨大的进步, 但仍受到部分因素的限制, 例如当同类物体中因较大的角度、尺寸等差异, 存在类内差异时, 将导致模型难以捕捉类别语义。因此, 本文提出基于多尺度先验信息融合的小样本语义分割模块, 通过提取高级特征中的多尺度先验掩码, 为模型补充类别相关信息, 降低类内差异的影响。为了融合多尺度特征及其先验信息, 本文提出权重自适应特征融合模块, 为各尺度添加信息交互, 并加权组合特征。最终通过消融实验以及与基类方法和其他经典算法的对比实验, 证明了本方法的有效性。

关键词

小样本语义分割, 深度学习, 多尺度特征, 微调

Few-Shot Semantic Segmentation Method Based on the Fusion of Multi-Scale Priors

Wenbo Tan

College of Electronic and Information Engineering, Tongji University, Shanghai

Received: Jan. 23rd, 2024; accepted: Feb. 22nd, 2024; published: Feb. 29th, 2024

Abstract

Few-shot semantic segmentation focuses on improving the model's generalization ability to provide pixel-level segmentation for unseen classes with limited samples. While existing methods have made significant progress in segmentation performance, they are still constrained by certain factors. For instance, when there are intra-class gaps due to significant variations in angles, sizes, etc., within the same class, the model struggles to capture category semantics. Therefore, this paper proposes a few-shot semantic segmentation module based on the fusion of multi-scale prior information. By extracting multi-scale prior masks from high-level features, this module sup-

plements category-related information for the model, reducing the impact of intra-class differences. To integrate multi-scale features and their prior information, the paper introduces a Weighted Adaptive Feature Fusion module, adding information interaction across various scales and combining features with weighted contributions. Ultimately, through ablation experiments and comparative experiments with baseline methods and other classical algorithms, the effectiveness of this approach has been demonstrated.

Keywords

Few-Shot Semantic Segmentation, Deep Learning, Multi-Scale Features, Fine-Tuning

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

一张数字图像中往往包括丰富的语义信息，包括不同种类的物体、背景等等，通过将图像中的每一个像素划分为不同的区域，来区分出图片中的不同主体类别，从而为后续的图像处理提供便利。这种为图像语义信息提供像素级标注的任务称为图像语义分割，其建立在更简单的视觉任务(例如图像分类和目标检测)的基础上，提供了更为精确的判别信息。

得益于大规模高精度标注数据集的出现和深度模型语义表征能力的不断增强，基于深度学习的图像语义分割方法已经取得了显著的性能提升，能够在标注样本充足的情况下实现准确分割。与传统的语义分割算法相比，现有的基于卷积神经网络的语义分割算法具有表征能力优越、分割性能良好等优势。然而，全监督图像语义分割算法对数据高度依赖，需要大规模、高质量的标注数据进行模型的监督训练，且现有的图像语义分割算法在泛化能力不足的问题，严重阻碍了分割算法在实际场景中的应用。

针对上述问题，基于多种学习方式的语义分割相继被提出，包括半监督、弱监督、小样本、零样本语义分割等等。其中小样本语义分割旨在利用少量的标注数据对未知类别实现预测，并且在训练和测试过程中都有少量标注数据可供学习，比较符合实际应用场景。然而可学习样本量较少，对模型的学习和泛化能力提出了很高的要求，小样本语义分割目前仍是一个充满挑战性的任务。

在目前的小样本分割研究中，研究人员利用元学习[1]，原型网络[2]等方法实现了较为可靠的分割结果。但目前的研究仍受限于两个因素，即支持集和查询集之间的类内差异，以及训练集和测试集之间的类间差异。较大的数据差异导致模型的类别表征能力，以及由支持集向查询集，训练集向测试集的知识迁移能力下降，削弱了分割性能。

本文提出了一种基于多尺度先验信息融合的小样本语义分割方法，通过利用高级特征提取多尺度先验掩码，并添加尺度间信息交互和加权组合，为模型学习补充类别相关信息，降低类内差异的影响，在主流数据集上取得了更好的分割性能。

2. 小样本语义分割算法基础

2.1. 元学习

为了提高小样本限制条件下模型的泛化能力，现有小样本语义分割算法通常采用元学习范式。元学习(Meta-Learning)也被称为“学会学习(Learning-to-Learn)”。元学习本质上是双层优化问题(Bilevel Op-

timization Problem), 其中一个优化问题嵌套在另一个优化问题中。其核心思想是在多个学习阶段改进学习算法的过程。

元学习有两个层面的含义。第一层指的是让模型学会如何学习, 从而具备分析和解决问题的能力, 提高在任务处理方面的能力。第二层指的是元学习旨在使机器学习模型更好地适应新领域, 使其能够成功处理具有较大差异性的新任务。因此元学习非常适合解决小样本语义分割问题。

