用于超像素分割的边缘增强状态空间模型

许云扬[1)] 房乐鑫[1)] 张彩明[1),2)] 李雪梅[1)]

[1)](山东大学软件学院 济南 250101)

[2)](山东省未来智能金融工程实验室 山东 烟台 264005)

摘 要 超像素分割算法通过聚合颜色及低级特征相似的像素,可以大幅减小计算机视觉任务中的处理对象数量,提高其计算效率。受限于CNN较小的感受野,现有的基于CNN的超像素分割方法在理解图像全局结构时存在一定的限制。此外,由于大多数方法依赖隐式学习来推断物体边界,导致其在复杂边界和弱边缘区域的分割效果不佳。并且,仅使用两个损失函数的简单加权来训练网络也限制了分割准确性和超像素形状规则性之间平衡的优化。在这项研究中,我们利用所提出的EE-SSM模型来解决长程空间依赖建模、复杂边界处理以及规则性和准确性二者间的平衡的挑战。通过一个基于状态空间模型构建的编码器和一个基于边界概率的自适应损失函数,EE-SSM实现了高精度的分割,能够在保持超像素规则性的同时维持边界的紧密贴合。论文的关键贡献是一个新颖的即插即用的轻量级边缘强化框架。该框架通过为编码过程提供显式的边缘特征,显著提升模型在处理复杂边界时的能力。通过在多个真实世界的图像数据集上进行广泛实验,EE-SSM展示了其卓越的有效性和鲁棒性。与近两年最先进的方法相比,EE-SSM在BR和UE指标上分别实现了2.41%-18.65%和14.00%-14.32%的显著提升。

关键词 超像素分割;Mamba;即插即用;显著性检测;边缘提取;边缘概率

中图法分类号 TP391 **DOI号** 10.11897/SP.J.1016.2025.02298

Edge-Enhanced State Space Model for Superpixel Segmentation

XU Yun-Yang[1)] FANG Le-Xin[1)] ZHANG Cai-Ming[1),2)] LI Xue-Mei[1)]

[1)](School of Software, Shandong University, Jinan 250101)

[2)](Shandong Provincial Laboratory of Future Intelligence and Financial Engineering, Yantai, Shandong 264005)

Abstract Superpixel segmentation algorithms significantly reduce the number of processing units in computer vision tasks by aggregating pixels with similar colors and low-level features, thereby improving their computational efficiency. Traditional superpixel segmentation methods typically rely on handcrafted features, which limits their performance in complex scenes or cross-domain images, resulting in lower accuracy and poor generalization. In recent years, deep learning-based superpixel segmentation algorithms have emerged, utilizing a convolutional neural network (CNN) to extract deep semantic features from images automatically. This approach has significantly enhanced both segmentation accuracy and cross-domain generalization. However, due to the limited receptive field of CNN, existing CNN-based superpixel segmentation methods face certain limitations when it comes to understanding the global structure of images. Moreover, as most methods rely on implicit learning to infer object boundaries without explicitly modeling the relationship between boundary information and semantic understanding, they struggle to differentiate regions that are semantically distinct but color-similar. This limitation becomes particularly pronounced when handling complex boundaries or low-contrast images, often

收稿日期:2025-04-21;在线发布日期:2025-07-22。本课题得到国家自然科学基金联合基金(Nos. U22A2033,U24A20219)资助。
许云扬,博士研究生,主要研究领域为图像超分辨率、图像分割。E-mail:xuyunyang@mail.sdu.edu.cn。**房乐鑫**,博士研究生,主要研究领域为医学图像分割。**张彩明**,博士,教授,主要研究领域为数字图像处理、时间序列分析。**李雪梅**(通信作者),博士,教授,主要研究领域为数字图像处理、时间序列分析。E-mail:xmli@sdu.edu.cn。

resulting in segmentation errors. Moreover, training the network with the simple weighting of two loss functions also limits the optimization of the balance between segmentation accuracy and superpixel shape regularity. In this study, we tackle the challenges of long-range spatial dependency modeling, complex boundary handling, and the balance between regularity and accuracy with our proposed framework called EE-SSM. The EE-SSM achieves high-precision segmentation by utilizing an encoder built on a state space model and an adaptive loss function based on boundary probability, which can preserve the regularity of superpixels while maintaining the adhesion of boundaries. Among them, the encoder based on the state-space model utilizes its selective scanning mechanism and hardware perception characteristics to effectively capture cross-region visual associations through serialized feature modeling, thereby greatly improving the semantic consistency of superpixel segmentation results. The key contribution is a novel plug-and-play lightweight edge enhancement framework. This framework significantly improves the model's ability to handle complex boundaries by providing explicit edge features during the encoding process. To maximize the utilization of boundary information, we design a feature fusion network that integrates cross-attention with a selective scanning mechanism, further enhancing the information consolidation process during encoding. Finally, to address the trade-off between segmentation accuracy and regularity, we propose a novel boundary-probability-guided adaptive loss function. This loss function dynamically adjusts the weighting of representational consistency loss and spatial compactness loss based on the boundary probability of each pixel. It strengthens representational consistency in boundary regions, improving boundary alignment, while prioritizing spatial compactness in non-boundary areas to preserve the regular shape of the superpixels. This loss function better balances the accuracy and regularity of superpixel segmentation. Through extensive experiments on multiple real-world image datasets, EE-SSM demonstrates its excellent effectiveness and robustness. The generated superpixel results are applied to the saliency detection task, which further verifies the effectiveness of the method in practical application scenarios and its adaptability to downstream visual tasks. Compared with the most advanced methods in the past two years, EE-SSM achieves significant improvements of 2.41%-18.65% and 14.00%-14.32% in BR and UE indicators, respectively.

Keywords superpixel segmentation; Mamba; plug and play; saliency detection; edge extraction; edge probability

1 引 言

超像素通过将颜色及其他低级属性相似的像素分组来提供紧凑的图像表示。这种方式能够显著提高视觉算法的计算效率[1-2]。自从超像素概念首次在[3]中提出以来，它已广泛应用于图像分割[4-5]、目标跟踪[6-7]、场景解析[8]、对象识别[9-10]、分类[11-12]和显著性检测[13-14]中。

受限于人工设计的特征，传统的超像素分割算法在复杂场景或跨域图像中往往表现出精度不足和泛化能力较差的问题。基于深度学习的算法（如SSN[15]、SSFCN[16]、AINet[17]和CDS[18]）通过卷积神经网络（CNN）自动提取深层语义特征，提升了分割精度及跨域泛化性能。但是受限于CNN固有的局部感受野特性，这些方法难以建模图像中的长程空间依赖关系，导致其在全局语义信息上的整合能力不足，易将相邻但语义不同的物体错误地划分到同一超像素中（例如图1中飞机的影子与地面被误分为同一部分）。

视觉Transformer（Vision Transformer，ViT）通过自注意力机制突破了CNN局部感受野的限制，但其自注意力机制的计算复杂度会随输入规模的增加呈二次增长。因此难以适用于需要快速响应的超像素分割场景中。为了解决全局信息建模能力与计算效率之间的矛盾，基于状态空间模型（SSM）

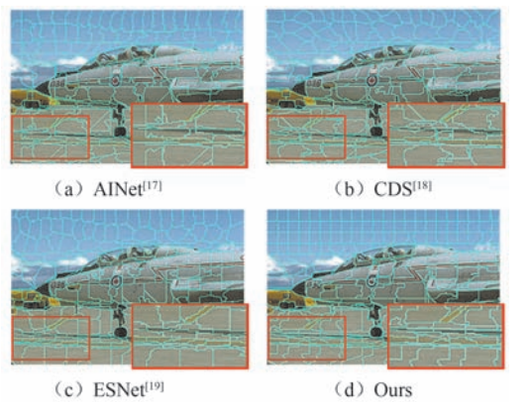


图1 不同超像素方法的分割结果

的方法[20-21]在计算机视觉领域受到了广泛关注。Mamba[22]通过引入的硬件感知和输入相关算法，进一步优化了SSM的计算效率和上下文建模能力。然而，SSM在超像素分割任务中的潜力尚未被充分挖掘。如何有效地将SSM应用于超像素分割任务，成为当前亟待探索的研究方向。

边界信息在图像分割任务中发挥了重要作用[23-25]，能够显著提升分割模型的语义一致性。现有的基于深度学习的超像素分割方法[17-19]通常侧重于从图像的颜色分布、纹理模式及局部对比度等视觉特征中隐式推断物体轮廓。由于未能显式建模边界信息与语义感知之间的关联，这些方法在处理复杂边界或低对比度图像时，难以区分语义不同但颜色相近的区域，从而引发分割错误（如图2中的鹿角部分）。因此，需要一种有效建模和利用边界信息的方法来增强超像素分割算法的语义感知能力。

