

# Activation Modulation and Recalibration Scheme for Weakly Supervised Semantic Segmentation

谢溥轩

## 摘要

图像级弱监督语义分割 (WSSS) 是一项基础但具有挑战性的计算机视觉任务, 有助于场景理解和自动驾驶。现有的大多数方法使用基于分类的类激活图 (CAM) 来充当初始伪标签, 受限于分类网络, 激活区域往往集中在具有判别性的区域上, 并且缺乏用于分割任务的定制特征。针对这个问题, 本文提出了一种新的激活调制和重新校准 (AMR) 方案, 该方案利用聚光灯分支和补偿分支来获得加权 CAM, 该加权 CAM 可以提供重新校准监督和特定于任务的概念。具体来说, 引入了注意力调制模块 (AMM) 从通道-空间顺序的角度并通过高斯函数重新排列特征重要性的分布, 可以显式地模拟信道间的相互依赖性和空间编码, 以自适应地调制面向分割的激活响应。此外, 我们引入了对偶分支的交叉伪监督, 这可以被视为一种语义相似的正则化, 以相互细化两个分支。此外, 本文提出的方法是即插即用的, 可以与其他方法相结合来提高它们的性能。实验证明, AMR 在 PASCAL VOC 2012 数据集上取得了优秀的性能, 甚至超越了一些依赖于更强监督的方法, 如显著性标签。

**关键词:** 弱监督; 语义分割; CAM; 高斯函数

## 1 引言

随着近年来深度学习算法的快速发展, 计算机视觉领域的应用范围越来越广泛, 语义分割作为计算机视觉领域的一项关键任务, 被应用于自动驾驶、工业自动化、视频特效等现实场景, 并且受到了大量学者的研究。语义分割主要执行的是像素级别的分类预测, 将图像中属于同一个对象类别的所有像素聚集在一起。在医学图像处理领域上, 可以通过语义分割将病人病灶进行分割, 从而让医生能够更快地对病情进行判断。在视频制作领域上, 对于危险场景的拍摄, 可以利用语义分割将演员从现实场景中分割出来, 再嵌入新的场景中。可见, 对于语义分割算法的研究, 具有重要的研究意义和研究价值。

早期的语义分割算法一般基于手工提取的纹理、颜色等特征进行分割, 这类分割方法不需要手工标注的标签, 由于缺乏足够的语义信息指导, 因此性能较差, 鲁棒性也比较差。近年来, 得益于卷积神经网络的蓬勃发展, 语义分割效果得到了显著提升, 但是这种提升往往依赖于手工标注的像素级标签, 即全监督的训练方式。众所周知, 获得手工标注的像素级标签是费时费力的, 因此, 为了解决这个问题, 越来越多的学者开始探索以更加简单的标注方式对网络进行训练, 如标注框、点标注、划线标注、图像级标注, 即弱监督的训练方式。其中, 图像级标注的方式因为其简单且最接近现实而最受研究者青睐。因此, 在本文的研究中, 同样将视线聚焦于以图像级标注作为标签的弱监督语义分割。

现有的对于弱监督语义分割任务的研究大多是利用分类网络生成的类别激活图 (CAM) 来进行的, 即将 CAM 作为初始的分割伪标签。然而, 由于 CAM 是基于分类网络生成的, 这也就不可避免地

导致生成的 CAM 会将注意力集中在最具有判别性的区域，例如马的头部、狗的头部等。这也就意味着，生成的 CAM 并没有针对语义分割这一任务进行特定的优化，会极大地影响分割网络的性能。为了解决这一突出问题，此前的一些研究者希望通过扩展激活区域、细化生成的 CAM 种子来提升模型的分割效果，如通过添加正则化项来拓展种子区域、通过迭代擦除的方式细化种子区域。但是，这些方法仍然没有很好地解决 CAM 的错误激活、欠激活的问题。因此，本文旨在探究一种更优的 CAM 生成优化模块，进而得到更好的伪标签来训练分割网络。

