分类号　　密级

中 国 地 质 大 学（北京）

专 业 硕 士 学 位 论 文

基于类激活映射的弱监督图像语义分割方法研究

研 究 生 陈若妍 学　　号 2104220035

专业学位类别 电子信息 专　　业 计算机技术

校内导师 牛云云 产业导师 张永虹

学习方式 全日制

2025 年 5 月

A Dissertation Submitted to
China University of Geosciences for Master of Professional Degree

A Weakly Supervised Image Semantic Segmentation Method based on Class Activation Mapping

Master Candidate: Ruoyan Chen

Professional Degree: Master of Electronic and Information

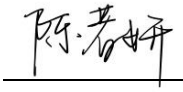
Dissertation Supervisor: Prof. Yunyun Niu

Associate Supervisor: Yonghong Zhang

China University of Geosciences (Beijing)

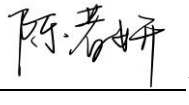
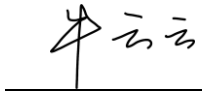
声 明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得中国地质大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

签　名：  日　期：2025.5.21

关于论文使用授权的说明

本人完全了解中国地质大学有关保留、使用学位论文的规定，即：学校有权保留送交论文的复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存论文。**延期公开的论文在到期公开后应遵守此规定。**

签　名：  导师签名：  日　期：2025.5.21

摘要

图像的语义分割是计算机视觉领域一个非常重要的研究内容，旨在通过提取图像特征来模拟人类视觉感知机制，从而完成对给定的图像进行逐像素分类。随着深度学习技术的发展，基于全监督的图像语义分割方法（Fully Supervised Semantic Segmentation, FSSS）迅猛发展，但其需要像素级人工标注，标注成本高昂、耗时较长，难以满足大规模数据训练的需求。因此，标注成本更低的基于弱监督的图像语义分割方法（Weakly Supervised Semantic Segmentation, WSSS）成为研究的重点。

目前，基于图像级标注的 WSSS 算法大多基于类激活映射（Class Avtication Mapping, CAM）算法展开研究，而其主要面临的问题在于 CAM 算法在激活目标前景类时存在的少活和误激活等挑战。为此，本文提出了一个基于标签阈值的一致性注意力架构（Consistency Attention Architecture for Label Thresholds, LTCAA）模型，针对上述存在的问题做了以下几点改进：

（1）针对 CAM 算法在激活前景目标类时不够准确的问题，本文设计了标签阈值模块，将图像级标注加入网络，并通过扩大前景和背景之间的差距，避免对目标前景类的错误激活。

（2）针对 CAM 算法对目标前景类激活不充分的问题，本文设计自注意力模块，通过建立上下文语义关联和通道间语义依赖关系，扩充监督信息，从而实现 CAM 的充分激活。

（3）为了使 CAM 算法在激活目标前景类时能够获得准确的边界，本文提出了一致性约束模块，通过双分支结构实现模型对仿射变换下图像特征的学习，强化模型对目标特征的捕捉能力，解决边界不够精细的问题。

为了验证 LTCAA 模型的性能，本文在 PASCAL VOC 2012 数据集上进行实验分析。为进一步探究 LTCAA 模型的实际应用意义，本文实现基于 LTCAA 模型的遥感图像语义分割任务，在 ISPRS postdam 数据集上进行实验分析。实验结果表明，LTCAA 模型在 PASCAL VOC 2012 数据集上能够取得很好的伪标注并获得很好的分割效果，并且本文设计的三个模块均为效果的提升做出贡献。与此同时，LTCAA 模型在 ISPRS postdam 数据集上能够获取高质量的伪标注，并获得很好的分割效果。

关键词：图像语义分割，弱监督，类激活映射，自注意力机制

Abstract

Semantic segmentation of images is a crucial research topic in the field of computer vision. It aims to simulate the human visual perception mechanism by extracting image features, thereby achieving pixel-wise classification of a given image. With the rapid advancement of deep learning techniques, fully supervised semantic segmentation (FSSS) methods have developed significantly. However, these methods require pixel-level annotations, which are costly and time-consuming, posing challenges for large-scale data training. Consequently, weakly supervised semantic segmentation (WSSS) methods, which demand lower annotation costs, have become a key focus of research.

At present, most WSSS methods based on image-level annotations rely on the Class Activation Mapping (CAM) algorithm. However, CAM-based methods suffer from insufficient and inaccurate activation of foreground classes. To address these issues, we proposes a Consistency Attention Architecture for Label Thresholds (LTCAA), which introduces several improvements.

(1) To address the problem of inaccurate activation of CAM, Label-based Specific Threshold Module (LSTM) is designed. LSTM adds image-level annotations to the network and widens the gap between foreground and background, avoiding false activation of the CAM.

(2) To address insufficient CAM activation for target foreground classes, a Self-Attention Module (SAM) is designed in this thesis. SAM enhance contextual and inter-channel semantic dependencies for fuller CAM activation.

(3) To refine CAM boundaries, this thesis proposes Consistency Regularization Module (CRM). The dual structure allows the model to learn image features after affine transformation. This strengthens LTCAA's ability to capture features and address boundary precision.

To validate the performance of LTCAA, we conducted comprehensive experiments on the PASCAL VOC 2012 dataset. Furthermore, to demonstrate the practical applicability of LTCAA, we extended the evaluation to the ISPRS Potsdam dataset for remote sensing image semantic segmentation tasks. The experimental results show that LTCAA achieves high-qaulity pseudo-label and good segmentation effect on the PASCAL VOC 2012 dataset. All the three proposed modules contribute to the improvement of the effect, as evidenced by ablation studies.

Meanwhile, LTCAA maintains robust generalization capability, attaining high-quality pseudo-label and good segmentation effect on the ISPRS Postdam dataset.

Key words: Image Semantic Segmentation, Weakly Supervision, Class Activation Mapping, Self-Attention Mechanism

目 录

1 引言

1.1 研究背景和意义

图像语义分割是计算机视觉领域一项基础且核心的研究课题，其主要目标是通过提取图像特征，模拟人类视觉的感知机制，为图像中每个像素赋予对应的语义标签，从而实现精细化的图像理解。该技术不仅关注图像整体的语义类别，还要求对像素级内容进行准确分类，以揭示图像更深层的语义结构。作为融合了机器学习（Machine Learning, ML）、人工智能（Artificial Intelligence, AI）、图像处理等多个领域的原理和方法，图像语义分割为更高级的计算机视觉任务提供了技术支持。随着数字化转型的持续推进，该技术凭借其在目标类别的像素级识别能力，已在多个应用场景中得到部署，并在跨学科融合中发挥关键作用，例如如医学影像中的肿瘤区域分割[1]，遥感图像中的目标提取[2]，以及城市交通中的自动驾驶感知系统[3] [4] 等。

自 20 世纪 70 年代数字图像处理技术兴起以来，图像语义分割逐渐成为研究者关注的重要课题。然而，受限于当时计算机硬件性能和图像处理算法的能力，早期的图像分割主要依赖于几何、拓扑等传统方法，此类方法多基于图像的底层特征，如灰度、颜色、空间纹理和形状等[5]，依赖于人为设定的先验知识来进行特征提取，难以有效捕捉图像中的深层语义信息，导致分割效果较为有限。近年来，随着计算能力的显著提升以及深度学习（Deep Learning, DL）技术的突破，语义分割取得了长足发展。DL 通过端到端的训练机制，显著增强了图像语义信息的建模能力，尤其是卷积神经网络（Convolutional Neural Network, CNN）在大规模标注样本的驱动下，能够实现精确的像素级分类。但是，CNN 在训练的过程中需要密集型像素级标注数据，这个过程十分困难且耗时耗力。

为了降低图像语义分割任务对像素级标注数据的依赖性，研究者逐渐转向利用更易取得的标注信息来实现图像的语义分割，因此基于弱监督学习的图像语义分割（Weakly Supervised Semantic Segmentation, WSSS）研究应运而生。WSSS 最重要的特征是使用弱标注作为监督信息，其获取难度低，但也存在信息量少的问题。目前，常用的弱标注包括边界框标注、涂鸦标注、点标注、图像级标注等，图 1-1 展示了几种常用的弱标注方式，其标注信息逐渐减少，取得难度从难到易。尽管弱标注在语义表达上不及精确的像素级标注，但其在标注效率与经济成本上的优势使其在实际应用与学术研究中具备重要的价值。特别是图像级标注因其仅需提供目标类别标签，成为弱监督方法中标注成本最低的形式，因此

利用图像级标注实现 WSSS 也成为计算机视觉领域的重要研究方向。目前，类激活映射（Class Activaction Mapping, CAM）方法是实现利用图像级标注进行 WSSS 的主流方法，因此基于类激活映射的弱监督语义分割成为一项充满挑战且意义重大的课题。

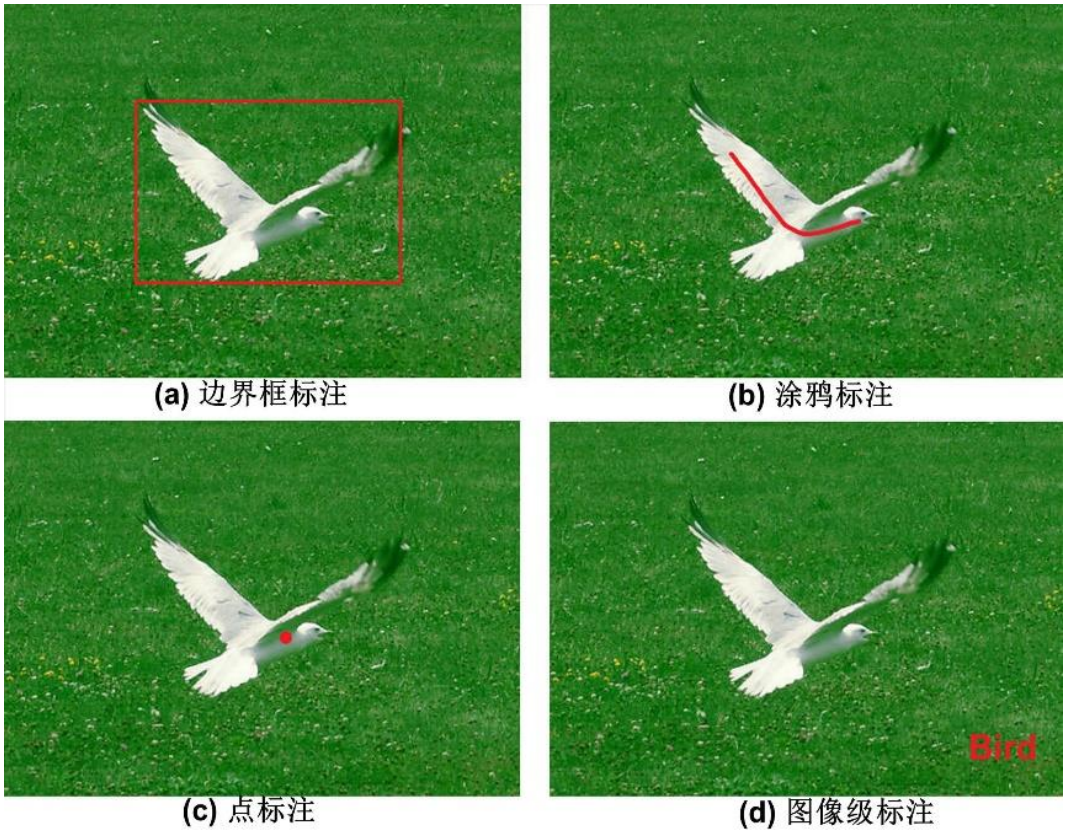


图 1-1 弱监督标注类型示例

1.2 国内外研究现状

随着 AI 的快速发展，大量研究者进行了基于深度学习的图像语义分割方法的探索。本节主要介绍图像语义分割方法的研究现状，包括全监督、弱监督以及基于类激活映射的图像语义分割。

1.2.1 全监督图像语义分割

全监督语义分割（Fully Supervised Semantic Segmentation, FSSS）采用大量像素级标注图像数据集来训练模型，从而准确地对图像中每个像素的语义类别进行分类。FSSS 主要利用 DL 方法实现，通过大规模标注样本驱动 CNN 模型进行端到端训练，学习图像的深层语义信息以完成像素级的语义分割。

2012 年，Hinton 研究组[6] 提出了结合 CNN 和 DL 方法的 AlexNet，创建了第一个深度神经网络（Deep Neural Networks, DNN）模型，并于同年在 ImageNet[7] 图像识别挑战

赛中获胜，验证了 DNN 的优越性。此后，VGG[8] 、GoogleNe[9] 、ResNet[10] 等一系列经典网络架构相继问世，进一步推动了 DL 技术在图像处理领域的广泛应用。伴随这一发展趋势，研究者开始探索将 DL 方法应用于图像语义分割任务。

2015 年，Long 等人[11] 提出全卷积神经网络（Fully Convolutional Network, FCN），首次利用深度学习有效地实现了图像的语义分割。如图 1-2 所示，FCN 提出构建一个全卷积网络，使网络可以接受任意尺寸的图像，并使用反卷积层将特征图映射回原始图像大小。然而，FCN 的反卷积过程实际上只是一个转置卷积运算，在重建图像细节方面表现有限，主要存在对局部像素间关系建模不足、上下文感知能力较弱、空间结构信息保持不充分等问题，导致分割结果在空间一致性和特征层次平衡等方面存在不足。为了克服上述局限，后续研究陆续提出了多种改进方法。

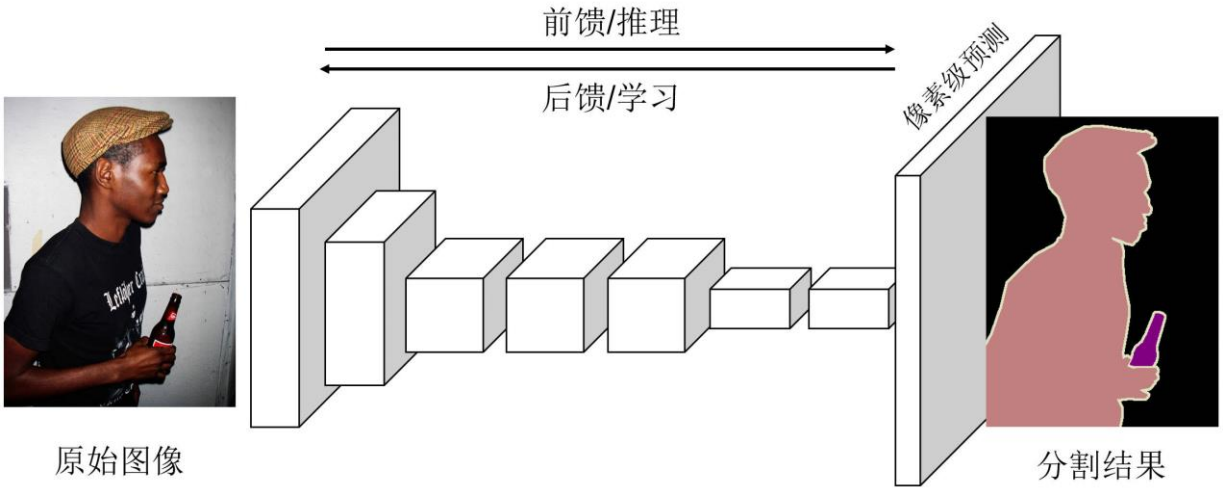


图 1-2 FCN 模型结构图

DeepLab 系列算法实现了全卷积图像语义分割方法，以 FCN 为基础，通过不断优化模型，提高神经网络的空间敏锐度，达到提高分割效果的目的。DeepLabV1[12] 引入了空洞卷积（Atrous Convolution）[13] ，在保持与普通卷积相同计算量的基础上扩大感受野，使得特征图在具有相同的下采样率时能够捕获更多全局信息。DeepLabV2[14] 在 DeepLabV1 的基础上做了部分改进，一方面采用残差网络（ResNet）的残差学习机制缓解深层网络训练中存在的细节丢失问题，提高信息完整性；另一方面设计了空洞空间金字塔池化（Atrous Spatial Pyramid Pooling, ASPP）模块，通过并行设置不同空洞率的卷积操作，实现对多尺度特征的融合，以适应场景中目标尺度多样的问题。DeepLabV3[15] 进一步改进了 DeepLabV2，将空洞卷积应用于级联模块，并在 ASPP 中引入批量归一化（Batch Normalization, BN）以提升训练稳定性，并在网络末端引入全局平均池化（Global Average

Pooling, GAP）层以增强模型对全局语义的建模能力。

基于编码器-解码器（Encoder-Decoder）结构的方法（如图 1-3 所示 SegNet 模型）通过构建对称的网络设计，有效缓解了传统 CNN 中因池化操作引发的空间分辨率下降和部分像素位置信息丢失问题。在该结构中，编码器通过卷积层与池化层构建下采样路径以获取像素的位置信息，而解码器则通过反卷积操作实现像素的位置信息还原，实现输入图像与分割结果的分辨率对齐。作为该架构最经典的工作，SegNet[16] 的编码器采用移除了全连接层（Fully Connected Layer, FC）的 VGG16 骨干网络进行特征提取，解码器对编码器生成的特征图作上采样，并通过记录最大池化索引实现精准上采样，显著提升了分割边界的精准度。U-Net[18] 通过特有的跨层连接机制改进了标准编码器-解码器结构，编码器通过卷积层和池化层提取输入图像的特征，解码器对特征图作上采样，将编码器与解码器对应层进行通道拼接，实现多层次融合，从而减少因为池化过程中的信息损失。DeepLabV3+则[19]则进一步整合了编码器-解码器思想，将 DeepLabV3 网络作为编码器，用以提取丰富的上下文信息，并引入轻量级解码器以增强空间分辨率的恢复能力。同时，DeepLabV3+在编码器与解码器中均引入深度可分离卷积，在保持分割精度的同时显著降低了模型参数量与计算开销，提升了整体运算效率。2018 年，Wu 等人[17] 提出双通道架构，扩展了编码器-解码器结构，其通过并行处理像素标注目标图像和所有源图像，在一定程度上解决了图像标注问题并实现精度提升。

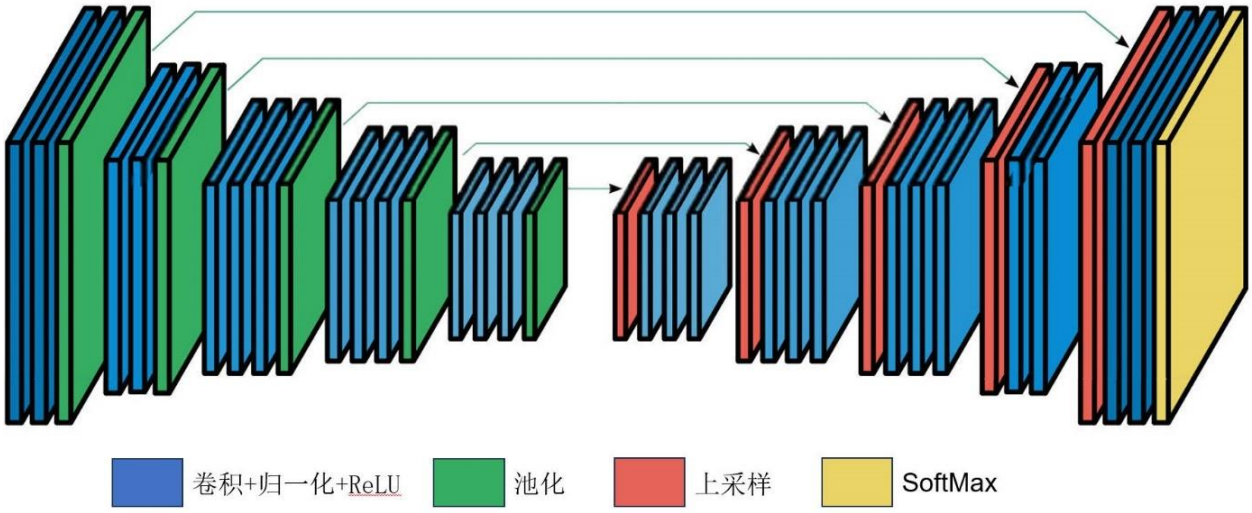


图 1-3 SegNet 模型结构图[16]

近年来，随着注意力机制（Attention Mechanism）[20] 的兴起为图像语义分割提供了新的技术路径。注意力机制的引入可以使网络更好地关注上下文信息，一定程度上解决了

FCN 因有限感受野难以捕获长距离依赖信息的问题。PSANet[21] 是首批将注意力机制成功应用于语义分割的模型之一，其通过生成预测注意力图，使每个像素根据其上下文环境进行动态信息聚合，从而实现更加精细的分割效果。随着自注意力机制（Self-Attention Mechanism）的引入，该领域取得了进一步突破。Wang 等人提出 non-local 模块[22]，首次实现了任意像素对之间的长距离建模，结构如图 1-4 所示。Non-local 模块通过全局信息交互机制增强了特征表达能力，显著提升了语义建模的广度和深度，但其计算复杂度随图像尺寸平方级增长的问题制约了其实用性。为提升效率，Huang 等人[23] 提出了 CCNet，构建了十字交叉注意力模块（Criss-Cross Attention, CCA），该机制通过仅沿水平与垂直方向聚合上下文信息，大幅降低了计算开销。DANet[24] 从多维度注意力建模出发，设计了并行的空间注意力与通道注意力模块，分别从空间位置和通道维度捕捉特征间的依赖关系，实现了局部与全局信息的融合，从而增强了语义分割模型的表达能力与泛化性能。