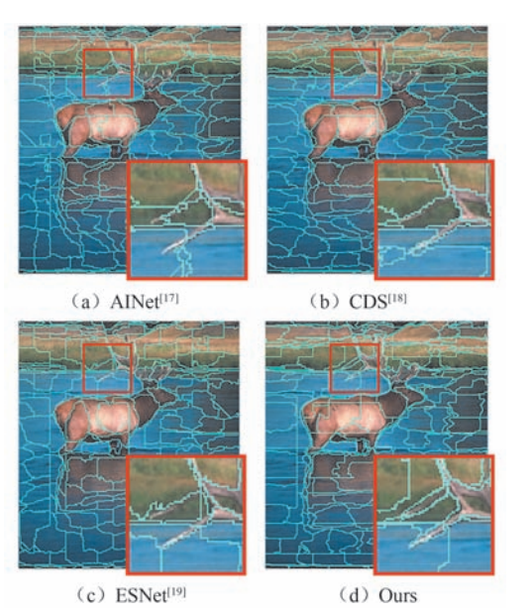


图2 不同方法在复杂区域的分割结果

此外，现有方法[17-19]通常采用双损失函数来优化超像素分割的准确性和规则性。其中，表征一致性损失通过鼓励模型将具有相似表征属性（如颜色、纹理等）的像素聚集在一起，从而确保超像素边界与物体边界紧密对齐；空间紧致性损失则将位置上相邻的像素分组在同一个超像素中以维持规则形状。特别是在非边界区域，保持超像素的规则性可为下游任务提供更可靠的支持[13,26]。然而，目前的方法采用固定的权重来加权二者，必然导致它们相互制约。例如，加强对表征一致性的依赖可以提高超像素的边界粘附，但易使非边界区域形状不规则。而强调规则性则会削弱边界的黏附能力。如何动态平衡分割精度与规则性，仍是该领域的核心挑战。

在本文中，我们提出了一种基于边界增强的状态空间模型（EE-SSM），从全局上下文建模、边界信息强化的语义感知以及动态损失优化三个层面入手，全面提升超像素分割的准确性与规则性。EE-SSM的关键贡献在于其即插即用的轻量级边缘强化框架。该框架为分割模型的推理阶段和后续优化提供显式的边缘信息，同时能够无缝集成至现有的超像素分割方法中，增强其对边界的区分能力。

具体而言，EE-SSM基于状态空间编码器-卷积解码器架构，以兼顾长程依赖建模与高效推理。其中，基于状态空间模型的编码器利用其选择性扫描机制和硬件感知特性，通过序列化特征建模有效捕获跨区域视觉关联，显著提升了超像素分割结果的语义一致性。在解码阶段则采用轻量级卷积网络提高模型分割效率。特别是，设计了一个用于超像素分割的即插即用的轻量级边缘强化框架为模型提供显式的边缘信息。为了确保边缘信息在编码过程中得到充分利用，我们还提出了一种融合交叉注意力与选择性扫描机制的特征融合网络。最后，使用新提出的基于边界概率的自适应损失函数对模型进行训练，以解决超像素分割准确性与规则性之间的平衡优化问题。该损失函数根据每个像素的边界概率动态调整表征一致性损失和空间紧致性损失的在总损失中的比重：在边界区域，模型更依赖表征一致性以增强边界对齐能力；在非边界区域，则更注重空间紧致性以保持超像素的规则形状。

我们在四个具有真实标注的基准数据集上进行了实验，并将生成的超像素应用于显著性检测任务，进一步验证了该方法在实际应用中的有效性与泛化能力。本研究的主要贡献如下：

（1）提出了EE-SSM，通过高效建模长程依赖

关系和显式利用边缘信息，显著提升了对结构复杂和低对比度区域的分割性能，同时在计算效率上具有明显优势。这一工作为超像素分割任务提供了一种性能优异的新范式。

(2)构建了一种轻量化的即插即用的边缘强化框架，专门用于提取图像的边缘特征。该框架可直接集成到现有超像素分割模型中，为推理阶段和后续优化提供显式的边缘信息。

(3)结合交叉注意力和选择性扫描机制设计了一个新颖的特征融合网络，通过将边缘信息有效地整合到编码过程中，增强了模型对复杂边界和弱边缘区域的语义理解能力，从而减少误分割并提高分割精度。

(4)设计了一种基于边界概率的自适应超像素分割损失函数。该损失函数通过像素的边界概率对表征一致性损失和空间紧致性损失进行自适应地加权组合，以更好地平衡分割准确性与规则性。

2　相关工作

在过去的二十年中，许多超像素算法被提了出来。在此简要回顾这些方法，并介绍状态空间模型。

2.1　传统的超像素方法

传统的超像素算法可分为两类：基于图的和基于聚类的。基于图的方法将超像素分割转化为图分割问题，通过图论算法划分图像。典型算法包括Random Walks[27]和ERS[28]，这类方法在分割边界准确性上表现较好，但规则性较差。基于聚类的方法受到经典聚类算法(如k-means)的启发，结合颜色和空间位置信息将像素聚类为超像素。SLIC[29]和LSC[30]等代表方法在规则性与准确性之间取得较好平衡。此外，SEEDS[31]和ETCCPS[32]等方法通过优化能量函数来提升超像素生成效果，在边界贴合性方面取得了进步。然而，受限于手工设计的特征，传统方法在精度和泛化能力方面存在一定的局限。

2.2　基于深度学习的超像素方法

传统的超像素分割算法通常基于离散计算，这使得它们难以与梯度反向传播机制兼容，限制了它们在深度学习框架中的应用。为了解决这一问题，Jampani等人提出了可微SLIC，并开发了第一个端到端可训练的超像素网络——SSN[15]，使得超像素分割能够与神经网络无缝整合，从而提升了分割精度。随后，SSFCN[16]将超像素分割转化为像素与邻近超像素的分类任务，提高了计算效率。在此基础上，ESNet[19]提出了一个由特征提取器和超像素生成器组成的高效网络，增强了对聚类友好特征的描述能力。PCNet[33]提出了一种新型深度学习框架，通过采用解耦的一致性学习策略，使得模型能够从低分辨率输入中预测高分辨率输出，从而高效生成高分辨率超像素结果，避免了GPU内存限制。CDS[18]则通过选择性地分离像素间的不变相关性和统计特性来优化超像素生成过程。ESNet和CDS是目前性能最优的SOTA算法。然而，当前基于深度学习的方法(如文献[17-19,34])大多使用CNN来处理图像，在捕捉大范围空间关系方面存在一定的局限性，同时尚未充分利用图像中的边界信息。

2.3　状态空间模型

为了更好地理解本文的工作，我们简要介绍状态空间模型(SSM)。SSM源于经典控制理论，近年来已成为深度学习中有效的序列建模方法。结构化状态空间序列模型(S4)[35]作为捕捉远程依赖关系的开创性工作，受到了广泛关注。基于S4的选择性机制，Mamba[22]进一步提升了性能，超越了Transform和其他先进架构，并被扩展到计算机视觉领域。例如，Vision Mamba[36]结合SSM与双向扫描机制，增强了图像各patch间的相互关联；VMamba[37]通过引入四个方向的扫描机制，进一步捕捉图像patch间的丰富关系。SSM已成功应用于医学图像分割[38-39]和图像恢复[40]等领域，取得了富有竞争力的结果。在这项工作中，我们探索了Mamba在超像素分割方面的应用潜力，并采用了强化边缘特征的设计，为未来的研究提供了一个简单而有效的基线方法。

3　预备知识

在本节中，我们将简要介绍状态空间模型以便更好地理解本文所提出的方法。

经典状态空间模型(State Space Model，SSM)表示一个连续系统，它将输入序列$\boldsymbol{x}(t)\in R^L$映射到潜在空间表示$\boldsymbol{h}(t)\in R^N$，然后根据此表示预测输出序列$\boldsymbol{y}(t)\in R^L$。由于其在建模长序列数据方面的高效性，SSM已成为处理长序列数据时极具潜力的架构，特别是在具有线性复杂度的场景中。从数学上讲，SSM可以描述如下：

$$\boldsymbol{h}'(t)=\boldsymbol{A}\boldsymbol{h}(t)+\boldsymbol{B}\boldsymbol{x}(t),\boldsymbol{y}(t)=\boldsymbol{C}\boldsymbol{h}(t) \quad (1)$$

其中，$\boldsymbol{h}'(t)$为$\boldsymbol{h}(t)$的时间导数，$\boldsymbol{A}\in R^{N\times N}$，$\boldsymbol{B}\in R^{N\times 1}$和$\boldsymbol{C}\in R^{1\times N}$是可学习参数，$N$为状态大小。为了处理图像和文本等离散序列，需要对连续的状态空间模型进行离散化操作。通过引入时间尺度参数$\Delta\in R$，并采用广泛使用的零阶保持[35]作为离散化规则，可以推导出$\boldsymbol{A}$和$\boldsymbol{B}$的离散化版本，分别表示为$\bar{\boldsymbol{A}}$和$\bar{\boldsymbol{B}}$：

$$\bar{\boldsymbol{A}}=\exp(\Delta\boldsymbol{A}),\bar{\boldsymbol{B}}=(\Delta\boldsymbol{A})^{-1}(\exp(\Delta\boldsymbol{A})-\boldsymbol{I})\Delta\boldsymbol{B} \quad (2)$$

