Salient Object Detection Enhanced Pseudo-Labels For Weakly Supervised Semantic Segmentation

**摘要**

本文针对弱监督语义分割任务中基于类激活图(CAM)生成伪标签的局限性,提出了一种新颖的显著性目标融合框架。该框架通过设计RGB-SOD网络捕获显著目标的完整轮廓和边缘细节,与CAM定位信息相互补充。我们还设计了显著性目标选择器,动态平衡CAM和SOD在生成单类伪标签时的权重,进一步提高伪标签质量。广泛实验表明,我们方法可精准获取高质量伪标签,在PASCAL VOC数据集上取得了最新的弱监督分割精度,显著提升了现有方法的性能瓶颈。该工作揭示了有效融合互补信息对于改进弱监督分割任务的重要性。

1. 介绍

语义分割是计算机视觉中一项重要且具有挑战性的任务，通常需要像素级语义标签以全监督的方式训练神经网络，然而，注释像素标签是一项耗时耗力的详尽工作，为了减轻工作量，我们考虑以仅需要图像级标签的弱监督方式学习分割模型。

当前的弱监督语义分割（WSSS）方法通常首先训练分类网络以生成类激活图（CAM）作为初始伪标签。然后在后续阶段对这些伪标签进行细化，最终导致基于细化伪标签的完全监督训练。然而，由于CAM不仅激活目标对象，还激活有助于类别识别的上下文信息，因此它们通常会导致目标对象和非目标对象或经常共同出现的事物之间的激活混淆，因此基于CAM的WSSS面临以下关键挑战： 1）定位图仅捕获一小部分目标对象，2）它遭受对象的边界不匹配，以及3）它几乎无法将共同出现的像素与目标对象分开（例如，铁路与火车）。

为了克服这三个挑战，我们思考到显着性图往往有着定位目标完整，边界清晰的特点，在本文中，我们首先提出了一种新的RGB-SOD网络，是的产生的silency map拥有着较好的边缘特征，以及对显著对象完整捕获的能力，能和CAM的优秀定位能力具有着互补关系。

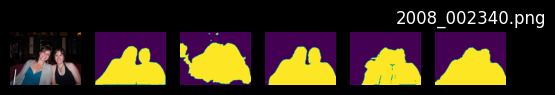
但如何将两者结合是我们遇到的另一个重要挑战，silency map本身没有类别的概念，于是我们先采取了只对ONE CLASS的数据进行了sod，然后通过基于CAM和silency map定义了一个合适的目标类别的定位可行度，选取合适阈值获得优质的可作为伪标签的silency map， 我们发现这部分生成的伪标签具有着极高的质量，在和传统方法生成的伪标签进行融合后，出现了很好的提点效果，从这个意义上，我们认为我们的方法使用两条互补的信息可以解决WSSS的性能瓶颈。

