引用格式：贺坤, 吴颖, 郁湧. 双特征引导超相关性的少样本分割网络 [J]. 电子科技大学学报, 2026, 55(3): 447-454.
HE K, WU Y, YU Y. Dual-feature guided hypercorrelation few-shot segmentation network[J]. Journal of University of Electronic Science and Technology of China, 2026, 55(3): 447-454.

双特征引导超相关性的少样本分割网络

贺 坤，吴 颖，郁 湧*

（云南大学软件学院，昆明 650504）

摘要：少样本语义分割旨在仅有少数支持图像注释样本情况下学习从给定类的查询图像中分割目标对象，近年来基于特征匹配的方法面对这一任务取得了显著成功。然而，多匹配关系带来更多的类间和类内噪声，为缓解高维卷积的运算压力而对权重进行稀疏化也导致了特征相关性的细粒度匹配精度损失。为了缓解上述问题，提出了双特征引导超相关性的少样本分割网络（DFGHNet）。DFGHNet 采用更加高效的 4D 卷积核以减少权重稀疏化导致的特征细粒度匹配精度损失，同时引入双特征掩码策略和采用不含可学习参数的非局部均值特征映射模块，在学习到的匹配模式进行相关性引导。在数据集 PASCAL-5^i、COCO-20^i 的标准少样本分割基准测试中，该方法的最大精度提升分别为 4.8% 和 6.5%，验证了其有效性。

关键词：元学习；少样本学习；语义分割；少样本语义分割
中图分类号：TP181 **文献标志码：**A DOI: 10.12178/1001-0548.2024270

Dual-feature guided hypercorrelation few-shot segmentation network

HE Kun, WU Ying, and YU Yong*

(School of Software, Yunnan University, Kunming 650504, China)

Abstract: Few-shot semantic segmentation aims to learn to segment target objects from query images of a given category when only a few samples support image annotation. Recently, matching-based methods establish correlation matching of dense features by applying 4D convolution and introducing background correlation. However, multiple matching relationships bring more inter-class noise and intra-class noise, and the sparse weights to alleviate the computational pressure of high-dimensional convolution also led to the loss of fine-grained matching accuracy of feature correlation. To alleviate the above problems, a dual-feature guided hypercorrelation few-shot segmentation network (DFGHNet) is proposed. DFGHNet uses a more efficient 4D convolution kernel to reduce the loss of feature fine-grained matching accuracy caused by weight sparsification. At the same time, DFGHNet introduces a dual feature mask strategy and adopts a non-local mean feature mapping module without learnable parameters to guide the correlation in the learned matching pattern. In the standard few-shot segmentation benchmarks of PASCAL-5^i and COCO-20^i datasets, the performance of the proposed method achieves the maximum accuracy improvement of 4.8% and 6.5% respectively, verifying the effectiveness of this method.

Key words: meta learning; few-shot learning; semantic segmentation; few-shot semantic segmentation

深度卷积神经网络（deep convolutional neural networks）[1-3] 的快速发展，极大地推动了计算机视觉领域的进步，如图像分类、目标检测、实例分割和语义分割等任务。然而，这些视觉任务的成功很大程度上依赖于庞大的数据集以及高质量的标签数据。获得这些高质量的标签数据，通常是一项很耗费人力和时间的工作。

相比之下，人类的视觉系统能在极其有限的样本容量下实现对新物体的外观辨别，受到人类快速学习能力的启发，研究人员希望机器学习模型能够在处理大量数据的某些类别后，对新类别仅需少量样本即可迅速学习泛化能力，这正是少样本学习（few-shot learning）[4-5] 旨在解决的问题。少样本学习是基于元学习模型[6-7] 的应用，元学习的核心思想是“学习如何学习”，即通过训练模型来掌握适应新任务的能力。元学习方法的出现，为计算机视觉任务中的少样本学习提供了新的解决方案。在元学习框架下，模型能够通过以不同已知类作为样

收稿日期：2024−09−28
作者简介：贺坤，主要从事少样本语义分割方面的研究。
*通信作者 E-mail: yuy1219@163.com

本来提升泛化到处理未知类的能力，使得它在未知类别下的模型处理依然能够表现出色。

根据视觉任务的不同，语义分割对数据量的依赖也不同，与图像分类和检测相比，它是一项更精细的任务。语义分割旨在将给定图像的像素映射到预定义的目标类别，需要给出图像中每个像素的分类概率，如果特征图太小，则会丢失大量信息。全卷积网络（fully convolutional networks, FCN）[8] 提出将不同层次的特征结合起来，逐渐获得相对精细的特征。另一种更直观的方法是通过编码器–解码器结构提取和恢复特征，U-Net[9] 利用结构中的残差结构引入低级特征的空间信息来提高分割精度。这些工作在语义分割任务中验证了其有效性，但它们均是依赖于大量标注数据才能取得优异的性能，在标注样本较少的情况下模型难以推广到不可知类。

少样本语义分割[10] 是少样本学习在语义分割领域的应用，旨在解决仅有少量标注样本的情况下实现对未知类的分割。目前的少样本语义分割方法主要是基于度量学习的元学习方法，度量学习的核心思想是通过学习一个度量空间，使得在这个空间中，相同类别的样本距离更近，而不同类别的样本距离更远。目前基于度量学习的少样本语义分割主流方法分为基于匹配网络[11-12] 和基于原型网络[13]。