其中，$\boldsymbol{I}$表示单位矩阵。将$\boldsymbol{A}$、$\boldsymbol{B}$离散为$\bar{\boldsymbol{A}}$、$\bar{\boldsymbol{B}}$后，公式(1)可改写为

$$\boldsymbol{h}(t)=\bar{\boldsymbol{A}}\boldsymbol{h}_{t-1}+\bar{\boldsymbol{B}}\boldsymbol{x}_t,\ \boldsymbol{y}_t=\boldsymbol{C}\boldsymbol{h}_t \quad (3)$$

公式(3)表示从$\boldsymbol{x}_t$到$\boldsymbol{y}_t$的序列到序列映射。这种设置允许将离散化的SSM作为RNN进行计算。然而，由于序列具有顺序性，这种离散化的循环SSM不适合用于训练。为了实现高效的并行训练，可以将该递归过程以全局卷积方式实现：

$$\bar{\boldsymbol{K}}=\left(\boldsymbol{C}\bar{\boldsymbol{B}},\boldsymbol{C}\overline{\boldsymbol{A}\boldsymbol{B}},\cdots,\boldsymbol{C}\bar{\boldsymbol{A}}^{L-1}\bar{\boldsymbol{B}}\right),\ y=x\odot\bar{\boldsymbol{K}} \quad (4)$$

其中，L表示输入序列x的长度，$\odot$表示卷积运算，$\bar{\boldsymbol{K}}\in R^{L}$是卷积核。虽然SSM对离散序列建模是有效的，但SSM中的参数相对于输入保持不变。为了解决这个问题，Mamba[22]引入的选择性状态空间(S6)机制，允许参数$\bar{\boldsymbol{B}}$、$\bar{\boldsymbol{C}}$和Δ根据输入序列x动态调整。这一改进显著增强了基于SSM的模型性能，使模型能够更有效地模拟长序列中复杂的相互作用。将$\bar{\boldsymbol{B}}$、$\bar{\boldsymbol{C}}$和Δ与输入相关后，公式(4)中的全局卷积核可以改写为

$$\bar{\boldsymbol{K}}=\left(\boldsymbol{C}_L\bar{\boldsymbol{B}}_L,\boldsymbol{C}_L\bar{\boldsymbol{A}}_{L-1}\bar{\boldsymbol{B}}_{L-1},\cdots,\boldsymbol{C}_L\prod_{i=1}^{L-1}\bar{\boldsymbol{A}}_i\bar{\boldsymbol{B}}_1\right) \quad (5)$$

4 方 法

在本节中，将详细介绍提出的用于超像素分割的边缘强化状态空间模型(EE-SSM)。如图3所示，该方法基于曼巴编码器-卷积解码器架构，通过建模长程依赖关系和显式利用边缘信息，实现了高精度的分割。

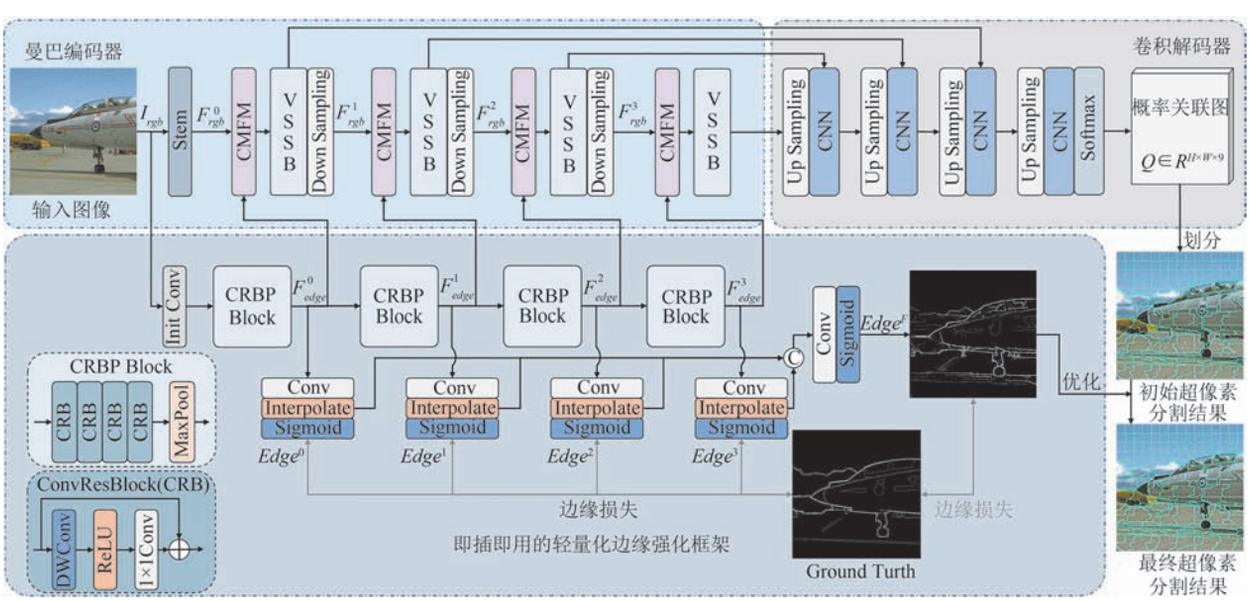


图3 模型结构图

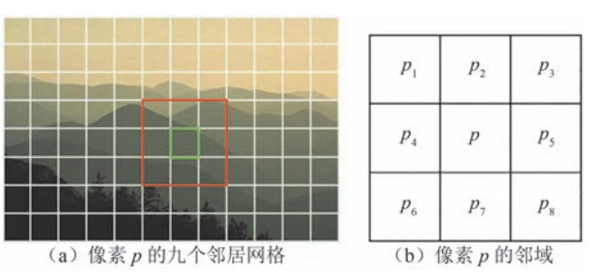
针对现有方法未能显式建模边界信息与语义感知之间关联的问题，EE-SSM首先引入轻量级边缘强化框架提取输入图像$\boldsymbol{I}_{\text{rgb}}\in R^{H\times W\times 3}$(其中$H$和$W$表示输入图像的高度和宽度)中的边缘信息，以为编码阶段提供显式的边缘特征$\boldsymbol{F}^n_{\text{edge}}$和后续优化过程提供边缘引导$Edge^F$，增强分割模型对边界的区分能力(第4.3节)。由于边缘强化框架引入了深度学习机理，并通过训练自动学习特征，能够提供比传统边缘检测方法(如Canny或Sobel)更精确的边缘信息，特别是在复杂边界和弱边缘区域，从而显著提升了边缘信息提取的准确性和鲁棒性。

在编码阶段，采用视觉状态空间块(Visual State Space Block，VSSB)和下采样模块(Down Sampling)，对输入图像$\boldsymbol{I}_{\text{rgb}}\in R^{H\times W\times 3}$进行逐层特征提取生成多尺度特征$\boldsymbol{F}^n_{\text{rgb}}$，并建模长距离依赖关系，使模型能够有效捕捉全局上下文信息(第4.1节)。

其中，VSBB将输入特征展平成序列，利用选择性扫描2D模块（Selective Scanning 2D Module，SS2DM）从四个方向分别建模远程依赖，并通过序列重组以融合多视角上下文信息，从而增强特征的表达能力。该设计有效克服了当前基于CNN方法在全局建模方面的局限。

接下来，来自VSSB的多尺度特征$\boldsymbol{F}^n_{rgb}$在交叉曼巴融合模块（Cross Mamba Fusion Module，CMFM）（第4.2节）中与边缘特征$\boldsymbol{F}^n_{edge}$进行自适应融合生成更具区分性的融合特征$\boldsymbol{F}^n_{fusion}$。为实现这一过程，CMFM首先通过联合选择性扫描模块（Joint Selective Scan Module，JSSM）对多尺度特征$\boldsymbol{F}^n_{rgb}$与边缘特征$\boldsymbol{F}^n_{edge}$进行联合建模。JSSM将$\boldsymbol{F}^n_{rgb}$和$\boldsymbol{F}^n_{edge}$分别序列化并拼接，通过SS2DM建模远程依赖关系，利用二者之间的语义互补性增强各自的表达能力。随后，CMFM引入交叉注意力机制，对增强后的特征进行动态加权融合，使模型能够根据分割需要自适应调整融合权重，从而生成更具区分度的融合特征表示$\boldsymbol{F}^n_{fusion}$。该融合特征随后被输入至下一个VSSB。以实现边缘信息在语义表示空间中的逐层累积和强化，提升模型对复杂结构的语义理解能力。