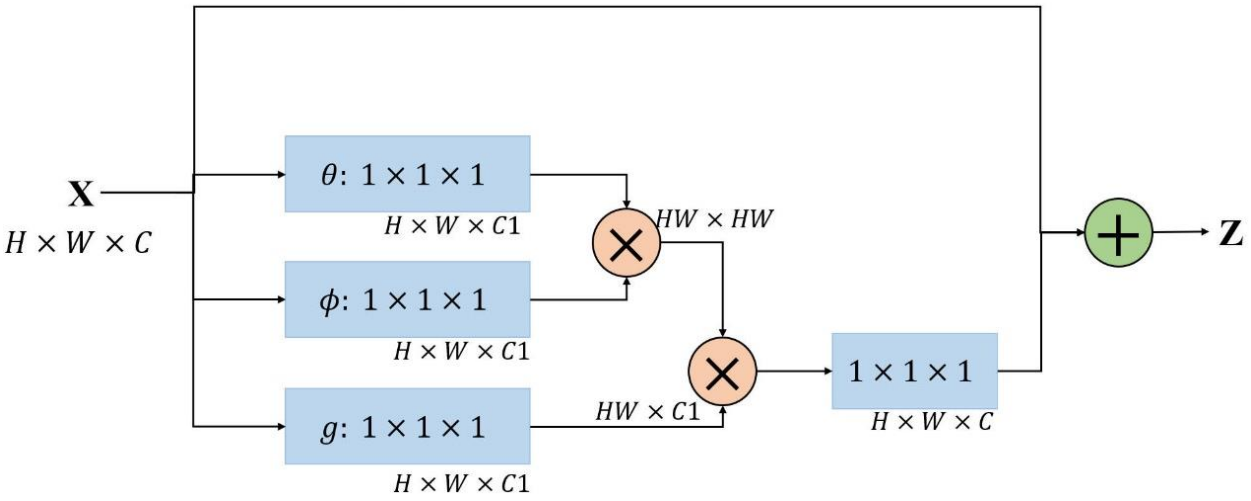


图 1-4 自注意力机制结构图

但是，CNN 的训练依赖大规模的标注数据，而语义分割所需的像素级标注成本高昂，获取过程十分困难且耗时耗力。为此，研究者们开始探索弱监督学习方法，旨在利用低成本的弱标注实现有效监督，当前这一方向已成为研究的重要分支。

1.2.2 弱监督图像语义分割

相较于全监督的语义分割方法，WSSS 采用弱标注来指导网络训练，显著降低了人工标注成本。典型的弱标注形式包括边界框标注、涂鸦标注、点标注以及图像级标注等。

边界框标注法通过矩形框来近似标注目标类的位置信息。与像素级标注相比，这种标注方法虽然造成了一些详细物体信息的遗漏，但大大降低了标注成本，并且在优化条件下分割性能可以做到接近 FSSS。Dai 等人[25] 提出了基于 FCN 的 BoxSup 网络，利用多尺

度组合分组[26] 生成初步候选区域。DeepCut[27] 进一步改进传统 GrabCut 算法[28] ，将分类问题转化为密集连接的条件随机场（Condition Random Field, CRF）[29] 内的能量最小化问题。Khoreva 等人[30] 将弱监督任务视为噪声标注问题，提出在不改变训练框架的前提下提升分割精度的策略。Ma 等人[31] 提出了“像素即实例”策略，采用类似于多实例学习[32] 的约束损失和标签平衡损失来训练弱监督语义分割模型，将每个像素视为加权的正实例或负实例，有效利用边界框标注信息。

涂鸦标注相较于边界框标注更具灵活性，能够适应复杂目标形状，但也其在位置和范围精度上存在不足。ScribbleSup 网络[33] 通过使用 GrabCut 算法[28] 标注训练数据，并利用 FCN 网络进行图像分割，实现从涂鸦标注到精准分割的转化。Tang 等人[34] 提出归一化切割损失函数，用以提升模型对涂鸦标签的判别能力，从而提高分割精度。BoundaryMix 方法[35] 通过生成伪训练图像-标注对来补足原始图像中缺失的语义边界。CC4S[36] 在网络中嵌入了随机游走（Randow Walk, RW）模块[54] ，并采用自监督训练方法，使性能进一步接近全监督方法。

点标注具有耗时少、操作简单的优点，但其提供的位置信息有限。Bell 等人[37] 首先提出了基于点标注的斑块级 CNN 分类器，并将其作为 CRF 中的一元势能项。在此基础上，Bearman 等人[38] 提出 WTP 算法，结合语义分割中常用的先验知识，实现对目标区域的粗略定位。

1.2.3 基于类激活映射的弱监督图像语义分割

作为弱监督学习中最简化的标注形式，图像级标注仅需将图像中目标对象的类别标签作为标注内容，具有最高的标注效率。但同样，由于图像级标注缺乏监督信息，导致利用图像级标注的 WSSS 效果与 FSSS 存在明显差距。为弥补图像级标注的不足，Zhou 等人[39] 提出了类激活映射（Class Activation Mapping, CAM）方法实现了从图像级标注导像素级伪标注的转换，为基于图像级标注的 WSSS 方法提供先验信息。CAM 算法已成为了利用图像级标注的 WSSS 方法的关键算法。然而，CAM 算法存在显著的局限性：其激活区域往往仅覆盖目标对象最具判别性的局部特征，导致关键语义信息缺失。因此，基于类激活映射的弱监督语义分割方法已成为该领域研究的主流。研究人员提出了许多 CAM 算法的改进方案，主要包括以下六类技术路径：

（1）基于区域生长与擦除的方法。该方法的核心目标是拓展前景目标，以此来解决由于标注信息过少而带来的激活区域过小的问题，主流的方法包括种子区域增长策略和对

抗性擦除策略。Kolesnikov 和 Lampert[40] 提出了一个经典框架 SEC，构建了包含种子损失、扩展损失和约束损失的三元复合损失，有效实现目标扩展。Huang 等人[41] 提出 DSRG 算法，在 SEC 方法的基础上迭代扩展了种子区域，实现动态监督，有效改善 SEC 方法存在的目标不完整问题。Jiang 等人[42] 引入了在线注意力积累机制，通过积累注意力来捕捉类激活图在分类网络不同训练阶段中的不同对象信息，从而获得扩展的类激活图。MDC 算法[43] 借助空洞卷积扩大感受野，使前景目标的激活范围得到有效扩大。EDGE 算法[45] 建立多尺度特征融合架构，融合浅层网络与深层网络，获得更完整的目标轮廓。擦除思想由 Chaudhry 等人[46] 首次提出，在 DCSP 算法中通过显著性目的部分擦除引导网络发现残余目标，进而融合生成更完整的伪标注。HaS 方法[47] 在训练时对图像进行切块并随机隐藏，促使网络学习到更全面的区域。Wei 等人[48] 基于 CAM 算法生成的类激活图，提出了 AdvEarsing，通过擦除显著性区域，驱使分类器挖掘剩余非显著性区域。Hou 等人[49] 设计了名为 SeeNet 的三分支网络，解决了擦除过程背景噪声被过度激活的问题。Redondo-Cabrera 等人[50] 构建了一个三分支网络，在某一分支中通过随机擦除图像区域来实现类激活图的扩展。Lee 等人[44] 提出的 AdvCAM 算法通过反对抗扰动策略干预显著性区域外部区域的特征学习，实现更完整的目标激活。Zhang 等人[51] 在分析了前向传输擦除的局限性后，设计由差异损失和交集损失构成的“分割与合并”优化策略，利用两个分类器进行对抗训练，yi 增强目标区域挖掘的完整性。Sun 等人[52] 的 InferCAM 网络在推理阶段引入擦除机制，通过分块与推理后擦除的方法挖掘剩余的目标区域，有效提升分割精度与效率。

（2）像素亲密度建模方法。基于像素亲密度的策略旨在利用像素间的语义相似性扩展初始激活区域。所谓像素亲密度，指的是图像中不同像素在高层特征空间中的语义关联特征。Ahn 等人[53] 提出 AffinityNet 建模高层特征的语义亲密度关系，基于类激活图生成正负样本对并定义相应损失函数，再生成转移矩阵并进行 RW[54] ，扩张类激活图。随后，Ahn 等人在此基础上又提出了 IRNet[55] ，将图像边界信息和语义亲密度相结合，通过训练网络预测边界，并借助 RW 扩大激活区域。Chen 等人[56] 再 IRNet 的基础上持续改进，采用滑动窗口的方式生成伪边界标注。Yi 等人[57] 利用超像素，在分割网络中交替地训练超像素区域分割网络和亲密度网络，以提高分割精度。

（3）基于显著性图的方法。基于显著性图的方法被广泛应用于 WSSS 任务，利用显著性图来区分前景区域和背景区域。一种典型的方法是直接将显著性图中提供的背景信息

用于伪标注生成，经典方法包括 OAA[58] ，DSRG 和 SeeNet 等。在此基础上，为进一步提升显著性图的利用效果，Wei 等人[59] 提出 STC 框架，实施渐进式训练策略，按照图像样本由易到难的顺序逐步优化模型。Yao 等人[60] 借助 Non-local 模块[22] 建模显著性图中像素间的语义亲密度，进而引导前景区域的逐步扩张。Xie 等人[61] 提出了 C²AM 方法，将前景区域作为正样本、背景区域作为负样本，构建对比损失函数以挖掘图像前后景信息，生成伪标注。

（4）基于多图像协同的方法。该方法的思想是通过联合建模多幅图像，挖掘跨图像的共享语义信息。Fan 等人[62] 认为，跨图像相关区域可以补充表征，提出端到端交叉图像亲和模块，并基于此设计了 CIAN 网络，通过像素级的跨图像关系增强前景区域的语义一致性和完整性。群组语义挖掘网络[63] 由高效协同机制表示一对图像之间的潜在关系，并提出图剔除层鼓励模型学习更准确、更完整的对象响应。Zhang 等人[64] 设计了 I² C 模型，基于不同对象间的像素相似性实现互补性学习，通过两种约束策略提升同类目标的特征一致性，从而获得更准确的物体定位结果。Qin 等人[65] 设计 AMR 方案，提出了双分支补偿策略，通过两条分支分别作激活调制和校准的不同处理后形成互补机制，并引入交叉伪监督来实现语义相似正则化。

（5）基于自监督学习的方法。随着注意力机制的提出和发展，自监督策略在 WSSS 中的应用逐渐受到研究者们的重视。Shimoda 等人[66] 认为伪标注与其映射的标注之间存在显著差异，因而提出 SSDD 网络来预测伪标注噪声，并通过去除噪声来提高准确度。Chang 等人[67] 引入自监督学习方法挖掘类别子类信息，在每个注释的父类中生成伪子类别信标签，借此提供更强的监督信号以增强前景目标的识别能力。Wang 等人[68] 提出了自监督等变注意机制，通过改变图像尺度来增加自监督信息，进一步增大了前景目标的激活区域。Chen 等人[69] 提出 SIPE 算法，利用特定图像原型探索形成捕捉完整区域特定类激活图，并提出一般特定一致性来构建原始类激活图和特定类激活图的一致性，为网络提供额外的监督。Lee 等人[70] 设计 AMN 网络，利用初始 CAM 生成的噪声伪标注作为额外自监督信号，通过重分配激活值来减少前景内部的不平衡性。Lei 等人[71] 提出扰动类激活图方案，通过在特征图中随机注入噪声来增强高置信度通道的鲁棒性，从而扩展目标激活范围。

（6）Transformer 方法。Transformer 结构[72] 利用多头自注意力机制实现对全局上下文的建模，近年来开始用于图像处理领域。Gao 等人[73] 设计了 TS-CAM 网络结构，依靠全局自注意力实现语义感知激活并避免部分激活，有效增大了前景面积。Ru 等人[74] 证实

了 Transformer 图像语义亲和性的自我注意之间的内在一致性，从而提出了 AFA 网络。Li 等人[75] 基于 Conformer 提出融合架构，通过结合 Transformer 分支的自注意力特征与卷积神经网络分支生成的类激活图，实现互补优化，从而显著提升伪标注的准确性和完整性。

1.3 本文主要工作

尽管上述研究在 WSSS 领域已取得积极进展，但现有基于类激活映射的 WSSS 与 FSSS 直接仍存在明显的性能差距。在当前的研究中，基于类激活映射的 WSSS 的缺陷主要是由于标注信息过少而导致的类激活图不够准确的问题。针对当前 CAM 算法存在的激活不平衡与激活不准确问题上的不足，本研究提出了一种新的模型——LTCAA，设计并实现了标签阈值模块、一致性约束架构，并引入自注意力机制，有效弥补了弱监督图像语义分割与全监督图像语义分割之间的性能差距。通过实验验证，本研究证明了 LTCAA 模型在 VOC12 数据集上的表现。通过系统的网络训练，LTCAA 模型成功生成了高质量的伪标注，并利用伪标注作为分割网络的像素级监督信息，得出了准确的最终分割结果。实验结果显示，本文设计的各个模块均为精度提升做出贡献，证明了本文各模块的有效性；LTCAA 模型在性能指标上优于对比算法，充分证明了其有效性和先进性。

在算法应用层面，本研究将 LTCAA 模型针对性地应用于遥感图像数据集。考虑到遥感图像数据集的特性，如大尺度、高分辨率以及复杂的地表覆盖类型等，本研究对遥感图像进行了精细的数据预处理工作，以确保模型能够充分发挥其性能。在实验研究阶段，将 LTCAA 模型应用于 ISPRS Postdam 遥感数据集，以进一步验证其在遥感图像分割任务中的性能。经过数据处理和网络训练，LTCAA 模型在 ISPRS Postdam 数据集上取得较高质量的伪标注，并表现出了较好的分割性能。

1.4 论文主要架构

本文共分为五章，各章节内容安排如下：

第一章为引言，主要介绍基于类激活映射的弱监督图像语义分割的研究背景及其意义，系统回顾国内外在全监督与弱监督图像语义分割方面的研究现状，并围绕不同类型弱标注信息的利用策略，对现有方法进行分类与比较，重点分析基于类激活映射方法的优势与局限。在此基础上，明确本文的研究内容与技术路线，概述所提出方法的主要创新点。

第二章为相关理论基础，介绍本文涉及的核心理论与技术体系。首先介绍了图像的语义分割网络相关的基本原理和算法模型。其次，着重对本文所采用的基于类激活映射的弱

监督图像语义分割算法进行介绍。

第三章提出了一种基于标签阈值的一致性注意力架构 LTCAA 模型。首先介绍本文设计的三个模块，包括标签阈值模块、一致性约束模块和自注意力模块，并有针对性地设计了损失函数，构成 LTCAA 模型，一定程度上解决了 CAM 算法中存在的对目标前景类激活不准确和激活不充分等问题。随后进行实验分析，通过消融实验证明模块有效性，通过对比实验验证 LTCAA 模型能够获取更高质量的伪标注，LTCAA 模型有助于图像语义分割任务。

第四章为 LTCAA 模型在遥感图像上的实际应用案例。首先介绍将 LTCAA 模型应用于遥感图像的意义和必要性。随后，对遥感数据集、数据预处理过程、图像级标注获取方法展开介绍。最后，进行实验分析，证明了 LTCAA 模型能够很好地获取遥感图像的伪标注，并在图像建筑物提取分割任务中取得好的结果。

第五章为总结与展望，主要回顾本文的核心研究内容与主要结果，并结合当前研究的不足之处，探讨未来的改进方向与研究前景。

2 相关理论基础

2.1 图像语义分割

2.1.1 全卷积网络语义分割

全卷积神经网络（FCN）作为深度学习领域的重要创新，首次将深度学习技术成功应用到图像语义分割任务中。与传统的卷积神经网络（CNN）仅能输出图像级分类结果不同，FCN 进行了关键性的架构创新。FCN 使用全卷积层结构，这使得网络能够灵活接受任意尺寸的输入图像，有效摆脱了传统模型对固定输入尺寸的限制；并且，FCN 引入反卷积层，通过构建上采样机制使网络最终生成与输入图像分辨率一致的输出特征图。FCN 的设计能够使网络精确到每个像素点，实现了像素级的类别预测，为后续图像语义分割任务的发展提供了技术支撑，FCN 语义分割网络架构图如图 2-1 所示。

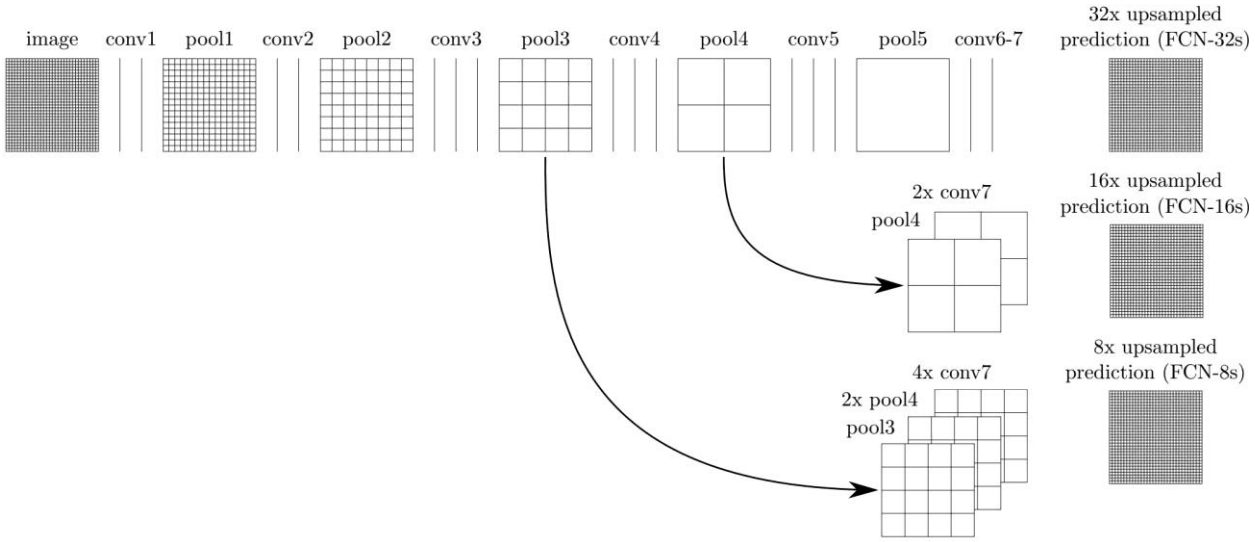


图 2-1 FCN 语义分割网络架构图[11]

从具体实现的角度上，FCN 主要包括全卷积化、上采样和特征融合三个核心思想。首先，全卷积化是 FCN 的基础，它用卷积层替换传统分类网络末端的全连接层，这一设计使网络能够接受任意尺寸的输入图像$x \in \mathbb{R}^{H\times W\times C}$，并且输出一个包含空间位置信息的空间特征图$F \in \mathbb{R}^{h\times w\times c}$。全卷积化策略不仅保留了卷积网络对局部特征的提取能力，还通过维持特征图空间信息从而支持密集预测任务。其次，FCN 构建上采样机制来弥补在下采样过程中的分辨率损失。FCN 利用公式（2-1）所示的反卷积运算对高层语义特征进行上采样，其中W是反卷积核权重，Y是输出特征图。反卷积层的权重可以初始化为双线性插值，并

在训练的过程中通过反向传播优化，从而自适应地学习特征的空间映射关系。与此同时，FCN 引入跳跃连接来实现多尺度特征的融合。深层网络输出的高维特征包含丰富的语义信息，但空间分辨率较低；而浅层特征保留了更多的细节，但其语义信息较弱。因此，FCN 通过公式（2-2）进行融合操作，将深层特征F_{deep}上采样，达到与浅层特征$F_{shallow}$相同的分辨率后，对其做逐元素相加或拼接，形成兼具全局语义与局部细节的融合特征F_{fused}。最后，再用公式（2-3）所示的 Softmax 函数来预测各像素点(i,j)属于每个类的概率p，将最大概率值的类别作为该像素点的分类。

$$Y = Deconv(F, W) \tag{2-1}$$

$$F_{fused} = F_{deep} \oplus F_{shallow} \tag{2-2}$$

$$p(i,j) = \frac{e^{F(i,j)}}{\sum e^{F(i,j)}} \tag{2-3}$$

FCN 的出现使得图像语义分割的算法研究进入了新的阶段，其核心的全卷积思想与特征融合机制为后续研究如 U-Net、DeepLab 等经典架构奠定了技术基础。但是，受到早期深度学习技术发展水平的限制，FCN 的局限性也很明显。FCN 的反卷积虽然能恢复特征图分辨率，但本质上只是转置卷积的操作，仍无法避免高频细节（如边界和小目标）容易丢失的问题；同时，FCN 缺乏对像素间长距离依赖关系的考虑，传统的卷积操作的局部感受野特性，使得深层特征图虽然可以捕获全局语义信息，却对局部上下文信息的建模能力不足，导致对复杂场景的理解有限。

2.1.2 DeepLab 系列图像语义分割网络

在 FCN 的基础上，DeepLab 分割网络被提出。目前，DeepLab 系列分割网络已经包括 DeepLabV1、DeepLabV2、DeepLabV3、DeepLabV3+等多种模型，针对图像分割领域的几个常见问题给出了相应的解决策略。并且，DeepLab 系列分割网络的结构较为简单，时间复杂度较低，既能保障分割精度，又兼顾计算效率，因而成为全监督图像语义分割研究中广泛采用的模型。

（1）DeepLabV1 在 FCN 的基础上引入了空洞卷积。如图 2-2 所示，空洞卷积通过在传统卷积核中插入“空洞”，实现了无需池化操作即可扩大感受野的效果。空洞卷积的引入有效避免了 FCN 中反卷积过程带来的特征图分辨率下降问题，同时使网络在不增加参数量的情况下提高了计算效率。网络架构层面，DeepLabV1 基于 VGG16 网络进行改进，将

原网络末端的全连接层替换为空洞卷积层，进一步强化了网络的密集预测能力。并且，DeepLabV1 在输出层后添加了条件随机场（CRF）模块，通过对像素间的空间关系进行建模来优化初始分割结果。在训练过程中，DeepLabV1 网络首先通过像素级交叉熵损失函数优化参数，然后在推理阶段应用 CRF 做后处理，优化边界，减少误分类，显著提升了分割精度。

然而，DeepLabV1 也存在显著的技术瓶颈。其一，网络对小目标的分割效果有限，深层特征对小目标不敏感。其二，网络缺乏多尺度上下文建模，这就限制了其在复杂场景下的性能。尽管如此，DeepLabV1 在细节保留和边界处理方面取得的突破，为后续 DeepLab 系列算法的策略改进奠定了理论基础，其在技术上存在的局限也为后续研究提供了演进方向。