基于原型网络的核心思想是通过掩码平均池化的方式，生成各类目标的代表性原型向量，这些原型向量可以被视为原型表示。具体而言，该方法通过掩码将背景像素滤除，以减少背景对目标像素的干扰，从而更加准确地提取目标区域的特征。接下来，通过计算查询图像中待分类像素与已知各类原型表示之间的欧氏距离，完成对查询图像中每个像素的分类，从而实现语义分割。由于平均池化的过程将特征信息压缩为单个向量，导致特征间的细节差异和局部特征相关性无法得到充分表达。

基于匹配网络的少样本语义分割方法提出进行多对多的特征匹配建立成对的特征相关性，通过建立特征相关性来执行支持–查询语义对应的集中式思想，该模式可以捕获更丰富的上下文相关性，文献 [14] 提出的语义对应方法表明，有效利用分布在所有中间层上的不同级别的卷积特征明显有利于匹配精度，趋势是在密集特征匹配上采用 4D 卷积，通过分析 4D 空间中的局部模式来识别空间一致的匹配。使用多级特征和使用 4D 卷积进行关系模式分析是视觉对应领域广泛采用的两种技术。HSNet[15] 采用了视觉对应中两种最有影响力的方法—多级特征和 4D 卷积来解决少样本分割问题，从而实现从粗到细聚合特征相关性。随后，FECANet[16] 提出引入背景相关性的支持，进一步丰富了多尺度信息相关性，以上工作都为本文研究提供了启发。

本文提出了双特征引导超相关性的少样本分割网络（dual-feature guided hypercorrelation few-shot segmentation network, DFGHNet），使用双特征掩码策略和无可学习参数的非局部均值映射模块对类间相似性和类内差异噪声有效抑制，同时降低模型复杂度以提升泛化性能；使用多级池化策略对 4D 卷积编码器进行参数稀疏化以提升多尺度相关的细粒度匹配精度。

1 网络整体架构

1.1 少样本分割问题定义

少样本语义分割的目标是仅在给出少量带注释示例的情况下执行分割。 为了避免由于训练数据不足而导致过度拟合的风险，本文采用广泛使用的元学习方法，称为情景训练，具体如图 1 所示。

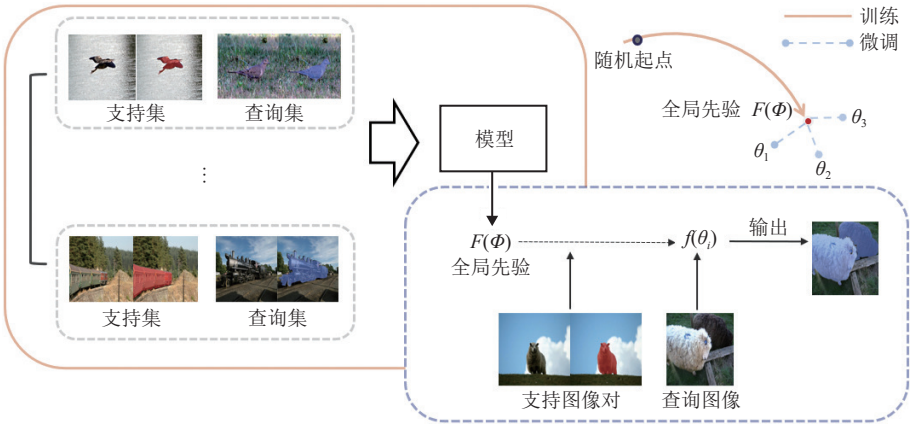


图 1 情景训练

将训练集和测试集表示为$\mathcal{D}_{\text{train}}$和$\mathcal{D}_{\text{test}}$，它们包含的样本类别互不相交，以确保对模型泛化能力的评估。两个集合都包含多个训练片段，每个片段由支持集$\mathcal{S}=(I^s,M^s)$和查询集$\mathcal{Q}=(I^q,M^q)$组成，其中I^*和M^*分别是图像及其相应的真实掩码标签。

在训练期间，模型迭代地从$\mathcal{D}_{\text{train}}$中采样一个片段，以学习从评估片段$(I^s,M^s,I^q)$到查询掩码$M^q$的映射。一旦模型训练完成即可作为新类泛化的全局先验，且无需进一步优化。从$\mathcal{D}_{\text{test}}$中随机采样$k$组类别$\theta_i$的评估片段进行模型全局先验的微调，得到对类别$\theta_i$微调后的全局先验$f(\theta_i)$来预测查询掩码，称为$k$-shot 评估。在后面的评估中，本文采用$k$=1 以及$k$=5 的设置进行模型评估，即 1-shot 和 5-shot，这也是目前主流方法采用的设置。

1.2　网络模型架构

少样本语义分割模型是在已知类别的标记数据上训练的，此外，当前主流方法通常采用具有大量可学习参数的模块来最大化利用有限的支持信息。一方面，这使得训练出的模型在少样本的情况下容易使得全局先验拟合到已知类，而致使类之间相似噪声过大从而影响新类的泛化。另一方面，对于利用 4D 卷积核捕捉多匹配关系的模型而言，大量的可学习参数会带来更大的运算压力。