在解码阶段，模型通过上采样操作（Up Sampling）和CNN逐步恢复特征图分辨率，并利用Softmax分类器生成概率关联图$\boldsymbol{Q}\in R^{H\times W\times 9}$。该图表示每个像素与其周围九个网格（图4(a)中红色框包围的9个网格）的归属概率。



（a）像素p的九个邻居网格　　（b）像素p的邻域

图4　像素p的邻居网格和其邻域

根据概率关联图$\boldsymbol{Q}$，像素被分配到概率最高的邻近网格，并更新超像素标签，得到超像素分割结果。最后，围绕超像素分割准确性与规则性之间平衡的优化问题，提出了一种基于边界概率的自适应损失函数。该损失函数能够根据像素的边界概率动态调整表征一致性损失和空间紧致性损失的权重，从而实现超像素分割准确性与规则性的优化平衡（第4.4节）。

4.1　曼巴编码器-卷积解码器架构

针对图像全局结构建模挑战，我们将状态空间模型引入超像素分割任务，设计了曼巴编码器（Mamba Encoder），突破了CNN方法在长程依赖建模的局限。该方法通过序列化特征建模，有效捕获跨区域视觉关联，显著提升了语义一致性。在解码阶段，则采用轻量级卷积网络，以高效地生成超像素分割结果。

曼巴编码器：Mamba Encoder由四个视觉状态空间块（VSSB）组成，如图3所示。首先，与ViT[41]类似，使用一个初始特征提取模块Stem对输入图像进行初步处理，将图像分解为多个patch，并生成初级特征图$\boldsymbol{F}^0_{rgb}$。将$\boldsymbol{F}^0_{rgb}$与边缘特征$\boldsymbol{F}^0_{edge}$经CMFM融合（第4.2节）后送入VSBB进行更深层的特征提取。注意：在整个特征提取过程中，边缘特征会逐尺度地与VSBB输出的特征$\boldsymbol{F}^n_{rgb}$进行融合。通过逐层融合机制，边缘信息得以在不同尺度上充分传递和增强，确保模型的编码器在不同尺度的特征层次上都能有效利用边界信息，从而提升模型对像素级语义的理解能力。编码过程如下：

$$\boldsymbol{F}^n_{fusion}=CMFM(\boldsymbol{F}^n_{rgb},\boldsymbol{F}^n_{edge}),\ n\leqslant 3 \tag{6}$$

$$\boldsymbol{F}^{n+1}_{rgb}=\begin{cases}Down\,Sampling(VSSB(\boldsymbol{F}^n_{fusion})), & n<3\\ VSSB(\boldsymbol{F}^n_{fusion}), & n=3\end{cases} \tag{7}$$

其中，使用选择性扫描2D模块（Selective Scanning 2D Module，SS2DM）[37]来实现VSBB。如图5所示，融合后的特征$\boldsymbol{F}^n_{\text{fusion}}$首先通过线性投影和深度可分离卷积处理以提取图像的初步特征$\boldsymbol{F}^n_{\text{init}}$。然后，使用具有残差连接的SS2DM来建模长距离空间信息。

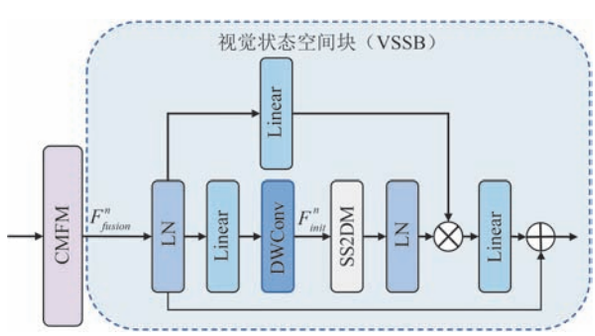


图5　视觉状态空间模块结构

由于Mamba是一个序列网络，无法直接处理二维图像数据，因此需要先将输入数据序列化，具体过程如图6所示。先从四个方向（包括：从左到右、从右到左、从上到下和从下到上）将输入的初步特征$\boldsymbol{F}^n_{init}$展平为四个序列。随后，利用四个SS2DM分别

处理每个方向的序列(以建模序列的远程依赖关系)。最后,将这四个处理过的序列重新对齐并合并,以有效整合来自不同视角的上下文信息。由于是按特定顺序展开序列,这使得每个特征在原始特征图中的空间位置得以保留,因此经过SS2DM增强后的序列能够准确还原为其对应的二维特征图,保证了四个扫描方向的增强特征图在二维空间中的严格对齐。当融合四个方向的增强特征图时,不会产生信息冲突或引入语义干扰。尽管多方向扫描可能在一定程度上引入信息冗余,但这种方式有助于在二维空间中建立更完整的全局感受野,从而提升模型的上下文建模能力与分割精度[37]。

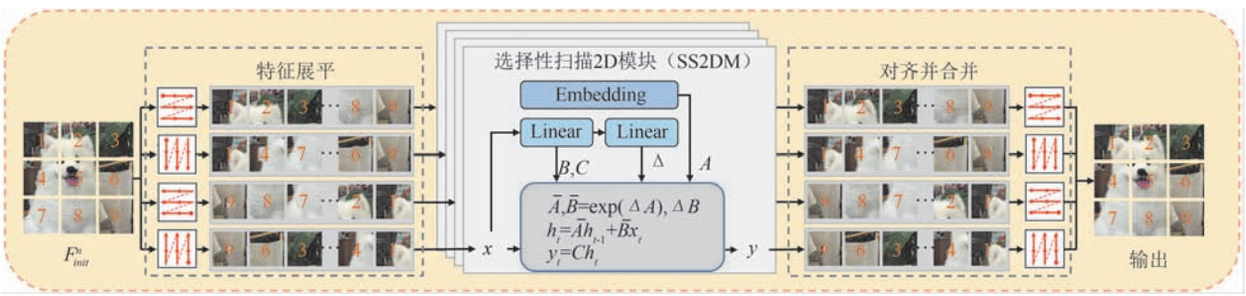


图6 择性扫描2D模块(SS2DM)

卷积解码器:在解码阶段,通过逐级上采样和CNN层将编码器中提取的深层次语义特征逐步恢复到原始的图像分辨率大小。每一级上采样(Up Sampling)由转置卷积实现。相较于线性插值,转置卷积通过可学习的权重参数实现上采样,具备更强的特征重建能力。在上采样后通过跳跃连接机制引入对应编码层的浅层特征。这些浅层特征与当前解码层上采样后的特征在通道维度上进行拼接,并通过卷积层(CNN)进行融合,以有效整合浅层细节与深层语义,从而增强模型对物体边界的感知能力和提高分割精度。在最后一级上采样完成后,解码器通过一个卷积层(CNN)对融合后的高维特征图进行通道压缩,并使用Softmax层输出每个像素在其九邻域内的归属概率,构建概率关联图$\boldsymbol{Q}\in R^{H\times W\times 9}$。根据$\boldsymbol{Q}$中的概率值,生成超像素分割结果。融合模块CMFM及边缘特征$\boldsymbol{F}^n_{edge}$的提取过程将在下文中详细讨论。

4.2 交叉曼巴融合模块

为了将边缘特征$\boldsymbol{F}^n_{edge}$与来自VSSB的多尺度特征$\boldsymbol{F}^n_{rgb}$进行融合,提升模型对复杂结构的语义理解能力,我们设计了一个基于选择性扫描机制和交叉注意力的交叉曼巴融合模块(Cross Mamba Fusion Module,CMFM),如图7所示。

由于多尺度特征$\boldsymbol{F}^n_{rgb}$和边缘特征$\boldsymbol{F}^n_{edge}$分别来自于不同的网络分支,其语义层次和特征分布可能存在明显差异,直接融合会导致信息错配。因此,CMFM首先使用联合选择性扫描模块(Joint Selective Scan Module,JSSM)对特征

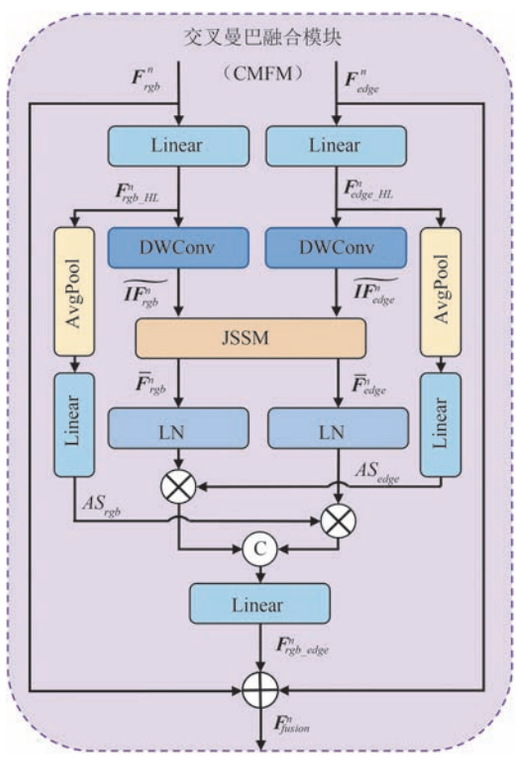


图7 交叉曼巴融合模块

$\boldsymbol{F}^n_{rgb}\in R^{H_n\times W_n\times C_n}$和$\boldsymbol{F}^n_{edge}\in R^{H_n\times W_n\times C_n}$进行联合建模,通过两种特征之间的相互补充和校准,增强各自的表达能力,从而得到增强后的特征$\bar{\boldsymbol{F}}^n_{rgb}$和$\bar{\boldsymbol{F}}^n_{edge}$。接着,采用交叉注意力机制对$\bar{\boldsymbol{F}}^n_{rgb}$和$\bar{\boldsymbol{F}}^n_{edge}$进行动态加权融合,使模型能够自动根据任务需求调整不同特征的权重,生成更具判别性的融合特征$\boldsymbol{F}^n_{rgb_edge}$,从而提高模型对复杂图像中不同物体的区分能力和边界的识别能力,增强超像素分割的准确性。