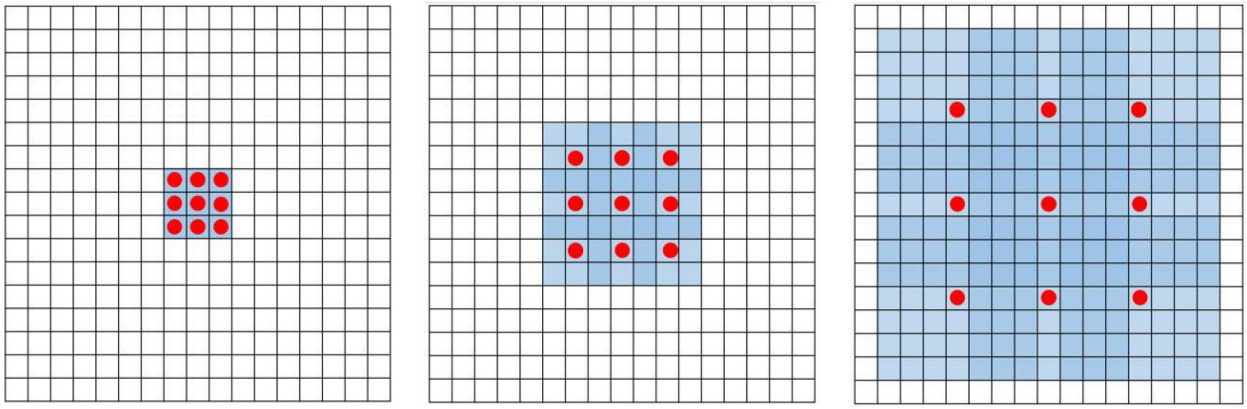


图 2-2 空洞卷积

（2）DeepLab V2 在 DeepLabV1 的基础上进行改进，引入了空洞空间金字塔池化（ASPP）。如图 2-3 所示，ASPP 并行使用多个不同空洞率的卷积层，构建了多尺度特征提取框架，同时捕捉局部细节和全局语义信息。这种分层级的特征提取机制使得网络能够适应不同尺度的物体，显著增强了对复杂场景的理解能力。网络架构层面，DeepLabV2 以 ResNet 网络替代原有的 VGG 网络，利用残差单元缓解深层网络训练中存在的梯度消失问题。将 ResNet 末端的全连接层替换为 ASPP 模块，网络在保留高分辨率特征图的同时，实现了多尺度上下文信息的融合。在最后推理阶段应用 CRF 进行后处理，进一步强化边界定位精度。

DeepLabV2 引入的 ResNet 显著的提升了特征表达能力，其跳跃连接结构使得网络能够学习到更复杂的语义层次。ASPP 模块通过不同空洞率的卷积核采样，有效捕获了多尺度上下文信息，这很好地提升了模型对复杂场景的适应能力。同时，DeepLabV2 保留空洞

卷积和 CRF，也使特征图分辨率问题和边界精度问题均能够得以解决。

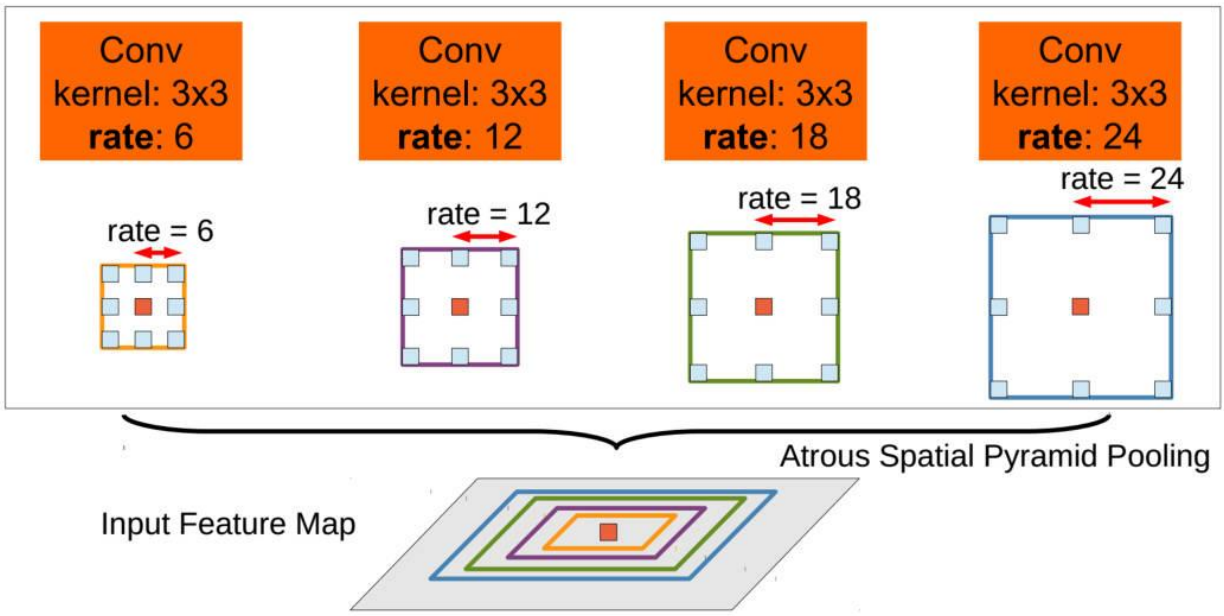


图 2-3 ASPP 结构图[14]

（3）DeepLabV3 进一步地对 DeepLabV2 做出改进，做出了系统性架构优化。DeepLabV3 沿用了 DeepLabV2 的骨干网络 ResNet 的同时，创新性地提出将空洞卷积应用在 ResNet 的级联模块，如图 2-4 所示。

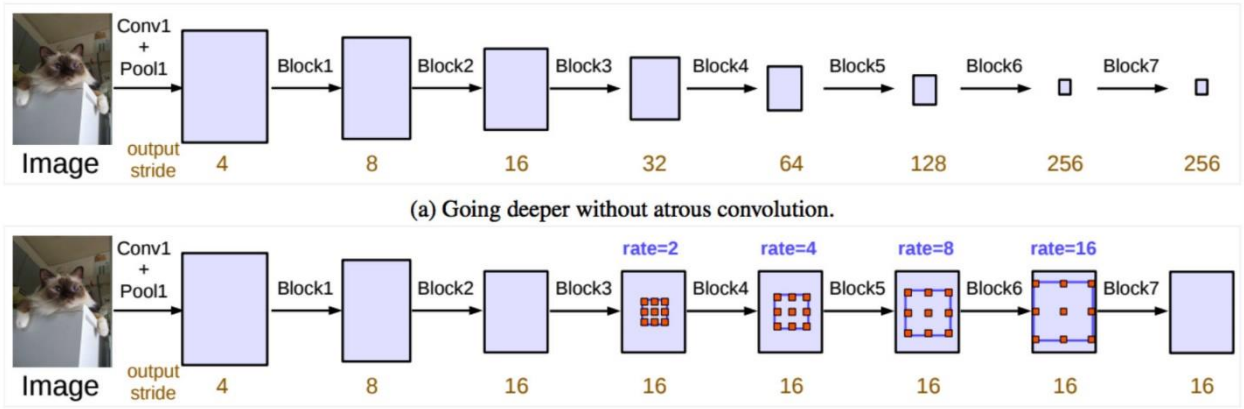


图 2-4 DeepLabV3 结构图[15]

DeepLabV3 通过复制 ResNet 末端的 Block 并将复制的 Block 进行级联堆叠，构建深层特征提取结构——每个 Block 有三个卷积层，除最后一个 Block 外，其余 Block 的磨蹭卷积步长都设置为 2，该设计在减少下采样次数的同时，显著扩大了感受野范围，使网络可以更容易地捕获深层块中的远程信息。并且，DeepLabV3 应用多重网格策略，进一步强化了多尺度上下文建模能力。通过在深层网络中的不同 Block 中应用不同的空洞率，网络针对小目标和大场景进行分层的特征提取，捕获不同尺度的上下文信息。与此同时，

DeepLabV3 进一步对 ASPP 模块进行改进，添加了 BN 层，进行归一化处理，加速模型收敛并缓解过拟合现象；在网络末端加入 GAP 层，实现对全局信息的获取。值得注意的是，DeepLabV3 舍弃了 CRF 的后处理环节，在不影响分割结果的同时提高了网络的处理速度。

DeepLabV3 在 DeepLabV2 做出的改进，使得分割网络更好地捕获深层特征的信息，进一步增强了多尺度特征提取能力；加入 BN 层更是使训练速度加快，而且在一定程度上避免过拟合。去掉了 CRF 的处理更是使得网络处理速度进一步提升。

（4）DeepLabV3+则是在 DeepLabV3 的基础上引入了编码器-解码器的结构，实现图像语义分割技术的关键突破。在编码器端，该模型基于 DeepLabV3，使用空洞卷积和 ASPP 模块来捕获多尺度上下文信息。而解码器模块则成为 DeepLabV3 + 的核心设计。如图 2-5 所示，解码器模块将编码器的特征做上采样，然后与来自编码器的浅层特征进行连接。这种特征融合机制充分发挥了浅层特征的高分辨率优势与深层特征的强语义表征能力，在物体边界定位和小目标捕捉方面表现突出，显著提升了分割结果的空间精准度。DeepLabV3+还提出深度可分离卷积策略，将一个传统标准卷积分解为深度卷积，然后再进行1×1卷积。深度卷积对每个输入通道独立执行一个空间卷积，专注于局部特征提取；逐点卷积用于合并深度卷积的输出，通过轻量级计算实现特征维度变换。该设计使模型在保持精度的同时，大幅削减计算量与参数量，提升了网络效率。此外，DeepLabV3+改用了 Xception 作为主干网络，进一步提高了网络的分割性能。编码器-解码器架构、深度可分离卷积与 Xception 骨干网络的创新和结合，DeepLabV3 +为语义分割任务提供了高性能解决方案。

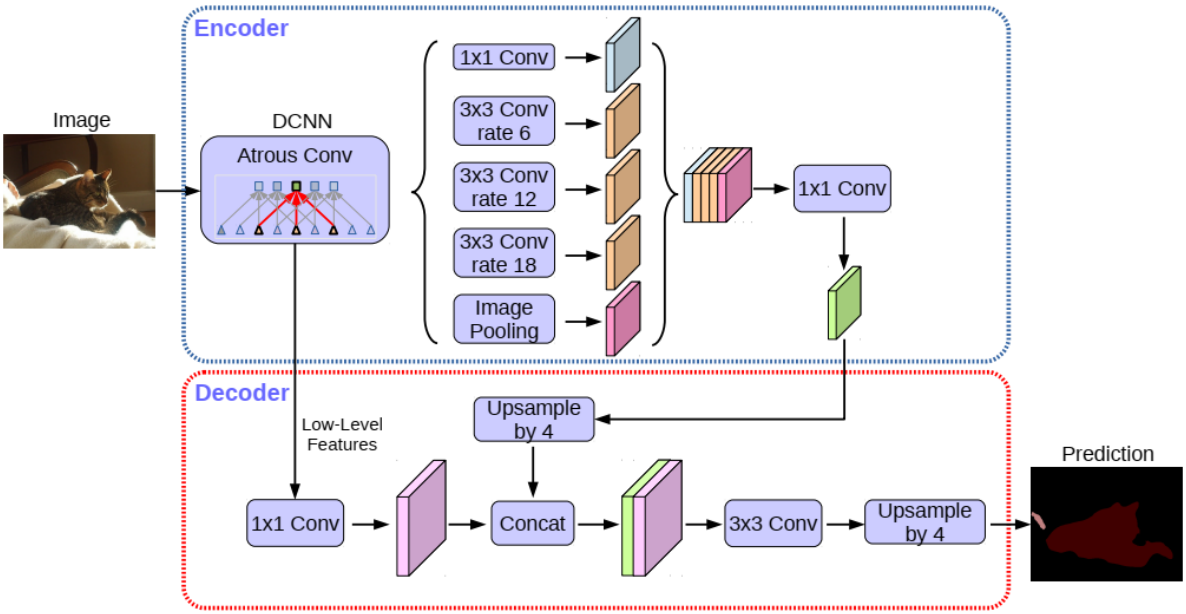


图 2-5 DeepLabV3+结构图[19]

2.2 基于类激活映射的弱监督图像语义分割

由于图像级标注仅包含物体的类别信息而不提供其空间位置信息，因此利用图像级标注开展 WSSS 在计算机视觉领域面临较大挑战。目前，主流方法普遍依赖于 CAM 算法，通过识别物体最显著的区域以实现目标初步定为，其流程如图 2-6 所示，整体流程分为三个部分进行。首先是类激活图的生成阶段：通过图像级标注训练分类网络来提取特征图，并结合 FC 权重生成类激活图；随后对类激活图进行上采样处理，使其与输入图像保持相同的分辨率，获得像素级初步预测结果；然后在原始类激活图的基础上进行优化，进一步激活前景中非显著的区域，同时抑制错误激活的背景区域，获取类激活映射。随后，利用类激活图生成伪标注：通过引入 AffinityNet 等模型提升伪标注的像素级质量。最后，进入图像分割网络训练阶段：在过程与 FSSS 的训练方法一致，不同之处在于将真实标注替换为生成的伪标注，以此完成像素级分割预测。以下将介绍类激活图的生成方法、伪标注生成和分割网络训练的具体内容。

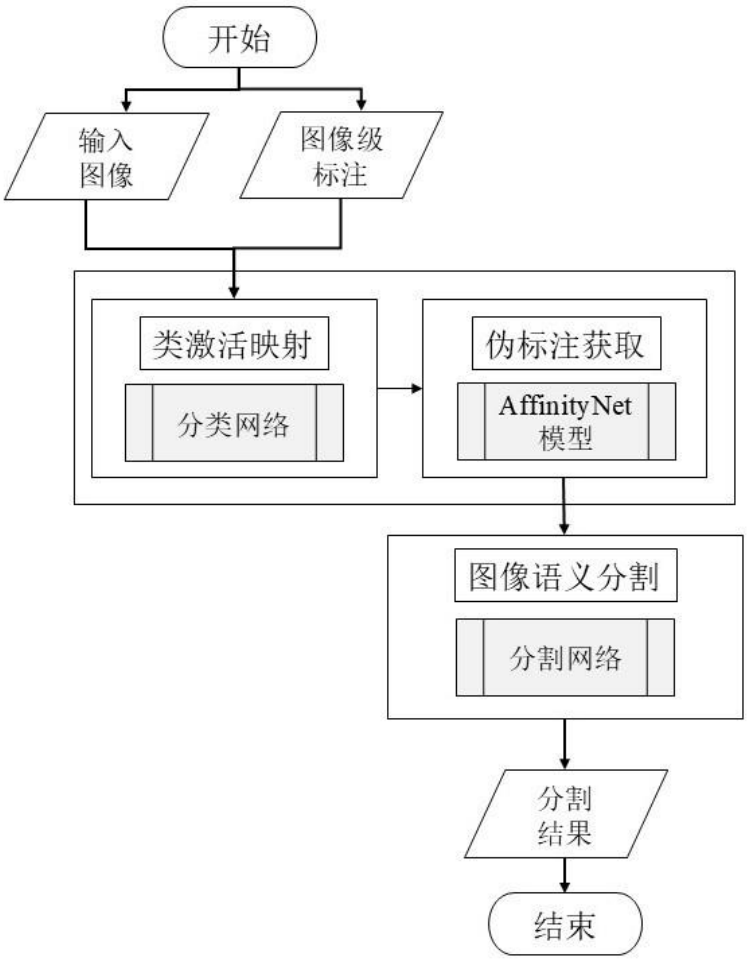


图 2-6 基于类激活映射的弱监督图像语义分割算法流程图

2.2.1 类激活映射

类激活映射（CAM）是利用图像级标注进行弱监督的图像语义分割的一种常用实现方式，由 Zhou 等人在 2016 年提出。CAM 算法是一种图像类问题的可解释性算法，利用 CNN 提取的特征图得到原始图像各个部分的重要性关系，通过为特征图赋予不同权重的方式生成类激活图。CAM 算法的架构图如图 2-7 所示。

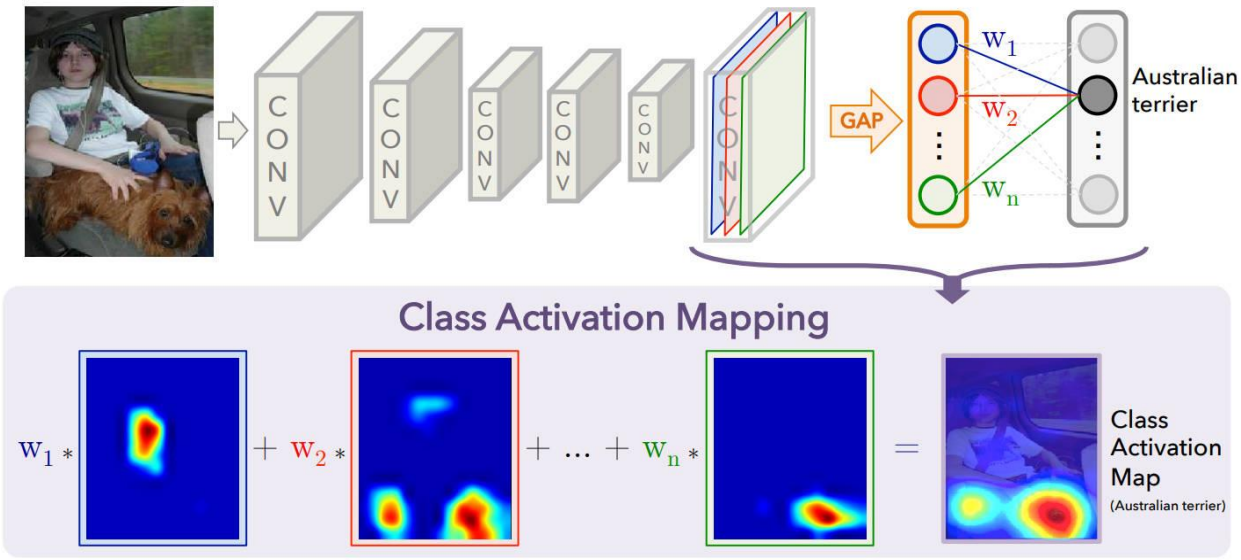


图 2-7 类激活映射[39]

CAM 算法通过全局平均池化（GAP）和全连接层（FC），将 CNN 提取的特征映射到类别激活图，从而定位图像中与特定类别相关的区域。CAM 算法的实现过程可以分为以下几个步骤。首先，输入的图像被定义为$x \in \mathbb{R}^{H\times W\times C}$，其中$H \times W$代表空间维度，$C$为通道数。CNN 通过卷积层提取特征图，随后将特征图送入 GAP 层，如图 2-7 所示，GAP 在最后一个卷积层输出每个单元的特征图的空间平均值，表示该特征图对目标类别的整体贡献。将平均池化后的特征图输入 FC，得到类别得分$Score$，其中$Score_k$表示第k个类别的得分。最后，通过加权求和特征图映射$F \in \mathbb{R}^{H\times W\times C}$和 FC 权重$\alpha_i^k$得到类激活图$CAM_k(x) \in \mathbb{R}^{H\times W\times C}$，公式如式（2-4），并在得到类激活图之后，通过归一化处理的方法生成热力图，突出显示与类别相关的区域。

$$CAM_k = \sum_{i=1}^{C} \alpha_i^k \cdot F_i \tag{2-4}$$

尽管 CAM 算法简单有效，但其性能高度依赖于 CNN 的特征提取能力，这导致其在复杂场景下的分割精度不同，成为一大局限性。为了克服这一缺陷，Selvaraju 等人[76] 在 CAM 算法的基础上提出了 Grad-CAM 算法，通过引入梯度信息生成更为精确的类别激活图。Grad-CAM 算法的实现过程与 CAM 算法大体上是类似的，但引入了梯度计算过程。Grad-CAM 算法通过计算特征图中每个通道对类别得分的梯度$\frac{\partial Score_k}{\partial F_i}$，捕捉特征图中对目标类别有贡献的区域。对每个通道的梯度进行全局平均池化，如公式（2-5），以得到通道权重α_i^k，该权重反映了每个通道对目标类别的贡献程度；随后，使用通道权重α_i^k对特征图F进行加权求和，得到类激活图$CAM_k(x) \in \mathbb{R}^{H\times W\times C}$，如公式（2-6）。其中，ReLU 函数用于过滤负响应，保留对类别k有正向贡献的区域。最后，对CAM_k进行归一化处理，生成热力图，突出显示与类别 k 相关的区域。

$$\alpha_i^k = \frac{1}{H \times W}\sum_{h}^{H}\sum_{w}^{W}\frac{\partial Score_k}{\partial F_i} \tag{2-5}$$

$$CAM_k = ReLU(\sum_{i=1}^{C}\alpha_i^k \cdot F_i) \tag{2-6}$$

Grad-CAM 算法通过引入梯度信息，改进了 CAM 的定位能力，同时保持了较强的可解释性。然而，当面对包含多个同类目标的图像时，Grad-CAM 算法的定位性能存在一定局限。为应对这一挑战，Chattopadhyay 等人[77] 提出了改进版本的 Grad-CAM++算法。Grad-CAM++算法在 Grad-CAM 算法的基础上，进一步引入二阶和三阶梯度信息，用于计算通道权重α_i^k，如公式（2-7）所示。

$$\alpha_i^k = \sum_{h}^{H}\sum_{w}^{W}\frac{\frac{\partial^2 Score_k}{\partial F_i^2}}{2\cdot\frac{\partial^2 Score_k}{\partial F_i^2} + \sum f_i(x)\cdot\frac{\partial^3 Score_k}{\partial F_i^3}} \cdot ReLU\left(\frac{\partial Score_k}{\partial F_i}\right) \tag{2-7}$$

此外，Grad-CAM++算法还结合空间位置的重要性，生成加权特征图CAM_k，计算方法如式（2-8）所示。最后，对CAM_k进行归一化处理，生成热力图，突出显示与类别 k 相关的区域。Grad-CAM++算法能够更精确地定位图像中与目标类别相关的区域，尤其是在多

目标场景中表现出更好的性能。

$$CAM_k = ReLU(\sum_{i=1}^{C} \alpha_i^k \cdot F_i) \tag{2-8}$$

2.2.2 伪标注的生成和获取

在 2.2.1 节中，通过分类网络训练初步生成的类激活图CAM_k能够直观地展示出图像中哪些区域对于分类结果具有较为重要的贡献。为了进一步提升类激活图精度和语义完整性，通常会结合 AffinityNet 算法和 RW 策略进行优化，最终生成高质量的伪标注，用于图像分割网络的训练。AffinityNet 算法的核心在于它能够精确地捕获像素与其邻域像素之间的语义相似性。通过分析像素间的特征相似性，将类激活图中局部响应的有效信息传播到同一语义类别的其他区域，从而使得整个类激活图对目标类别的表示更加完整和准确。在 AffinityNet 的基础上，结合 RW 策略进一步优化类激活图。RW 策略从全局的角度对类激活图进行调整和改进，使得类激活图中的信息更加的合理有效。最后，将优化后的类激活图转化为伪标注形式，整体流程如图 2-8 所示。