为了缓解以上问题，本文提出了双特征引导超相关性的少样本分割网络（DFGHNet），模型结构如图 2 所示。

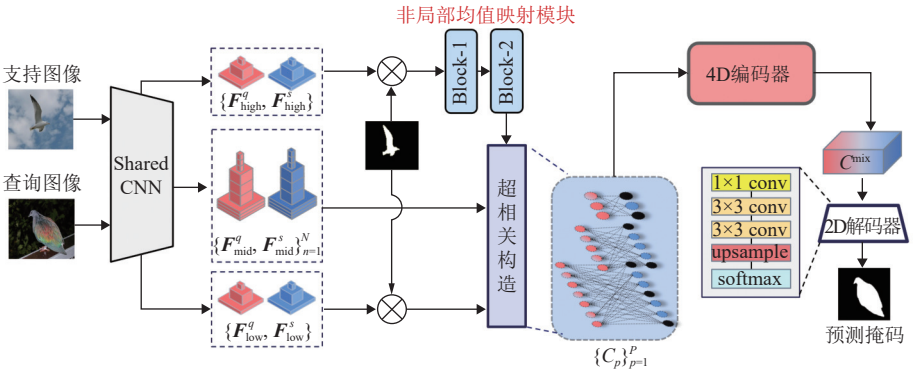


图 2　模型架构图

该模型在骨干网络提取多个中间层级的特征作为相关性构建的主要对象，并提取少量富含语义信息的高级特征和富含空间信息的低级特征，然后对各个层级大小相同的特征进行聚合，再通过 4D 编码器来引导中层密集特征的前景和背景的特征相关性捕捉，最后通过 2D 解码器进行特征相关解码得到查询图像的预测掩码。

1.3　双特征掩码策略

为了更好地捕捉上下文相关性，本文将提取出的中层特征作为主要特征相关性来源，并且不使用支持掩码策略以捕捉特征全局相关。即给定一对查询和支持图像$I^q,I^s\in\mathbb{R}^{3\times H\times W}$，通过主干网络生成$N$对中间特征图$\left\{\left(\boldsymbol{F}^q_{\text{mid}},\boldsymbol{F}^s_{\text{mid}}\right)\right\}_{n=1}^{N}$以及单对的低级特征$\left\{\boldsymbol{F}^q_{\text{low}},\boldsymbol{F}^s_{\text{low}}\right\}$和高级特征$\left\{\boldsymbol{F}^q_{\text{high}},\boldsymbol{F}^s_{\text{high}}\right\}$，共计$L$对特征。

本文使用支持掩码$\boldsymbol{M}^s\in\{0,1\}^{H\times W}$掩蔽低级和高级特征的支持特征图$\boldsymbol{F}^s_{\text{low}}$和$\boldsymbol{F}^s_{\text{high}}$以丢弃不相关的激活得到$\hat{\boldsymbol{F}}^s_{\text{low}}$和$\hat{\boldsymbol{F}}^s_{\text{high}}$：

$$\hat{\boldsymbol{F}}^s=\boldsymbol{F}^s\odot\zeta(\boldsymbol{M}^s)\tag{1}$$

式中，$\odot$是 Hadamard 积；$\zeta(\cdot)$是一个函数，它将支持掩码$\boldsymbol{M}^s$双线性插值到特征图$\boldsymbol{F}^s$的每个层级的空间大小，再进行 Hadamard 积以丢弃不相关的激活。然后沿通道维度扩展，在第l层使得$\zeta_l:\mathbb{R}^{H\times W}\rightarrow\mathbb{R}^{C_l\times H_l\times W_l}$。

1.4　超相关性构造

特征超相关构造。即对每一层成对的查询图像特征和经过不同处理的支持图像特征利用余弦相似构造一个 4D 相关张量$\hat{\boldsymbol{C}}_l\in\mathbb{R}^{H_l\times W_l\times H_l\times W_l}$：

$$\hat{\boldsymbol{C}}_l(x^q,x^s)=\text{ReLU}\left(\frac{\boldsymbol{F}^q_l(x^q)\hat{\boldsymbol{F}}^s_l(x^s)}{\|\boldsymbol{F}^q_l(x^q)\|\ \|\hat{\boldsymbol{F}}^s_l(x^s)\|}\right)\tag{2}$$

式中，x^q和x^s分别表示特征图$\boldsymbol{F}^q_l$和$\hat{\boldsymbol{F}}^s_l$的二维空间位置，利用 ReLU 得到抑制噪声后的相关性分数。在得到的一组 4D 相关性张量集合$\left\{\hat{\boldsymbol{C}}_l\right\}_{l=1}^{L}$中，如果它们具有相同的空间大小，就将其在通道维度进行聚合，并将子集表示为$\left\{\hat{\boldsymbol{C}}_l\right\}_{l\in\mathcal{L}_p}$，其中$\mathcal{L}_p$是某个金字塔层$p$处 CNN 层索引$\{1,2,\cdots,L\}$的子集，其中在

子集 l=1 和 l=L 处分别为浅层相关张量 $\hat{\boldsymbol{C}}_{\text{low}}$ 和深层相关张量 $\hat{\boldsymbol{C}}_{\text{high}}$。

对 $\{\hat{\boldsymbol{C}}_l\}_{l\in\mathcal{L}_p}$ 中的所有 4D 相关张量沿通道维度连接，以形成超相关张量 $\boldsymbol{C}_p\in\mathbb{R}^{|\mathcal{L}_p|\times H_p\times W_p\times H_p\times W_p}$，其中 (H_p,W_p,H_p,W_p) 表示金字塔层 p 处超相关的空间分辨率。给定 P 个金字塔层即 P 个不同大小的超相关张量层，超相关金字塔表示为 $\mathcal{C}=\{\boldsymbol{C}_p\}_{p=1}^{P}$，表示来自多个特征层级聚合后的特征相关性集合。