联合建模和加权融合的具体过程如下。首先对特征$\boldsymbol{F}^n_{rgb}$和$\boldsymbol{F}^n_{edge}$进行初步的特征提取：

$$\begin{cases}\boldsymbol{F}^n_{rgb\_HL}=Linear(\boldsymbol{F}^n_{rgb})\\ \boldsymbol{F}^n_{edge\_HL}=Linear(\boldsymbol{F}^n_{edge})\\ \widetilde{\boldsymbol{IF}^n_{rgb}}=DWConv(\boldsymbol{F}^n_{rgb\_HL})\\ \widetilde{\boldsymbol{IF}^n_{edge}}=DWConv(\boldsymbol{F}^n_{edge\_HL})\end{cases}\tag{8}$$

其中，$\boldsymbol{F}^n_{rgb_HL}$和$\boldsymbol{F}^n_{edge_HL}$是经线性层升维后的特征，可为后续的融合提供丰富的信息基础。随后，将经由深度卷积层初步提取的特征$\widetilde{\boldsymbol{IF}^n_{rgb}}$和$\widetilde{\boldsymbol{IF}^n_{edge}}$输入至联合选择性扫描模块（Joint Selective Scan Module，JSSM），其结构如图8所示。

在JSSM内，上述两个特征首先被按“从左到右”的顺序展平成为一维序列$\widetilde{\boldsymbol{F}^n_{rgb}}\in R^{H_k\times W_k\times C_k}$和$\widetilde{\boldsymbol{F}^n_{edge}}\in R^{H_k\times W_k\times C_k}$，并在序列长度维度上拼接这两个特征序列，形成联合特征序列$\boldsymbol{S}^n_{Cat}\in R^{(2\times H_k\times W_k)\times C_k}$。同时，生成其逆序列$\boldsymbol{S}^n_{Inv}\in R^{(2\times H_k\times W_k)\times C_k}$，使用两个SS2DM分别处理$\boldsymbol{S}^n_{Cat}$和$\boldsymbol{S}^n_{Inv}$，分别建模长程依赖关系，以实现全局范围的信息交互与增强，从而获得更具判别力的特征表示$\widehat{\boldsymbol{S}^n_{Cat}}$与$\widehat{\boldsymbol{S}^n_{Inv}}$（公式(9)）。使用公式(10)将$\widehat{\boldsymbol{S}^n_{Inv}}$翻转以和$\widehat{\boldsymbol{S}^n_{Cat}}$进行对齐，并和$\widehat{\boldsymbol{S}^n_{Cat}}$融合，得到两个方向融合的联合特征拼接序列$\boldsymbol{S}^n_{fusion}\in R^{(2\times H_k\times W_k)\times C_k}$。翻转处理使序列$\widehat{\boldsymbol{S}^n_{Inv}}$与$\widehat{\boldsymbol{S}^n_{Cat}}$对齐，可确保两者中对应位置的特征均来源于图像的同一空间位置。这种空间位置的一致性有效避免了后续特征融合时的信息冲突。虽然融合可能带来一定的信息冗余，但这种来自不同方向的上下文信息，有助于丰富特征表示，从而增强特征的表达能力[42]。最后，将$\boldsymbol{S}^n_{fusion}\in R^{(2\times H_k\times W_k)\times C_k}$拆解为两个子序列，并还原以得到增强后的特征表示$\bar{\boldsymbol{F}}^n_{rgb}\in R^{H_k\times W_k\times C_k}$和$\bar{\boldsymbol{F}}^n_{edge}\in R^{H_k\times W_k\times C_k}$。

$$\widehat{\boldsymbol{S}^n_{Cat}},\ \widehat{\boldsymbol{S}^n_{Inv}}=SS2DM\left(\boldsymbol{S}^n_{Cat}\right),\ SS2DM\left(\boldsymbol{S}^n_{Inv}\right)\tag{9}$$

$$\boldsymbol{S}^n_{fusion}=\widehat{\boldsymbol{S}^n_{Cat}}+Inverse(\widehat{\boldsymbol{S}^n_{Inv}})\tag{10}$$

$$\bar{\boldsymbol{F}}^n_{rgb},\bar{\boldsymbol{F}}^n_{edge}=Separate(\boldsymbol{S}^n_{fusion})\tag{11}$$

其中，$Inverse(\cdot)$表示翻转，$Separate(\cdot)$表示将序列进行拆解。之后，采用交叉注意力机制加权融合特征$\bar{\boldsymbol{F}}^n_{rgb}$和$\bar{\boldsymbol{F}}^n_{edge}$，目的是使得模型能够灵活地调整不同特征的权重，生成更具区分度的融合特征$\boldsymbol{F}^n_{rgb_edge}$。其中，为了提高注意力计算的精度，使用高维空间中的特征$\boldsymbol{F}^n_{rgb_HL}$和$\boldsymbol{F}^n_{edge_HL}$来计算注意力分数$AS_{rgb}$和$AS_{edge}$。具体而言，$AS_{edge}$由$F^n_{rgb_HL}$计算得到，$AS_{rgb}$由$F^n_{edge_HL}$计算得到。这种交叉计算方式强制模型显式地建模两个特征之间的依赖关系，避免特征融合陷入局部最优。随后，利用注意力分数对特征$\bar{\boldsymbol{F}}^n_{rgb}$和$\bar{\boldsymbol{F}}^n_{edge}$加权，并在通道维度上串联后通过线性层完成融合得到融合特征$\boldsymbol{F}^n_{rgb_edge}$。过程如下：

$$AS_{rgb}=FC(AvgPool(\boldsymbol{F}^n_{edge\_HL}))\tag{12}$$

$$AS_{edge}=FC(AvgPool(\boldsymbol{F}^n_{rgb\_HL}))\tag{13}$$

$$\boldsymbol{F}^n_{r\_e}=Conat(AS_{edge}*\bar{\boldsymbol{F}}^n_{rgb},AS_{rgb}*\bar{\boldsymbol{F}}^n_{edge})\tag{14}$$

$$\boldsymbol{F}^n_{rgb\_edge}=Linear(\boldsymbol{F}^n_{r\_e})\tag{15}$$

其中，采用$AvgPool(\cdot)$操作以压缩冗余空间信息并保留通道级语义分布；$FC(\cdot)$为全连接操作，用于生成注意力分数。最终，融合特征$\boldsymbol{F}^n_{rgb_edge}$通过残差连接进行调整，以确保信息流的稳定性，并输出优化后的最终特征表示$\boldsymbol{F}^n_{fusion}$。

4.3　即插即用的轻量化边缘强化框架

基于深度学习的超像素分割方法[17-19]往往忽视了边界信息在提升语义理解中的重要作用，导致这些方法在处理复杂边界或低对比度图像时难以准确区分物体边界。为了解决这一问题，我们提出了一种即插即用的轻量化边缘强化框架。该框架通过提供额外的边缘信息，增强了模型在推理阶段以及后续分割结果优化过程中的边界感知能力，从而有效提升超像素分割的性能。