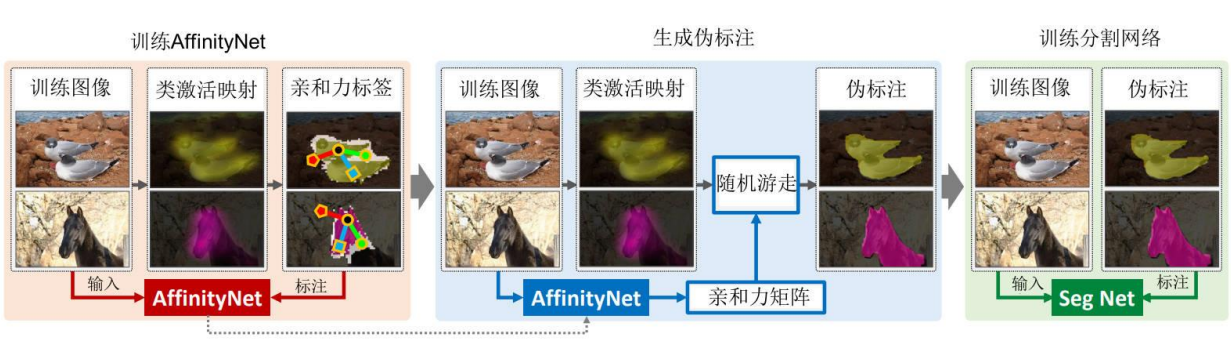


图 2-8 AffinityNet 算法流程[53]

为了生成高质量的伪标注，首先需要从从类激活图CAM_k中选择高置信度的像素对作为正样本，这些像素对在分类过程中具有较高的确定性，能够准确代表目标类别；同时选择低置信度的像素对作为负样本，这些像素对在分类过程中存在较大的不确定性，可能属于背景或其他类别。通过二元交叉熵损失函数对这些样本进行训练，衡量预测值与真实标注之间的差异，输出像素对相似性值。量化像素间的语义关系可以为后续信息传播提供依据。随后，基于像素对的相似性值，构建语义相似性矩阵A，$A_{i,j}$表示像素i和像素j之间的相似性，为像素之间的信息传播提供量化依据。利用语义相似性矩阵A，将类激活图中高

置信度区域的关键信息传播到低置信度区域，把高置信度区域中包含的目标类别信息传播到低置信度区域中以提升低置信度区域对于目标类别的表示能力。经过t次迭代传播，最终得到收敛的置信度图CAM_{new}，如公式（2-9）。

$$CAM_{new}^{t+1} = A \cdot CAM_{new}^{t} \tag{2-9}$$

最后在传播后的置信度图CAM_{new}的基础上，通过阈值分割、连通区域分析、类别分配以及边界优化等操作，生成最终的伪标注。伪标注作为图像分割网络的训练数据，其质量和准确性极大程度上决定了图像分割的效果。高质量的伪标注能够为图像分割网络提供准确的监督信息，使得网络能够学习到更有效的特征表示，从而提高图像分割的精度和性能。结合 AffinityNet 算法和 RW 策略优化类激活图，并生成高质量的伪标注，可以显著提升弱监督图像分割的性能，为图像分割任务提供了重要的技术支持

2.2.3 图像分割网络模型的训练

在本研究中，我们将上述生成的伪标注作为像素级监督信息，用于训练全监督的图像语义分割模型。所采用的分割模型为基于 ResNet-50 主干网络的 DeepLabV2。ResNet-50 借助残差结构，有效缓解了 DNN 训练中常见的的梯度消失和梯度爆炸问题。这种架构不仅提高了模型的训练稳定性，还使模型能够挖掘并学习更为丰富的图像特征。在训练过程中，模型使用交叉熵损失（Cross Entropy Loss）函数，用于度量预测值和真实标注之间的差异。通过最小化交叉熵损失，模型逐步调整参数，使得预测结果更加接近真实值，从而获得更精确的分割效果。此外，DeepLabV2 分割网络采用空洞卷积，通过在标准卷积核中引入空洞（如图 2-2），在不增加参数数量和计算量的前提下扩大感受野。更大的感受野使得模型能够更广泛地获取上下文信息，从而更好地理解图像中不同区域之间的关系。同时，空洞卷积的设计在提升模型感知能力的同时，也降低了网络复杂度，提高了训练效率和泛化性能。

与此同时，DeepLabV2 模型采用多尺度融合的方式以预测更精细化的结果。不同尺度的图像包含不同层次的信息，通过融合这些多尺度信息，模型可以实现对图像中物体更为精确的分割。模型在不同尺度下提取特征，并将这些特征进行融合，从而在最终的分割结果中兼顾全局语义和局部细节。最后，DeepLabV2 模型利用双线性插值将像素级预测的尺寸恢复至输入图像大小，对于待插值点(x, y)，其像素值$g(x, y)$的计算方式如公式（2-9）

所示。

$$g(x,y)=\left(1-\frac{x-x_0}{x_1-x_0}\right)\left(1-\frac{y-y_0}{y_1-y_0}\right)g(x_0,y_0)+\left(1-\frac{x-x_0}{x_1-x_0}\right)\frac{y-y_0}{y_1-y_0}g(x_0,y_1)+\frac{x-x_0}{x_1-x_0}\left(1-\frac{y-y_0}{y_1-y_0}\right)g(x_1,y_0)+\frac{x-x_0}{x_1-x_0}\frac{y-y_0}{y_1-y_0}g(x_1,y_1) \tag{2-9}$$

为了进一步优化分割结果，恢复尺寸后的像素级预测结果通过条件随机场（CRF）建立概率图。CRF 是一种基于概率图模型的方法，能够考虑相邻像素之间的相关性，从而有效缓解目标物体边缘定位不准确的问题。通过引入 CRF 方法，模型能够在像素级预测的基础上，进一步优化分割边界，使得分割结果更加准确和精细；并且考虑相邻像素之间的相关性，有效解决目标边缘定位不准确的问题，从而实现更精确的分割。

2.3 本章小结

本章首先对基于 CNN 的经典全监督图像语义分割网络方法进行介绍，其中 DeepLabV2 网络是本文实验中所采用的分割网络。其次，本章介绍了基于类激活映射的弱监督图像语义分割的流程进行介绍，详细介绍了类激活映射（CAM）算法的具体实现方式，并展开阐述了伪标注生成的过程和图像分割网络模型的训练过程。

3 基于标签阈值的一致性注意力架构模型

3.1 引言

在实际应用中，FSSS 依赖于大量精确标注的训练数据，然而全监督所需的数据标注成本高昂且耗时，往往难以满足。因此，WSSS 算法逐渐成为研究热点。目前，WSSS 算法研究多以 CAM 算法为核心展开，通过借助分类网络生成类激活图，从而精准有效地定位目标物体的位置和形状信息。然而，尽管 CAM 算法在弱监督学习中表现出一定的潜力，但仍存在着局限性，亟需进一步改进和优化。

首先，CAM 算法生成的类激活图通常仅能激活目标物体中最具判别性的区域，难以覆盖完整的物体轮廓。这导致类激活图可能仅激活目标物体的部分边缘，而忽略了整体结构，这导致分割结果的不完整性。其次，CAM 算法还会出现错误激活背景类或非目标类的情况，即类激活图图可能会错误地激活与目标物体无关的区域，从而影响分割的准确性。这些问题限制了 CAM 算法的性能，对后续的分割任务产生了一定不良影响。进一步而言，WSSS 模型在构建过程中使用类别标签作为监督信号，这种监督方式在处理图像仿射变换的时候存在固有缺陷，当输入图像发生旋转、缩放等仿射变换后，类别标签无法同步映射，导致网络缺乏对变换前后的语义约束，成为弱监督与全监督图像语义分割性能差异的关键因素之一。

为应对上面的挑战，本文设计并提出了一种基于标签阈值的一致性注意力架构（Consistency Attention Architecture for Label Thresholds, LTCAA）模型，旨在解决 CAM 算法在弱监督图像语义分割中的局限性。为了解决 CAM 算法对目标前景类激活不准确的问题，本文提出了标签阈值模块。该模块通过动态调整激活阈值，减少了目标区域错误激活的情况。针对 CAM 算法生成的显著图边界不清晰的问题，本文设计了一致性约束模块。该模块通过引入几何一致性约束，确保在不同变换下生成的 CAM 图具有一致性，从而提高了目标前景类激活的准确性。为了解决 CAM 算法激活不充分的问题，本文设计了自注意力模块。该模块通过扩充监督信息，增强了模型对目标物体整体结构的理解，从而解决了激活不充分的问题。为了验证 LTCAA 模型的有效性和先进性，本文在 PASCAL VOC 2012 数据集上进行了实验验证。实验结果表明，LTCAA 模型在弱监督图像语义分割任务中表现出色，优于现有的 CAM 算法。

3.2 算法框架

本研究提出的 LTCAA 模型的算法框图如图 3-1 所示。LTCAA 模型基于卷积神经网络（CNN）构建，模型引入了三个关键模块：标签阈值模块（Label-based Specific Threshold Module，LSTM），一致性约束模块（Consistency Regularization Module，CRM）和自注意力模块（Self-Attention Module, SAM），分别针对 CAM 算法在弱监督学习中的不同局限性进行了改进，显著提升了弱监督图像语义分割的性能。在本节中，我们将详细介绍这三个模块。

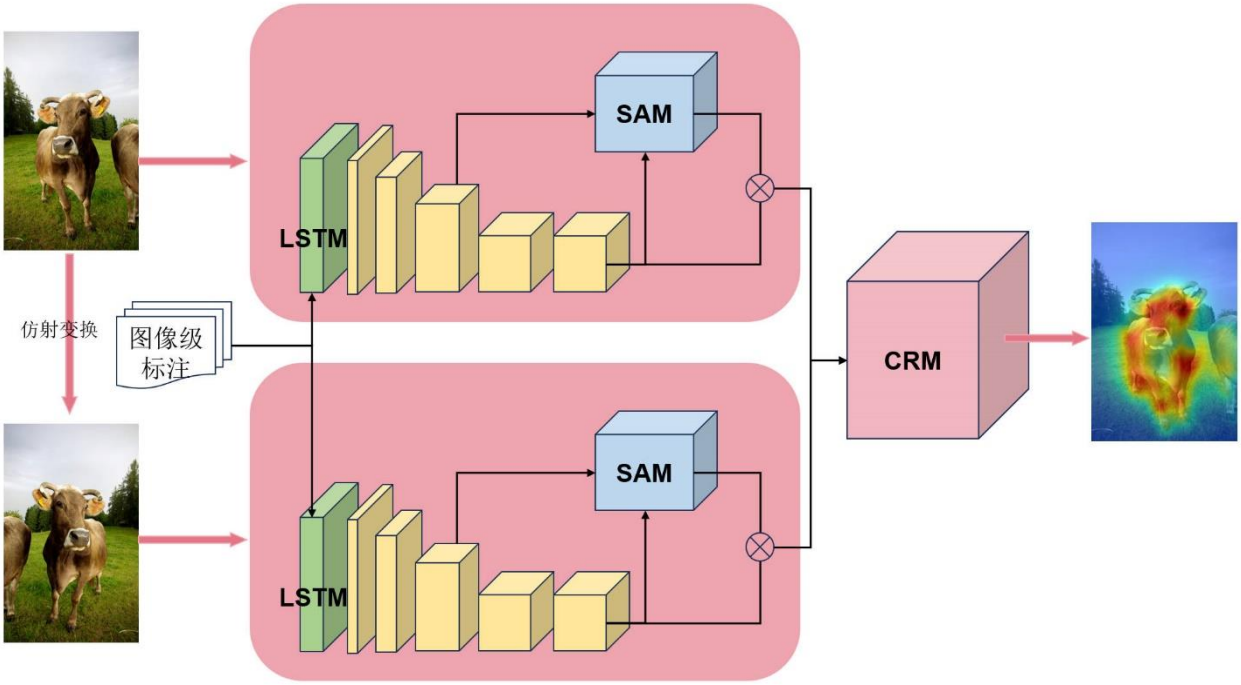


图 3-1 LTCAA 模型算法框图

3.2.1 标签阈值模块

原始的 CAM 算法通过全局阈值来划分目标前景类与背景类，并基于前景区域生成伪标注。然而这种方法存在明显的局限性，全局阈值的单一性难以适应不同图像的特征分布，导致类激活图在激活目标区域时存在显著问题：一方面，全局阈值可能无法准确反映每张图像的最佳分割阈值，从而导致激活区域过小或过多；另一方面，GAP 层的使用进一步加剧了这一问题。GAP 层通过对特征图进行全局平均操作，使得不同类激活图可能产生相同的分类得分，而忽略其内部激活分布的差异性。这种特性导致类激活图在目标区域激活时缺乏精确性，尤其是在复杂场景中，目标与背景的区分变得尤为困难。然而，为每张图像寻找最佳阈值通常依赖于像素级标注，这与弱监督学习的初衷相悖。因此，如何在弱监督

条件下有效丰富监督信息，成为提升模型性能的关键挑战。针对这些问题，本研究设计了一个标签阈值模块（LSTM），旨在提升类激活图对目标前景类激活的准确性与鲁棒性。LSTM 的架构图如图 3-2 所示。

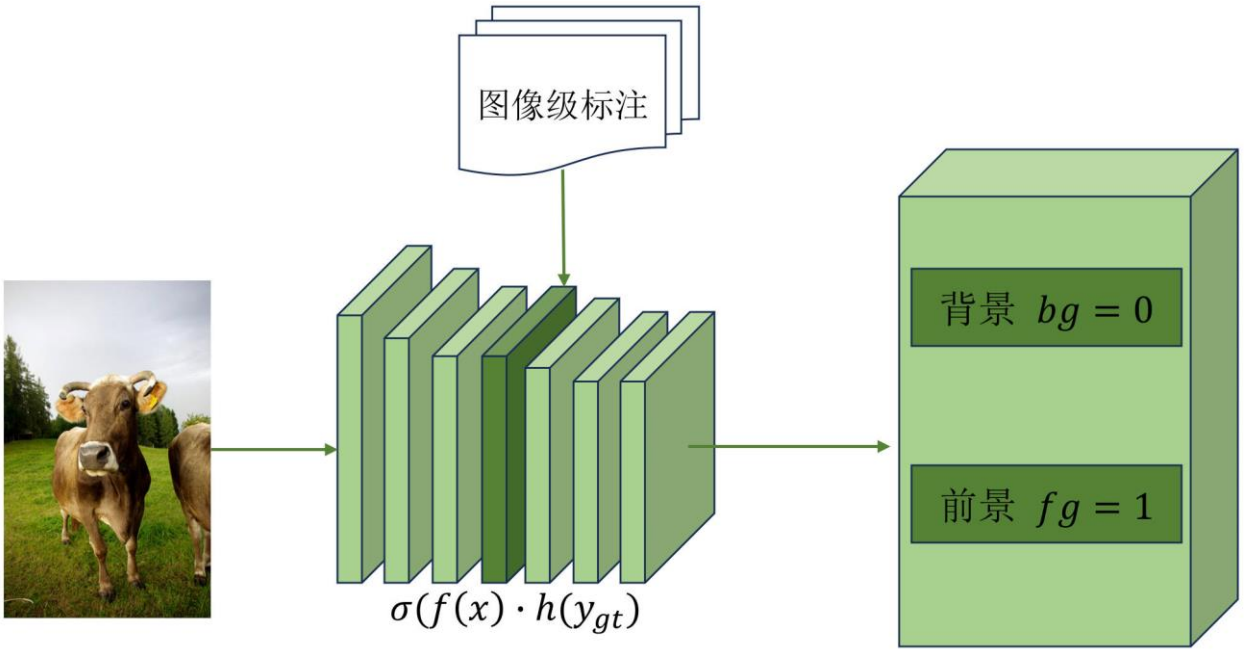


图 3-2 标签阈值模块

LSTM 通过引入图像级标注信息，并调整目标与背景之间的激活水平，从而在弱监督条件下有效优化显著图的生成过程。首先，LSTM 建立了图像级标注和显著图之间的映射关系，将每幅图像所对应的真实图像级标注编码为一个特征向量，并与特征图逐向量相乘，从而为网络提供更丰富和精确的学习信号。这一过程通过引入额外的线性层实现，如公式（3-1）所示：

$$CAM = \sigma(f(x)\cdot h(y_{gt}) \tag{3-1}$$

其中，σ代表激活函数，用于确保输出的非线性；f 是一系列卷积层，用于预测像素级的伪标注；$f(x)$是通过卷积层提取的特征图；h是 LSTM 额外引入的一个线性映射层，其作用是将图像级标注y_{gt}映射到一个特征向量空间，实现图像级语义与局部特征的对齐。通过这种方式，LSTM 将图像级标注信息直接融入类激活图的生成过程，增强了对目标类别的识别能力。

在此基础上，为进一步增强前景类目标的显著性表达，同时抑制背景区域的错误激活，LSTM 引入了一种两端激活映射策略，为目标前景类和背景类设定不同的激活级别。该策

略通过保留目标前景类与背景类之间的最大激活间隙，减少目标前景类内部的激活不平衡现象，从而扩大目标前景类和背景类之间的差异。两端激活映射策略的具体实现如公式（3-2）所示。

$$CAM_{opt} = CAM^t - lr \cdot \nabla_{CAM}(-y_{gt} \cdot logCAM^t + (1 - y_{gt}) \cdot log(1 - CAM^t)) \quad (3\text{-}2)$$

其中t表示迭代次数，CAM^t表示第t次迭代的类激活图，∇_{CAM}表示对类激活图的梯度，lr为学习率。通过两端激活映射策略，LSTM 能够准确定位并强化前景区域，同时有效抑制非目标区域的激活。

LSTM 的设计不仅解决了全局阈值的局限性，还通过整合图像级标签信息与两端激活映射策略，显著提升了类激活图对目标前景类激活的准确性与鲁棒性。LSTM 通过将图像级标注直接引入显著图的生成过程，实现对激活的精准调控，增强了模型对目标类别的识别能力，促进了不同类别间激活的重新分配。其两端激活映射策略更是为前景与背景的精确区分提供了强有力的支持，准确定位并强化前景区域的同时抑制非目标区域的激活。LSTM 的设计具有较强的适应性，能够在不同的图像特征分布下保持稳定的性能。通过动态调整激活阈值，LSTM 减少了目标区域错误激活的情况，提高了分割的准确性，为弱监督条件下的图像语义分割提供了一种高效、稳定且具鲁棒性的解决方案，展示出较强的泛化能力与应用前景。

3.2.2 一致性约束模块

传统的 CAM 方法在显著区域定位中常出现边界模糊与细节缺失的问题，其根本原因在于分类网络与分割网络在参数优化方式上的本质差异。分类网络侧重于图像的全局语义不变性，即无论输入图像经过何种变换，其分类结果应保持一致；而分割网络则需要对图像中的每一个像素进行精确分类，因此更加关注图像的局部结构、边缘轮廓与细节信息。这种目标差异导致 CAM 方法在生成显著图时难以兼顾全局一致性与局部精确性，尤其是在处理复杂场景时，无法有效捕捉物体边界与细微结构，从而出现类激活图的前景定位不准确、边界激活不连续等问题。为解决这一问题，本研究提出一致性约束模块（CRM），架构如图 3-3 所示。

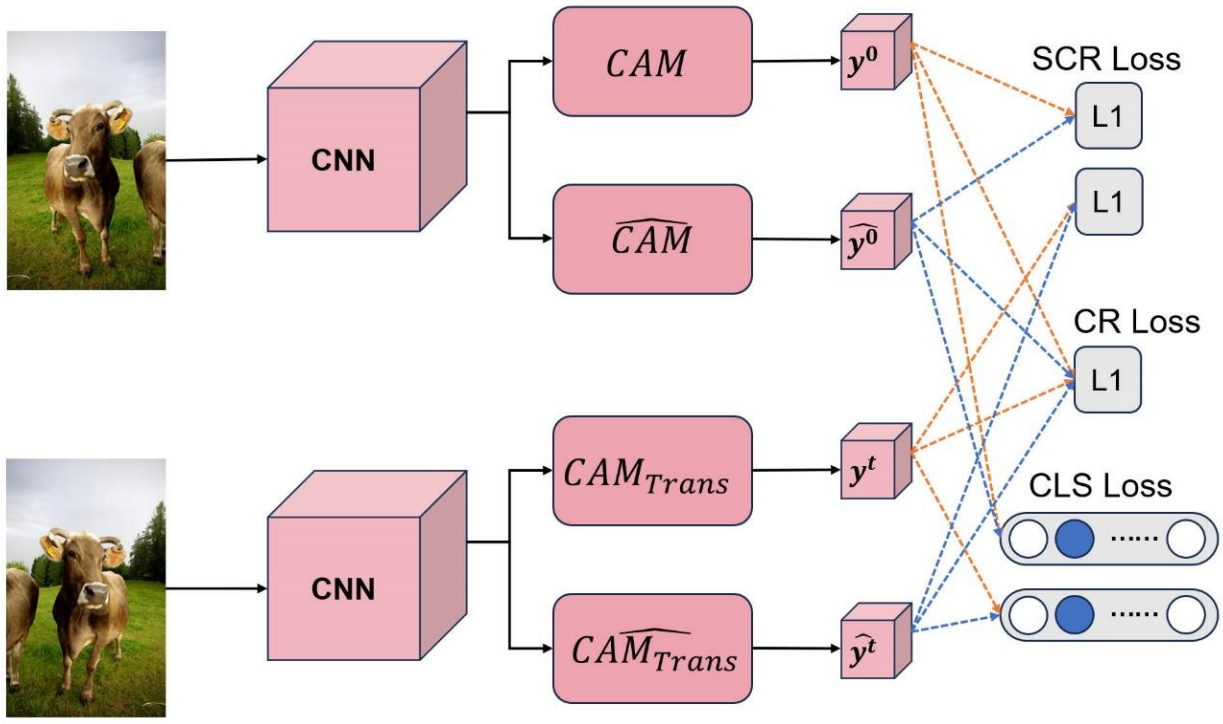


图 3-3 一致性约束模块

CRM 通过引入一致性正则化（Consistency Regularization, CR）策略，增强模型对图像本质特征的捕捉能力。CRM 在不增加额外标注成本的前提下，通过约束模型在不同输入变换下生成的一致性特征，强化对物体边界和形状等关键信息的识别能力，进而提升分割精度。CRM 采用双分支结构，其中两个分支共享相同的网络权重，但分别处理不同的输入变换：第一条分支为原始分支，输入原始图像x；第二条分支为变换分支，输入原始图像经过仿射变换$Trans$后的图像$x_{transform}$。两条分支分别生成对应的类激活图，并对原始分支的类激活图施加与变换分支相同的仿射变换，以确保特征图的一致性。在训练过程中，CRM 算法通过最小化两条分支输出之间的一致性损失，鼓励模型学习在不同变换下保持稳定的特征。这些特征往往与物体的边界、形状等关键信息紧密相关，从而能使得显著图的边界清晰度与细节保留能力得到提升。一致性正则化的具体实现如公式（3-3）所示。

$$CR = \left\| CAM_{transform} - Trans(CAM) \right\|_1 \tag{3-3}$$

其中，$CAM_{transform}$表示变换分支生成的 CAM 图，$Trans(CAM)$表示对原始分支 CAM 施加相同仿射变换后的结果，$\|\cdot\|_1$表示 L1 范数，用于衡量两条分支之间的差异。通过最小化一致性损失，CRM 能够有效约束模型在不同输入变换下生成的一致性特征，从而提升显