在任务上, 与一般深度学习过程类似, 元学习主要分为元训练和元测试两个阶段。元学习的基本单位是样本集(Episode)。元训练集、元验证集和元测试集都是由抽样样本组成的样本集集合。在元训练阶段, 利用元训练集和元验证集去学习类别变化的情况下模型的泛化能力。在元测试阶段, 面对全新的类别, 无需变动已有的模型, 只需要通过一部或者少数几步训练, 就可以完成需求。

2.2. 任务设定

小样本语义分割和一般的语义分割的任务设定不同, 一般的语义分割任务是分割出多种类别的物体。而当前小样本语义分割任务的研究重点是二类分割, 只分割前景类别和背景类别。

基于元学习的小样本分割算法设计中, 引入大量训练类别的带标签图片样本构造训练任务集。在每个训练任务上通过训练阶段和测试阶段用带标签图片来训练和测试模型。这里的测试图片和训练图片(也被称为未见过类和已见过类)来自同一数据集 C 。按照未见过类和已见过类划分数据集 C 为两个数据集, 分别为 C_{unseen} 和 C_{seen} , 且二者交集为空。

与小样本图像分类任务类似, 小样本语义分割任务采用 N-way K-shot 设定, 即在数据集中采样选取 N 个类别, 每个类别包含 K 个样本的样本集来作为一个基本单位。来自 C_{unseen} 和 C_{seen} 的采样分别作为训练集 D_{train} 和测试集 D_{test} 。其中, $D_{train} = \{(S_i, Q_i)\}^{N_{train}}$, $D_{test} = \{(S_i, Q_i)\}^{N_{test}}$ 和 $S, Q, N_{train}, N_{test}$ 分别代指支持集、查询集、训练阶段和测试阶段的需要抽取的批次数。每一个支持集查询集对 (S_i, Q_i) 作为一个基本的分割样本集, 每个 S_i 中包含 K 个带标注的图片, 而每个 Q_i 只包含一个带标注图片。

在训练过程中, 模型利用支持集中的信息来指导查询集上的分割任务, 多次不同的采样使得训练过程中包含多个不同的分割任务, 因此可以泛化到其他分割任务中。模型训练完成后, 在测试集中进行类似的过程, 根据测试集对其进行微调, 使得模型可以完成新类中的分割任务。

3. 基于多尺度先验融合的小样本语义分割算法

3.1. 网络架构

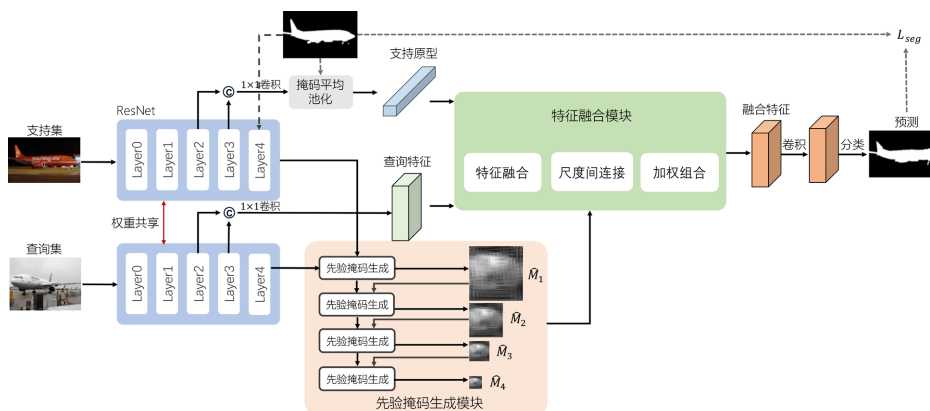


Figure 1. Overview of the few-shot semantic segmentation model proposed in this paper
图 1. 本文提出的小样本语义分割模型架构

本文模型主要包括特征提取骨干网络，掩码平均池化，多尺度权重自适应特征融合模块，多尺度先验掩码生成模块，分类器。

其中，模型以数据划分的样本集作为元训练或元测试的基本单位，样本集中包含的支持集图片和查询集图片分别使用基于特征提取骨干网络进行特征提取，并将中层特征保留；支持集图片的中层特征经过掩码平均池化提取原型，用以后续指导分割；支持集图片的高层特征作为多尺度先验掩码生成模块的输入，计算其与查询集图片高层特征的相似度，得到多个尺度的先验掩码；最后，查询集图片中层特征、支持集原型、多尺度先验掩码共同送入多尺度特征融合模块，综合信息；最后使用分类器对融合特征进行分类，得到查询图片预测，并与查询集掩码计算交叉熵损失。整体网络架构如图 1 所示。