1.5 非局部均值映射模块

类内差异噪声是指由于同一类别内样本间的差异或多样化而产生的噪声。类内差异的存在会影响模型的性能，特别是对于语义分割这种密集预测任务。因此，降低类内差异、增强类内相似性是提高分割精度的有效方法之一。

目前大部分方法仍使用注意力机制或大步长的卷积进行特征捕捉和特征增强，过多的可学习参数使得模型的全局先验在训练样本上快速拟合，这样可以获得更快的模型拟合速度，也导致了模型的泛化性能受损。

受到非局部响应[17]的启发，本文提出了一个不含可学习参数的非局部均值映射模块（non-local mean map module）来完成特征相关性的增强，模块设计如图 3 所示。

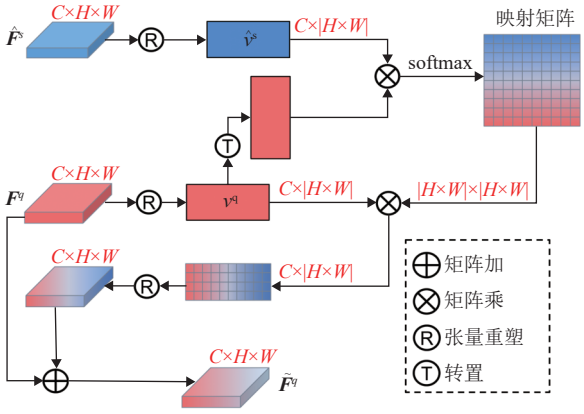


图 3 非局部均值映射模块

本文从骨干网络中的深层提取深层特征，再进行支持特征掩码得到 $\{\boldsymbol{F}^q_{\text{high}},\hat{\boldsymbol{F}}^s_{\text{high}}\}$，然后将其输入，得到从查询特征前景到支持特征全局特征的映射，再与查询特征融合以进行特征增强。

首先，本文对特征进行重塑，得到特征向量：

$$\hat{\boldsymbol{v}}^s=\mathcal{F}_{\text{reshape}}(\hat{\boldsymbol{F}}^s)\in\mathbb{R}^{C\times|H\times W|} \tag{3}$$

随后对它们进行矩阵点积，得到支持图像背景的全局特征到查询图像前景特征的特征相关映射矩阵，再和转置后的支持特征向量进行矩阵乘积和相关重塑操作，得到映射后的支持特征 $\tilde{\boldsymbol{F}}^q$：

$$\tilde{\boldsymbol{F}}^q=\boldsymbol{F}^q+\sum_{k=1}^{H\times W}\frac{\exp\left(\hat{\boldsymbol{v}}^s_k\text{transpose}(\boldsymbol{v}^q)\right)}{\sum_{l=1}^{H\times W}\exp\left(\hat{\boldsymbol{v}}^s\boldsymbol{v}^q_l\right)}\boldsymbol{v}^q \tag{4}$$

本文将掩码后的深层特征对 $\{\boldsymbol{F}^q_{\text{high}},\hat{\boldsymbol{F}}^s_{\text{high}}\}$ 经过 2 个非局部均值映射块作用以得到映射后的深层支持特征对 $\{\tilde{\boldsymbol{F}}^q_{\text{high}},\hat{\boldsymbol{F}}^s_{\text{high}}\}$，配合双特征掩码策略将特征聚合输入 4D 编码器，以对提取的中层特征对进行相关性信息引导。

1.6 多级混合池化 4D 卷积核

在最近的视觉对应任务研究中，验证了 4D 卷积捕捉特征多尺度对应关系的有效性，同时也给模型带来了巨大的参数运算压力。目前，针对传统 4D 卷积核的优化有特征裁剪和权重剪枝。而由于语义分割任务关于特征图分辨率的高敏性，特征裁剪造成的性能损失较大，HSNet 采用卷积核权重裁剪有效缓解了传统 4D 卷积的运算压力。具体而言，对于 $\mathbb{R}^{H\times W}$ 做稀疏比例为 prune_s 的稀疏操作，HSNet 按步长 prune_s 保留响应，将其余响应作为非相关响应直接丢弃得到 $\mathbb{R}^{H/\text{prune}_s\times W/\text{prune}_s}$，该 4D 卷积核采用的响应丢弃权重裁剪方案导致了边界的细粒度匹配精度损失。面对这一问题，本文提出了多级混合池化 4D 卷积核编码器，如图 4 所示。

相比目前主流的方法[18-19]，DFGHNet 采用更大的特征分辨率，以及在对特征进行局部小步长、整体多级混合池化的同时进行对应的多级分离 4D 卷积。

为了便于处理，本文在通道维度上对输入的超相关张量 $\boldsymbol{C}_p\in\mathbb{R}^{|\mathcal{L}_p|\times H_p\times W_p\times H_p\times W_p}$，首先在第 1 和第 2 维度进行压缩，以降低维度并形成 2D 相关张量 $\boldsymbol{C}_{p\{3,4\}}\in\mathbb{R}^{|\mathcal{L}_p\times H_p\times W_p|\times H_p\times W_p}$：