图一：对比效果图

JPEG , GT , SEAM , silency map（ours），QA-CILMS, RS+EPM









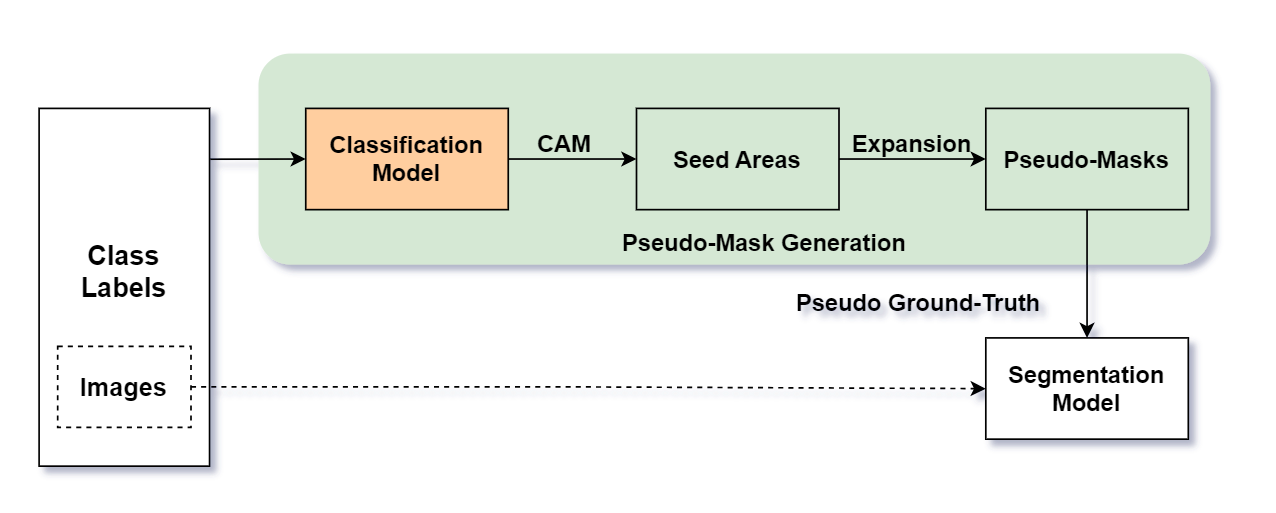
2. 相关工作

2.1 WSSS

一阶段 WSSS 方法直接利用图像级标签作为监督来训练端到端分割网络。早期的作品 [43, 44] 将这个问题表述为多实例学习。后来，帕潘德里欧等人。 [42]提出了一种期望最大化（EM）方法，利用中间预测来监督分割网络。张等人。 [62]利用图像分类分支生成注意力图并构造伪分割标 签来监督并行分割分支。阿拉斯拉诺夫等人。 [3]提出了一种自监督机制，在训练期间应用图像外观先验来生成伪分割标签。陈等人。 [7]构建了一个端到端框架，使用编码器-解码器网络来探索对象边界。与两阶段 WSSS 方法相比，一阶段方法通常性能较差且吸引力较小。

两阶段 WSSS 方法依赖于注意力图来生成伪分割标签，然后将其用于训练分割网络。两阶段 WSSS 方法的核心是生成高质量的注意力图。为了实现这一目标，最近提出了很多工作。魏等人。提出了对抗性擦除策略，该策略迭代地遮挡挖掘的对象区域以驱动分类网络发现新的对象区域。侯等人。 [22]通过使用自擦除策略来改进对抗性擦除策略，以防止注意力扩散到背景。科列斯尼科夫等人。 [28]引入了种子扩展的思想，它从预先计算的注意力图中扩展初始种子区域，并约束扩展的区域与对象边界对齐。后来，江等人。 [26]提出了利用不同训练阶段的注意力图的在线注意力积累策略。张等人。 [6]利用子类别信息来突出非歧视性语义区域。

另一系列工作尝试细化注意力图以获得具有珍贵边界的完整对象区域。安等人。 [2] 学习像素亲和力将注意力图中强烈响应的语义传播到相邻像素。陈等人。 [8] 和 Ahn 等人。 [1]通过显式学习类边界进一步改进了该方法。李等人。 [32]利用现成的显着性图作为监督来指导区域学习生成高质量的注意力图。