## 2 相关工作

### 2.1 弱监督语义分割

随着语义分割研究的细化，一方面，AutoML 的技术被用来提高分割质量<sup>[1]</sup>。另一方面，具有轻量级注释成本的训练被广泛探索，图像级 WSSS 近年来被广泛研究。现有的先进方法通常依赖于分类网络生成的类别激活图（CAM）的种子区域<sup>[2]</sup>。现有的一些方法大多数可以分为两个方面：生成高质量的 CAM 种子和提炼伪标签。一方面，一些方法直接扩展了 CAM 的响应区域，因为原始激活图只突出了图像的辨别区域。Wei 等人使用具有不同扩张率的扩张卷积来增加目标区域<sup>[3]</sup>。OAA 模型在不同的训练过程中融合多注意力图<sup>[4]</sup>。SEAM 通过分类网络中的等方差正则化从变换图像中捕获不同区域<sup>[5]</sup>。常等人探讨了带注释类子类别的特征学习能力<sup>[6]</sup>。Cian 捕获跨图像语义相似性和差异性的信息<sup>[7]</sup>。孙等人将两个神经共同注意合并到分类器中，用于发现一对训练图像中的共享或非共享语义。这有助于从分类器中提取更完整的初始 CAM<sup>[8]</sup>。荣格贝姆等人提出了一种对抗性的方式，以在激活图中发现目标对象的更多区域<sup>[9]</sup>。另一方面，一些工作集中在基于初始 CAM 的伪标签的细化上。SEC 探讨了提炼种子的三个原则，即种子、扩展和约束<sup>[10]</sup>。Ahn 和 Kwak 设计了一个深度神经网络，称为 AffinityNet，来预测一对相邻图像坐标之间的语义亲和力，然后应用这种语义亲和力来将生成的初始 CAM 细化为伪地面真实掩模<sup>[11]</sup>。此外，几种方法将 CAM 作为前景线索，将显著图<sup>[12]</sup>作为背景线索。姚等人引入了基于图的全局推理单元来发现非显著区域中的对象<sup>[13]</sup>。FickleNet 随机选择特征地图中的隐藏单元，以发现对象的另一部分。然而，这些方法是以迭代和随机的方式制定的，这可能会丢失必要的信息<sup>[14]</sup>。为了缓解这个问题，本文提出了一种激活调制和重新校准方案来生成高质量的 CAM。

### 2.2 注意机制

注意机制已广泛应用于分割网络-致力于构建图像的全局上下文关系<sup>[15-16]</sup>。Nonlocal 首次考虑了要素图中每个空间点之间的相关性<sup>[17]</sup>。然后，Asymmet 提出了一种非对称非局域网络，以加强非局域网络的连接<sup>[18]</sup>。SE 通过计算通道之间的相互作用来了解通道特征的重要性<sup>[19]</sup>。在这项工作之后，ECA-Net 使用基于通道的卷积来学习相互作用<sup>[20]</sup>。CBAM 利用空间和通道的注意力来突出通道和空间维度中的重要线索<sup>[21]</sup>。Gcnet 将长期依赖性纳入基本注意力模块<sup>[22]</sup>。在本文中，我们介绍了一个注意力调制模块，以增强分割任务的次要但重要的特征。

### 3 本文方法

#### 3.1 方法概述

现有的工作大多希望通过扩展判别响应区域和细化初始的 CAM 种子来提升模型的分割性能。然而，这些方法通常将扩展过程描述为一个复杂的训练阶段，也就是说，这些模型通常需要经过多个训练阶段。而 AMR 重新将目光放回到了单阶段的训练以及基于 CAM 本身的优化上来。具体来说，AMR 设计了两个分支——聚光灯分支和补偿分支，二者为互补的关系。其中，聚光灯分支是一个生成初始 CAM 的基本分类网络，会注意到图像中最具有判别性的区域，例如狗的头部。而补偿分支会挖掘图像中的关键区域，即那些对目标分类很重要但往往被分类网络所忽略的区域，如狗的身体。为实现补偿分支的功能，AMR 在补偿分支中引入了一个注意力调制模块（AMM）。具体来说，在 CAM 中，拥有高激活的区域往往对应着目标最具有判别性的区域，拥有中等激活的区域对应着目标的次要特征，而拥有最低激活的区域对应着背景区域。因此，AMM 模块通过对输入特征进行高斯函数的映射，从而增强次要特征，抑制最敏感和最不敏感的区域，从而实现对次要特征的挖掘。最后，再将聚光灯分支得到的 CAM 与补偿分支得到的 CAM 进行加权和，即可得到同时注意到判别区域和次要特征区域的 CAM，从而提高分割网络的性能。整体网络框架如图 1 所示。