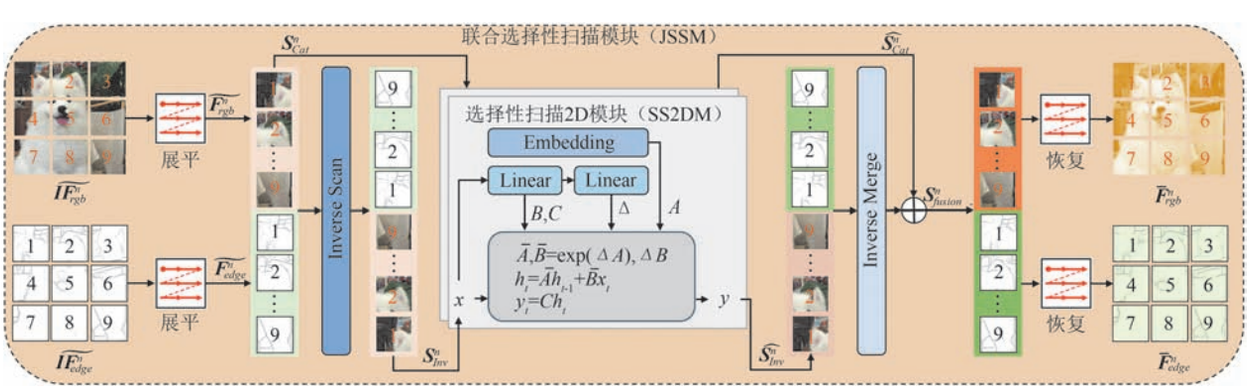


图8　联合选择性扫描模块

边缘强化框架：考虑到超像素分割结果需要快速生成，我们采用了深度可分离卷积（DWConv）来构建此框架，以加速推理过程。整个边缘强化框架包含四个CRBP模块，以从粗到细地提取边缘特征（如图3所示）。每个CRBP模块包含四个卷积残差块（CRB）和一个最大池化层。在CRB结构中，依次是深度可分离卷积（DWConv）、ReLU激活层和点式卷积层（1x1 Conv），分别用于提取边界特征、引入非线性变换和融合通道信息。这种结构不仅通过残差连接增强了特征传递，还有效缓解了训练过程中的梯度消失问题。边缘特征$\boldsymbol{F}^n_{edge}$提取过程如下：

$$\boldsymbol{F}^n_{edge}=\begin{cases}CRBP(InitConv(\boldsymbol{I}_{rgb})), & n=0\\ CRBP(\boldsymbol{F}^n_{edge}), & n>0\end{cases} \tag{16}$$

为了使边缘强化框架能够学习到丰富的层次化边缘特征[43]，从每个CRBP的输出中生成边缘图，并通过与真实边缘图计算损失来实现深度监督。具体来说，先通过卷积层将边缘特征$\boldsymbol{F}^n_{edge}$压缩至单通道映射$edge^n$，并将其插值至原始图像大小。随后，使用Sigmoid函数生成最终的边缘映射$\boldsymbol{Edge}^n\in R^{H\times W\times 1}$。

$$edge^n=\text{Interpolate}(CNN(\boldsymbol{F}^n_{edge}),H,W) \tag{17}$$

$$\boldsymbol{Edge}^n=\text{Sigmoid}(edge^n) \tag{18}$$

接下来，采用交叉熵损失函数和$\boldsymbol{Edge}^n$来计算边缘损失。第n个边缘图$\boldsymbol{Edge}^n$中的第i个像素p^n_i的损失l^n_i的计算公式如下：

$$l^n_i=\begin{cases}\alpha\cdot\log(1-p^n_i), & \text{if } y_i=0\\ 0, & \text{if } 0<y_i<\eta\\ \beta\cdot\log(p^n_i), & \text{otherwise}\end{cases} \tag{19}$$

其中，y_i是真实边缘概率，η是预定义的阈值。如果少于η个注释者将某个像素标记为正，则在计算损失时将这个像素丢弃并且不将其视为样本，以避免混淆。β是负类像素样本的百分比，$\alpha=\lambda\cdot(1-\beta)$。边缘强化框架的总损失为

$$L_{edge}=\sum_{n,i}l^n_i \tag{20}$$

此外，为了生成一个更加全面、精确的边缘结果用于引导分割结果的优化，将四个不同颗粒度的单通道特征图$edge^n$进行拼接，并通过卷积和Sigmoid函数处理，以融合多尺度边缘信息，最终生成边缘结果$\boldsymbol{Edge}^F\in R^{H\times W\times 1}$。随后，采用式子（19）对$\boldsymbol{Edge}^F$进行监督，以提升$\boldsymbol{Edge}^F$的准确性。过程如下：

$$\boldsymbol{Edge}^C=\text{Concat}(edge^1,edge^2,edge^3,edge^4) \tag{21}$$

$$\boldsymbol{Edge}^F=\text{Sigmoid}(Conv(\boldsymbol{Edge}^C)) \tag{22}$$

推理阶段：在推理过程中，边缘特征$\boldsymbol{F}^n_{edge}$通过特征融合模块CMFM输入到分割模型中，以强化模型对边缘区域的识别和划分能力（第4.1节）。

分割结果优化：为了优化被误分类的边界像素，我们利用边缘图$\boldsymbol{Edge}^F$对分割结果进行调整。首先，移除分割结果中与边缘图$\boldsymbol{Edge}^F$中的边缘像素对应的像素的超像素标签，将这些像素重新标记为较小的超像素。随后，通过合并操作将这些较小的超像素重新分配给属性最相近的邻接超像素，从而提升边界的黏附效果。超像素间的相似性由颜色和空间的欧式距离综合计算。距离越小，表明两个超像素越相似，更有可能被归并到一起。合并操作从像素数量最少的超像素开始，逐步进行，直到图像中的超像素总数达到预定的目标数量为止。引入边缘图$\boldsymbol{Edge}^F$后提升了边界的黏附效果，如图9所示。

（a）优化前的结果 （b）引入边界图像后的优化结果

图9 优化后的超像素分割结果

此外，现有深度学习方法通过调整输入图像的尺寸来控制超像素数量。但由于图像分辨率必须以整数表示，缩放后的尺寸无法精确匹配预定的超像素数量。相比之下，我们的方法在重标记后通常会生成超过预定数量的超像素，通过合并操作可以精确地生成所需的超像素总数。

4.4 损失函数

现有方法通常采用固定权重对表征一致性损失和空间紧致性损失进行加权，在不同类型的图像和复杂场景中，难以自适应地权衡边界黏附性与超像素规则性，导致二者相互制约。为了解决这一问题，我们提出了一种基于边界概率的自适应超像素分割损失函数，该方法能够根据像素的边界置信度自适应地分配权重，在保持超像素边界高黏附性的同时，提升了超像素的规则性。

由于超像素分割任务缺乏可用的基准真值，因此网络的训练需要通过间接方式进行。通过概率关联图$\boldsymbol{Q}$，利用像素的原始属性（如语义标签、位置向量等）来计算超像素的中心属性。接着，根据超像素的中心属性重建像素的属性。最后通过最小化像素

的原始属性与重建属性之间的差异来优化超像素分割结果。

设$\boldsymbol{f}(p)$为像素p的表征属性(采用语义标签的N-维单热编码向量表示，N为类别数)，$\boldsymbol{c}(p)$为像素p的位置属性$[x,y]^T$，$\boldsymbol{f}'(p)$和$\boldsymbol{c}'(p)$分别为像素p重建的表征属性和位置属性。损失函数定义如下：

$$L(Q)=\sum_p w(p)d_f(\boldsymbol{f}(p),\boldsymbol{f}'(p))+(2-w(p))d_c(\boldsymbol{c}(p),\boldsymbol{c}'(p)) \tag{23}$$

其中，$d_f(.,.)$和$d_c(.,.)$分别为基于交叉熵的表征一致性损失和采用L2范数的空间紧致性损失。权重函数$w(p)$依据像素p位于边界的概率动态调节两者的比重。当像素p处于边界区域时，$w(p)$值较高，此时增加$d_f(.,.)$的权重可使模型在分割过程中更加侧重于表征一致性损失，从而更准确地划分边界像素。相反，若$w(p)$较小，则表示像素p可能位于非边界区域，增加$d_c(.,.)$的权重则有助于增强非边界区域超像素的规则性。

权重函数$w(p)$的定义如下：首先，在CIELab颜色空间的L通道上计算边界概率[24-25]。为了在图像的弱边缘区域获取更可靠的边界值，将Ground Truth边缘数据与原始图像的L通道相结合，以此增强L通道中的边缘信息。随后，在边缘增强后的L通道上计算边界概率。边界概率通过像素的邻域(如图4(b)所示)梯度值来定义。

$$t=\sqrt{t_h^2+t_v^2} \tag{24}$$

$$w(p)=\begin{cases}0, & t<T\\ t, & \text{otherwise}\end{cases} \tag{25}$$

其中，t是像素p的邻域梯度值。$t_h=p_4-p_5$和$t_v=p_2-p_7$分别表示像素p在水平和垂直方向的梯度变化。为保留物体轮廓并抑制假边缘，设定阈值$T=5$。若$t<T$，则认为像素位于非边界区域，其边界概率设0。随后，对t进行归一化处理，将其转换为概率值。为避免$w(p)=0$时的梯度消失问题，对归一化后的值加1，将其范围调整至(1，2)。最后，为使边界概率与边缘Ground Truth对应，仅保留边缘Ground Truth中非零值对应的边界概率。