图二：传统的二阶段WSSS过程

**2.2 SOD**

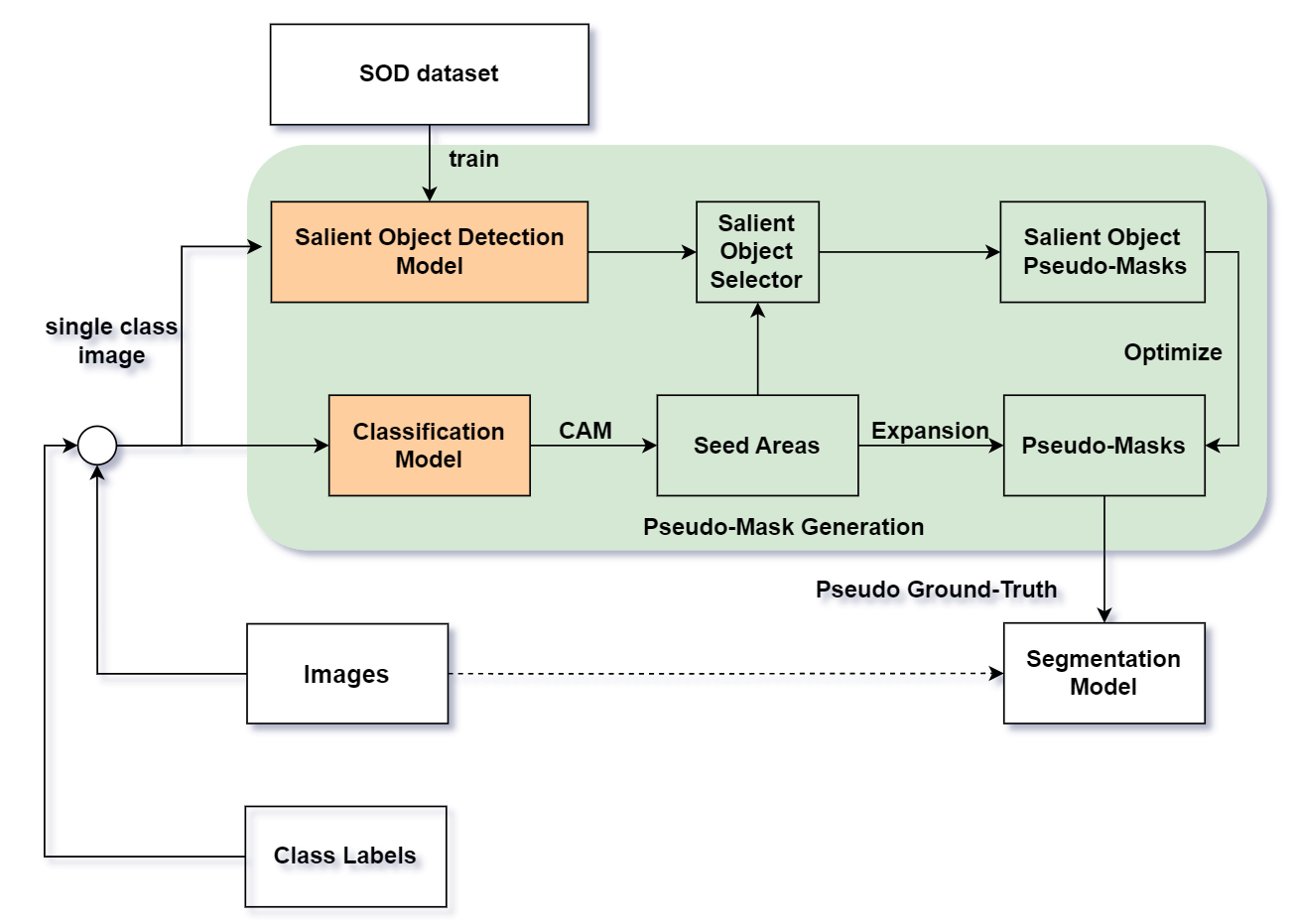
显著性检测，就是使用图像处理技术和计算机视觉算法来定位图片中最“显著”的区域。显著区域就是指图片中引人注目的区域或比较重要的区域，例如人眼在观看一幅图片时会首先关注的区域。显著与否是一种很主观的感觉，根据环境的不同，显著的对象往往不同，显著性往往不能被数学公式很好的表达。

SOD领域的早期工作主要基于低层次的视觉特征，如颜色、对比度和纹理。Itti等人提出的模型是最具代表性的早期工作之一，它模拟了人类视觉系统的底层视觉特性，利用中心-周围机制来预测注意力区域。随着深度学习技术的发展，近年来的研究已经转向使用深度神经网络来识别显著区域，这些方法能够利用高层语义信息，并在多个基准数据集上取得了显著进步。

具体来说，基于卷积神经网络（CNN）的方法已经成为SOD的主流，因为它们可以学习到更复杂的特征表示，从而在各种复杂场景中准确地检测显著区域。此外，一些研究集成了注意力机制，使网络能够更加专注于显著的区域而忽略不相关的背景。

我们在本文中提出的RGB-SOD网络正是在这样的背景下，旨在更进一步地推动SOD技术的发展，特别是在弱监督语义分割的任务中，通过精准地捕获显著区域边界，从而为生成高质量的伪标签提供支持。