在特征提取上，本研究使用基于 ImageNet [3] 预训练的 ResNet 骨干网络提取支持集和查询集图片特征。参考 CANet [4] 和 PANet [5] 的设置，在元训练和元测试阶段，本研究将使用 ImageNet 预训练后的 ResNet 骨干网络的参数冻结，使用权重共享的网络对支持集和查询集进行特征提取，从而避免模型过拟合，偏向于基类。对于 K-shot 场景，本方法将多张支持特征求和取平均值作为输出特征。

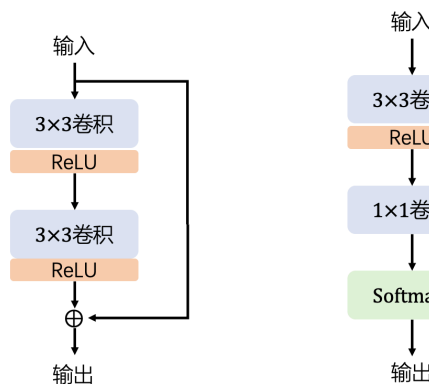


Figure 2. Convolutional classifier network structure
图 2. 卷积分类器网络结构

首先，对于支持图片 I_s 和查询图片 I_q 进行中层特征提取，获得支持特征 f_s^m 和查询特征 f_q^m 。骨干网络为使用经过 ImageNet 预训练的 ResNet 骨干网络进行中层特征提取。该中层特征是将 ResNet 的 layer2 和 layer3 的输出特征按通道维度串联得到的。

使用掩码平均池化获得支持集前景原型后，由多尺度先验掩码生成模块利用支持集和查询集的高层特征生成多尺度先验掩码，然后共同送入多尺度特征融合模块，获得融合特征。

经过上述过程处理后，模型得到了融合查询集特征、支持集特征、多尺度类别信息的融合特征，需要最后对该特征进行像素级分类，得到最终的分割结果。

本研究中选择以使用卷积作为分类工具。该卷积分类器由一个卷积块和一个分类头组成。卷积块由两个 3×3 卷积组成，并添加了残差连接。分类头依次由 3×3 卷积、 1×1 卷积、Softmax 激活函数组成。该卷积分类器如图 2 所示。

使用卷积分类器对特征进行分类后，得到最终的查询集预测图 M_{pre} 。将该预测图 M_{pre} 与数据集中的查询集掩码 M_q 计算交叉熵损失，作为模型训练的主要损失函数。

3.2. 多尺度先验掩码生成

出现类内差异的很大原因之一是同一类物体的大小尺寸差异，为了模型在类内差异较大的场景下的性能，需要提高模型对于多尺度信息的学习能力。此外，使用 CNN 网络提取的高层特征能够更好的反应

物体的类别信息，因此为模型增加对高层特征的学习能够增强类相关信息，增强对同类物体的识别能力。基于上述问题，本章节提出了多尺度先验掩码生成模块。该模块的流程如图 3 所示。

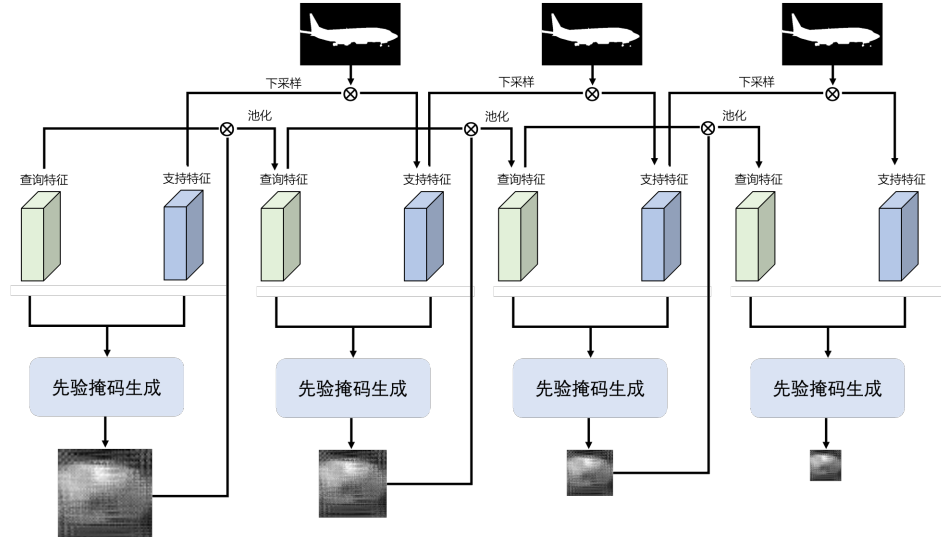


Figure 3. Multi-scale prior mask generation module structure
图 3. 多尺度先验掩码生成模块结构

具体而言，为了增强模型对于前景类别的信息的辨别能力，本研究选择在 ResNet 的最后一层前为支持集特征增加掩码过滤，其公式表达如下：

$$f_s^h = R^4 \left(f_s^m \otimes \mathcal{J}(M_s, f_s^m) \right) \quad (1)$$

其中， R^4 为 ResNet 网络的 layer-4， $\mathcal{J}(M_s, f_s^m)$ 通过插值将 M_s 上采样或下采样至与 f_s^m 相同的大小。