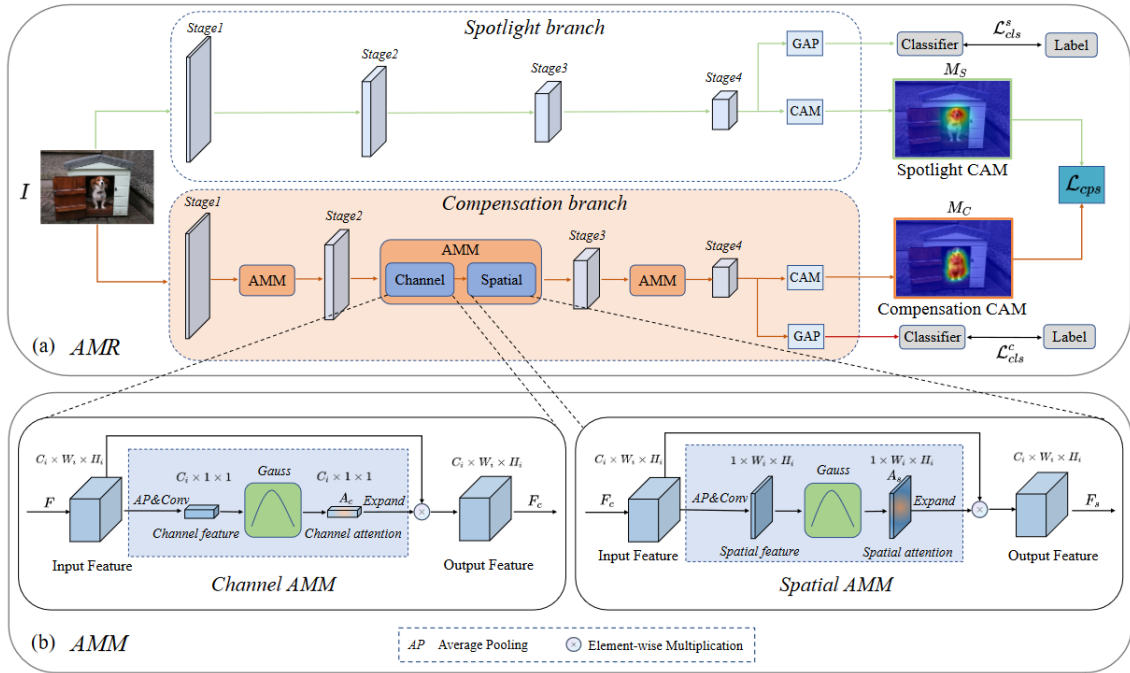


图 1: AMR 整体框架图

#### 3.2 聚光灯分支

聚光灯分支是一个生成初始 CAM 的基本 CNN 分类网络，因此仍然使用分类网络提取到的特征图和最后一层分类器的权重相乘得到初始的 CAM 图。此外，由于使用了图像级的标签，因此将特征图输入一个分类器，得到预测的标签，再与输入的真实标签计算损失。

#### 3.3 补偿分支

补偿分支与聚光灯分支其实是对偶的一个网络，二者的网络结构基本相似。但是对于补偿分支来说，为了能够挖掘出图像中的次要特征，在分类网络的每一个阶段后嵌入 AMM 模块（在章节 3.4 进行介绍），再通过与聚光灯分支相同的操作得到补偿分支的 CAM。由于经过 AMM 模块的调制，补偿

分支得到的 CAM 图能很好地注意到目标的次要特征。同样的，由于拥有对应的图像级标签，补偿分支也通过分类器得到的预测标签与输入的真实标签计算损失。

### 3.4 注意调制模块

如模块图所示，AMM 模块由通道注意力调制和空间注意力调制两部分组成，这两个部分是几乎一样的。通过通道 AMM 模块，可以更好地捕获到图像中的通道特征。如图所示，左图代表初始的通过分类网络得到的 CAM。正如上面所述的那样，最敏感的特征对应于判别区域，次要的特征表示重要但容易被忽略的区域，平淡的特征可能表示背景概念。而我们往往希望模型能够同时注意到那些容易被忽略的特征。因此，对输入特征进行简单的高斯函数调制即可达到这个目的。如右图所示，通过高斯函数的调制后，中间的次要特征的激活值被大幅度地提高了，而背景区域和判别区域被很好地抑制了。高斯函数的调制计算如下所示。

$$V_A = G(V_{A_f}) \quad (1)$$

其中， $G$  为高斯调制函数， $V_{A_f}$  为经过平均池化和卷积操作后的特征。高斯调制函数所用的均值和方差计算公式如下所示。

$$\mu = \frac{1}{M} \sum_{i=1}^M (V_{A_f}^i), \quad \sigma = \sqrt{\frac{1}{M} \sum_{i=1}^M (V_{A_f}^i - \mu)^2} \quad (2)$$