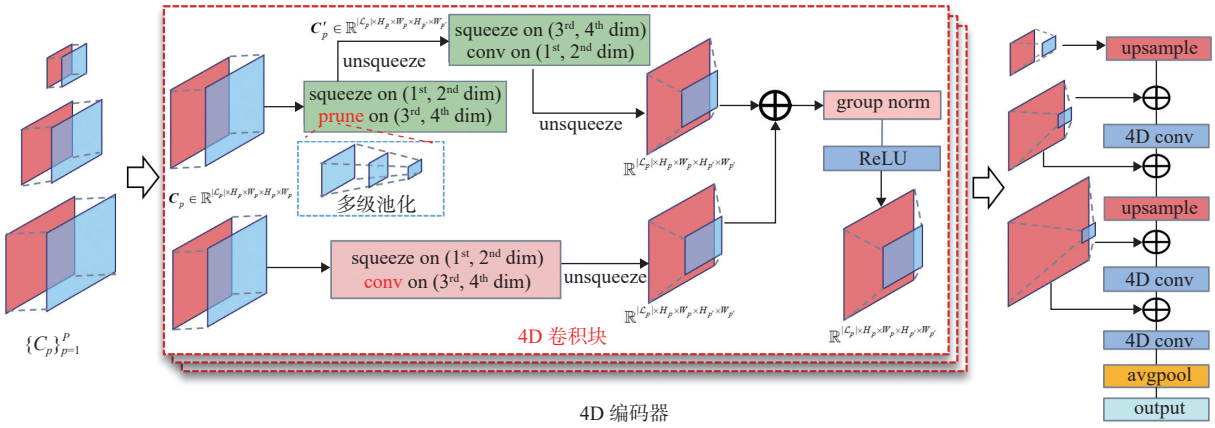
$$\boldsymbol{C}_{p\{3,4\}}=\mathcal{F}_{\text{squeeze}}\left(C_p(1^{\text{st}},2^{\text{nd}}\text{dim})\right) \tag{5}$$

本文将 $\boldsymbol{C}_p\in\mathbb{R}^{|\mathcal{L}_p|\times H_p\times W_p\times H_p\times W_p}$ 作为同源张量，对其进行第 1、2 维度的压缩，得到 $\boldsymbol{C}_{p\{3,4\}}\in\mathbb{R}^{|\mathcal{L}_p\times H_p\times W_p|\times H_p\times W_p}$ 以作为后续不同处理的同源 2D 相关张量。对于稀疏处理，即多级小步长混合池化，其对步长为 s 的卷积稀疏公式的为

$$y^{k+1}_{ij}=\lambda\max_{(m,n)\in\mathcal{R}_{ij}}x^k_{mn}+(1-\lambda)\frac{1}{|\mathcal{R}_{ij}|}\sum_{(m,n)\in\mathcal{R}_{ij}}x^k_{mn} \tag{6}$$

式中，$\mathcal{R}_{ij}$ 表示响应在二维位置 (i,j) 的大小为 $s\times s$ 的邻域范围；x^k_{mn} 表示经过 k 次步长 s 次混合池

化后在该邻域的(m,n)点的响应；y_{ij}^{k+1}表示经过 k 次混合池化对应二维位置(i,j)的输出响应。由于不同模型和参数对池化操作的敏感性不同，加入自定义权重参数λ以适应不同模型的调整需求。



4D 编码器

图 4　4D 编码器结构

对于分体 4D 卷积具体操作为对同源 2D 相关张量$C_{p\{3,4\}}$的不同处理后的响应对齐，即在第 3 和第 4 维度上的多级池化稀疏操作和卷积操作后的维度对齐，进一步强化特征的多尺度分离能力。具体来说，对于同源张量$\boldsymbol{C}_{p\{3,4\}}\in\mathbb{R}^{|\mathcal{L}_p\times H_p\times W_p|\times H_p\times W_p}$的第 1 种处理是将稀疏处理后的 2D 相关张量进行维度解压，重塑后的张量大小为$\mathbb{R}^{\mathcal{L}_p\times H_p\times W_p\times H_{p'}\times W_{p'}}$，其中$p'=p/(s\times k)$，$s$ 为单次池化步长，k 为池化次数。将稀疏并且解压缩重塑后的张量$\boldsymbol{C'}_p\in\mathbb{R}^{|\mathcal{L}_p|\times H_p\times W_p\times H_{p'}\times W_{p'}}$再进行第 1、2 维度上的压缩操作，得到$\boldsymbol{C'}_{p\{1,2\}}\in\mathbb{R}^{|\mathcal{L}_p\times H_{p'}\times W_{p'}|\times H_p\times W_p}$以进行大小不变卷积。

对同源张量$\boldsymbol{C}_{p\{3,4\}}\in\mathbb{R}^{|\mathcal{L}_p\times H_p\times W_p|\times H_p\times W_p}$的第 2 种处理为卷积操作，卷积操作的具体参数取决于第 1 种处理中同源张量在第 3、4 维度的稀疏步长 s 和次数 k，以及$\boldsymbol{C'}_{p\{1,2\}}\in\mathbb{R}^{|\mathcal{L}_p\times H_{p'}\times W_{p'}|\times H_p\times W_p}$的卷积核大小，即通过对应大小的卷积步长和填充以进行同源张量在第 3、4 维度上的响应数量对齐。

在以上过程中，每进行 k 次步长为 s 的池化操作，执行 1 次分体 4D 卷积，以实现稀疏特征的有效建模。处理完成后，2D 张量被重新解压缩为 4D 超相关张量，由于更大特征分辨率带来更大的计算压力，需要减小 batchsize 以缓解显存溢出问题。本文采用不依赖于 batch 内样本统计信息的分组归一化[20]（group normalization）技术。分组归一化能够适应较小的 batchsize，同时保持模型的精度。

最后通过 ReLU 激活函数激活特征响应。不同大小的张量通过上采样与残差融合操作相结合，形成多尺度特征关联的融合，每一次特征融合后都通过 4D 卷积块处理，经过最后一次 4D 卷积后再经过第 3、4 维度的平均池化以输出聚合语义信息的 2D 相关张量，输入 2D 解码器以解码得到预测查询掩码。