得到支持集输入特征 f_s^h 和查询集输入特征 f_q^h 之后，需要计算二者的余弦相似度作为 f_q^h 的先验掩码。此时， $f_s^h, f_q^h \in \mathcal{R}^{C \times H_1 \times W_1}$ ，将二者按照像素划分为 $H_1 \times W_1$ 个向量，则对于向量 x^q 和 x^s 计算相似度后取平均值，其公式表达如下：

$$c^q = \text{mean}_{s \in \{1, 2, \dots, H_1 \times W_1\}} \cos(x^q, x^s) \quad (2)$$

然后将相似度组合拼接为相似度图，其公式表达如下：

$$\hat{C}_1 = [c^1, c^2, \dots, c^{H_1 \times W_1}] \in \mathcal{R}^{H_1 \times W_1 \times 1} \quad (3)$$

其中 $c^q \in [-1, 1]$ ，为了使先验掩码的像素数值归一化，使用 min-max 标准化对概率图进行处理，使其像素数值在 [0, 1] 范围内，其公式表达如下：

$$\hat{M}_1 = \frac{\hat{C}_1 - \min(\hat{C}_1)}{\max(\hat{C}_1) - \min(\hat{C}_1) + \epsilon} \quad (4)$$

此处 ϵ 为一个极小值，防止分母为 0 导致计算失败，在本实验中 $\epsilon = 1e-7$ 。

得到第一层的先验掩码后，为了拓展到下一个尺度的先验掩码生成，首先对上一层的查询集输入特征使用池化操作进行下采样。同时为了增强不同尺度之间的信息交互，减少池化操作带来的信息损失，本研究在池化操作之前使用上一层的先验掩码对查询集特征进行过滤，其公式表达如下：

$$f_q^h = \mathcal{F}_{pool} (f_q^h \otimes \hat{M}_1) \in \mathcal{R}^{C \times H_2 \times W_2} \quad (5)$$

获得下一层的查询集输入特征之后,使用双线性下采样将上一层的支持集输入特征变换为相同尺寸,与第一层的操作类似,同样为支持集特征进行掩码过滤,其公式表达如下:

$$f_s^h = f_s^h \otimes \mathcal{J}(M_s, f_s^h) \in \mathcal{R}^{C \times H_2 \times W_2} \quad (6)$$

得到第二层的输入特征后,采用相同的操作进行后三层的先验掩码生成,合并第一层的结果后,输出最终的多尺度先验掩码 $\hat{M} = \{\hat{M}_i\}_{i=1}^4$ 。

3.3. 多尺度权重自适应特征融合

为了进一步增强模型对于多尺度物体的学习能力,融合所得特征以及先验掩码,直接利用原型进行度量匹配不能满足小样本语义分割的需求。在提控多尺度空间信息上,为模型添加 ASPP (Atrous Spatial Pyramid Pooling)模块[6]是一个合理的选择,然而直接使用这些模块处理特征存在问题: 1) 这些模块可以为合并后的融合特征提供空间信息,但无法在每个尺度内进行特定的细化过程。2) 没有考虑到不同尺度之间的层次关系。3) 没有考虑到不同尺度信息的权重问题。

为了解决上述这些问题,本研究借鉴 PFENet [7]的特征丰富模块,融合多尺度先验掩码,并为每个尺度增添自适应权重,提出了多尺度权重自适应特征融合模块。模块的整体架构如图 4 所示。