**3.our works**



**图三：算法示意图**

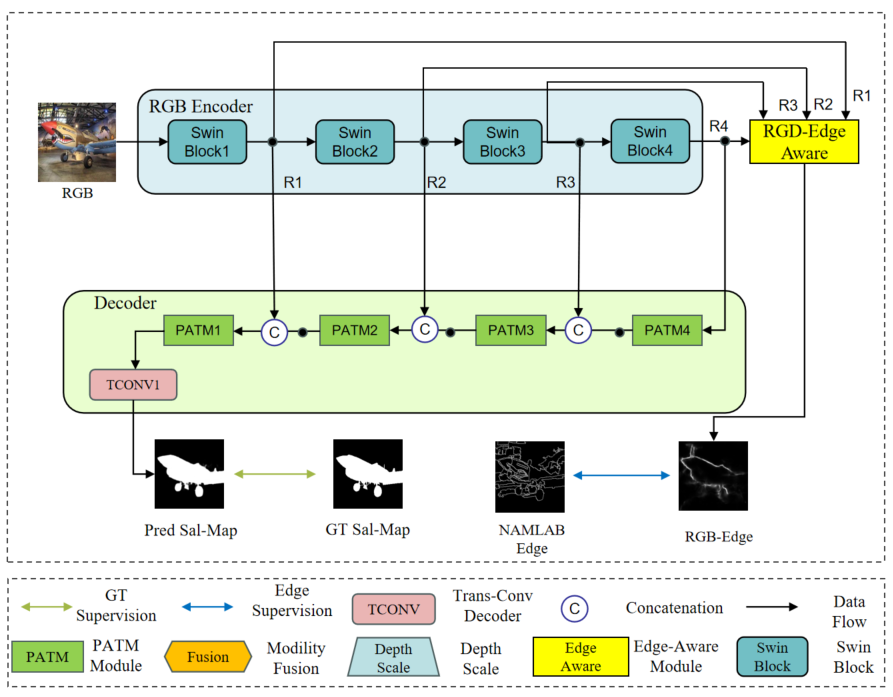
在本节中，我们详细介绍了我们方法的整个框架。在描述框架之前，我们首先对注意力图生成进行一些基本介绍。

**3.1.先决条件**

我们首先介绍生成注意力图的方法。给定输入图像 I，令 y 为图像级标签。最后一个卷积层的输出特征 F 有 C 个通道，与类的数量相同。最后一个卷积层后面是全局平均池化层，其中特征 F 被池化为大小为 C 的向量 f。我们通过应用 sigmoid 交叉熵损失函数来计算分类损失，其公式如下：

其中 σ 是 sigmoid 函数。注意力图可以从最后一个卷积层的输出生成。对于某些类 c，注意力图 是从 F 的 c 通道导出的，可以表示为

**3.2 显著性物体检测net**

我们搭建了以上RGB-SOD网络，它是我们框架的核心组成部分之一，负责显著性物体检测。该网络由两个主要部分组成：RGB编码器和解码器。

RGB编码器：包含四个顺序排列的Swin Transformer块（Swin Block1至Block4），它们依次处理输入的RGB图像，逐步提取和细化特征。每个Swin块都与相应的边缘感知模块连接，标记为R1至R4。其中，R4块特别设计为RGB-Edge Aware，以增强网络对图像边缘的感知能力。

解码器：由四个PATM（Pixel Attention Module）组成，旨在处理和集成来自编码器各阶段的特征。解码器通过转置卷积层（TCONV1）实现特征图的上采样，并通过串联（Concatenation）操作与PATM模块相连。PATM模块专注于保留显著区域的空间细节，提升边界的清晰度。

解码器最终生成的是预测显著图（Pred Sal-Map），该显著图与地面真实显著图（GT Sal-Map）对齐，并受到两个监督信号的共同指导：显著图监督和边缘监督，保证模型训练的正确性和有效性。此外，NAMLAB边缘模块被集成到网络中，进一步优化边缘特征，提高了RGB边缘的辨识能力。

整体来看，RGB-SOD网络利用了显著性检测在定位完整性和边缘清晰度方面的优势，与CAM提供的定位信息互补。这样的设计使得网络能够更精确地捕获目标对象的完整轮廓，同时处理共现物体的分离问题，显著提升了伪标签的质量，为弱监督语义分割任务提供了更可靠的标注信息。

**3.3 总体框架**

我们提出的显著性目标融合框架贯穿了基于类激活映射(CAM)和显著目标检测(SOD)的两条并行路径,旨在为弱监督语义分割任务生成高质量的伪标签。具体流程如下:

给定输入的单类图像及其类别标签,首先通过分类模型获取CAM注意力图,初步估计目标对象的粗略位置,作为种子区域(Seed Areas)。与此同时,我们将单类图像输入到专门设计的SOD模型中,利用其显著性检测能力捕获完整的目标对象轮廓与精细边界信息,生成Salient Object映射。