下面讨论如何计算重建属性$\boldsymbol{f}'(p)$和$\boldsymbol{c}'(p)$。给定预测的概率关联图$\boldsymbol{Q}\in R^{H\times W\times 9}$，可以计算任意超像素$s$的中心$\boldsymbol{c}_s=(\boldsymbol{p}_s,\boldsymbol{l}_s)$，其中$\boldsymbol{p}_s$为表征向量，$\boldsymbol{l}_s$为位置向量。

$$\boldsymbol{p}_s=\frac{\sum_{p:s\in N_p}\boldsymbol{f}(p)\cdot q_s(p)}{\sum_{p:s\in N_p}q_s(p)} \tag{26}$$

$$\boldsymbol{l}_s=\frac{\sum_{p:s\in N_p}\boldsymbol{c}(p)\cdot q_s(p)}{\sum_{p:s\in N_p}q_s(p)} \tag{27}$$

其中，N_p是p周围超像素的集合，而$q_s(p)$表示网络预测像素p与超像素s相关联的概率(来自Q)。根据上述公式(26)(27)，每个超像素的中心由所有可能被分配给该超像素的像素加权确定。因此，任意像素p的重建表征属性$\boldsymbol{f}'(p)$和位置属性$\boldsymbol{c}'(p)$可通过其所属超像素的中心属性进行表示。重建表征属性$\boldsymbol{f}'(p)$可以表示为

$$\boldsymbol{f}'(p)=\sum_{s\in N_p}\boldsymbol{p}_s\cdot q_s(p) \tag{28}$$

对于重建的位置属性$\boldsymbol{c}'(p)$，现有方法[17-19]通常通过超像素s的中心来计算像素p的重建位置属性(公式(29))。

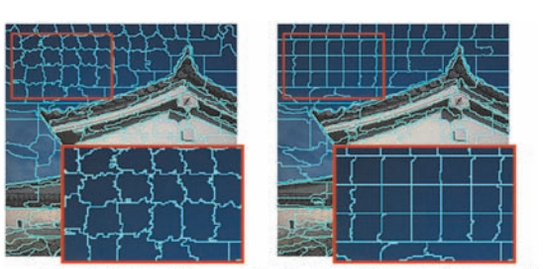
$$\boldsymbol{c}'(p)=\sum_{s\in N_p}\boldsymbol{l}_s\cdot q_s(p) \tag{29}$$

接下来，最小化像素的重建位置与其原始位置之间的差异，来保证超像素的规则性。然而，这种基于像素间误差最小化的策略忽略了超像素的整体形态，可能导致超像素内部像素分布较为离散。相比之下，我们直接使用与像素p空间距离最近的超像素s的位置属性来表示像素p重建的位置属性$\boldsymbol{c}'(p)$(公式(30))。

$$\boldsymbol{c}'(p)=\boldsymbol{l}_s,\left(s=\arg\min_{s\in N_p}d(\boldsymbol{l}_s,\boldsymbol{c}(p))\right) \tag{30}$$

然后，通过最小化像素与超像素中心之间的空间距离来优化规则性，可以使像素在超像素中心周围分布更均匀。如图10所示，生成的超像素形状更加规则。最后，结合边缘强化框架的损失公式(20)，超像素分割模型得的总损失函数为L。

$$L=L(\boldsymbol{Q})+L_{edge}. \tag{31}$$



(a) 像素间最小化　　(b) 像素与超像素中心间的最小化

图10　不同最小化方法的分割结果

5 实 验

在本节中，我们将详细介绍所使用的数据集、模型实施细节和评估指标等内容。并通过与现有先进方法的对比，展示所提模型在不同场景中的优越性。

5.1 数据集

我们在四个常用数据集上评估了所提出的方法，分别为BSD500[44]、PASCAL-S[45]、FASH[46]和DRIVE[47]。其中，BSD500包含200张训练图像、100张验证图像和200张测试图像。PASCAL-S数据集包含850张来自不同类别与尺寸的自然图像，包含覆盖遮挡、多物体和复杂背景等典型场景。该数据集提供两类标注：(1)图像分割标注，用于评估超像素方法的分割性能；(2)二值化的显著性区域标注，广泛应用于显著性检测任务。该数据集中的图像结构复杂、边界模糊度高，对超像素分割方法的边缘分割能力、区域一致性保持能力以及在下游显著性检测任务中的适应性提出了极大的挑战。FASH数据集包含685张时尚服饰图片；DRIVE数据集包含40张尺寸为565×584的视网膜医学图像，提供了清晰的血管结构标注，该数据集的图像中常表现出结构细长，边界模糊以及类间对比度低等医学图像的典型特征，适用于检验模型在存在较大领域偏移下的泛化能力。超像素方法需具备对弱边缘区域更准确的语义理解能力以及强大的上下文感知能力，才能有效分割细微且连贯的血管区域。这对分割模型在医学图像场景下的边缘保持能力及泛化性能提出了更高要求。每个数据集都提供了真实标签，支持模型训练和评估。

5.2 实施细节

EE-SSM模型基于PyTorch实现，并在BSD500数据集上进行训练。为确保公平比较，我们遵循先前研究[16,18-19]的设置，将BSD500训练集中的每个标注视为独立样本，最终获得1087个训练样本、546个验证样本和1063个测试样本。在训练过程中，采用数据增强策略，包括随机调整图像尺寸、水平/垂直翻转和颜色抖动，同时将图像裁剪为256×256作为输入，batch size设为24，总共训练240个epoch。优化策略采用AdamW优化器，初始学习率设为$6e-4$，权重衰减系数为0.01。此外，利用VMamba[37]在ImageNet-1K[48]上预训练的SigmaTiny模型作为编码器的初始化权重，以提升训练稳定性和模型性能。所有方法的推理均在相同的计算环境下进行，具体为：NVIDIA H800 GPU、Intel(R) Xeon(R) Platinum 8468V CPU、128 GB内存。模型训练也在同一环境中完成。

5.3 评估指标

采用了标准的超像素评估指标，以全面衡量超像素分割算法的性能。这些指标包括边界召回(BR)、可实现分割精度(ASA)、欠分割错误(UE)以及紧凑性(CO)。给定一个图像的标准分割结果，将图像分为M个互不重叠的部分，将其表示为$g_i(i=1,2,\cdots,M)$。$s_l(l=1,2,\cdots,K)$表示第s_l个超像素。图中的像素被定义为$p_n(n=1,2,\cdots,N)$，其中N是像素点的总数。

(1)边界召回(Boundary recall，BR)：BR衡量了超像素分割结果对真实边界的贴合程度。BR计算的是人工标注的边界点中，有多少比例能够落在至少两个相邻超像素的边界上。较高的BR值表明超像素的边界可以更好地对齐于真实边界。

(2)可实现分割精度(Achievable Segmentation Accuracy，ASA)：ASA用于评估图像中的物体是否被正确分割。通过用拥有最大重叠区域的标准分割为每个超像素贴上标签来计算最高的识别率。ASA值越大，表示图像中正确分割的物体数量越多。其定义如下：

$$ASA=\frac{\sum_l \arg_i \max\left|s_l\cap g_i\right|}{\sum_i\left|g_i\right|} \tag{32}$$

(3)欠分割错误(Under-segmentation Error，UE)：UE用于衡量算法在分割图像时产生的错误，具体通过计算从真实轮廓泄漏到超像素中的像素百分比来评估。如果一个超像素与多个真实轮廓有有效重叠，UE值会增加。UE的计算公式如下：

$$UE=\frac{1}{N}\left[\sum_{i=1}^{M}\left(\sum_{\{s_l\mid\left|s_l\cap g_i\right|>B\}} area(s_l)\right)-N\right] \tag{33}$$

其中，$area(s_l)$是超像素s_l的面积，B是最小数量的重叠区域面积。B被设置为$area(s_l)$的5%。越小的UE表明图像中更多的物体被识别了出来。

(4)紧凑性(Compactness，CO)：引入CO[49]来评估超像素的规则性。CO将每个超像素面积A_S与周长相同的圆的面积A_C进行比较。CO值越大意味着超像素的规则性越好，如下定义：

$$CO=\sum_{S\in Q} Q_S\frac{|S|}{|I|},\ Q_S=\frac{A_S}{A_C}=\frac{4\pi A_S}{L_S^2} \tag{34}$$

其中，S是图像I的超像素集合，$|S|$是集合的大小，

L_S为超像素的周长。

5.4　与最先进的方法进行对比

我们与当前最先进的11种超像素算法进行了全面比较，包括SLIC[50]，WSBB[51]（2023年），CRTrees[52]（2022年），VSS[53]（2023年），LNSnet[54]（2021年），FGSLT[34]（2022年），AINet[17]（2021年），ESNet[19]（2023年），CDS[18]（2024年），SSFCN[16]（2020年）和SSN[15]（2018年）。其中，WSBB、CRTrees和VSS为最先进的传统方法，SSFCN、LNSnet、FGSLT、AINet、SSN、ESNet和CDS为深度学习方法，且ESNet和CDS代表当前最先进的深度学习方法。所有算法的实验结果基于公开源代码，未修改作者提供的代码和预训练模型。在评估深度学习方法时，我们遵循既有的研究标准[18-20]，使用BSD500训练集进行训练并在多个独立测试集上进行评测（不进行微调）。