2　实验

本节对 DFGHNet 的模型训练细节和参数进行详细介绍。对于骨干网络，采用在 ImageNet 上预训练的 VGG16 和 ResNet50。

对于 VGG16 骨干网络，本文在最后 2 个构建块中的每个卷积层之后提取特征作为中层密集特征。对于 ResNet50 骨干网络，本文在第 3～5 层的 ReLU 函数激活之前提取每个瓶颈末端的特征作为中层密集特征，对于深层特征和浅层特征的提取也是基于以上方案作为边界，即在 ResNet50 骨干网络，取第 2 层和第 6 层。该特征提取方案从每个骨干网络提取特征以生成 3 个不同特征大小的金字塔层（P=3）。该网络将训练和评估的支持图像和查询图像的空间大小设置为 463×463，权重参数λ设置为 0.5。

该网络在 PyTorch 中实现并使用 SGD 优化器进行优化。PyTorch 版本为 1.5.1，CUDA 版本为 10.1，batchsize 大小设置为 8，学习率 lr 设置为 0.005。对于 VGG16 骨干网络，设置训练 epoch=100，对于 ResNet50 骨干网络，设置训练 epoch=200。该模型冻结预先训练的骨干网络，以防止它们学习到训练数据的特定类的表示。

本文在 2 个主流的少样本分割数据集上评估提出

的网络：PASCAL-5^i[10] 和 COCO-20^i [21]。PASCAL-5^i 由来自 20 个对象类的掩码注释图像组成，均匀地分为 4 个文件夹：$\{5^i : i \in \{0,1,2,3\}\}$。COCO-$20^i$ 由来自 80 个对象类的掩码注释图像组成，也均匀地分为 4 个文件夹：$\{20^i : i \in \{0,1,2,3\}\}$。遵循目前 FSS 领域常见的训练评估模式。

本文采用平均交并比（mean intersection over union, MIoU）和前景−背景交并比（toreground-background intersection over union, FB-IoU）作为评估指标。MIoU 指标对目标文件夹中所有类的 IoU 值进行平均，FB-IoU 忽略对象类别并计算前景和背景 IoU 的平均值。由于 mIoU 比 FB-IoU 更好地反映了模型的泛化能力和预测质量，因此本文在上述指标中主要关注 MIoU。

本文对所有文件夹进行交叉验证：即对于每个文件夹i，使用其他文件夹的样本作为训练样本，并从目标文件夹 i中随机抽取 1 000 个片段进行模型性能评估。对于每一个文件夹，都使用交叉验证的原则进行全局先验模型训练和性能检验，以确保训练和评估的类不相交。该网络分割损失 L 定义为查询图像预测掩码 P 和查询图像真实掩码 M 之间的交叉熵：

$$L = -\sum_{H,W}\sum_{c\in C} M^{(H,W,C)} \log P^{(H,W,C)} \tag{7}$$

3 结果

本文在 PASCAL-5^i、COCO-20^i 数据集上评估所提出的模型，并将结果与此前的最优方法进行比较。 表 1 和表 2 总结了 DFGHNet 在 PASCAL-5^i 和 COCO-20^i 上的 1-shot 和 5-shot 评估结果，粗体数字表示最佳性能，-表示数据暂缺。可见 DFGHNet 在 VGG16 和 ResNet50 两种不同骨干网络上的模型都实现了最佳性能。

表 1 DFGHNet 在数据集 PASCAL-5^i 中与现有方法的性能对比

%

骨干网络	方法	1-shot						5-shot					
		5^0	5^1	5^2	5^3	Mean	FB-IoU	5^0	5^1	5^2	5^3	Mean	FB-IoU
VGG16	OSLSM[10]	33.6	55.3	40.9	33.5	40.8	61.3	35.9	58.1	42.7	39.1	43.9	61.5
	PFENet[18]	56.9	68.2	54.4	52.4	58.0	72.0	59.0	69.1	54.8	52.9	59.0	72.3
	HSNet[15]	59.6	65.7	59.6	54.0	59.7	73.4	64.9	69.0	64.1	58.6	64.1	76.6
	DPCN[22]	58.9	**69.1**	63.2	55.7	61.7	73.7	63.4	70.7	**68.1**	59.0	65.3	77.2
	FECANet[16]	66.5	68.9	63.6	58.3	**64.3**	76.2	68.6	70.8	66.7	60.7	66.7	77.6
	DFGHNet	**66.6**	68.6	**63.9**	**59.2**	**64.5**	**77.1**	**69.8**	**71.5**	66.0	**63.4**	**67.6**	**78.9**
ResNet50	PFENet[18]	61.7	69.5	55.4	56.3	60.8	73.3	63.1	70.7	55.8	57.9	61.9	73.9
	HSNet[15]	64.3	70.7	60.3	60.5	64.0	76.7	70.3	73.2	67.4	67.1	69.5	80.6
	RePRI[23]	59.8	68.2	62.1	48.5	59.7	-	64.6	71.4	**71.1**	59.3	66.6	-
	ASR[24]	55.2	70.3	53.3	53.6	58.1	-	58.3	71.8	56.8	55.7	60.9	-
	NTRENet[25]	65.4	72.3	59.4	59.8	64.2	77.0	66.2	72.8	61.7	62.2	65.7	78.4
	DPCN[22]	65.7	71.6	**69.1**	60.6	66.7	78.0	70.0	73.2	70.9	65.5	69.9	80.7
	ASNet[26]	68.9	71.7	61.1	62.7	66.1	77.7	72.6	74.3	65.3	67.1	70.8	80.4
	FECANet[16]	69.2	72.3	62.4	**65.7**	67.4	78.7	72.9	74.0	65.2	67.8	70.0	80.7
	SCCAN[27]	68.2	72.5	66.8	59.8	66.8	77.7	72.3	74.1	69.1	65.6	70.3	81.8
	DAM[28]	67.3	72.0	62.4	59.9	65.4	-	73.6	74.6	69.9	67.2	71.3	-
	DFGHNet	**69.7**	**72.9**	65.1	65.3	**68.2**	**79.7**	**74.2**	**75.1**	71.0	**71.3**	**72.9**	**82.2**