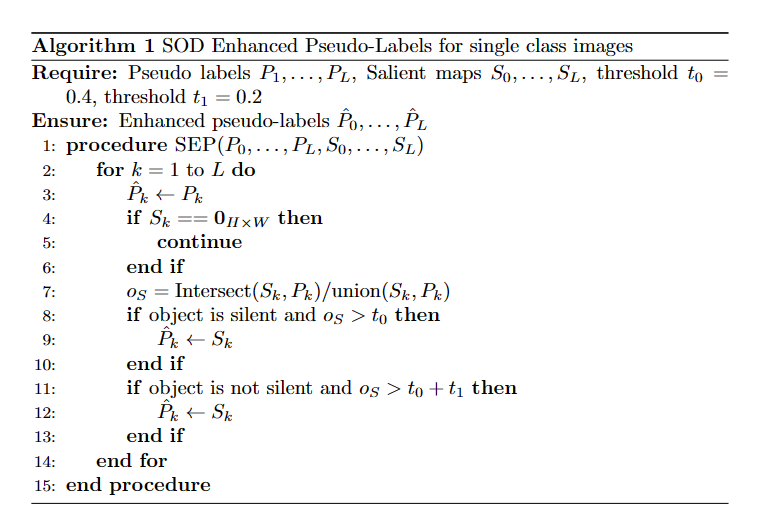
接下来,我们的关键环节是Salient Object Selector模块,它对SOD映射进行合理的过滤与优选,选取与CAM注意力图在定位一致性(α\*)较高的优质Salient Object Pseudo-Masks伪掩码。

在伪标签生成阶段,CAM注意力图被视为初始种子,并利用Expansion扩展策略进行区域扩张,生成初步粗糙的伪标签。然后我们将优化过的Salient Object Pseudo-Masks并入其中,通过有监督的优化策略融合两种互补信息,最终生成高精度的Pseudo-Masks伪标签。

最后,我们将生成的高质量Pseudo-Masks伪标签整合进Pseudo Ground-Truth伪真值集中,作为监督信号训练目标语义分割模型,从而显著提升分割性能。

该框架的核心创新在于巧妙地集成了CAM和SOD两种互补的目标定位途径,并设计了一系列模块在不同阶段协调融合两者,生成高质量监督信号,突破了基于单一路径的局限性,实现了弱监督语义分割的有效优化。

**3.4 显著性物体selector（SOS）**



在本研究中，我们探讨了一种利用类激活映射（CAM）和显著性目标检测（SOD）信息来单类图像的伪标签准确度的策略，即显著性伪标签策略。此策略的核心在于精选出与CAM定位一致的SOD图像作为伪标签，进而生成用于分割模型训练的高质量标签。

我们引入了一个关键参数 ，旨在量化SOD图像在目标类别定位上的准确性。具体而言， 通过计算CAM和SOD图像之间在特定阈值条件下的交并比（IoU）来表征它们的相互关系及其在目标定位上的一致性。这不仅揭示了CAM与SOD之间的相互作用，还允许我们通过动态调节阈值来细化伪标签的生成过程。公式表达如下：

同时我们还要综合考虑了物体本身的固有显著性。为此，我们在选择器中引入了一个范数度量，用于表示单个物体的显著性强度，这对于识别本身显著性较弱的对象，如椅子，沙发等，是至关重要的。

为了整合物体的固有显著性，我们增加了每个类在显著性图上的范数，以此作为其显著性的量度。通过结合这一范数和判断参数获得新的，

这里， 和 分别代表由CAM和SOD技术产生的伪标签，而 是预设的阈值，用于调节CAM和SOD图像之间的选择关系。借助此方法，成为一个动态调整的参数，平衡CAM和SOD在最终伪标签生成中的影响力。

最后我们得到的单类伪标签为：

当CAM和SOD之间的重叠区较大，表示它们在目标区域定位上高度一致时，SOD的信息在伪标签生成过程中获得更显著的权重；相反，当二者重叠较少，表明SOD在目标定位上更为独立时，其信息则在伪标签中占据主导地位。通过这种方法，我们不仅提升了伪标签的质量，还增强了分割模型训练过程中标签数据的准确性和可靠性。

**4，expriment**

**4.1. Experimental Setup**

**数据集。**

我们首先在广泛使用的Pascal VOC2012分割基准上评估所提出的框架。它包含21个具有像素级注释的类。训练集中有1,464张图像，验证集中有1,449张图像，左边的1,456张图像用于测试。按照比较方法中的相同设置，我们用来自SBD[21]的数据扩充训练集。因此，训练集中有10,582张图像，验证集中有1,449张图像。

**评估指标。**

我们用PASCAL VOC 2012上的验证和测试集上来验证我们的方法。PASCAL VOC 2012测试集上的评估结果来自PASCAL VOC官方评估服务器。此外，我们采用平均intersection-over-union（mIoU）来衡量分割模型的准确性。

**数据增强。**

为了进行数据增强，输入图像的短尺寸被调整为512。全局视图的分辨率为448×448，是从输入图像中裁剪出来的。从全局视图中裁剪分辨率为 320×320 的局部图像块。