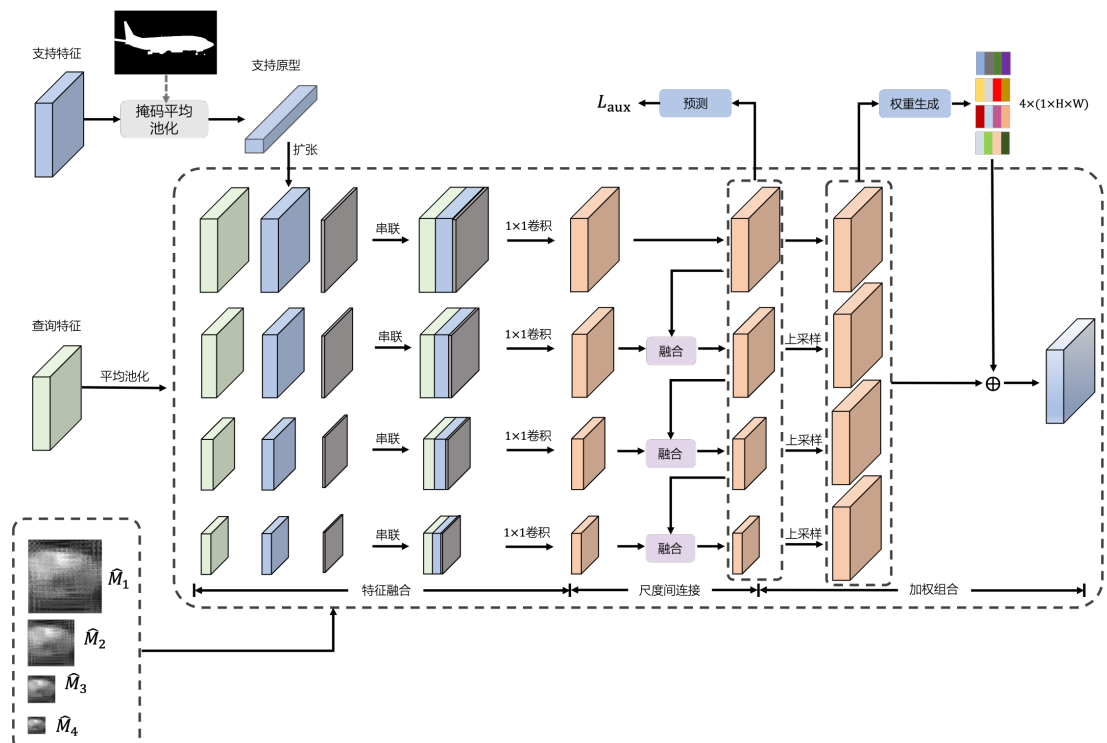


Figure 4. Multi-scale weighted adaptive feature fusion module structure.

图 4. 多尺度权重自适应特征融合模块结构

具体而言,本模块的算法流程如下:

1) 特征融合

为了进行多尺度的信息融合,并对每一个尺度信息进行细化和加权,首先需要对每一个尺度的特征

分别进行融合。对于查询特征 $f_q \in \mathcal{R}^{C \times H \times W}$ 首先使用平均池化生成对应尺度的查询特征 $\hat{f}_q = \{\hat{f}_q^i\}_{i=1}^4$ ，其公式表达如下：

$$\hat{f}_q^i = \mathcal{F}_{pool}(f_q) \in \mathcal{R}^{C \times H_i \times W_i}, i \in [1, 4] \quad (7)$$

其中 H_i 和 W_i 为第 i 层多尺度特征的高度和宽度。

对于支持集前景原型 $p \in \mathcal{R}^{C \times 1 \times 1}$ 将其扩张为对应的尺度大小 $\hat{p} = \{\hat{p}^i\}_{i=1}^4$ ，其中 $\hat{p}^i \in \mathcal{R}^{C \times H_i \times W_i}$ 。

将拓展至四个尺度的查询特征 \hat{f}_q 、支持集前景原型 \hat{p} ，以及由多尺度先验掩码生成模块提取的多尺度先验掩码 \hat{M} 三者按照通道维度串联，并使用 1×1 卷积进行特征融合，其公式表达如下：

$$f_{mix} = \mathcal{F}_{1 \times 1}(\text{concat}(\hat{f}_q, \hat{p}, \hat{M})) \quad (8)$$

其中 $f_{mix} = \{f_{mix}^i\}_{i=1}^4$ 为融合后的多尺度特征，其中 $f_{mix}^i \in \mathcal{R}^{C \times H_i \times W_i}$ ， concat 为按通道维度串联。

通过扩展特征到多个尺度，这种自上而下的合成路径能够自适应地将信息从精细特征传递到粗略特征，可以在特征融合模块中更好的建立层次关系，减少不同尺寸物体带来的类间差异影响。

2) 尺度间连接

特征融合模块将多个尺度的特征信息水平的融合在一起，为了增加尺度间信息的交互能力，即在垂直方向上进行信息融合，需要进行尺度间的信息连接。

首先将第一层融合 f_{mix}^1 特征使用两层 3×3 卷积进行转换，其公式表达如下：

$$\hat{f}_{mix}^1 = \text{conv}_\alpha(f_{mix}^1) \in \mathcal{R}^{C \times H_1 \times W_1} \quad (9)$$

其中：

$$\text{conv}_\alpha(x) = \text{ReLU}(F_{3 \times 3}(\text{ReLU}(F_{3 \times 3}(x)))) \quad (10)$$

其中 $F_{3 \times 3}$ 为 3×3 卷积， \hat{f}_{mix}^1 为跨尺度融合的中间结果。

接下来将后续的每一尺度的融合特征与上一层次的特征进行融合，其公式表达如下：

$$\hat{f}_{mix}^i = \text{conv}_\alpha(\text{concat}(f_{mix}^i, \hat{f}_{mix}^{i-1})), i \in [2, 4] \quad (11)$$