PASCAL-5^i 数据集的性能评估结果如表 1 所示，DFGHNet 比此前最优方法取得了更优越的性能，在 1-shot 设置中，DFGHNet 相比此前最优性能对于骨干网络 VGG16 和 ResNet50 实现了 0.2% 和 0.8% 的 MIoU 性能提升。在 5-shot 设置中，DFGHNet 性能较此前最优性能对于骨干网络 VGG16 和 ResNet50 实现了 0.9% 和 1.6% 的 MIoU 性能提升。

COCO-20^i 数据集的性能评估结果如表 2 所示，在 COCO-20^i 数据集上该方法明显优于其他先进方法。在 1-shot 设置中，DFGHNet 相比此前最优性能对于骨干网络 VGG16 和 ResNet50 实现了 3.1% 和 4.1% 的性能提升。在 5-shot 设置中，DFGHNet 性能较此前最优性能于骨干网络 ResNet50 实现了 1.2% 的 MIOU 性能提升，而于 VGG16 骨干网络实现了 5.3% 的 MIOU 性能超越和 2.8% 的 FB-IoU 性能提升。验证了 DFGHNet 在少样本语义分割任务中的优越性和有效性。

表 2　DFGHNet 在数据集 COCO-20^i 中与现有方法的性能对比　　%

骨干网络	方法	1-shot						5-shot					
		20^0	20^1	20^2	20^3	Mean	FB-IoU	20^0	20^1	20^2	20^3	Mean	FB-IoU
VGG16	FWB[29]	18.4	16.7	19.6	25.4	20.0	-	20.9	19.2	21.9	28.4	22.6	-
	PFENet[18]	35.4	38.1	36.8	34.7	36.3	-	38.2	42.5	41.8	38.9	40.4	-
	FECANet[16]	34.1	37.5	35.8	34.1	35.4	65.5	39.7	43.6	42.9	39.7	41.5	67.7
	DFGHNet	**36.9**	**42.2**	**42.4**	**36.2**	**39.4**	**67.8**	**42.4**	**50.5**	**49.7**	**44.9**	**46.8**	**70.5**
ResNet50	PFENet[18]	36.5	38.6	34.5	33.8	35.8	-	36.5	43.3	37.8	38.4	39.0	-
	RePRI[23]	32.0	38.7	32.7	33.1	34.1	-	39.3	45.4	39.7	41.8	41.6	-
	ASR[24]	30.6	36.7	32.6	35.3	33.8	-	33.1	39.5	34.1	36.2	36.7	-
	HSNet[15]	36.3	43.1	38.7	38.7	39.2	68.2	43.3	51.3	48.2	45.0	46.9	70.7
	NTRENet[25]	36.8	42.6	39.9	37.9	39.3	68.5	38.2	44.1	40.4	38.4	40.3	69.2
	FECANet[16]	38.5	44.6	42.6	40.7	41.6	69.6	44.6	51.5	48.4	45.8	47.6	71.1
	DAM[28]	39.8	41.0	40.1	40.7	40.4	-	**50.1**	51.0	50.4	49.6	50.3	-
	DFGHNet	**42.5**	**48.7**	**46.1**	**45.5**	**45.7**	**70.8**	47.3	**55.1**	**54.1**	**49.6**	**51.5**	**72.6**

此外，本文还进行了骨干网络为 ResNet50、数据集为 PASCAL-5^i 的模型参数性能对比，如表 3 所示，对比基准模型 HSNet，在 1-shot 和 5-shot 设置下 DFGHNet 分别实现了 4.2% 和 5.8% 性能提升的同时拥有最少的可学习参数。

表 3　ResNet50 为骨干网络的性能参数比较结果

方法	MIoU/%		可学习参数量/M
	1-shot	5-shot	
PFENet	60.8	61.9	10.8
HSNet	64.0	67.1	2.6
FECANet	67.4	70.0	3.5
DFGHNet	**68.2**	**72.9**	**2.5**

4　消融实验

本节进行了消融研究，以调查模型中主要组成部分的影响：双特征掩码策略（dual-feature mask strategy, DFMS）、非局部均值映射模块（non-local mean mapping module, NMM）和多级池化 4D 卷积核稀疏策略（multi-level pooling 4D convolution kernel sparse strategy, Mul-4D）。下述所有消融研究实验均在 PASCAL-5^i 数据集上使用 ResNet50 作为骨干网络进行，以所有文件夹的平均交并比 mIoU 的平均值（Mean）为评估指标。