**分割。**

我们选择 DeepLab-v2 和 DeepLabv3+ [21]作为我们的分割网络。我们基于 ResNet-101 [46] 和 resnest101，resnest269 报告性能。对于 ResNet-101，我们使用 COCO 预训练模型。按照[32]，我们使用相同的方式生成伪标签。给定注意力图，我们为背景通道分配一个固定阈值，并使用 argmax 函数来获取每个像素的标签。

**4.2. Comparisons with State-of-the-art Methods**

我们使用了常规用于比较的resnest101+deeplabv2的分割网络来训练的同时，为了进一步提高网络性能，我们又采用了ResNeSt架构，该架构普遍改进了学习到的特征表示，以提升图像分类、目标检测、实例分割和语义分割的性能。

在表1中，我们展示了最终训练网络的性能。与其他SOTA相比，我们的伪标签在具有相同训练设置的val和测试集上的性能都有显着提高。你可以在 <http://host.robots.ox.ac.uk:8080/anonymous/HYF7A0.html>和<http://host.robots.ox.ac.uk:8080/anonymous/ZA3UVZ.html>

上分别看到我们采用ResNeSt269+deeplabv3+在voc服务器上的水平。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Method** | **Public** | **Backbone** | **Sup.** | **val** | **test** |
| BCM[1] | CVPR19 | ResNet101 | I+B | 70.2 | - |
| BBAM[2] | CVPR21 | ResNet101 | 73.7 | 73.7 |
| EPS[3] | CVPR21 | ResNet101 | I+S | 71 | 71.8 |
| L2G[4] | CVPR22 | ResNet101 | 72.1 | 71.7 |
| SEAM[5] | CVPR20 | ResNet38 | I | 64.5 | 65.7 |
| AdvCAM[6] | CVPR21 | ResNet101 | 68.1 | 68 |
| OC-CSE[16] | ICCV21 | ResNet38 | 68.4 | 68.2 |
| CPN[7] | ICCV21 | ResNet38 | 67.8 | 68.5 |
| VWE[15] | IJCV22 | ResNet101 | 70.6 | 76.7 |
| CLIMS[14] | CVPR22 | ResNet101 | 70.4 | 70 |
| MCTformer[8] | CVPR22 | ResNet38 | 71.9 | 71.6 |
| SIPE[9] | CVPR22 | ResNet101 | 68.8 | 69.7 |
| W-OoD[10] | CVPR22 | ResNet38 | 70.7 | 70.1 |
| AMN[11] | CVPR22 | ResNet101 | 69.5 | 69.6 |
| ViT-PCM[12] | ECCV22 | ResNet101 | 70.3 | 70.9 |
| Yoon et al.[17] | ECCV22 | ResNet38 | 70.9 | 71.7 |
| ToCo[13] | CVPR23 | ViT-B | 69.8 | 70.5 |
| CLIP-ES[19] | CVPR23 | ResNet101 | **73.8** | **73.9** |
| ClsuterCAM[20] | IEEE Access24 | DeiT-Se | 70.3 | 70.7 |
| SFC[18] | AAAI24 | ResNet101 | 71.2 | 72.5 |
| (Ours) | - | ResNet101 | I+S | 74.9 | 74.59 |
| (Ours) | - | ResNeSt101 | 76.55 | 77.06 |
| (Ours) | - | ResNeSt269 | 77.52 | 77.73 |

Table 1: Comparison of our method with existing state-of-the-art methods on the PASCAL VOC 2012 dataset for val and test sets. I, image-level labels; B, bounding box labels; S, external saliency model.

|  |  |  |  |
| --- | --- | --- | --- |
| Method | Pub. | Sup. | mIoU (%) |
| PSA | CVPR 2018 | I | 58.4 |
| ICD | CVPR 2020 | I | 62.2 |
| SubCat | CVPR 2020 | I | 63.4 |
| SEAM | CVPR 2020 | I | 63.6 |
| A^2GNN | TPAMI 2021 | I | 65.3 |
| QA\_CLIMS | ACM MM 2023 | I | 71.8 |
| L2G | CVPR 2022 | I+S | 69.8 |
| ESEPM(baseline) | - | I | 72.5 |
| (ours) | - | I+S | 73.8 |

Table2：不同α值下模型在各项性能指标上的表现，说明显著性物体选择器参数α在调整CAM与SOD间信息平衡方面的作用。

**4.3.消融研究**

我们设计了多个消融实验来对我们的方法进行参数选择和健全性检查。所有消融实验均在 PASCAL VOC 2012 数据集上进行。由于我们的显著性物体selector只作用在单类图像上，所以在与评测伪标签相关的算法我们只采取了voc训练数据集中的单类图像进行评测。