将尺度信息由上自下完成后，为每一个尺度的融合特征添加残差连接，其公式表达如下：

$$\ddot{f}_{mix}^i = f_{mix}^i + \hat{f}_{mix}^i \in \mathcal{R}^{C \times H_i \times W_i} \quad (12)$$

其中， $\ddot{f}_{mix}^i (i \in [1, 4])$ 为该模块第 i 个尺度的最终输出结果。

融合后的特征可以看作代表了各个尺度的完整特征信息，对这些特征进行分类后得到的预测图与数据集掩码进行对比，设置辅助损失函数，可以更好地约束模型在各个尺度的特征生成质量。在这里，本文对每个尺度使用与前文相同网络架构的分类头进行预测，其公式表达如下：

$$L_{aux} = \frac{1}{4} \sum_{i=1}^4 \ell(\mathcal{I}(\text{cov}_{cls}(\ddot{f}_{mix}^i), M_q), M_q) \quad (13)$$

3) 加权组合

在得到上一个模块的多尺度特征 \ddot{f}_{mix}^i 后，为了进行后续融合和分割，需要首先将各尺度特征统一至查询集图片的大小。经过上采样后得到融合特征 $\hat{f}_{mix} = \{\hat{f}_{mix}^i \in \mathcal{R}^{C \times H \times W}\}_{i=1}^4$ 。本研究认为，对于不同的物体，不同尺度的特征的重要性不同，因此需要对每个尺度的特征赋予自适应权重，作为一种注意力机制，使

模型更加关注于重要的信息。因此，为了生成权重，首先将各层特征沿通道维度串联，其公式表达如下：

$$f_{all} = \text{concat} \left(\left\{ \hat{f}_{mix}^i \right\}_{i=1}^4 \right) \in \mathcal{R}^{4C \times H \times W} \quad (14)$$

接下来使用 1×1 卷积降低通道数，融合各层特征，其公式表达如下：

$$\dot{f}_{all} = \text{ReLU} \left(\mathcal{F}_{1 \times 1} (f_{all}) \right) \in \mathcal{R}^{C \times H \times W} \quad (15)$$

之后使用 3×3 卷积进一步降低通道数，并使用 sigmoid 函数进行标准化处理，为每一层特征生成权重矩阵，其公式表达如下：

$$W = \text{sigmoid} \left(\mathcal{F}_{3 \times 3} (\dot{f}_{all}) \right) \in \mathcal{R}^{4 \times H \times W} \quad (16)$$

得到的权重矩阵的每一个通道都代表着对其中一个维度的权重图，我们将其划分为四张权重图 $W = \{W_i \in \mathcal{R}^{1 \times H \times W}\}_{i=1}^4$ ，使用矩阵乘法依次为每一层特征加权，最后直接将各层特征相加，得到最终的输出结果，即将查询集特征，支持集前景原型以及代表了类别信息的查询集多尺度先验掩码三者融合，并在尺度内和尺度间丰富了特征信息后的融合特征。

4. 实验结果

4.1. 对比实验

为了证明算法的改进效果，本次实验中在 PASCAL-5i 和 COCO-20i 数据集上分别完成训练和测试，通过实际分割将改进前后的模型效果进行对比。在本实验中，本文采用 ResNet-50 [8] 作为特征提取模块。本实验使用 SGD 作为优化器，动量设置为 0.9，权重衰减设置为 $1e-4$ 。对于 PASCAL-5i，本实验对每个子集训练 200 个 epoch，初始学习率为 $5e-3$ ，批处理大小设置为 4。对于 COCO-20i，本实验对每个子集训练 50 个 epoch，初始学习率为 $5e-3$ ，批处理大小设置为 8。

为了确定基线模型，本实验将所提出的多尺度先验掩码模块替换为单一尺度先验掩码生成，并将该掩码通过下采样方式插入至多尺度权重自适应特征融合模块中的特征融合模块，最后将多尺度权重自适应特征融合模块中特征融合模块意外的部分替换为 ASPP 模块[4]以获取多尺度信息。该基线模型的基本架构如图 5 所示。