总的来说，通道 AMM 模块对输入的特征进行空间维度的平均池化和卷积操作，得到通道特征，再通过高斯调制得到调制后的通道特征，将通道特征拓展后与输入特征相乘，得到在通道维度上重激活后的特征。通道 AMM 模块的计算公式如下所示。

$$F_c(I) = \text{Expand}(G(H(P_s(F(I)))) \otimes F(I) \quad (3)$$

其中， $F(I)$  为输入通道 AMM 的特征图， $P_s$  为空间平均池化操作， $H$  为卷积操作， $G$  为高斯调制函数， $F_c(I)$  为通道 AMM 模块的输出特征。

为了进一步模拟特征在空间维度上的空间关系，引入了与通道 AMM 模块极其相似的空间 AMM 模块。唯一不同的是，空间 AMM 模块运用的是通道平均池化操作而非通道 AMM 模块中使用的空间平均池化操作。空间 AMM 模块的计算公式如下所示。

$$F_s(I) = \text{Expand}(G(H(P_c(F_c(I)))) \otimes F_c(I) \quad (4)$$

其中， $P_c$  为通道平均池化操作， $F_s(I)$  为空间 AMM 模块的输出特征图。

由于通道 AMM 模块和空间 AMM 模块的输入特征图和输出特征的大小没有发生改变，因此整个 AMM 模块是一个即插即用的模块，可以很轻易地嵌入在任何分割网络中用于提升模型的分割性能。

### 3.5 损失函数

由于我们拥有对象的图像级标签，因此可以直接对两个分支分别计算类别损失，最终的类别损失为两个分支类别损失的平均。计算公式如下所示。

$$L_{cls}^{s/c} = -\frac{1}{N} \sum_{i=1}^N \left( \bar{Y}_i \log \left( \frac{1}{1 + e^{-Y_i}} \right) + (1 - \bar{Y}_i) \log \left( \frac{e^{-Y_i}}{1 + e^{-Y_i}} \right) \right), \quad L_{cls} = \frac{1}{2} (L_{cls}^s + L_{cls}^c) \quad (5)$$

其中， $N$  为类别数， $Y_i$  为预测类别标签， $\bar{Y}_i$  为真实类别标签。

此外，为了充分利用聚光灯分支和补偿分支互补的 CAM，因此对聚光灯分支生成的 CAM 和补偿

分支生成的 CAM 采用了交叉伪监督。它可以被视为每个分支的语义相似正则化：

$$L_{cps} = \|M_S - M_C\|_1 \quad (6)$$

其中， $M_S$  为聚光灯分支生成的 CAM， $M_C$  为补偿分支生成的 CAM。因此，整体的损失函数如下所示。

$$\mathcal{L}_{all} = \mathcal{L}_{cls} + \mathcal{L}_{cps} \quad (7)$$

通过上述的损失函数不仅可以对补偿分支进行正则化处理，并且将判别区域和容易被忽略的区域拉近。因此，我们可以获得两个互补区域作为种子来重新校准初始 CAM。最终生成的 CAM 的计算公式如下所示。

$$M_{weighted}(I) = \xi M_S(I) + (1 - \xi) M_C(I) \quad (8)$$

其中， $\xi$  为两个分支生成的 CAM 的均衡系数。

## 4 复现细节

### 4.1 与已有开源代码对比

本文在复现过程中没有使用 AMR 开源代码，骨干网络部分均为自己动手实现。对于用于重新细化 CAM 时所使用的 gradCAM 算法为参考 github 源代码实现。其余部分均为自己实现。

### 4.2 实验环境

本文在 PASCAL VOC2012<sup>[23]</sup>数据集上对本文方法进行了验证。该数据集包含了 20 个前景对象类和 1 个背景类。遵循常用方法<sup>[24-25]</sup>，我们使用 10582 张图像进行训练，1449 张图像进行验证，1456 张图像进行测试。在整个训练过程中，我们只采用图像级的班级标签进行监督。每个图像可以包含多类标签。为了评估实验的性能，我们计算所有类的平均并集交集（mIoU）作为模型的性能评估指标。