表 4 展示了 DFGHNet 各模块的消融实验结果。Base 代表 HSNet，通过采用双特征掩码策略，模型的性能提升至 65.1%。相比 HSNet 中针对支持背景的密集掩码策略，DFGHNet 双特征掩码策略带来了 1.1% 的性能提升。引入非局部均值映射模块（NMM）以捕捉更多全局相关性后，性能进一步提升至 66.3%。加入多级混合池化 4D 卷积核稀疏策略（Mul-4D）后，模型的最佳性能达到了 68.2%，证明了 DFGHNet 各模块的有效性。

表 5 展示了特征掩码策略的消融实验。此实验将 DFGHNet 中的双特征掩码策略替换为其他对应的掩码策略。在加入非局部均值映射和多级混合池化 4D 卷积后，相较于 HSNet 中的密集特征掩码策略，即对所有提取的支持特征进行掩码。本文的双特征掩码策略性能增益从 1.1% 提升到了 2.2%。相比 FECANet 中的单特征掩码策略，DFGHNet 的性能增益为 0.8%。

表 4　DFGHNet 各模块消融实验结果　　%

Base	DFMS	NMM	Mul-4D	Mean
√	-	-	-	64.0
√	√	-	-	65.1
√	√	√	-	66.3
√	√	√	√	**68.2**

表 5　掩码策略消融实验结果　　%

掩码策略	Mean
密集特征掩码策略	66.0
单特征掩码策略	67.5
双特征掩码策略**(本文)**	**68.2**

5　结束语

针对目前少样本分割方法对捕捉多级特征中抑制类间、类内噪声方案的不足以及高维卷积稀疏导致的精度损失问题，本文提出了一种双特征引导超相关性的少样本分割网络（DFGHNet），并在 PASCAL-5^i、COCO-20^i 数据集上验证了该方法的优越性和有效性。同时进行了消融实验，证明了本文提出的各个模块的有效性，并以最少的可学习参数实现了最优越的性能。

参 考 文 献

[1] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2016: 770-778.

[2] HUANG G, LIU Z, VAN D, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2017: 4700-4708.

[3] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[4] SUNG F, YANG Y, ZHANG L, et al. Learning to compare: Relation network for few-shot learning[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2018: 1199-1208.

[5] SUN Q, LIU Y, CHUA T S, et al. Meta-transfer learning for few-shot learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2019: 403-412.

[6] HOSPEDALES T, ANTONIOU A, MICAELLI P, et al. Meta-learning in neural networks: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(9): 5149-5169.

[7] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[C]// International Conference on Machine Learning. New York: PMLR, 2017: 1126-1135.

[8] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2015: 3431-3440.

[9] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C]//Medical Image Computing and Computer-Assisted Intervention. Munich: Springer, 2015: 234-241.

[10] SHABAN A, BANSAL S, LIU Z, et al. One-shot learning for semantic segmentation[C]//British Machine Vision Conference. London: British Machine Vision Association, 2017: 72.

[11] ROCCO I, ARANDJELOVIĆ R, SIVIC J. Efficient neighbourhood consensus networks via submanifold sparse convolutions[C]//European Conference on Computer Vision. Glasgow: Springer, 2020: 605-621.

[12] ROCCO I, CIMPOI M, ARANDJELOVIĆ R, et al. Neighbourhood consensus networks[J]. Advances in Neural Information Processing Systems, 2018, 31: 1651-1662.

[13] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few-shot learning[J]. Advances in Neural Information Processing Systems, 2017, 30: 4077-4087.

[14] MIN J, LEE J, PONCE J, et al. Learning to compose hypercolumns for visual correspondence[C]//European Conference on Computer Vision. Glasgow: Springer, 2020: 346-363.

[15] MIN J, KANG D, CHO M. Hypercorrelation squeeze for few-shot segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2021: 6941-6952.

[16] LIU H, PENG P, CHEN T, et al. Fecanet: Boosting few-shot semantic segmentation with feature-enhanced context-aware network[J]. IEEE Transactions on Multimedia, 2023, 25: 8580-8592.

[17] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2018: 7794-7803.

[18] TIAN Z, ZHAO H, SHU M, et al. Prior guided feature enrichment network for few-shot segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(2): 1050-1065.

[19] WU Z, SHI X, LIN G, et al. Learning meta-class memory for few-shot semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2021: 517-526.

[20] WU Y, HE K. Group normalization[C]//European Conference on Computer Vision. Glasgow: Springer, 2018: 3-19.

[21] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context[C]//European Conference on Computer Vision. Glasgow: Springer, 2014: 740-755.

[22] LIU J, BAO Y, XIE G S, et al. Dynamic prototype convolution network for few-shot semantic segmentation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2022: 11553-11562.

[23] BOUDIAF M, KERVADEC H, MASUD Z I, et al. Few-shot segmentation without meta-learning: A good transductive inference is all you need? [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2021: 13979-13988.

[24] LIU B, DING Y, JIAO J, et al. Anti-aliasing semantic reconstruction for few-shot semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2021: 9747-9756.

[25] LIU Y, LIU N, CAO Q, et al. Learning non-target knowledge for few-shot semantic segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2022: 11573-11582.

[26] KANG D, CHO M. Integrative few-shot learning for classification and segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2022: 9979-9990.

[27] XU Q, ZHAO W, LIN G, et al. Self-calibrated cross attention network for few-shot segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2023: 655-665.

[28] CHEN H, DONG Y, LU Z, et al. Dense affinity matching for few-shot segmentation[J]. Neurocomputing, 2024, 577: 127348.

[29] NGUYEN K, TODOROVIC S. Feature weighting and boosting for few-shot segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Computer Society, 2019: 622-631.

责任编辑 叶 芳