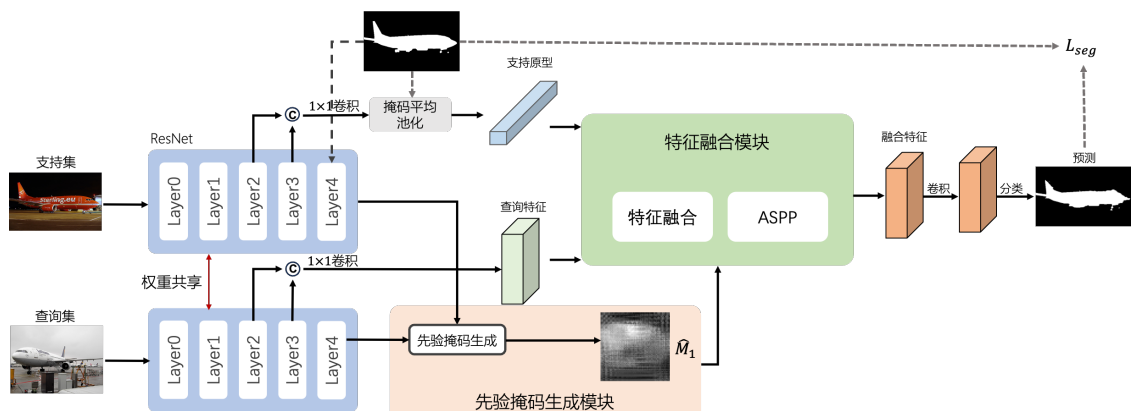


Figure 5. Baseline model structure using the ASPP module

图 5. 使用 ASPP 模块的基线模型架构

为了量化所提出模型的改进效果，本章节首先在 PASCAL-5ⁱ 数据集 1-shot 场景下完成了训练和测试，

使用 ResNet-50 作为特征提取骨干网络, 使用 mIoU 作为评价指标, 并与当今主流方法相对比。具体的实验数据如表 1 所示, 其中“-”表示原论文未给出数据。

Table 1. Comparative experiments on PASCAL-5ⁱ

表 1. PASCAL-5ⁱ 数据集上的分割对比实验

方法	Split-0	Split-1	Split-2	Split-3	平均
CANet [4]	52.5	65.9	51.3	51.9	55.4
CRNet [9]	-	-	-	-	55.7
PPNet [10]	48.58	60.58	55.71	46.47	52.84
CWT [11]	56.3	62.0	59.9	47.2	56.4
BRiNet [12]	56.54	67.20	51.56	53.02	57.08
PFENet [7]	61.7	69.5	55.4	56.3	60.8
ASGNet [13]	58.84	67.86	56.79	53.66	59.29
基线方法	56.77	65.78	51.60	53.14	56.82
所提方法	62.19	69.03	56.16	57.39	61.19

根据本文还在 COCO-20ⁱ 数据集上完成测试, 相比于 PASCAL-5ⁱ 数据集该数据集包含的类别更多, 更具有挑战性。具体实验结果如表 2 所示。

Table 2. Comparative experiments on COCO-20ⁱ

表 2. COCO-20ⁱ 数据集上的分割对比实验

方法	Split-0	Split-1	Split-2	Split-3	平均
RPMMs [14]	29.53	36.82	28.94	27.02	30.58
HFA [15]	28.65	36.02	30.16	33.28	32.03
RePRI [16]	32.00	38.70	32.70	33.10	34.13
CANet [4]	34.25	34.44	30.87	31.21	32.69
CWT [11]	32.2	36.0	31.6	31.6	32.9
BRiNet [12]	32.88	36.20	37.44	30.93	34.36
ASGNet [13]	-	-	-	-	34.56
基线方法	33.45	36.40	33.12	32.76	33.93
所提方法	35.34	38.07	36.45	34.54	36.10

4.2. 消融实验

为了证明本文提出的各个模块的实际作用, 本章节针对模块的每个组成部分在 PASCAL-5ⁱ 数据集的第 0 子集上进行小样本语义分割测试, 对比不同设置以及功能应用的情况下对分割性能的影响。

对于基于多尺度先验信息合成的小样本语义分割模型, 本文首先对多尺度先验掩码生成模块和多尺度权重自适应特征融合的使用效果进行消融实验。具体的实验数据如表 3 所示, 其中 ASPP [4] 模块为基线方法中替代多尺度特征融合模块的方法并去除辅助损失。不使用多尺度掩码生成模块时, 使用单尺度掩码生成进行替代。

Table 3. Ablation experiments on each module**表 3.** 各模块消融实验

方法	多尺度掩码	多尺度特征融合	ASPP	mIoU
基线方法			✓	56.77
		✓		61.79
所提方法	✓		✓	61.63
	✓	✓		62.19

在先验掩码生成模块上, 本文添加尺度间连接, 在多尺度权重自适应特征融合模块上, 本文添加了尺度纵向和横向信息交互, 并为各尺度生成权重。为了验证这些信息交互的作用, 本文为其进行了消融实验。具体的试验结果如表 4 所示。其中, 去除尺度权重时, 本文选择用直接相加的方式进行最终特征融合。从实验结果可以看出, 同时使用多种尺度交互, 并为其赋予权重时分割效果最好。

Table 4. Ablation experiments on inter-scale information interaction**表 4.** 尺度间信息交互消融实验