著图的鲁棒性与精确性。这种设计不仅解决了传统 CAM 方法在边界模糊和细节丢失方面的局限性，还为后续的分割任务提供了更为可靠的特征表示。

CRM 的设计很好地弥补了分类网络与分割网络之间的差距：通过让分类网络经历一致性正则化的训练，模型能够学习到对图像变换不敏感但又能准确反应物体边界的特征。CRM 通过引入一致性正则化策略，有效解决了传统 CAM 方法在边界模糊和细节丢失方面的局限性，显著提升了显著图的边界清晰度与细节保留能力，为后续分割任务提供了更为可靠的特征表示。此外，CRM 还具有较高的灵活性与可扩展性。由于两个分支共享网络权重，因此 CRM 可以在不显著增加计算负担的情况下，轻松地集成到现有的神经网络架构中。同时，通过调整输入变换的类型和强度，可以进一步控制一致性正则化的程度，从而实现对模型性能的精细调优。

3.2.3 自注意力模块

由于仅依赖图像级的类别标注作为监督信号，CAM 算法在生成类激活图时往往存在激活不充分的问题。类激活图通常仅覆盖目标物体的最具判别性区域，而难以完整捕捉物体的整体轮廓。CAM 算法的这一局限性主要源于图像级标注的监督信息不完整，缺乏像素级的空间信息，导致模型难以学习到目标物体的全局信息。为解决这一问题，本研究引入了自注意力机制，设计了自注意力模块（SAM），其模块架构图如图 3-4 所示。

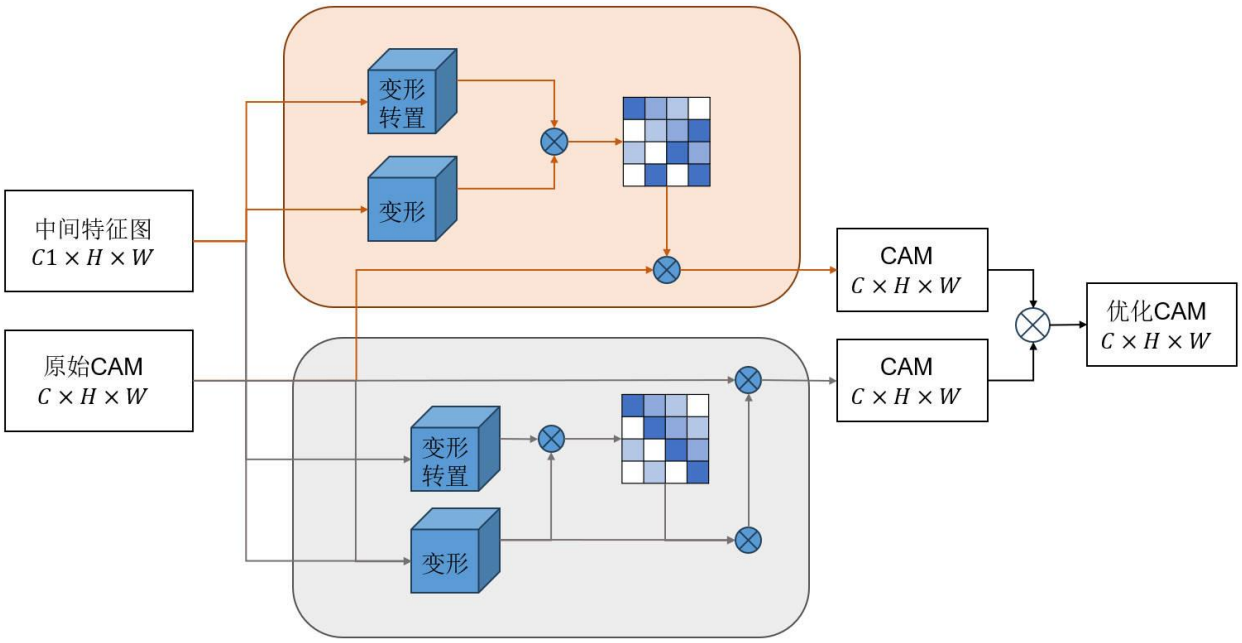


图 3-4 自注意力模块

SAM 通过捕捉图像中不同空间位置之间的上下文关系，增强对目标物体全局结构的理

解，从而优化类激活图的生成。首先，SAM 使用1×1的卷积层实现嵌入函数θ，对输入特征图进行变换，将像素特征映射到高维空间。随后，通过计算特征图中任意两个像素X_i和X_j之间的余弦相似性，评估其相关性，如公式（3-4）所示。

$$g(X_i, X_j) = \frac{\theta(X_i)^T \theta(X_j)}{\|\theta(X_i)\| \cdot \|\theta(X_j)\|} \tag{3-4}$$

其中，$\theta(X_i)$和$\theta(X_j)$分别代表像素X_i和X_j的嵌入特征，$\|\cdot\|$表示向量的 L2 范数。通过计算像素间的相似性，SAM 能够捕捉图像中的长距离依赖关系，从而增强对目标物体整体轮廓的建模能力。

为进一步优化类激活图的生成，SAM 采用 ReLU 激活函数抑制负值，屏蔽无关像素的影响，并通过归一化操作生成亲和注意力图。具体实现如公式（3-5）所示：

$$Y = \frac{1}{C(X_i)} \sum_{\forall j} ReLU\left(\frac{\theta(X_i)^T \theta(X_i)}{\|\theta(X_i)\| \cdot \|\theta(X_i)\|}\right) Y_j \tag{3-5}$$

其中，Y_j表示原始 CAM 中像素j的值，Y表示输出的修正后的 CAM 图，$C(x_i)$是归一化因子，$(\cdot)^T$表示转置操作。通过引入亲和注意力图，SAM 能够在相关区域生成更平滑的显著图，从而扩展类激活图的激活范围。

此外，由于中间特征图的每一个通道都可以看作是某个类别的相关响应或不同语义之间的相互关联。因此，SAM 还通过通道关系矩阵捕捉中间特征图与类激活图之间的通道依赖关系。利用 Softmax 函数处理类激活图，随后将类激活图与转置处理的中间特征图进行矩阵相乘，生成通道关系矩阵 M，如公式（3-6）所示：

$$M = Softmax(Y) X^T \tag{3-6}$$

通道关系矩阵模拟了类激活图与中间特征图任意通道间的依赖关系。随后，为进一步增强特征表示，SAM 采用基于高斯函数的嵌入式应用方式处理输入特征，对中间特征图和通道关系矩阵进行降维，并通过矩阵乘法生成像素点权重矩阵。最终，利用余弦距离归一化相似度，生成修正后的 CAM 图$\underline{Y}$，如公式（3-7）所示。

$$\underline{Y} = ReLU\left(\frac{\Phi(X)^T\Psi(M)}{\|\Phi(X)\| \cdot \|\Psi(M)\|}\right)Y \tag{3-7}$$

其中，Φ和Ψ表示嵌入式函数，用于对特征图和通道关系矩阵进行降维处理；$\underline{Y}$表示输出的修正后的类激活图。最终，通过叠加原始类激活图Y和修正后的类激活图$\underline{Y}$，SAM 生成最终优化后的类激活图$\widehat{CAM}$，如公式（3-8）所示。

$$\widehat{CAM} = Y + \underline{Y} \tag{3-8}$$

SAM 通过捕捉空间和通道两个维度的上下文信息，扩展了类激活图的激活范围，提升了类激活图的细节精度。SAM 在经典自注意力机制的基础上去除了残差连接，保持激活强度的一致性，简化嵌入函数以减少参数数量，避免过拟合。实验结果表明，SAM 模块有效解决 CAM 算法激活不充分的问题，为弱监督图像语义分割提供更可靠的特征表示。

3.2.4 损失函数设计

在 LTCAA 模型的网络训练过程中，损失函数的设计是确保模型能够有效学习并优化弱监督语义分割性能的关键。本研究的损失函数由三部分构成：分类损失（$Loss_{cls}$）、一致性正则化损失（$Loss_{CR}$）和双分支交叉约束损失（$Loss_{SCR}$）。这三部分损失函数从不同角度为网络提供监督信号，共同作用以提升模型的整体性能。最终的损失函数是这三部分的加权和，如公式（3-9）所示：

$$Loss = Loss_{cls} + Loss_{CR} + Loss_{SCR} \tag{3-9}$$

（1）分类损失。分类损失是基于图像级标注的唯一人工监督信号，用于指导模型正确预测图像中存在的目标类别。由于本研究中使用了双分支的结构，因此总的分类损失是对每个分支的分类损失结果取平均值。以z表示分类网络输出的预测向量，K表示图像级标注中包含的所有目标前景类别总数，对于分类损失$Loss_{cls}$的具体定义如公式(3-10)所示。

$$Loss_{cls} = avg\left(-\frac{1}{K}\sum_{k}^{K}\left(y_{gt}log\left(\frac{1}{1+e^{-z_k}}\right) + (1-y_{gt})log\left(\frac{e^{-z_k}}{1+e^{-z_k}}\right)\right)\right) \tag{3-10}$$

其中，y_{gt}表示图像级标注的真实标签，z_k表示第k个类别的预测得分。分类损失通过

最小化预测值与真实标签之间的差异，确保模型能够准确识别图像中的目标类别。

（2）一致性正则化损失。一致性正则化损失基于一致性约束模块提出，模型在不同输入变换下生成的 CAM 图具有一致性。$Loss_{CR}$通过计算原始分支生成的 CAM 图与变换分支生成的 CAM 图之间的 L1 范数差异，约束模型在仿射变换下保持特征的一致性。其定义如公式（3-12）所示。

$$Loss_{CR} = \left\| Trans(CAM) - CAM_{transform} \right\|_1 \tag{3-12}$$

其中，$Trans(CAM)$表示对原始分支 CAM 图施加与变换分支相同的仿射变换，$CAM_{transform}$表示变换分支生成的 CAM 图。通过最小化$Loss_{CR}$，模型能够学习到对图像变换不敏感但又能准确反映物体边界的特征，从而提升显著图的鲁棒性。

（3）双分支交叉约束损失。双分支交叉约束损失（$Loss_{SCR}$）用于处理自注意力模块（SAM）前后 CAM 图之间的关系，防止优化过程中模型陷入局部最小值，导致所有像素被预测为同一类别。$Loss_{SCR}$通过计算原始 CAM 图与经 SAM 处理后的 CAM 图之间的 L1 范数差异，约束模型在优化过程中保持显著图的多样性。其定义如公式（3-13）所示：

$$Loss_{SCR} = \left\| Trans(CAM) - \widehat{CAM_{transform}} \right\|_1 + \left\| Trans\left(\widehat{CAM}\right) - CAM_{transform} \right\|_1 \tag{3-13}$$

其中，$\widehat{CAM}$和$\widehat{CAM_{transform}}$分别表示经SAM处理后的原始分支和变换分支的CAM图。通过引入$Loss_{SCR}$，模型能够在优化显著图的同时，避免过度平滑或局部最优解的问题，从而生成更具判别性的显著图。

3.3 实验与分析

3.3.1 数据集介绍

为了全面评估 LTCAA 模型的性能，本文选择使用 PASCAL VOC 2012 增强数据集[78]对模型性能进行评估。PASCAL VOC 2012 数据集是计算机视觉领域广泛使用的标准数据集之一，因其具图像资源丰富、标注信息精确以及场景内容多样化等特点，成为图像分类、目标检测、语义分割以及动作识别等核心任务的常用测试数据。而且，PASCAL VOC 2012 数据集的多样性和复杂性在能够有效验证模型性能的基础上，还能确保结果的可靠性和泛化能力，这使其成为弱监督图像语义分割研究的常用数据集。

PASCAL VOC 2012 数据集涵盖 20 个前景类别（例如猫、狗、船、桌子、自行车等）和 1 个背景类别，涵盖了日常生活中常见物体和场景，如图 3-5 所示。原始 PASCAL VOC 2012 数据集的训练集包含 1464 张图像，测试集包含 1456 张图像，验证集包含 1449 张图像，训练的样本量相对有限，直接用于训练高性能的网络模型可能会导致模型训练不充分，从而影响模型的性能和泛化能力。因此，我们选择了 PASCAL VOC 2012 的增强版本作为训练数据。

PASCAL VOC 2012 增强数据集通过整合 SBD 数据集[79] 中的 11355 张图像到 PASCAL VOC 2012 原始训练集中，扩充了训练集的规模。在去除重复图像后，增强数据集的训练样本总数达到 10582 张，而验证集和测试集保持不变，分别为 1449 张和 1456 张。值得注意的是，本文中所有提及的 PASCAL VOC 2012 数据集，均指经过 SBD 数据扩充后的版本，以确保模型训练与评估的充分性和准确性。

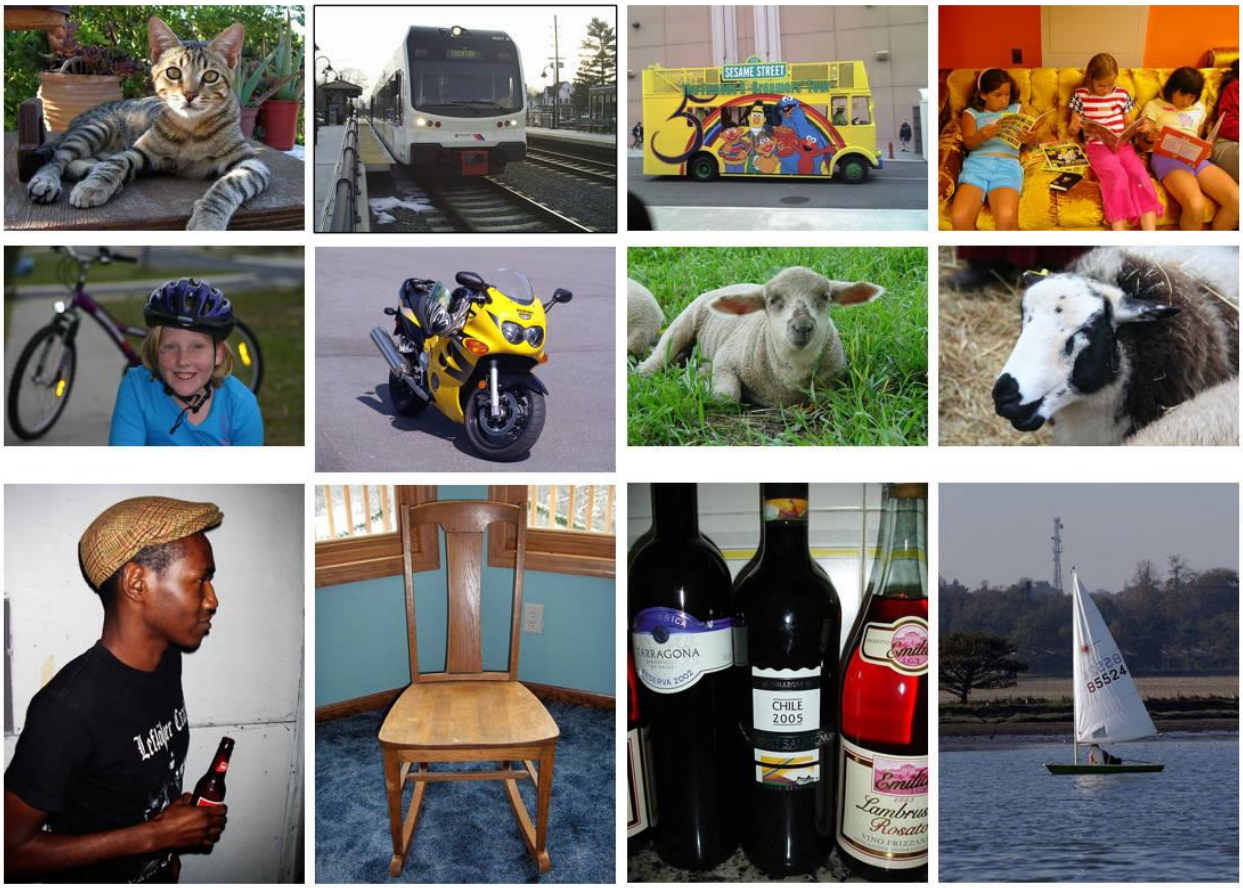


图 3-5 PASCAL VOC 2012 增强数据集

3.3.2 评价指标

在图像语义分割任务中，评估模型性能的关键指标之一是平均交并比（Mean Intersection over Union, mIoU）。为了更好地理解 mIoU 的计算过程，首先介绍混淆矩阵

（Confusion Matrix），如图 3-6。

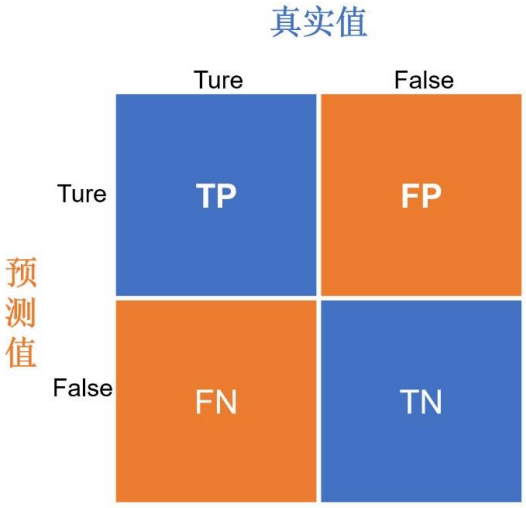


图 3-6 混淆矩阵

混淆矩阵是统计分类模型结果工具，通常使用 0（负例）和 1（正例）来表示预测值和真实值，并通过四个基本指标来描述模型的分类结果，分别是真正例（True Positives, TP）、假正例（False Positives, FP）、假负例（False Negatives, FN）、真负例（True Negatives, TN）。TP 表示真实值和预测值都标记为目标类别的像素数量；FP 表示真实值为其他类别但被错误预测为目标类别的像素数量；FN 表示真实值为目标类别但被错误预测为其他类别的像素数量；TN 表示真实值和预测值均被标记为其他类别的像素数量。

在图像语义分割任务中，mIoU 是衡量分割结果准确性的重要工具，它综合考虑了模型的精度和召回率，为评价模型的整体性能提供了一个全面且直观的视角。mIoU 通过计算预测结果与真实标注之间的重叠程度，能够有效反映模型在像素级别上的分类准确性，在语义分割领域被广泛采用。交并比（Intersection over Union, IoU）是针对单个类别计算的，表示真实值与预测值之间交集和并集的比值，其中交集部分指的是在真实值和预测值中都标记为同一类别的像素数量，即 TP；而并集部分则包括了真实值中该类别的所有像素（即真实值中的正例，无论是否被正确预测）和预测值中该类别的所有像素（即预测值中的正例，无论是否真正属于该类），并减去交集部分以避免重复计算。IoU 的公式如式（3-14）。

$$IoU = \frac{TP}{TP + FN + FP} \tag{3-14}$$

其中，IoU 的值域为$[0,1]$，值越接近 1，表示模型在该类别上的分割性能越好。

为了全面评估模型在所有类别上的性能，我们计算所有类别的 IoU 的平均值，即平均

交并比（mIoU）。mIoU 是所有类别 IoU 值的和除以类别数，它提供了一个综合的性能指标，能够反映模型在不同类别上的平均分割效果。公式表示如下：

$$mIoU = \frac{1}{K}\sum_{i=1}^{K}\frac{TP_i}{TP_i + FP_i + FN_i} \tag{3-15}$$

其中，K表示类别总数，TP_i、FP_i和FN_i分别表示第i个类别的 TP、FP 和 FN。mIoU 的值域同样为$[0,1]$，值越接近 1，表示模型在所有类别上的平均分割性能越好。

在实际应用中，mIoU 常被用作比较不同分割模型性能的主要依据，因为它能够直观地反映模型在像素级别上的分类准确性。与单一指标（如精度或召回率）相比，mIoU 综合考虑了模型在不同类别上的表现，避免了单一指标可能带来的偏差。同时，mIoU 在处理类别不平衡问题时也表现出色，在语义分割任务中，某些类别的像素数量可能远多于其他类别（例如背景类别通常占据较大比例）。

3.3.3 实验设置

本文的实验均在硬件配置为 16 vCPU Intel(R) Xeon(R) Gold 6430 处理器、GPU 配置为 NVIDIA RTX 4090（24GB 显存）的高性能计算平台上进行，操作系统为 Linux，CUDA 版本为 11.0。LTCAA 模型的实现基于 Python 3.8 编程语言，并依托 PyTorch 1.7.0 深度学习框架构建。

在分类网络的训练阶段，LTCAA 模型使用 ResNet50 作为主干分类网络，并在 ImageNet 数据集上对其进行预训练。CAM 的实现细节遵循了 Ahn 等人的研究配置。在模型训练过程中，本实验采用 Adam 优化器进行参数优化，其学习率设置为$5e-6$，训练批次$batchsize$设置为8，权重衰减设置为 5e-4，动量设置为0.9，模型共训练的 epoch 数为8。

在伪标注生成阶段，本文采用 AffinityNet 算法对 CAM 图进行进一步优化。AffinityNet 通过 RW 增强显著图的连通性与一致性，从而生成更精确的伪标注。相关参数设置与 AffinityNet 算法的原始配置保持一致。

对于分割网络的训练，本文使用 DeepLabV2 算法与 ResNet101 主干网络进行实验。分割网络的训练设置遵循 AdvCAM[44] 的默认配置，包括学习率、优化器参数和数据增强策略等。

3.3.4 消融实验

本研究提出的 LTCAA 模型，主要针对 CAM 算法中存在的对目标前景类激活不充分和激活不准确的问题进行改进。为此，LTCAA 模型共提出了三个改进策略，分别是标签阈值模块（LSTM）、一致性约束模块（CRM）和自注意力模块（SAM）。为了验证模这些模块的有效性，本节设计了一系列消融实验，通过逐步引入 LSTM、CRM 和 SAM 等模块，来分析它们对模型性能的具体贡献。消融实验以结合了 IRN 的 CAM 算法为基线，并逐步引入 LSTM、CRM 和 SAM，通过对比不同模块组合下的性能表现，评估各模块对目标前景类激活的改进效果。