本文使用了 ResNet50<sup>[26]</sup>作为 AMR 网络模型的骨干网络。受限于实验室算力，与原文所设置的大小为 16 的 batch size 不同，在复现过程中本文的 batch size 大小为 8，共迭代训练 8 次。初始学习率设置为 0.01，动量为 0.9。我们利用随机梯度下降算法进行权重衰减为 0.0001 的网络优化。我们还对训练图像进行了一些典型的数据扩充，如随机缩放和水平翻转。此外，本文在获得的 CAM 上利用 gradCAM 算法<sup>[27]</sup>来细化伪标签。

### 4.3 创新点

经过复现后，发现本文还存在着两个比较明显的问题。首先，是前景对象之间的歧义混淆。因为在数据集中，某些对象总是一起出现在图像中，比如火车和铁轨，人和自行车。而对于分类网络来说，网络总是会理所当然地将这些一起出现的对象联系起来，这就导致了对于某一个类别的激活区域会覆盖到另一个类别所在的区域，因为网络分不清这两个类别的具体边界。而第二个问题是背景的错误激活。这个问题其实是第一个问题的延伸，因为背景中的某些语义部分会和前景对象产生联系，所以分类网络也是自然而然地把背景中的一些区域和前景联系起来，导致前景对象的激活区域覆盖到了背景区域上。例如图 2 中的沙发和地板，地板往往和沙发会产生联系，网络就会自然而然地认为地板也是沙发的一部分。



图 2: 背景误激活示意图

针对上述问题，本文在 AMR 模型的基础上引入了高斯阈值调制函数以及像素级的区域正则化，从而对激活区域进行更好的约束，从而产生更好的 CAM 作为伪标签。

#### 4.3.1 高斯阈值调制模块

在 AMR 中，作者使用了高斯函数来对提取到的特征进行映射，从而实现对背景区域的激活抑制以及加强对次要特征所在区域的激活。但是 AMM 模块仍然存在一定的问题——背景的错误激活。由于 AMM 模块中的高斯函数只是对背景区域的激活值进行了一定的抑制，但是并没有完全去除掉这种错误的激活，因此仍然存在对背景区域的错误激活。

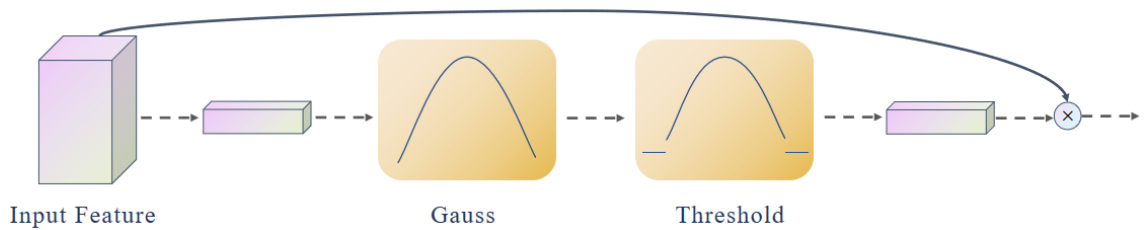


图 3: 高斯阈值调制模块结构图

针对上述存在的问题，本文提出了一个高斯阈值调制模块，即通过将阈值法和高斯函数映射结合起来，能够在一定程度上缓解这种对背景错误激活的问题。高斯阈值调制模块结构如图 3 所示。

具体来说，与 AMM 模块中先前进行的操作一致，先将输入的特征图通过空间平均池化操作和卷积操作得到通道特征，将通道特征通过高斯函数进行映射。对于映射后得到的特征图来说，已经将背景区域的激活值抑制到了较低的值。此时，再对特征图应用一个阈值函数，对于特征图中低于阈值的激活值来说，直接将对应位置的激活值设置为 0；对于特征图中高于阈值的激活值来说，则不进行任



何操作。因此，上述过程可以由下式表示。

$$F_{thres\_gaussian}(x, y) = \begin{cases} F_{gaussian}(x, y), & \text{if } F_{gaussian}(x, y) \geq \text{threshold} \\ 0, & \text{if } F_{gaussian}(x, y) < \text{threshold} \end{cases} \quad (9)$$