**4.3.1 确定显著性物体selector的alph参数**

我们详细记录了在不同α值下的实验结果，包括准确率（Acc）、类别准确率（Acc\_class）、平均交并比（mIoU）和频率加权交并比（fwIoU）。通过调整α值，我们能够细致地观察到显著性图和CAM之间的平衡如何影响最终分割性能。这个实验不仅揭示了α参数在我们方法中的重要性，还指导了如何选择最优的α值以达到最佳分割效果。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **alph(6249)** | **Acc** | **Acc\_class** | **mIoU** | **fwIoU** |
| 0 | 94.12% | 85.30% | 77.57% | 89.10% |
| 0.05 | 94.35% | 85.89% | 78.35% | 89.48% |
| 0.1 | 94.41% | 86.18% | 78.66% | 89.58% |
| 0.2 | 94.49% | 86.63% | 79.04% | 89.75% |
| 0.3 | 94.54% | 87.13% | 79.44% | 89.89% |
| **0.4** | **94.44%** | **87.36%** | **79.45%** | **89.74%** |
| 0.5 | 94.25% | 87.39% | 79.15% | 89.46% |
| 0.6 | 94.10% | 87.72% | 79.08% | 89.29% |
| 0.7 | 93.86% | 87.71% | 78.83% | 88.92% |
| 0.8 | 93.61% | 87.43% | 78.33% | 88.51% |
| 0.9 | 93.52% | 87.35% | 78.15% | 88.36% |
| 1 | 93.51% | 87.33% | 78.13% | 88.35% |

Table3：

**4.3.2 探索不同算法产生的伪标签上显著性物体selector的通用性**

**我们**对比了在不同的基线伪标签生成策略下，显著性物体选择器应用前后分割性能的变化。我们分别考察了RS+EPM、QA-CLIMS、RCA、L2G和SEAM这些策略。该表格展示了显著性物体选择器对伪标签质量的提升，无论是在准确率、类别准确率、平均交并比还是频率加权交并比上。这些结果表明了我们提出的显著性物体选择器具有良好的通用性，能够有效提升不同基线方法生成的伪标签的质量。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | sod | **Acc** | **Acc\_class** | **mIoU** | **fwIoU** |
| RS+EPM | × | 93.51% | 87.33% | 78.13% | 88.35% |
|  | √ | 93.86% | 87.71% | 78.83% | 88.92% |
| QA-CLIMS | × | 93.92% | 86.46% | 77.88% | 88.72% |
|  | √ | 94.34% | 87.09% | 78.91% | 89.47% |
| RCA | × | 90.52% | 68.30% | 64.72% | 82.46% |
|  | √ | 91.91% | 72.07% | 68.03% | 84.95% |
| L2G | × | 93.36% | 87.90% | 76.86% | 87.91% |
|  | √ | 93.76% | 88.13% | 77.69% | 88.59% |
| SEAM | × | 85.40% | 78.77% | 60.60% | 75.62% |
|  | √ | 87.28% | 80.01% | 63.38% | 78.47% |

Table4：

**6.结论**

我们提出了一种新颖的弱监督分割框架，即显著性目标融合。受基于CAM的定位图和显着图之间互补关系的激励，我们的显著性目标融合从伪像素反馈中学习，结合显着图和定位图。由于我们的联合训练方案，我们成功地补充了两边的噪声或缺失信息。因此，我们的显著性目标融合可以对显著性特征满足要求的目标捕获精确的对象边界，显着提高了伪掩码的质量。广泛的评估和各种案例研究证明了我们的显著性目标融合的有效性以及出色的性能，WSSS在PASCAL VOC 2012上的最新精度。