实验在 PASCAL VOC 2012 数据集上进行，通过计算生成伪标注的 mIoU 值来评估模型性能。表 3-1 展示了 LTCAA 模型在 PASCAL VOC 2012 训练集上的消融实验结果，反映了各模块对伪标注生成质量的影响。实验结果表明，随着各模块的逐步引入，模型的性能得以提升。基线模型采用结合 IRN 的 CAM 算法，其生成的伪标注 mIoU 为 66.31%。在基线模型基础上引入 CRM 模块后，伪标注的 mIoU 提升至 68.15%，较基线模型提高了 1.84%。这一改进验证了 CRM 模块通过双分支结构引入额外监督信息的有效性，增强了对目标边界的捕捉能力，从而提升了分割精度。在基线模型和 CRM 模块的基础上进一步引入 LSTM 模块后，模型的 mIoU 显著提升至 72.06%，较只引入 CRM 模块的模型提高了 3.91%。这一结果证明了 LSTM 模块通过将图像级标注加为辅助监督信息并根据其调节阈值方法的有效性，使模型对目标前景类的激活更加准确。在基线模型、CRM 模块和 LSTM 模块的基础上引入 SAM 模块后，模型的 mIoU 进一步提升至 72.38%，较未引入 SAM 模块的模型提高了 0.32%。这一改进表明，SAM 充分地利用全局上下文信息，融合空间、通道维度信息建立语义依赖，生成更高质量的类激活图。

表 3-1 LTCAA 各模块消融实验对比

baseline	CRM	LSTM	SAM	mIoU(%)
√				66.31
√	√			68.15
√	√	√		72.06
√	√	√	√	72.38

图 3-7 展示了 LTCAA 模型各模块叠加后生成的 CAM 图可视化结果，直观反映了各模块对目标前景类激活的改进效果。其中，(a)为原始输入图像，(b)是基线方法生成的 CAM

图，(c)-(e)是分别为逐步加入 CRM、LSTM 和 SAM 后生成的 CAM 图。由图 3.7 可以观察到，随着各模块的逐步引入，模型对目标前景类的激活范围逐渐扩大，且激活边界更加清晰，错误激活背景的情况显著减少。这表明 LTCAA 模型通过结合 LSTM、CRM 和 SAM 模块，能够有效解决 CAM 算法在目标激活不充分和不准确方面的问题。

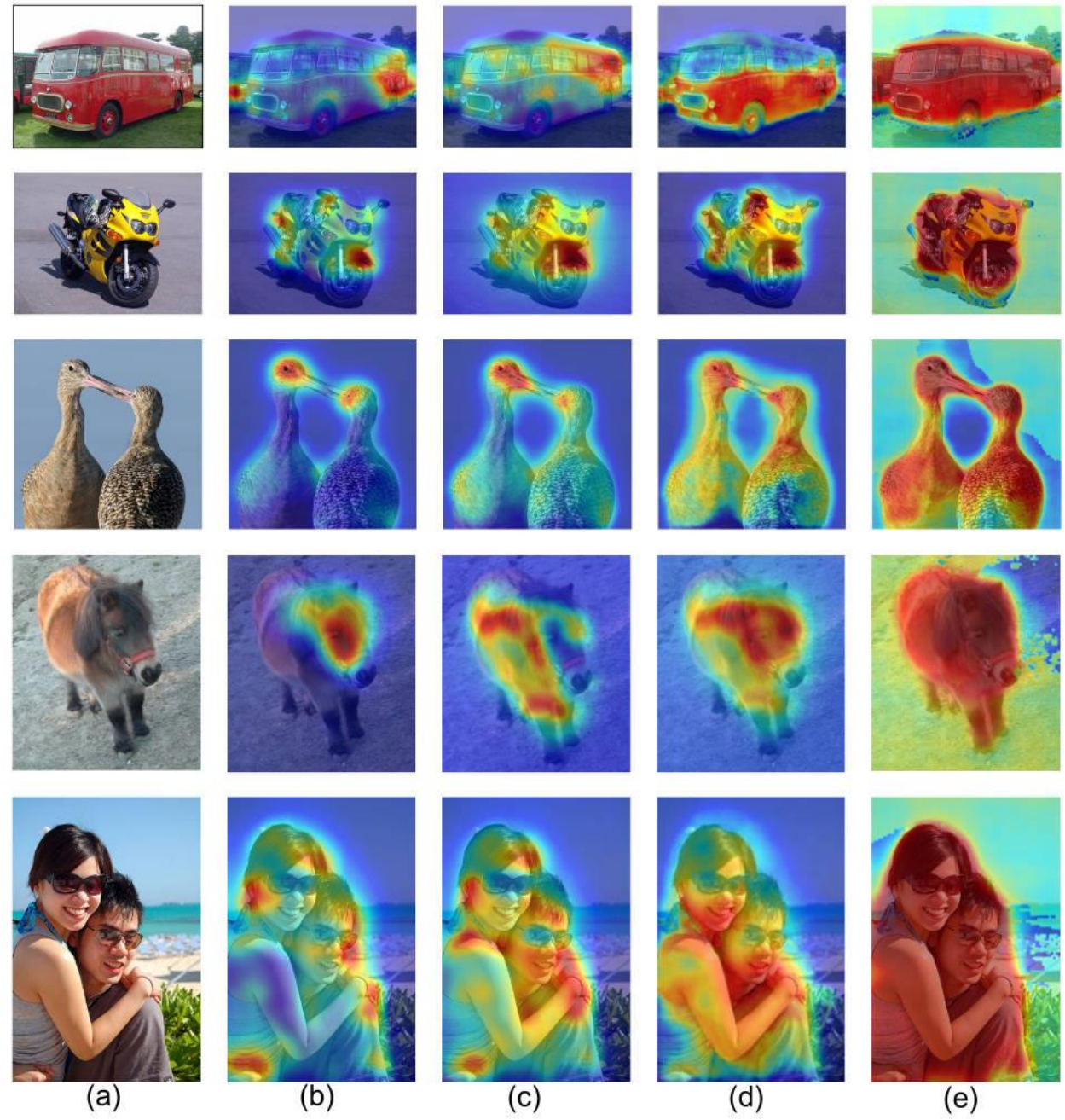


图 3-7 CAM 图可视化对比图

3.3.5 算法对比分析

为了全面评估本文所提出的基于标签阈值的一致性注意力架构（LTCAA）模型在弱监督语义分割任务中的性能，本节设计并实施了一系列对比实验。通过将 LTCAA 模型与当

前现有的多种先进的弱监督分割方法进行比较，从伪标注生成质量和最终分割性能两个维度，通过定量和定性分析相结合的方法进行评估，全面验证 LTCAA 模型的有效性和先进性。

伪标注生成质量是衡量弱监督语义分割模型性能的重要指标之一，其直接决定了后续分割网络的训练效果和最终分割精度。因此，本节对伪标注的生成质量进行评估，通过将生成的伪标注与真实标注进行对比的方式得到 mIoU。表 3-2 展示了 LTCAA 模型与其他现有模型在 PASCAL VOC 2012 数据集上生成的伪标注的 mIoU 对比结果。实验结果表明，LTCAA 模型生成的伪标注 mIoU 达到 72.4%，显著优于其他现有方法。LTCAA 模型相较于可视为基线的 IRN 模型提高了 6.1%，相较于其他先进方法如 AMN（72.2%）和 AdvCAM（69.9%）也有不同程度的提升。这一结果充分证明了 LTCAA 模型在伪标注生成任务中的优越性。

表 3-2 LTCAA 模型在 VOC12 数据集上获取伪标注的结果

模型	mIoU(%)
IRN[55]	66.3
SEAM[68]	63.6
CONTA[80]	67.9
AdvCAM[44]	69.9
DCAM[71]	68.7
AMR[65]	69.7
AMN[70]	72.2
LTCAA(ours)	**72.4**

为进一步验证 LTCAA 模型在实际语义分割任务中的综合性能，本节在 PASCAL VOC 2012 数据集的验证集和测试集上分别评估了各方法的最终分割结果。所有方法均基于第 3.3.3 节描述的实验配置进行训练和测试，以保证结果的可比性与公正性。表 3-3 展示了 LTCAA 模型与其他现有模型在验证集和测试集上的 mIoU 对比结果。从表 3.3 可以观察到，LTCAA 模型在测试集上的 mIoU 达到 69.9%，较可视为基线的 IRN 模型提升了 5.1%，且同时优于其他对比方法。同时，在验证集上，LTCAA 模型取得了 69.4%的 mIoU，相较于基线的 IRN 模型提升了 5.9%，相较大部分的对比方法也有显著提升。这一结果充分表明，LTCAA 模型不仅在伪标注生成阶段展现出优势，高质量的伪标注也可有效转化为分割精

度的提升，在最终分割任务中展现出显著的性能优势，也进一步验证了本文所提出方法的有效性。

表 3-3 LTCAA 模型在 VOC12 数据集上的图像语义结果

模型	骨干网络	验证集 mIoU(%)	测试集 mIoU(%)
SEC[40]	VGG16	50.2	51.1
AdvErasing[48]	VGG16	55.0	55.7
CONTA[80]	ResNet38	66.1	66.7
SEAM[68]	ResNet38	64.5	65.7
IRN[55]	ResNet50	63.5	64.8
OAA[58]	ResNet101	65.2	66.4
AdvCAM[44]	ResNet101	68.1	68.0
DCAM[71]	ResNet101	69.2	69.4
AMR[65]	ResNet101	68.8	69.1
PuzzleCAM[81]	ResNet101	66.9	67.7
AMN[70]	ResNet101	**69.5**	69.6
LTCAA(ours)	ResNet101	69.4	**69.9**

为了进一步细化性能分析，本节还对 LTCAA 模型在不同类别物体上的分割性能进行了逐类别评估。表 3-4 展示了 LTCAA 模型在 20 个前景类别和 1 个背景类别上的分割精度（mIoU），结果表明，LTCAA 模型在绝大多数类别上均表现出显著的性能优势。其中，LTCAA 模型在背景类别的分割精度达到了 89.93%，显著优于基线模型的 79.21%，以及其他对比方法的分割效果。这一提升表明 LTCAA 模型在处理复杂背景时具有较强的辨别和分割能力。在其他前景类别中，LTCAA 模型在飞机、自行车、船、瓶子、巴士、汽车、椅子、餐桌、摩托车、人物、植物、羊、沙发以及电视等类别的分割任务中，分割精度有着显著地提升。这一结果表明，LTCAA 模型不仅能够提升整体的分割性能，还能在多类别复杂场景中保持较好的稳定性和鲁棒性。然而，在部分物体形态多变、背景复杂的类别上，例如鸟、猫、狗、马等，LTCAA 模型的分割表现略低于部分对比方法，这可能与目标类别在图像中姿态变化大、尺寸变化大、背景环境复杂有关，从而对伪标注的生成和分割网络训练提出更高挑战。此外，火车的分割精度也低于其他对比方法，这可能与火车与轨道难以在激活时进行区分有关。

表 3-4 LTCAA 模型在 PASCAL VOC12 数据集上各类别的分割性能

类别	baseline(%)	AffinityNet(%)	SEAM(%)	DCAM(%)	LTCAA(ours) (%)
Background	79.21	88.21	88.81	88.35	**89.93**
Aeroplane	46.40	68.23	68.50	67.25	**74.22**
Bicycle	28.58	30.62	33.28	44.00	**45.18**
Bird	40.99	81.09	**85.73**	77.57	83.93
Boat	33.83	49.57	40.36	55.97	**69.02**
Bottle	43.01	61.02	63.31	60.27	**65.24**
Bus	55.97	77.79	78.88	79.91	**86.06**
Car	48.39	66.14	76.30	65.86	**76.92**
Cat	37.05	75.14	79.81	**83.10**	81.92
Chair	24.40	29.01	29.10	36.30	**37.18**
Cow	53.68	65.99	75.47	86.82	84.90
Dining Table	30.82	40.18	48.09	49.88	**59.53**
Dog	41.95	80.36	79.92	**85.27**	79.12
Horse	44.77	62.00	73.83	**84.21**	81.35
Motorbike	54.61	70.42	71.43	79.80	**80.30**
Person	48.92	73.66	73.18	72.70	**75.77**
Potted Plant	39.32	42.48	48.88	54.20	**60.94**
Sheep	56.39	70.71	79.82	88.26	**89.11**
Sofa	38.96	42.57	40.94	60.56	**60.76**
Train	47.00	**68.12**	58.18	61.86	65.24
Tv/Monitor	42.02	51.59	53.04	57.45	**57.61**

总体而言，LTCAA 模型在伪标注生成阶段有效提升了类激活图的语义完整性和空间一致性，在分割阶段实现了更高精度的前景定位与背景抑制。实验证据表明，LTCAA 不仅在平均性能上优于现有主流弱监督方法，在多数单类别分割任务中亦具备明显优势，具有良好的泛化性与应用潜力。

3.3.6 语义分割结果可视化

为了更好地展示 LTCAA 模型的分割性能，本节中以可视化的形式直观展示了 LTCAA

模型的图像语义分割结果，对 PASCAL VOC 2012 数据集中包含的 20 个目标前景类和 1 个背景类的分割效果全部进行了展示，如图 3-8 和图 3-9 所示。

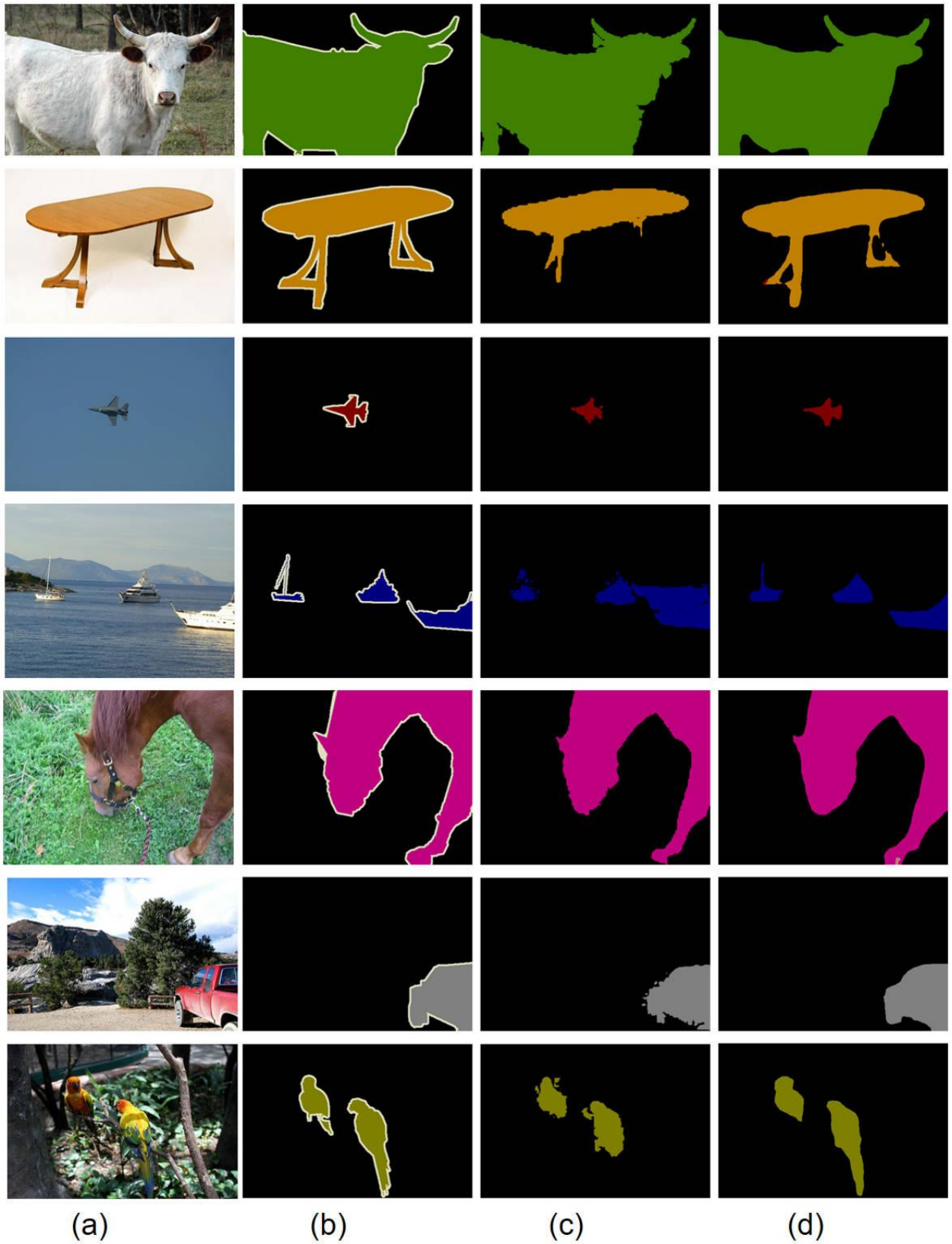


图 3-8 LTCAA 模型的图像语义分割结果（单类别目标）

图 3-8 展示了本文算法对单类别物体的语义分割情况，其中(a)是原始图像，(b)为真实标注，(c)为经 LTCAA 模型处理生成的伪标注，(d)是最终分割结果。分别展示了牛、桌子、飞机、船、马、汽车和鸟等七个类别的图像语义分割结果，涵盖了单类别分割的多种目标类别存在情况和场景。其中，第一行和第二行的牛和桌子，展示了模型对尺寸较大、形状规则的目标前景的分割能力，在该类目标存在情况下，由 LTCAA 模型生成的伪标注与真实标注几乎重合，取得的分割结果良好，说明本文的模型对尺寸较大、形状规则的目标前景的分割能力较强。第三行对飞机的分割结果和第四行对船的分割结果可以看出，LTCAA 模型对尺寸较小的目标同样能够进行较好的分割，但是由于其目标过于小且细节部分过于精细，因此生成的伪标注在细节上有细微的欠缺。在第五行和第六行中，展示了目标前景类别的形状不完整（汽车）或存在动态变化（马），在这种情况下 LTCAA 模型仍能对目标前景类进行准确的提取并生成效果优秀的伪标注，这也表明 LTCAA 模型对类别的特征提取能力效果强。第七行展示了一个背景杂乱且存在多个单类别目标的例子，多只鸟处于存有杂乱树木、树叶的背景中，较好的伪标注和分割结果也证明 LTCAA 模型能够很好地抑制错误背景类的激活，准确、完整的获取全部目标前景类。

图 3-9 展示了 LTCAA 模型在多类别图像语义分割任务中的性能表现，(a)是原始图像，(b)为真实标注，(c)为经 LTCAA 模型生成的伪标注，(d)是最终分割结果。与单类别图像相比，图 3.9 中的原始图像展示了更为复杂的分割场景，不仅包含多个类别的目标，且同一类别中可能存在单个或多个目标物体。从整体上看，LTCAA 模型在多类别图像语义分割任务中的准确性较高，表现出较好的泛化能力。例如，在第一行和第七行的图像中，LTCAA 模型对猫、狗和羊的分割表现较为出色，生成的伪标注以及最终分割结果均有比较清晰的轮廓，且类别标注准确，表明了 LTCAA 模型能够有效捕捉目标物体的整体特征并实现精确分类。对于形状规整的大型物体，如第二行和第五行的沙发、第四行的桌子与椅子以及第二、四、五行的电视，LTCAA 模型能够大致分割出这些物体的轮廓，但在细节的部分仍存在不够清晰的问题。对于形状不规则但特征明确的物体，如第三行的摩托车和第五行的自行车，LTCAA 模型能够比较准确地分割出目标的整体轮廓，并正确地进行类别标注。在包含人物类别的场景中（如第三行、第五行和第七行），LTCAA 模型能够在整体上分割出人物轮廓，但由于人物的姿态和动作变化较为复杂，在动作变化较大或存在遮挡较为严重的时候，模型的识别精度有所下降。对于数量多且尺寸较小的物体，如第四行和第七行中的瓶子，LTCAA 在生成伪标注时能进行部分的识别，但分割准确率较低，可能会存在

遗漏或错误标注的情况。此外，在第六行的图像中，火车与火车轨道通常同时出现，导致 LTCAA 模型在识别时难以准确区分两者，从而错误地将轨道标注为火车类别。

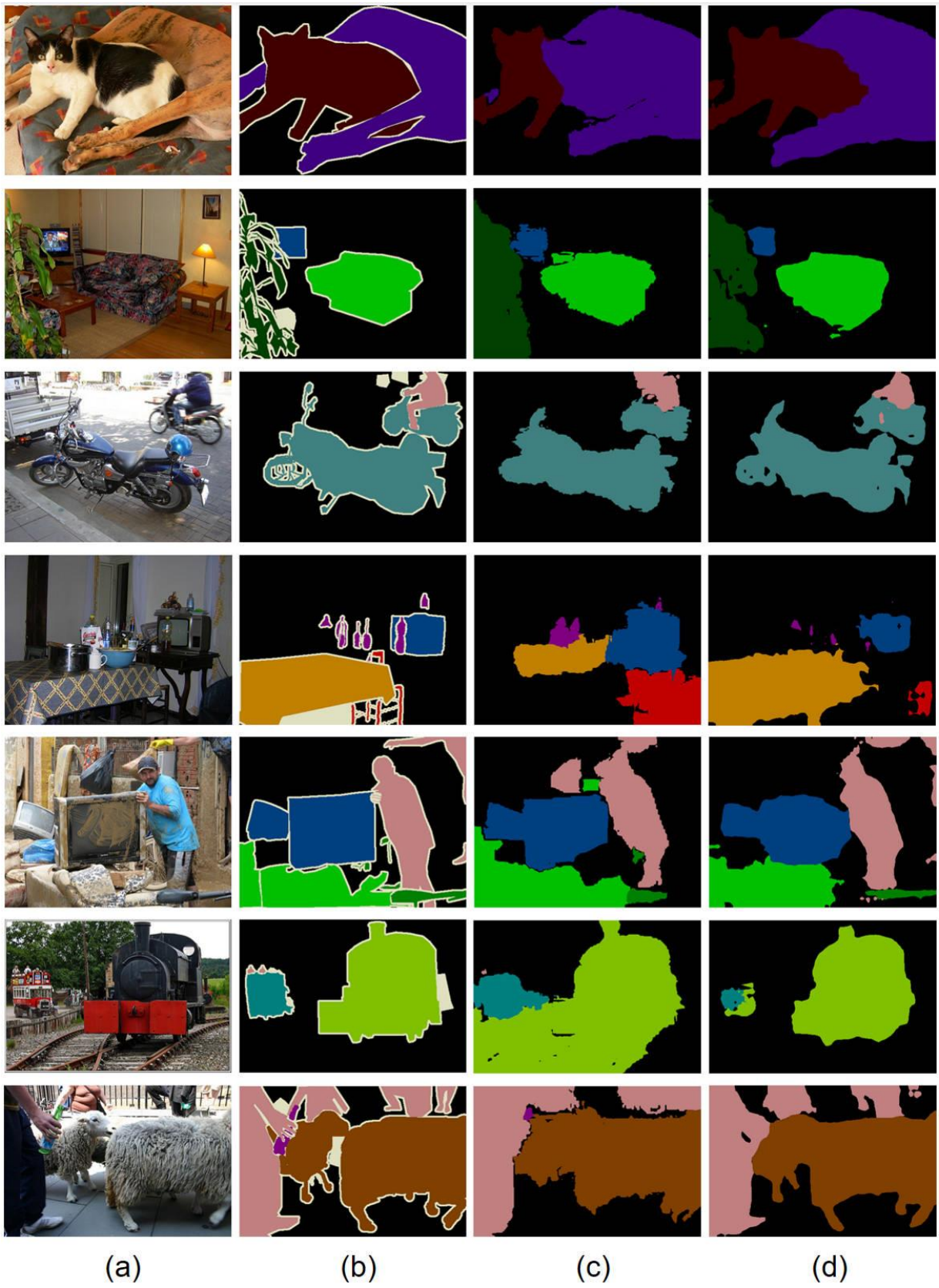


图 3-9 LTCAA 模型的图像语义分割结果（多类别目标）

3.4 本章小结

本章针对基于类激活映射的弱监督图像语义分割任务中 CAM 算法存在的激活不充分和激活不准确等问题，提出了一种基于标签阈值的一致性注意力模型 LTCAA。从设计思路、实现路径、模型特点等方面介绍了该模型包括的标签阈值模块 LSTM、一致性约束模块 CRM 和自注意力模块 SAM 等三种改进策略，并根据模型设计了对应的损失函数。其次，本章通过消融实验验证了 LTCAA 模型中各模块的有效性，实验结果表明 LSTM、CRM 和 SAM 模块的引入均对模型性能提升起到了重要作用。最后，本章通过对比实验评估了 LTCAA 模型的综合性能，实验结果表明 LTCAA 模型在 PASCAL VOC 2012 数据集上的伪标注生成质量和最终分割性能均优于现有的多种先进方法。