方法	尺度间连接	尺度横、纵向交互	尺度权重	mIoU
基线方法				57.26
	✓			60.06
		✓		55.27
所提方法	✓	✓		61.00
	✓	✓	✓	62.19

对于尺度间连接模块中使用的辅助损失函数, 本文也进行了消融实验。实验结果如表 5 所示。从实验结果中可以看出, 同时使用分割损失和辅助损失时的分割效果最好。

Table 5. Ablation experiments on loss functions**表 5.** 损失函数消融实验

方法	分割损失	辅助损失	mIoU
基线方法	✓		60.23
所提方法	✓	✓	62.19

5. 总结

为了降低类内差异对小样本语义分割模型的影响, 本文在基于原型提取和卷积分类的基础架构上提出了两个模块。多尺度先验掩码生成模块生成多尺度先验掩码, 代表类别相关信息, 增强类内的信息交互。为了融合多尺度特征, 多尺度权重自适应特征融合模块为各个尺度进行加权融合, 并增加尺度间和尺度内信息交互。在多个对比实验和消融实验中, 证明了本文提出方法的有效性。

参考文献

- [1] Wu, Z., Shi, X., Lin, G. and Cai, J. (2021) Learning Meta-Class Memory for Few-Shot Semantic Segmentation. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, 10-17 October 2021, 517-526. <https://doi.org/10.1109/ICCV48922.2021.00056>

-
- [2] Zhang, X., Wei, Y., Yang, Y. and Huang, T.S. (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>
- [3] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., *et al.* (2015) ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, **115**, 211-252. <https://doi.org/10.1007/s11263-015-0816-y>
- [4] Zhang, C., Lin, G., Liu, F., Yao, R. and Shen, C. (2019) CANet: Class-Agnostic Segmentation Networks with Iterative Refinement and Attentive Few-Shot Learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15-20 June 2019, 5217-5226. <https://doi.org/10.1109/CVPR.2019.00536>
- [5] Wang, K., Liew, J.H., Zou, Y., Zhou, D. and Feng, J. (2019) PANet: Few-Shot Image Semantic Segmentation with Prototype Alignment. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, 27 October 2019 - 2 November 2019, 9197-9206. <https://doi.org/10.1109/ICCV.2019.00929>
- [6] Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L. (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>
- [7] Tian, Z., Zhao, H., Shu, M., Yang, Z., Li, R. and Jia, J. (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>
- [8] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. *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>
- [9] Liu, W., Zhang, C., Lin, G. and Liu, F. (2020) CRNet: Cross-Reference Networks for Few-Shot Segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 4165-4173. <https://doi.org/10.1109/CVPR42600.2020.00422>
- [10] Liu, Y., Zhang, X., Zhang, S. and He, X. (2020) Part-Aware Prototype Network for Few-Shot Semantic Segmentation. *Computer Vision—ECCV 2020: 16th European Conference*, Glasgow, 23-28 August 2020, 142-158. https://doi.org/10.1007/978-3-030-58545-7_9
- [11] Lu, Z., He, S., Zhu, X., Zhang, L., Song, Y.Z. and Xiang, T. (2021) Simpler Is Better: Few-Shot Semantic Segmentation with Classifier Weight Transformer. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Montreal, 10-17 October 2021, 8741-8750. <https://doi.org/10.1109/ICCV48922.2021.00862>
- [12] Yang, X., Wang, B., Chen, K., Zhou, X., Yi, S., Ouyang, W. and Zhou, L. (2020) Brinet: Towards Bridging the Intra-Class and Inter-Class Gaps One-Shot Segmentation. arXiv: 2008.06226.
- [13] Li, G., Jampani, V., Sevilla-Lara, L., Sun, D., Kim, J. and Kim, J. (2021) Adaptive Prototype Learning and Allocation for Few-Shot Segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 8334-8343. <https://doi.org/10.1109/CVPR46437.2021.00823>
- [14] Yang, B., Liu, C., Li, B., Jiao, J. and Ye, Q. (2020) Prototype Mixture Models for Few-Shot Semantic Segmentation. *Computer Vision—ECCV 2020: 16th European Conference*, Glasgow, 23-28 August 2020, 763-778. https://doi.org/10.1007/978-3-030-58598-3_45
- [15] Liu, B., Jiao, J. and Ye, Q. (2021) Harmonic Feature Activation for Few-Shot Semantic Segmentation. *IEEE Transactions on Image Processing*, **30**, 3142-3153. <https://doi.org/10.1109/TIP.2021.3058512>
- [16] Liu, Y., Liu, N., Cao, Q., Yao, X., Han, J. and Shao, L. (2022) Learning Non-Target Knowledge for Few-Shot Semantic Segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, 18-24 June 2022, 11573-11582. <https://doi.org/10.1109/CVPR52688.2022.01128>