**引用文献**

1. **Huang Yongchuang, Liu  Junxiu, Harkin  Jim, et al. An memristor-based synapse implementation using BCM learning rule In Neurocomputing, Volume 423, Pages 336-342, January 29, 2021.**
2. **Lee Jungbeom, Yi  Jihun, Shin  Chaehun, et al. BBAM: Bounding Box Attribution Map for Weakly Supervised Semantic and Instance Segmentation In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Pages 2643-2651, 2021.**
3. **Lee Minhyun, Lee Seungho, Lee Jongwuk,et al. Saliency as Pseudo-Pixel Supervision for Weakly and Semi-Supervised Semantic Segmentation In IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 45, Issue 10, Pages 12341-12357, October 1, 2023.**
4. **Jiang Peng-Tao, Yang Yuqi, Hou Qibin, et al. L2G: A Simple Local-to-Global Knowledge Transfer Framework for Weakly Supervised Semantic Segmentation Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Volume 2022-June, Pages 16865-16875, 2022.**
5. **Wang Yude, Zhang Jie, Kan Meina, et al. Self-Supervised Equivariant Attention Mechanism for Weakly Supervised Semantic Segmentation Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Pages 12272-12281, 2020.**
6. **Jungbeom Lee, Eunji Kim, and Sungroh Yoon. Anti-adversarially manipulated attributions for weakly and semi-supervised semantic segmentation. In IEEE Conf. Comput. Vis. Pattern Recog., pages 4071-4080, 2021.**
7. **Fei Zhang, Chaochen Gu, Chenyue Zhang, and Yuchao Dai. Complementary patch for weakly supervised semantic segmentation. In IEEE ICCV, pages 7242-7251, 2021. 2, 7, 11**
8. **Lian Xu, Wanli Ouyang, Mohammed Bennamoun, Farid Boussaid, and Dan Xu. 2022. Multi-class token transformer for weakly supervised semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 4310-4319.**
9. **Qi Chen, Lingxiao Yang, Jian-Huang Lai, and Xiaohua Xie. 2022. Self-supervised image-specific prototype exploration for weakly supervised semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 4288-4298.**
10. **Jungbeom Lee, Seong Joon Oh, Sangdoo Yun, Junsuk Choe, Eunji Kim, and Sungroh Yoon. 2022. Weakly supervised semantic segmentation using out-of distribution data. In Proceedings of the IEEE/CVF Conference on Computer Visionand Pattern Recognition. 16897-16906.**
11. **Minhyun Lee, Dongseob Kim, and Hyunjung Shim. 2022. Threshold matters in WSSS: manipulating the activation for the robust and accurate segmentation model against thresholds. In Proceedings of the IEEE/CVF Conference on Computer**
12. **Rossetti, Simone, Zappia, Damiano, Sanzari, Marta, et al. Max Pooling with Vision Transformers Reconciles Class and Shape in Weakly Supervised Semantic Segmentation Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Volume 13690 LNCS, Pages 446-463, 2022**
13. **Lixiang Ru, Heliang Zheng, Yibing Zhan, and Bo Du. Token contrast for weakly-supervised semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3093-3102, 2023. 1, 2, 6, 8**
14. **Jinheng Xie, Xianxu Hou, Kai Ye, and Linlin Shen. Clims:Cross language image matching for weakly supervised semantic segmentation. In CVPR, pages 4483–4492, 2022. 1,2, 6, 7, 8**
15. **Lixiang Ru, Bo Du, Yibing Zhan, and Chen Wu. Weaklysupervised semantic segmentation with visual words learning and hybrid pooling. International Journal of Computer Vision, 130(4):1127–1144, 2022. 6, 7, 14**
16. **Hyeokjun Kweon, Sung-Hoon Yoon, Hyeonseong Kim,Daehee Park, and Kuk-Jin Yoon. Unlocking the potential of ordinary classifier: Class-specific adversarial erasingframework for weakly supervised semantic segmentation. InProceedings of the IEEE/CVF International Conference on Computer Vision, pages 6994–7003, 2021. 6, 7, 14, 15**
17. **Sung-Hoon Yoon, Hyeokjun Kweon, Jegyeong Cho, Shinjeong Kim, and Kuk-Jin Yoon. Adversarial erasing framework via triplet with gated pyramid pooling layer for weakly supervised semantic segmentation. In European Conference on Computer Vision, pages 326–344. Springer, 2022. 6, 7, 14**
18. **SFC: Shared Feature Calibration in Weakly Supervised Semantic Segmentation (AAAI24)**
19. **Yuqi Lin, Minghao Chen, Wenxiao Wang, Boxi Wu, Ke Li, Binbin Lin, Haifeng Liu, and Xiaofei He. Clip is also an efficient segmenter: A text-driven approach for weakly supervised semantic segmentation. In CVPR, pages 15305–15314, 2023. 1, 2, 3, 4, 5, 6, 7, 8**
20. **Y. W. Kim and W. Kim, "Clustering-Guided Class Activation for Weakly Supervised Semantic Segmentation," in IEEE Access, vol. 12, pp. 4871-4880, 2024, doi: 10.1109/ACCESS.2024.3350176.**
21. **Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In ECCV, pages 1–18, 2018. 6, 8**