4 LTCAA 模型在遥感图像的应用案例研究

4.1 引言

遥感图像语义分割作为遥感技术的重要研究方向，在土地利用分类、环境监测、城市规划等领域具有广泛的应用价值。然而，遥感图像的空间范围广、地物类型复杂、尺度多样性显著等特点，使得获取其像素级标注信息成为一项耗时耗力的工作，极大地限制了全监督学习方法在实际应用中的推广。因此，基于图像级标注的弱监督语义分割方法在遥感图像处理中具有重要的研究意义。目前，弱监督语义分割的研究主要集中在通用数据集（如PASCAL VOC 2012）上，而针对遥感图像的弱监督分割研究相对较少。由于遥感图像的特殊性，传统 CAM 算法在应用于遥感图像时表现出更加明显的激活不准确和激活不充分等问题，例如目标区域激活不完整、边界模糊以及错误激活背景等，这些问题限制了模型在遥感图像分割中的性能。

因此，本文将 LTCAA 模型应用于遥感图像数据集，旨在验证其在复杂场景下的有效性和鲁棒性，并探索其在遥感图像语义分割中的实际应用价值。LTCAA 模型通过引入标签阈值模块（LSTM）、一致性约束模块（CRM）和自注意力模块（SAM），有效解决了传统 CAM 算法在目标激活不充分和不准确方面的问题。为了使 LTCAA 模型更好地适应遥感图像的特点，本章对遥感数据集进行了额外的预处理。由于遥感图像通常包含多类别地物信息，且同一图像中可能同时出现多种地物类型，这使得 CAM 图的生成更具挑战。LSTM 模块通过阈值调整和多类别标注信息的融合，有效解决了多类别地物同时激活的问题；CRM 模块引入的一致性约束策略增强了模型对不同尺度和变换条件下图像的适应能力；SAM 模块则通过全局上下文信息的捕捉，进一步挖掘了地物之间的语义关系，从而提升了模型在复杂场景下的分割性能。

本章通过将 LTCAA 模型应用于遥感图像数据集，展示了其在多类别、多尺度场景下的优异性能，进一步证明了该模型在实际应用中的有效性和鲁棒性。实验结果表明，LTCAA 模型不仅能够有效解决传统 CAM 算法在遥感图像分割中的局限性，还为弱监督遥感图像语义分割任务提供了解决方案。

4.2 数据准备

4.2.1 遥感数据集介绍

本章实验采用的 ISPRS Postdam 数据集是一个遥感图像分析与地理信息系统

（Geographic Information System, GIS）领域中的一个重要基准数据集。该数据集具有空间分辨率高、覆盖区域广泛、地物类别丰富等特点，广泛应用于遥感图像语义分割、地物分类、变化检测等任务中。数据集以德国波茨坦市及其周边地区为研究对象，覆盖了约 6 平方公里的区域，涵盖了城市、农村、森林等多种复杂场景，为遥感图像语义分割任务提供了多样化的数据支持。

ISPRS Postdam 数据集包含多种遥感数据源，包括高分辨率光学遥感图像、激光雷达（LiDAR）数据以及高精度数字地表模型（DSM），这些多源数据的结合为研究者提供了多维度的信息，有助于提升语义分割模型在精度、边界细节恢复及遮挡处理等方面的综合表现。ISPRS Postdam 数据集由 38 幅高精度正校正图像构成，每幅图像的尺寸为 6000×6000像素，空间分辨率高达 5 厘米，具备了极高的图像细节还原能力。图 4-1 展示了部分 ISPRS Postdam 数据集的部分原始样本图像，第一行为标准的 RGB 格式图像，第二行为包含近红外通道的 IRRG 格式图像。其中，IRRG 图像使用近红外通道替代常规 RGB 图像中的蓝色通道，增强了对植被、建筑等地物类别的光谱区分能力。值得注意的是，ISPRS Postdam 数据集中的 IRRG 图像均为 8 位无符号整数格式，即像素值范围为[0, 255]，与 LTCAA 模型的输入要求一致，无需进行数据动态范围的额外处理。

图 4-1 ISPRS Postdam 数据集

ISPRS Postdam 数据集提供了完整详细的像素级真实标注，涵盖六大类地物目标：不透水表面（如道路、停车场等）、建筑物、低植被（如草地、灌木丛等）、树木、汽车以及杂波背景，其中前 5 个伪前景目标类，1 个背景类别用于区分非目标区域。这些类别覆盖了城市和自然环境中主要的地物类型，反映了遥感图像中常见的复杂场景。

4.2.2 数据预处理

PASCAL VOC12 数据集作为图像语义分割任务的常用数据集，包含各类常见生活场景，场景相对简单，且图像分辨率较低。相比之下，ISPRS Postdam 数据集的图像分辨率高，覆盖城市区域范围大，场景复杂且目标类别分布不均匀。直接应用自然图像的处理方法难以满足遥感图像的需求，因此需要进行针对性的数据预处理，以使得遥感数据集能更好地符合 LTCAA 模型对输入的要求并提高训练的效率。首先，由于 ISPRS Postdam 数据集的原始图像数量有限，且具有极高的分辨率和过大的尺寸，直接输入模型会导致网络训练不充分、计算资源消耗过高等问题。因此，本文首先进行图像的裁剪分块处理，将原始图像裁剪为512×512像素的图像块，使大尺寸图像被分解成易于处理的小块，使模型在训练中更加专注于局部区域特征的学习。其次，ISPRS Postdam 数据集中的场景复杂，通常各个类别会同时存在于同一图像中，并且存在遮挡、阴影、目标尺度变换等问题。因此，本文在裁剪分块时设计相邻图像块之间存在 30%的重叠区域，即每个图像块与相邻的图像块之间有部分的像素是重复的，以增加样本多样性和覆盖范围，使模型能够更好地适应复杂场景，提高识别能力和泛化能力。此外，在 ISPRS Postdam 数据集中存在目标类别分布不均衡的问题，例如筑物和道路可能占据较大比例，而汽车和植被等类别占比较小。对此本文进行了数据增强和样本筛选操作，在一定程度上缓解类别不平衡的问题，提升模型对各个类别的识别能力。与此同时，本文还特别注意避免引入干扰模型学习的样本，即舍弃在裁剪后前景类占比小于 25%的图像块。原因是这些图像块中建筑物的特征不够显著，大部分区域仍属于背景。如果将这些图像块作为训练样本，可能会误导模型，使其难以准确提取建筑物的特征，从而降低模型的性能。

4.2.3 数据集准备

为了使数据集满足 LTCAA 模型的输入要求，本文进一步对 ISPRS Postdam 数据集进行处理。本文按照 8:2 的比例将裁剪后的图像块划分为训练集和验证集。训练集用于模型的参数优化和学习过程，共 24908 张图像块；而验证集则用于实时监控模型的性能，共 6228

张图像块。而测试集则直接从特定测试图像中提取，并保留其像素级标注，以全面评估模型性能。后续本文提到的 ISPRS Postdam 数据集均为经本节处理后的数据集。

由于 ISPRS 波茨坦数据集并非基于图像级标注的弱监督语义分割任务的常用数据集，因此我们需要从像素级真实标注中获取图像级标注信息。ISPRS 波茨坦数据集的真实标注如图 4-2 所示，其中不透水表面的 RGB 颜色为(255,255,255)，建筑物为(0,0,255)，低矮植被为(0,255,255)，树木为(0,255,0)，汽车为(255,255,0)，背景类别的 RGB 颜色为(255,0,0)。为此，本研究设计以下的标注获取规则：如果一个图像块中建筑物像素的比例超过 25%，则将其标注为“building”，这表明该图像块中建筑物的特征较为显著，适合用于训练模型识别建筑物；而当图像块中完全不含建筑物像素时，则标注为“non-building”，用于训练模型识别背景或其他非建筑物地物。图像块中的其他类别如不透水表面、低矮植被、树木、汽车以及背景类等也均以上述方式进行图像级标注的获取。

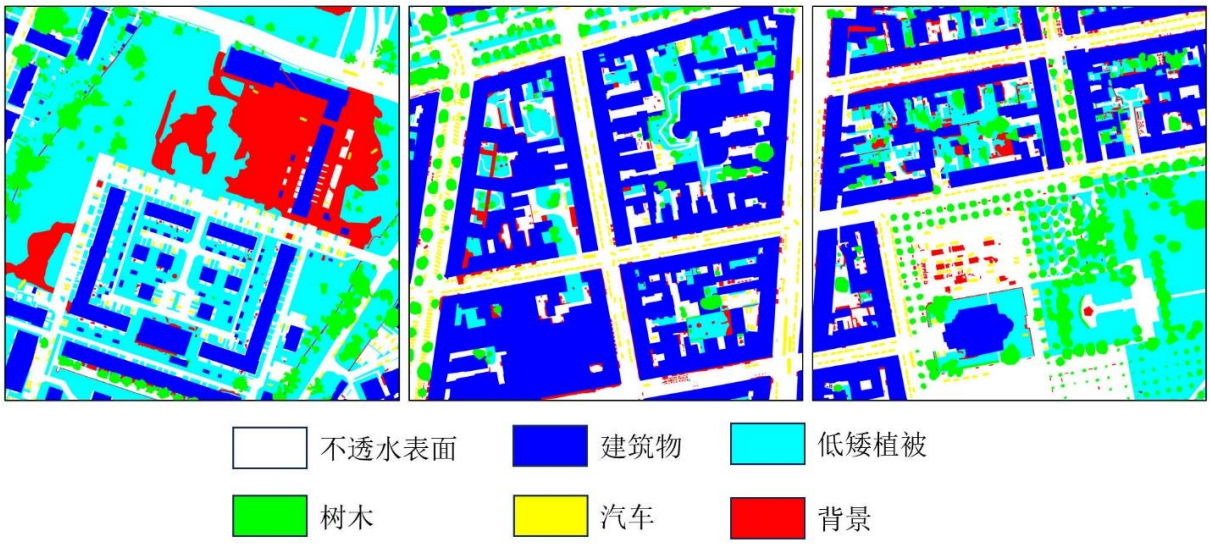


图 4-2 ISPRS Postdam 数据集真实标注

4.3 实验与分析

4.3.1 实验设置

本章的实验在 GPU 配置为 RTX 4090(24GB)，CUDA 版本为 11.0 的 Linux 平台下进行，基于 Python3.8 的 P yTorch1.7.0 的框架实现。分类网络采用 LTCAA 模型，即采用在 ImageNet 数据集上对预训练 ResNet50 作为主干分类网络，并嵌入了 LSTM、CRM 和 SAM 模块。在训练过程中采用 Adam 优化器，学习率设置为 5e-6，训练批次 batchsize 设置为 8，权重衰减设置为 5e-4，动量设置为 0.9，模型共训练的 epoch 数为 8。在伪标注生成阶段，本文使用 AffinityNet 算法将改善的 CAM 图通过随机游走策略进一步优化，相关参数设置同

AffinityNet 算法保持一致。对于分割网络，本研究使用 DeepLabV2 算法与 ResNet101 主干网络进行实验，并遵循 AdvCAM 的默认训练设置。本章的实验评价指标采用 mIoU，具体介绍详见 3.3.2 节。

4.3.2 实验结果和对比分析

为了评估 LTCAA 模型在遥感图像弱监督图像语义分割任务中的性能，本节设计进行了实验并展示了实验结果。

首先，在基于类激活映射的弱监督图像语义分割任务中，生成的伪标注质量对最终分割结果具有决定性的影响，而 CAM 算法生成的激活图是生成伪标注过程中的核心环节。由于图像的语义分割模型是使用伪标注进行训练的，伪标注的改进能够显著提升模型的建筑物提取性能。表 4-1 展示了 LTCAA 模型在 ISPRS Postdam 数据集上生成的伪标注结果，并与其他弱监督图像语义分割方法进行了对比。实验结果表明，本研究提出的 LTCAA 模型在 ISPRS Postdam 数据集上表现优异，生成的伪标注效果良好，在对目标前景类的激活准确性和激活充分性上取得了平衡。LTCAA 获取的伪标注 mIoU 达到 67.1%，显著优于原始 CAM 算法，与 IRN、AdvCAM 和 SEAM 等对比方法也有着不同程度的提升。值得一提的是，由于遥感图像中各类别分布密集，图像级标注信息几乎均为多类别标注，而 LTCAA 模型很好地实现了多类别标注的处理和使用。这一结果验证了 LTCAA 模型有效地应用于遥感图像数据集，同样能够成功缓解 CAM 算法存在的激活不准确和激活不充分等问题。

表 4-1 LTCAA 模型在 ISPRS Postdam 数据集上获取伪标注的结果

模型	mIoU(%)
CAM[39]	63.4
IRN[55]	64.5
PuzzleCAM[81]	66.2
SEAM[68]	66.9
LTCAA(ours)	**67.1**

表 4-2 给出了 LTCAA 模型在 ISPRS Postdam 数据集上的的建筑物提取结果，并与其他弱监督图像语义分割方法做了对比。实验结果表明，本研究提出的 LTCAA 模型在 ISPRS Postdam 数据集上取得了最佳的建筑物提取性能，mIoU 达到 83%，不同程度上优于其他的对比方法。由于本研究 LTCAA 模型着重提升了激活目标前景类时的准确性和完整性，并追求二者平衡，因而取得了较好的分割效果；而部分倾向于某一特定能力的方法，在遥感

数据集上会由于特征信息提取不足而导致分割效果不佳，如 IRN、CONTA、ReCAM 等。

表 4-2 LTCAA 模型在 ISPRS Postdam 数据集上的图像语义分割结果

模型	骨干网络	mIoU(%)
CAM[39]	ResNet38	73.4
SEAM[68]	ResNet38	79.9
CONTA[80]	ResNet38	77.1
IRN[55]	ResNet50	81.3
AdvCAM[44]	ResNet101	76.5
PuzzleCAM[81]	ResNet101	77.9
OME[82]	ResNet101	82.7
LTCAA(ours)	ResNet101	**83.0**

4.3.3 遥感图像语义分割结果可视化

图 4-3 展示了 LTCAA 模型在 ISPRS Postdam 数据集上生成的激活图可视化结果，直观反映了 LTCAA 模型对各个目标前景类别的激活情况。

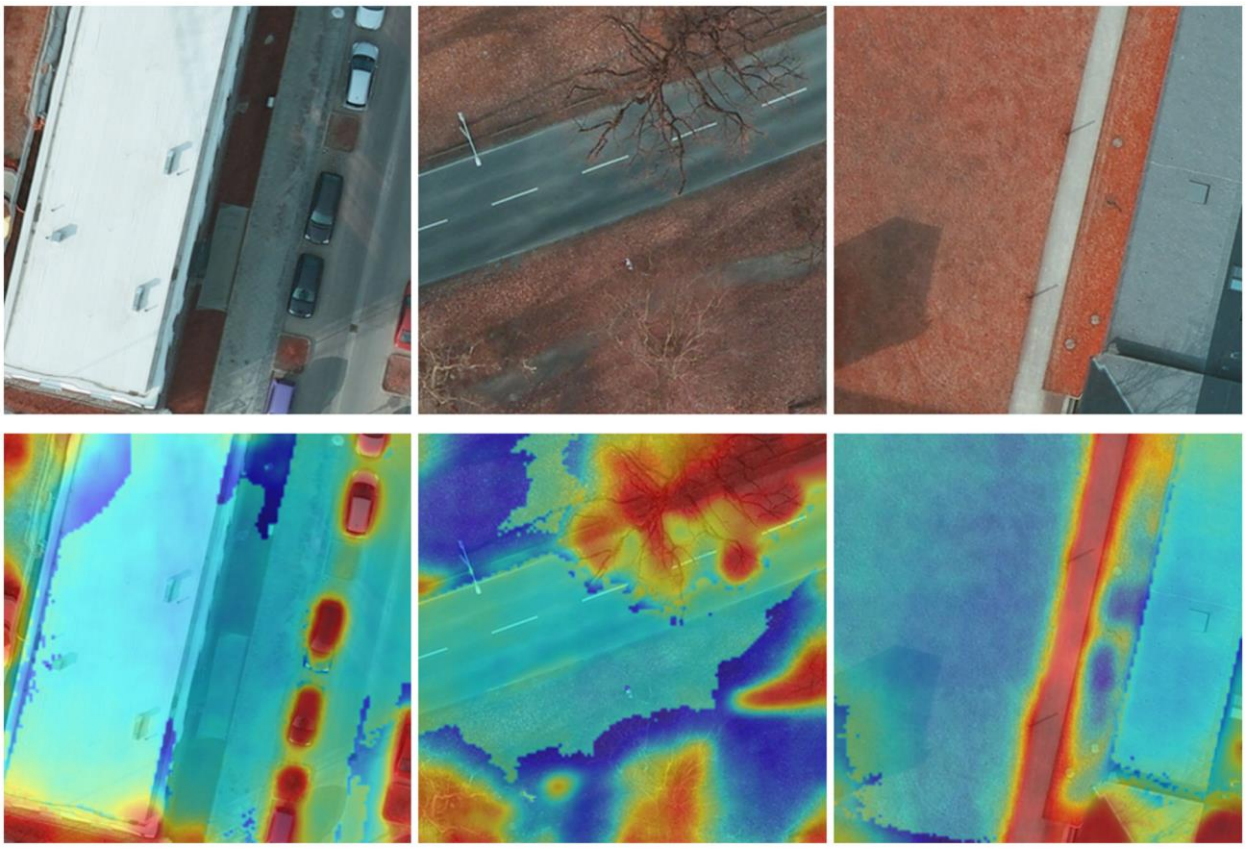


图 4-3 ISPRS Postdam 数据集的 CAM 图可视化结果

其中，第一行为输入图像，第二行是 LTCAA 模型处理后生成的激活图；(a)为对汽车类的激活图，(b)为对树木类的激活图，(c)为对不透水表面的激活图。由图 4-3 可以观察到，LTCAA 模型能够较为充分地激活各目标前景类，且对目标前景类的激活范围准确。为了进一步展示 LTCAA 模型在 ISPRS postdam 数据集上的建筑物提取性能，图 4-4 展示了 LTCAA 模型在 ISPRS Postdam 数据集上的分割结果可视化。其中(a)是原始图像，(b)为真实标注，(c)是 LTCAA 模型生成的最终分割结果，从图中可以明显看到真实标注与最终分割结果的对比情况。本研究提出的 LTCAA 模型在处理不同场景的遥感图像时表现出较好的鲁棒性。对于图像中包含单个类别，需要准确分割目标类的内部结构和外部轮廓，LTCAA 模型能够对轮廓形状复杂的单体简直实现很好的分割。然而通常情况下，遥感图像中目标类并非单独出现，LTCAA 模型在此场景下仍然表现出色，能够有效区分不同类别的目标。此外，LTCAA 模型在处理遥感图像中的多尺度目标时也表现出优异的性能，对于图像中同时存在的大型建筑和小规模建筑，LTCAA 模型能够准确识别并分割不同尺度的目标并完成建筑物提取。