#### 4.3.2 区域正则化

正如上文所提到的那样，分类损失和交叉伪监督损失并不能很好地对一些靠近前景对象边缘的背景区域的激活进行抑制。因此，本文引入一个像素级的区域正则化项来约束激活图的大小，能够在一定程度上将一些不相关的背景排除在激活区域之外。

$$\mathcal{L}_{area\_regularization} = \frac{1}{KHW} \sum_{k=1}^K \sum_{h=1}^H \sum_{w=1}^W \mathbf{P}_k(h, w) \quad (10)$$

## 5 实验结果分析

我们使用训练好的 AMR 模型以及经过高斯阈值调制模块和像素级区域正则化改进后的 AMR 模型在 PASCAL VOC2012 的测试集上进行性能测试，以 mIoU 作为评价指标，实验结果如表 1 所示。

Method	CAM
AffinityNet	48.0
IRNet	48.3
CONTA	48.8
SEAM	55.4
Original-AMR	<b>56.8</b>
Recurrence-AMR	53.2
Improvement-AMR	54.5

表 1: 在 VOC2012 测试集上的效果

从表中可以看出，本文并没有完全复现出原文中 AMR 所达到的效果，经过比较与分析推断，batch size 大小以及 GPU 型号对模型训练效果存在一定影响。此外，经过高斯调制函数以及像素级的区域正则化改进后的 Improvement-AMR 能够在 Recurrence-AMR 基础上提升了 1.3%，也证明了本文所提出的创新模块的有效性。为了能够进一步凸显出改进后 AMR 模型的效果，将 CAM 进行可视化，可视化结果如图 4 所示。

从图中可以看出，当前景对象为飞机时，模型能够将激活区域很好地包围在整个前景对象周围。当前景对象为狗时，模型并不会像以往普通的分类网络那样，仅仅将注意力集中在狗的头部——这一最具判别性的区域，而是会将注意力开始往狗的身体移动，从而产生更好的激活效果。对于马和牛也是如此，模型都能将注意力放在除了头以外的更多区域上。而对于汽车而言，AMR 模型也能正确地将注意力放在汽车的车窗上，而不是将其认为是背景。



图 4: 实验结果示意

此外，针对最终 CAM 生成过程中使用到的加权均衡系数  $\xi$ ，通过消融实验最终确定其取值，消融实验结果如表 2 所示。

$\xi$	0.1	0.3	0.5	0.7	0.9
mIoU(%)	49.2	53.4	<b>56.8</b>	54.5	50.7

表 2: 超参数  $\xi$  消融实验结果

## 6 总结与展望

AMR 网络模型利用聚光灯分支和即插即用补偿分支来获得加权 CAM，并提供更多面向语义分割的概念。其中 AMM 模块从通道-空间顺序的角度重新排列特征重要性的分布，这有助于突出分割任务中一些重要但容易被忽略的区域。

本文对 AMR 网络模型进行了复现，并且在原文模型的基础上提出了两点相应的创新。一是高斯阈值调制模块，该模块作为一个即插即用的模块，能够很好地与任意一个网络模型相结合，而不显著地增加整体模型的参数量，并且能够在一定程度上提升模型的分割性能。通过将原文中的高斯函数与阈值相结合，能够更好地将背景区域的错误激活进一步抑制，从而迫使网络模型能够将注意力放在对图像分割十分重要但是很容易被网络模型忽略的次要特征上。二是像素级别的区域正则化。从原文模型产生的激活效果图可以看出，由于分类网络对背景区域和前景对象的错误关联，背景区域中仍然存在被错误激活的情况。而本文引入的像素级别的区域正则化能够对整体的激活区域大小进行限制，从而在一定程度上抑制背景的误激活。

然而，本文在创新点的实现上仍然存在不足。对于高斯阈值调制模块而言，虽然能够在一定程度上抑制了背景区域的激活，但是仍然没有完美地解决对于前景对象中容易被忽视区域激活不足的问题，生成的伪标签效果还是不足以很好地监督分割网络的训练。此外，高斯阈值调制模块中所使用的阈值很难确定。阈值过大，可能会影响到前景对象区域的激活；阈值过小，则对背景区域激活的抑制效果可能不足。对每张图像而言，整张图像中的像素分布都是不一样的，而我们也很难为每一张图像设定一个单独的阈值，这也就导致了高斯阈值调制模块在不同图像上表现的效果不一。在算法实现上，需要对特征图中每一个坐标的激活值都与阈值进行一次比较判断，可能会因此导致训练时间的增加。