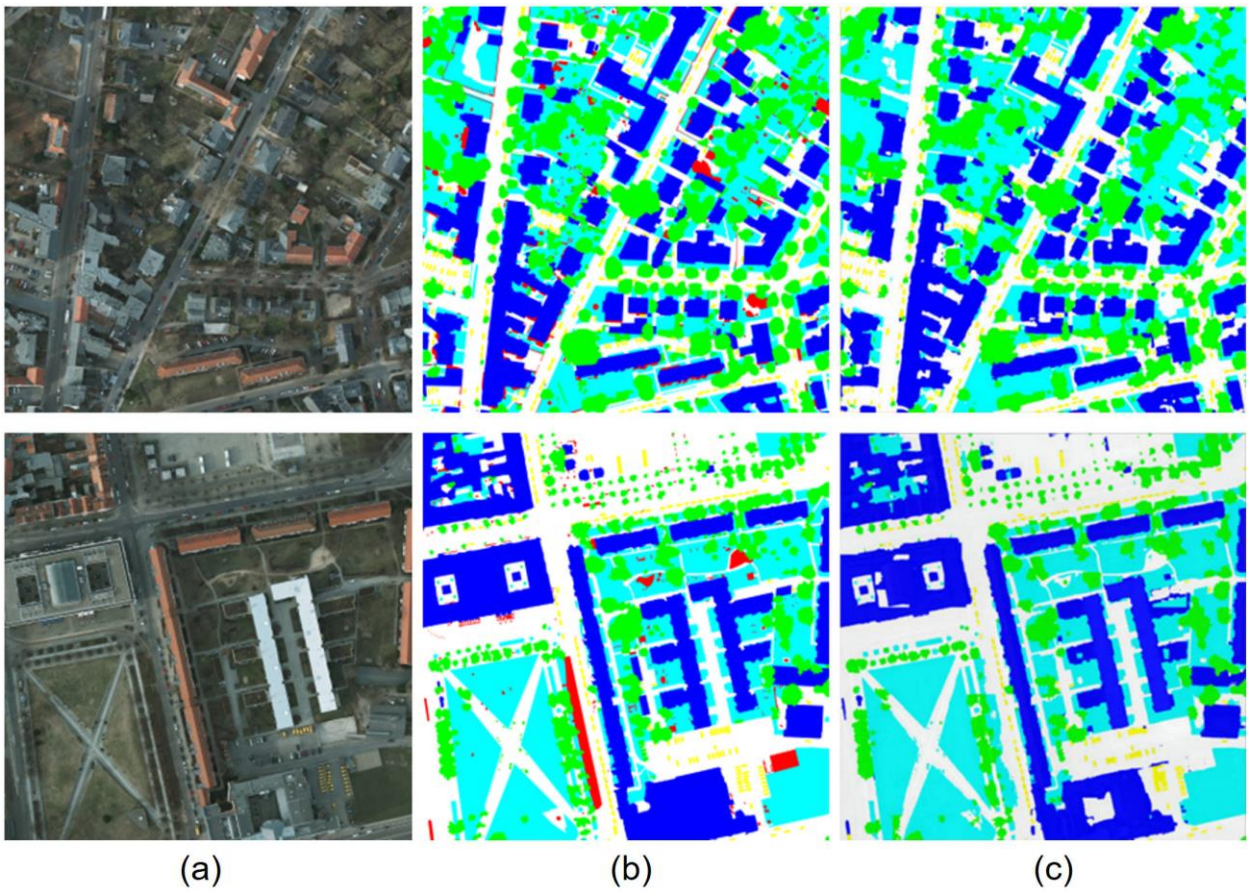


图 4-4 ISPRS Postdam 数据集的图像语义分割结果

4.4 本章小结

为了展示 LTCAA 模型在实际应用上的意义，进一步证明 LTCAA 模型的有效性和鲁棒性，本章将 LTCAA 模型应用到遥感图像数据集，进行了一系列实验研究。为了使 LTCAA 模型能够在遥感数据集上发挥最佳效果，我们首先对数据做预先处理，并获取数据的图像级标注信息。随后，我们将 LTCAA 模型在 ISPRS postdam 数据集上进行实验，并在伪标注结果和分割结果两个维度上进行实验对比，同时分析 LTCAA 模型对遥感数据集中不同场景的分割效果，表明了 LTCAA 模型能够很好地应用于遥感数据集的语义分割任务。

5 总结与展望

5.1 工作总结

图像的语义分割使计算机视觉领域中的一个重要任务，其目标是对图像中的每个像素进行分类，从而理解图像的场景和内容。基于图像级标注的弱监督图像语义分割方法是含有最少监督信息的方法，具有较高的研究价值和挑战性。为此，本文提出了一种基于标签阈值的一致性注意力架构 LTCAA 模型，采用获取成本低的图像级标注为监督信息，针对类激活映射 CAM 算法中存在的对目标前景类激活不充分和激活不准确的缺陷提出了改进策略。并且，为了证明 LTCAA 模型的实际应用意义，本文将其应用于遥感图像中，探究了 LTCAA 模型对遥感图像的分割效果。本文的具体工作总结如下：

（1）为了解决 CAM 算法现存的对目标前景类激活不充分和激活不准确等问题，本文提出了一种基于标签阈值的一致性注意力架构模型。该方法设计了标签阈值模块、一致性约束模块和自注意力模块三种改进策略。为了提高目标激活的准确性，标签阈值模块将图像级标注与显著图之间建立映射，同时通过两端映射策略扩大前景和背景之间的差距，避免对目标类的错误激活。为了解决目标激活不充分的问题，自注意力模块通过建立上下文语义关联和通道间语义依赖关系，扩充监督信息，提高了对目标类激活的充分性。同时提出了一致性约束模块，通过双分支结构实现模型对仿射变换下图像特征的学习，强化模型对目标特征的捕捉能力，解决了边界不够精细的问题。实验证明，本文设计的三个模块体现出了其有效性，LTCAA 模型能够有效提高伪标注的质量，并获取较好的图像语义分割结果。

（2）为了探究 LTCAA 模型的实际应用意义，本文实现了基于 LTCAA 模型的遥感图像语义分割。针对 LTCAA 模型对实际输入图像的要求，并结合 ISPRS postdam 数据集和 VOC12 增强数据集存在的差距，本文设计了对 ISPRS postdam 数据集的预先处理，并获取图像级标注信息。实验证明，LTCAA 模型处理遥感图像时能够获取高质量伪标注，实现了对遥感图像的语义分割任务，并且对于不同场景下的图像均能进行很好的处理，具备有效性和鲁棒性。

5.2 工作展望

本文提出的基于标签阈值的一致性注意力架构在弱监督语义分割任务中取得了不错的效果，但因为时间和个人能力的限制，本文目前的工作仍有改进的空间，具体如下：

（1）数据集增强。目前的模型在处理同时出现的高相关像素区域（如火车和铁轨等）时，仍易产生混淆，常出现误激活背景类别的现象，影响分割精度。未来可尝试将模型生成的伪标注结果重新注入训练数据，通过引入额外的伪监督样本打破原始数据集中固有的上下文依赖，提升模型泛化能力与抗干扰能力。

（2）弱监督语义分割流程优化。弱监督语义分割目前主流的方法是双阶段处理方式，即首先获取伪标签，然后再训练分割模型，这种方法虽然能够提高模型精度，但是无疑增加了弱监督语义分割任务的复杂性，因此，未来考虑优化算法流程，设计端到端的弱监督语义分割方法。

参考文献

[1] Wang J, Han Z, Chen X, et al. A fast and accurate 3D lung tumor segmentation algorithm[J]. Pattern Analysis and Applications, 2025, 28(2): 1-14.

[2] Song X, Erhao G. YGNet: A Lightweight Object Detection Model for Remote Sensing[J]. IEEE Geoscience and Remote Sensing Letters, 2025, 22: 1-5.

[3] Lyu H, Fu H, Hu X, et al. Esnet: Edge-based segmentation network for real-time semantic segmentation in traffic scenes[C]// IEEE International Conference on Image Processing (ICIP). 2019: 1855-1859.

[4] Xu Q, Guo J. Multi-Target Detection Method of Intelligent Driving Traffic Scene Based on Faster R-CNN++[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2025, 39(1).

[5] Bomans M, Hohne K H, Tiede U, et al. 3-D segmentation of MR images of the head for 3-D display[J]. IEEE Transactions on Medical Imaging, 1990, 9(2): 177-183.

[6] Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks[J]. Advances in neural information processing systems, 2017, 60(6): 84-90.

[7] Russakovsky O, Deng J, Su H, et al. ImageNet Large Scale Visual Recognition Challenge[J]. International Journal of Computer Vision, 2015, 115(3):211-252.

[8] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. CoRR, 2014, abs/1409.1556.

[9] Szegedy C, Liu W, Jia Y, et al. Going Deeper with Convolutions[C]// Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR). 2015: 1-9.

[10] 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 (CVPR). 2016: 770-778.

[11] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]// Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR). 2015: 3431-3440.

[12] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2018, 40(4): 834-848.

[13] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[C]// Proceedings of the International Conference on Learning Representations (ICLR). 2016.

[14] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: Semantic Image Segmentation with Deep

Convolutional Nets, Atrous Convolution, and Fully Connected CRFs[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(4): 834-848.

[15] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 7268–7277.

[16] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(12): 2481 2495.

[17] Wu Z，Han X，Lin Y L，et al. DCAN: Dual Channel-wise Alignment Networks for Unsupervised Scene Adaptation[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018.

[18] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation [J]. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015, 2015, 9351: 234-241.

[19] Chen L, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018: 801-818.

[20] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in neural information processing systems. 2017: 5998-6008.

[21] Zhao H, Zhang Y, Liu S, et al. Psanet: Point-wise spatial attention network for scene parsing[C]// Proceedings of the European conference on computer vision (ECCV). 2018: 267-283.

[22] Wang X, Girshick R, Gupta A, et al. Non-local neural networks[C]// Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR). 2018:7794-7803.

[23] Huang Z, Wang X, Huang L, et al. Ccnet: Criss-cross attention for semantic segmentation[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 603-612.

[24] Fu J, Liu J, Tian H, et al. Dual attention network for scene segmentation[C]// Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition (CVPR). 2019: 3146-3154.

[25] Dai J, He K, Sun J. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation[C]// Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2015: 1635-1643.

[26] Arbelaez P, Pont-Tuset J, Barron J T, et al. Multiscale Combinatorial Grouping[C]// Proceedings of the

IEEE conference on Computer Vision and Pattern Recognition (CVPR). 2014: 328-335.

[27] Rajchl M, Lee M C H, Oktay O, et al. Deepcut: Object segmentation from bounding box annotations using convolutional neural networks[C]// IEEE transactions on medical imaging. 2016, 36(2): 674-683.

[28] Boykov Y Y, Jolly M P. Interactive graph cuts for optimal boundary & region segmentation of objects in ND images[C]// Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2001(1): 105-112.

[29] Lafferty J, McCallum A, Pereira F C N. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence[C]// Proceedings of the 18th International Conference on Machine Learning 2001 (ICML 2001). 2001: 282-289.

[30] Khoreva A, Benenson R, Hosang J, et al. Simple does it: Weakly supervised instance and semantic segmentation[C]// Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR). 2017: 876-885.

[31] Ma T, Wang Q, Zhang H, et al. Delving Depper into Pixel Prior for Box-Supervised Semantic Segmentation[C]// IEEE Transactions on Image Processing. 2022(31): 1406-1417.

[32] Carbonneau M A, Cheplygina V, Granger E, et al. Multiple instance learning: a survey of problem characteristics and applications[J]. Pattern Recognition, 2018, 77(1): 329-353.

[33] Lin D, Dai J, Jia J, et al. Scribblesup: Scribble-supervised convolutional networks for semantic segmentation[C]// Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR). 2016: 3159-3167.

[34] Tang M, Djelouah A, Perazzi F, et al. Normalized Cut Loss for Weakly-Supervised CNN Segmentation[C]// Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition (CVPR). 2018: 1818-1827.

[35] Lu W, Gong D, Fu K, et al. Boundarymix: Generating pseudo-training images for improving segmentation with scribble annotations[J]. Pattern Recognition, 2021(117): 107924.

[36] Pan Z, Sun H, Jiang P, et al. CC4S: Encouraging Certainty and Consistency in Scribble-Supervised Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12): 1-18.

[37] Bell S, Upchurch P, Snavely N, et al. Material Recognition in the Wild with the Materials in Context Database[C]// Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR).

2015(14).

[38] Bearman A, Russakovsky O, Ferrari V, et al. What's the point: Semantic segmentation with point supervision[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2016: 549-565.

[39] Zhou B, Khosla A, Lapedriza A, et al. Learning deep features for discriminative localization[C]// Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR). 2016: 2921-2929.

[40] Kolesnikov A, Lampert C H. Seed, expand and constrain: Three principles for weakly-supervised image segmentation[C]// Proceedings of the European Conference on Computer Vision (ECCV). Springer, Cham, 2016: 695-711.

[41] Huang Z, Wang X, Wang J, et al. Weakly-Supervised Semantic Segmentation Network with Deep Seeded Region Growing[C]// Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR). 2018: 7014-7023.

[42] Jiang P, Hou Q, Cao Y, et al. Integral Object Mining via Online Attention Accumulation[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2019:2070-2079.

[43] Wei Y, Xiao H, Shi H, et al. Revisiting dilated convolution: A simple approach for weakly-and semi-supervised semantic segmentation [C]// Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR). 2018:7268-7277.

[44] Lee J, Kim E, Yoon S. Anti-Adversarially Manipulated Attributions for Weakly and Semi-Supervised Semantic Segmentation[C]// Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition (CVPR). 2021:4071-4080.

[45] Yu M, Wei J, Wang C, Jiang H, Yu J, Zhang R, Li X, Yu R. EDGE Enhancement Network for Weakly Supervised Semantic Segmentation[C]// IEEE International Conference on Multimedia and Expo (ICME). 2021:1-6.

[46] Chaudhry A, Dokania P K, Torr P H. Discovering class-specific pixels for weakly supervised semantic segmentation[C]// Procedings of the British Machine Vision Conference. 2017.

[47] Singh K K, Lee Y J. Hide-and-seek: Forcing a network to be meticulous for weakly supervised object and action localization[C]// Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017:3544-3553.

[48] 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 (CVPR). 2017: 1568-1576.

[49] Hou Q, Jiang P, Wei Y, et al. Self-Erasing Network for Integral Object Attention[C]// Advances in Neural Information Processing Systems 31 (NeurIPS 2018). 2018: 547–557.

[50] Redondo-Cabrera C, Baptista-Rios M, López-Sastre R J. Learning to exploit the prior network knowledge for weakly supervised semantic segmentation[J]. IEEE Transactions on Image Processing, 2019, 28(7): 3649-3661.

[51] Zhang T, Lin G, Liu W, et al. Splitting vs. merging: Mining object regions with discrepancy and intersection loss for weakly supervised semantic segmentation[C]// Proceedings of the European Conference on Computer Vision (ECCV). Springer, Cham, 2020: 663-679.

[52] Sun W, Zhang J, Barnes N. Inferring the Class Conditional Response Map for Weakly Supervised Semantic Segmentation[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022:2878-2887.

[53] Ahn J, Kwak S. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation[C]// Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition (CVPR). 2018: 4981-4990.

[54] Bertasius G, Torresani L, Yu S, et al. Convolutional random walk networks for semantic image segmentation[C]// Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR). 2017: 858-866.

[55] Ahn J, Cho S, Kwak S. Weakly supervised learning of instance segmentation with inter pixel relations[C]// Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition (CVPR). 2019:2209-2218.

[56] Chen L, Wu W, Fu C, et al. Weakly supervised semantic segmentation with boundary exploration[C]// Proceedings of the European Conference on Computer Vision (ECCV). Springer, Cham, 2020: 347-362.

[57] Yi S, Ma H, Wang X, et al. Weakly-supervised semantic segmentation with superpixel guided local and global consistency[J]. Pattern Recognition, 2022, 124: 108504.

[58] Jiang P, Han L, Hou Q, et al. Online Attention Accumulation for Weakly Supervised Semantic Segmentation[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2019: 10539-10548.

[59] Wei Y, Liang X, Chen Y, et al. STC: A Simple to Complex Framework for Weakly-supervised Semantic

Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(11). 2314-2320.

[60] Yao Q, Gong X. Saliency Guided Self-Attention Network for Weakly and Semi-Supervised Semantic Segmentation[J]. IEEE Access, 2020, 8, 14413–14423.

[61] Xie J, Xiang J, Chen J, et al. C^2 AM: Contrastive learning of Class-agnostic Activation Map for Weakly Supervised Object Localization and Semantic Segmentation[C]// Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR). 2022: 979-988.

[62] Fan J, Zhang Z, Tan T, et al. CIAN: Cross-Image Affinity Net for Weakly Supervised Semantic Segmentation[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020:10762-10769.

[63] Li X, Zhou T, Li J, et al. Group-wise semantic mining for weakly supervised semantic segmentation[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2021:1984-1992.

[64] Zhang X, Wei Y, Yang Y. Inter-Image Communication for Weakly Supervised Localization[C]// Proceedings of the European Conference on Computer Vision (ECCV). Springer, Cham, 2020:271-287.

[65] Qin J, Wu J, Xiao X, et al. Activation Modulation and Recalibration Scheme for Weakly Supervised Semantic Segmentation[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2022.

[66] Shimoda W, Yanai K. Self-Supervised Difference Detection for Weakly-Supervised Semantic Segmentation[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 5208-5217.

[67] Chang Y, Wang Q, Hung W, et al. Weakly-Supervised Semantic Segmentation via Sub-Category Exploration[C]// Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR). 2020: 8988-8997.

[68] Wang Y, Zhang J, Kan M, et al. Self-Supervised Equivariant Attention Mechanism for Weakly Supervised Semantic Segmentation[C]// Proceedings of the IEEE/ CVF Computer Vision and Pattern Recognition (CVPR). 2020: 12272-12281.

[69] Chen Q, Yang L, Lai J, et al. Self-supervised Image-specific Prototype Exploration for Weakly Supervised Semantic Segmentation[C]// Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition (CVPR). 2022: 4278-4288.

[70] Lee M, Kim D, Shim H. Threshold Matters in WSSS: Manipulating the Activation for the Robust and Accurate Segmentation Model Against Thresholds[C]// Proceedings of the IEEE/CVF Computer Vision and Pattern Recognition (CVPR). 2022: 4320-4329.

[71] Lei J, Yang G, Wang S, et al. DCAM: Disturbed class activation maps for weakly supervised semantic segmentation[J]. Journal of Visual Communication and Image Representation, 2023(94): 103852.

[72] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]// Advances in Neural Information Processing Systems. 2017: 5998-6008.

[73] Gao W, Wan F, Pan X, et al. TS-CAM: Token Semantic Coupled Attention Map for Weakly Supervised Object Localization[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2021: 2866-2875.

[74] Ru L, Zhan Y, Yu B, et al. Learning Affinity from Attention: End-to-End Weakly-Supervised Semantic Segmentation with Transformers [C]// Proceedings of the IEEE/CVF Computer Vision and Pattern Recognition (CVPR). 2022: 16846-16855.

[75] Li R, Mai Z, Trabelsi C, et al. TransCAM: Transformer Attention based CAM Refinement for Weakly Supervised Semantic Segmentation[J]. Journal of Visual Communication and Image Representation, 2023(92): 103800.

[76] Selvaraju R R, Das A, Vedantam R, et al. Grad-CAM: Why did you say that? Visual Explanations from Deep Networks via Gradient-based Localization[J]. arXiv e-prints, 2016.

[77] Chattopadhay A, Sarkar A, Howlader P, et al. Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks[C]// 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2018: 839-847.

[78] Everingham M, Winn J. The pascal visual object classes challenge 2012 (voc2012) development kit[J]. Pattern Analysis, Statistical Modelling and Computational Learning, Tech Rep, 2011(8): 5.

[79] Hariharan B, Arbeláez P, Bourdev L, et al. Semantic contours from inverse detectors[C]// 2011 International Conference on Computer Vision, Barcelona, Spain, 2011: 991-998.

[80] Zhang D, Zhang H, Tang J, et al. Causal intervention for weakly-supervised semantic segmentation[J]. Advances in Neural Information Processing Systems, 2020, 33: 655-666.

[81] Jo S, Yu I J. Puzzle-cam: Improved localization via matching partial and full features[C]// 2021 IEEE International Conference on Image Processing (ICIP). IEEE, 2021: 639-643.

[82] Li Z, Zhang X, Xiao P. One Model Is Enough: Toward Multiclass Weakly Supervised Remote Sensing Image Semantic Segmentation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 66: 1-13.

致谢

时光如白驹过隙，转眼间三年的研究生生涯已近尾声，研究生生活即将画上句点。读研这三年，课程学习、科学研究以及组会分享的过程令我收益良多，培养了我的独立思考能力、逻辑思维能力、文档撰写能力和语言表达能力。研究生的成长与蜕变离不开师长和同门的帮助，此时此刻，毕业论文即将完成之际，我谨向他们表达我最诚挚的感谢。

首先我要感谢我的导师牛云云教授。自第一次见面，就感受到了牛老师的认真负责、温和耐心、平易近人，在我整个研究生期间提供了悉心的指导和支持。在学术研究上，牛老师给予我宝贵的指导和建议，引导我逐渐学会独立思考。在这篇论文的研究和撰写期间，您更是事事关心，为论文的完成提供了许多宝贵意见。在生活以及个人发展上，您也给予了我许多关心和鼓励，在您的教诲中了解到全面发展对个人发展的重要意义，也在文献阅读过程中认识到厚积薄发的重要性。教诲如春风，师恩似海深，我会将老师您的教诲刻入心底，伴随我今后的人生，在此真诚地祝福您身体健康，愿您桃李满天下，春晖遍四方！

其次我要特别感谢我的同门杨雨晴、阮龙、张小豪、徐泽彦和师兄朱国栋。在日常的学习中你们给了我许多的帮助，耐心解答我在科研上遇到的许多疑问；在找工作的时候你们无私分享了许多求职信息，让我能够顺利找到满意的工作。然后我还要感谢朝夕相处的室友王潇娴、刘振莉、李源琳和来自临近宿舍的杨雨晴、陈华熔、刘嘉薇、夏广宁，在学习和生活上给予我的帮助，相互扶持和包容的日子历历在目，感谢你们陪伴着我跨过一个又一个难关，让我感受到了家乡之外的温暖。未来的日子里，愿我们依然互相陪伴，携手共进，希望大家步入社会后各自灿烂，在更高处相见！

最后我要感谢我的父母。读研三年，岁月也在你们的脸上留下了痕迹。有了你们对我的默默支持与付出，我才能见识到更广阔的世界。生活中，父亲的爱如大山般坚定，寡言少语的背后是对我殷勤的期盼，在我困顿之时您总能给我指点迷津，让我更有勇气面对生活的挑战。母亲的爱如春日里的暖阳，衣食住行面面俱到，您的每一次通话都让我感受到了家人的温暖，更加强了我奋勇向前的决心。二十五载的养育之恩，愿自己能够不断努力成长，成为你们的骄傲！衷心祝愿父母身体健康，开心过好每一天！

人生无不散之筵席，我即将告别学生时代，正式踏入社会。在今后的工作和生活中，我会铭记老师的教诲，在今后的学习工作中加倍努力。再次感谢上述提到或未提到的所有帮助过我的人，谢谢！

附录

一、个人简介

陈若妍，女，1999 年 12 月出生，河北张家口人，中共党员。

2018 年 9 月~2022 年 6 月于华中农业大学信息学院攻读学士学位，专业为计算机科学与技术；2022 年 9 月~2025 年 6 月于中国地质大学（北京）信息工程学院攻读硕士学位，专业为计算机技术，研究方向为人工智能与模式识别。

二、研究生期间获奖情况

于 2022-2023 学年取得中国地质大学（北京）研究生新生奖学金。

于 2023-2024 学年取得中国地质大学（北京）研究生学业一等奖学金。

于 2024-2025 学年取得中国地质大学（北京）研究生学业二等奖学金。

二、研究生期间学术成果

计算机软件著作权《基于图像级标注的弱监督学习图像语义分割算法系统 v1.0》。