针对上文所提到的对象之间的歧义混淆问题，其实是上下文先验信息误导分类网络的结果。模型会把一些正相关的信息直接认为是强关联信息，比如，我们在看到地板的时候会想到附近可能会有沙发，但是不应该认为地板是沙发的一部分。因此，未来可以寻找一种方法来阻断这种错误的关联信息，从而更好地解决对象之间的歧义混淆问题。

## 参考文献

- [1] LI S, WU J, XIAO X, et al. Revisiting Discriminator in GAN Compression: A Generator-discriminator Cooperative Compression Scheme[C]// . 2021.
- [2] ZHOU B, KHOSLA A, LAPEDRIZA A, et al. Learning Deep Features for Discriminative Localization [Z]. 2016.
- [3] WEI Y, XIAO H, SHI H, et al. Revisiting Dilated Convolution: A Simple Approach for Weakly- and Semi- Supervised Semantic Segmentation[J]. IEEE, 2018.
- [4] JIANG P T, HOU Q, CAO Y, et al. Integral Object Mining via Online Attention Accumulation[C]// International Conference on Computer Vision.
- [5] WANG Y, ZHANG J, KAN M, et al. Self-Supervised Equivariant Attention Mechanism for Weakly Supervised Semantic Segmentation[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
- [6] CHANG Y T, WANG Q, HUNG W C, et al. Weakly-Supervised Semantic Segmentation via Sub-category Exploration[C]// . 2020.
- [7] FAN J, ZHANG Z, TAN T, et al. CIAN: Cross-Image Affinity Net for Weakly Supervised Semantic Segmentation[C]//National Conference on Artificial Intelligence. 2020.
- [8] SUN G, WANG W, DAI J, et al. Mining cross-image semantics for weakly supervised semantic segmentation[C]//European conference on computer vision. 2020: 347-365.
- [9] LEE J, KIM E, YOON S. Anti-Adversarially Manipulated Attributions for Weakly and Semi-Supervised Semantic Segmentation[Z]. 2021.
- [10] KOLESNIKOV A, LAMPERT C H. Seed, Expand and Constrain: Three Principles for Weakly-Supervised Image Segmentation[C]//European Conference on Computer Vision. 2016.
- [11] AHN J, KWAK S. Learning Pixel-level Semantic Affinity with Image-level Supervision for Weakly Supervised Semantic Segmentation[Z]. 2018.
- [12] ZHANG P, SU L, LI L, et al. Training Efficient Saliency Prediction Models with Knowledge Distillation [C]//the 27th ACM International Conference. 2019.
- [13] YAO Y, CHEN T, XIE G, et al. Non-Salient Region Object Mining for Weakly Supervised Semantic Segmentation[J]., 2021.

- [14] LEE J, KIM E, LEE S, et al. FickleNet: Weakly and Semi-supervised Semantic Image Segmentation using Stochastic Inference[J]. IEEE, 2019.
- [15] WU J, HU H, YANG L. Pseudo-3D Attention Transfer Network with Content-aware Strategy for Image Captioning[J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2019.
- [16] WU J, HU H, WU Y. Image Captioning via Semantic Guidance Attention and Consensus Selection Strategy[J]. ACM Transactions on Multimedia Computing Communications and Applications, 2018, 14(4): 1-19.
- [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. 2018: 7794-7803.
- [18] XU M, BAI S, HUANG T, et al. Asymmetric Non-local Neural Networks for Semantic Segmentation [Z]. 2019.
- [19] JIE H, LI S, GANG S, et al. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, PP(99).
- [20] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks[C] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020.
- [21] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C] // Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
- [22] CAO Y, XU J, LIN S, et al. Gcnet: Non-local networks meet squeeze-excitation networks and beyond [C] // Proceedings of the IEEE/CVF international conference on computer vision workshops. 2019: 0-0.
- [23] EVERINGHAM M, ESLAMI S, VAN GOOL L, et al. The pascal visual object classes challenge: A retrospective[J]. International journal of computer vision, 2015, 111(1): 98-136.
- [24] WEI Y, FENG J, LIANG X, et al. Object region mining with adversarial erasing: A simple classification to semantic segmentation approach[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1568-1576.
- [25] WANG Y, ZHANG J, KAN M, et al. Self-supervised equivariant attention mechanism for weakly supervised semantic segmentation[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 12275-12284.
- [26] 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. 2016: 770-778.
- [27] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization[C] // Proceedings of the IEEE international conference on computer vision. 2017: 618-626.