5.4.1　分割精度

如图11、图12和图13所示，我们的方法在分割准确性方面超越了目前的传统超像素算法和深度学习方法。在超像素数量介于100至1000之间时，我们的模型在BR和ASA指标上均取得最佳表现，表明其能够更紧密地贴合边界。此外，当超像素数量处于100至700之间时，我们的方法在UE指标上的表现也更优。这首先得益于长程依赖建模，增强了模型对全局结构的理解能力。此外，在编码阶段显式融合边界信息，使模型能更好地处理复杂边界和低对比度区域。最后，基于边界概率的自适应损失函数使模型在边界区域更注重语义一致性。

5.4.2　紧凑性

紧凑性得分衡量了超像素的形状规则性。现有的基于深度学习的方法通常采用固定权重对损失项进行加权，难以在提高CO的同时兼顾边界贴合性，尤其是在非边界区域维持超像素的规则性，并在边界附近增强黏附性。图11、图12、图13和图14显示，我们的方法在确保超像素分割质量的同时，实现了更优的紧凑性。在分割准确性达到最优的情况下，其规则性仍优于SLIC、CRTrees、VSS和LNSnet等算法。这主要归功于基于边界概率的自适应损失函数：在边界区域，模型更关注表征一致性，以提升边界对齐效果；而在非边界区域，则优先优化空间紧致性，以维持超像素的形状规则。

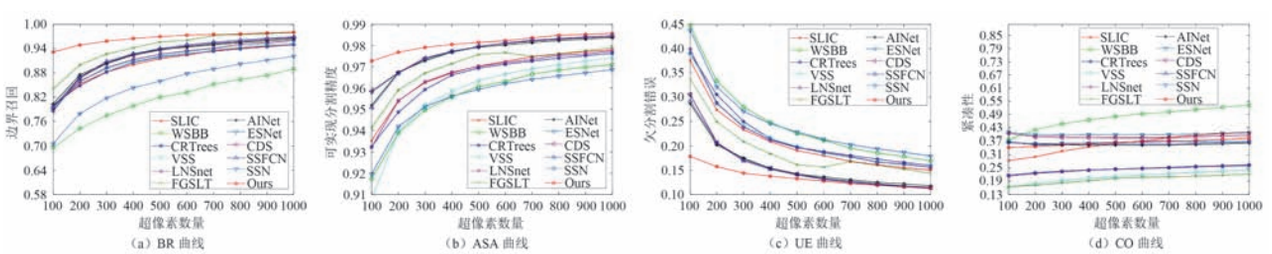


图11　不同方法在PASCAL-S数据集上的数值比较结果

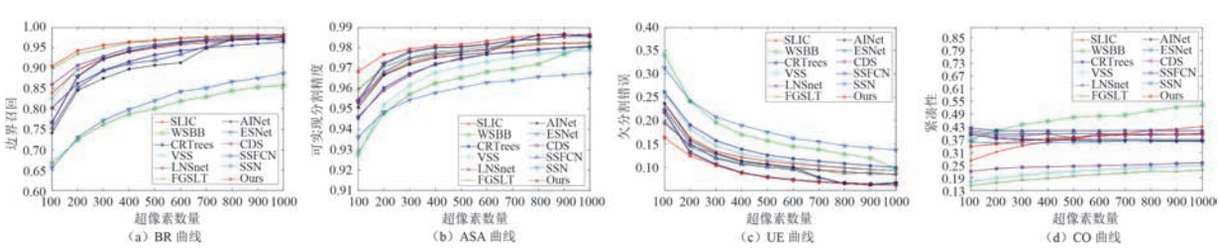


图12　不同方法在FASH数据集上的数值比较结果

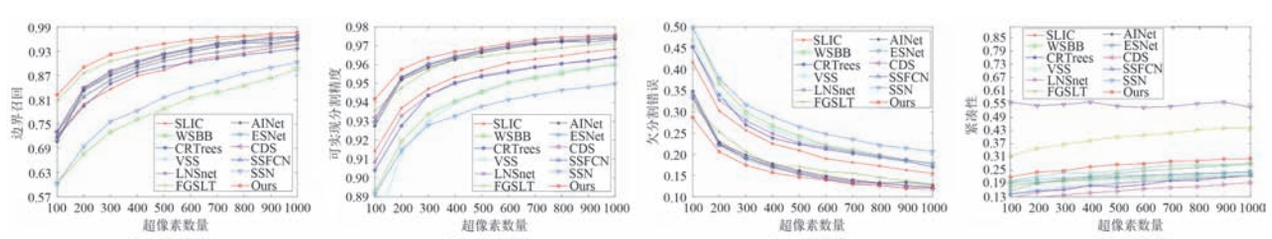


图13　不同方法在BSD500数据集上的数值比较结果

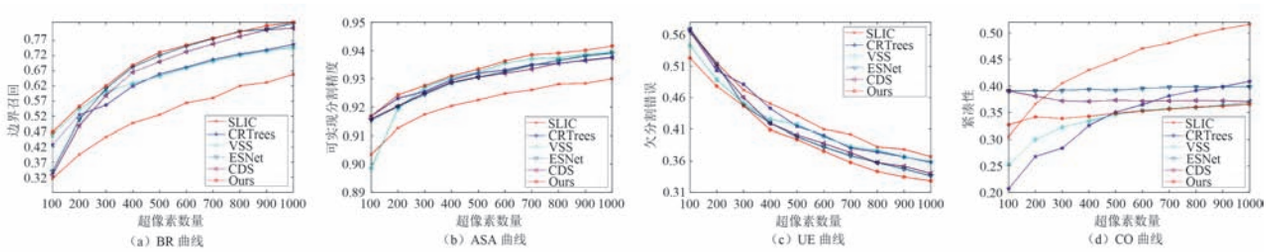


图14 不同方法在DRIVE数据集上的数值比较结果

5.4.3 计算效率

表1，表2和表3对比了不同超像素分割方法的运行时间。实验结果表明，我们的方法在速度上优于WSBB、VSS、LNSNet、FGSLT和SSN，并且与最新的ESNet和CDS方法相比，具有较强竞争力。这表明，我们的方法在确保分割质量的同时，依然具备较高的计算效率。这一优势主要源于两方面的设计：首先，采用线性时间复杂度的SSM进行建模；其次，引入了深度可分离卷积构建边缘强化框架，提升了模型的推理效率。

表1 当超像素数量为300时，不同方法在PASCAL-S数据集上的数值结果

	SLIC	WSBB	CRTrees	VSS	LNSnet	FGSLT	AINet	ESNet	CDS	SSFCN	SSN	Ours
BR	0.8820	0.7739	0.8818	0.8893	0.8928	0.9267	0.9040	0.9025	0.9046	0.9081	0.8173	**0.9584**
ASA	0.9625	0.9495	0.9592	0.9519	0.9629	0.9670	0.9728	0.9733	0.9742	0.9738	0.9511	**0.9791**
UE	0.2320	0.2793	0.2493	0.2400	0.2373	0.2089	0.1751	0.1735	0.1714	0.1692	0.2700	**0.1439**
CO	0.3261	0.4477	0.2360	0.1923	0.2332	0.1818	0.3554	**0.3982**	0.3888	0.3587	0.3618	0.3539
Time (s)	0.0715	52.315	**0.0089**	2.5483	0.6215	0.5672	0.0510	0.0750	0.0815	0.0680	0.3294	0.0906

注：粗体表示最优值。

表2 当超像素数量为300时，不同方法在BSD500数据集上的数值结果

	SLIC	WSBB	CRTrees	VSS	LNSnet	FGSLT	AINet	ESNet	CDS	SSFCN	SSN	Ours
BR	0.8360	0.7302	0.8496	0.8613	0.8583	0.9153	0.8700	0.8800	0.8802	0.8747	0.7561	**0.9215**
ASA	0.9471	0.9338	0.9437	0.9298	0.9438	0.9570	0.9584	0.9605	0.9598	0.9598	0.9279	**0.9635**
UE	0.2566	0.2998	0.2691	0.2900	0.2798	0.2062	0.1983	0.1877	0.1926	0.1929	0.3162	**0.1748**
CO	0.2980	0.3677	0.2174	0.1701	0.2017	0.1861	0.3556	**0.3889**	0.3596	0.3793	0.3624	0.3300
Time (s)	0.0621	30.215	**0.0078**	1.4121	0.5491	0.5396	0.0439	0.0675	0.0735	0.0440	0.2884	0.0824

注：粗体表示最优值。

表3 当超像素数量为300时，不同方法在Fash数据集上的数值结果

	SLIC	WSBB	CRTrees	VSS	LNSnet	FGSLT	AINet	ESNet	CDS	SSFCN	SSN	Ours
BR	0.9234	0.7608	0.8943	0.9284	0.9287	0.9484	0.8744	0.9213	0.9235	0.8933	0.7713	**0.9564**
ASA	0.9728	0.9570	0.9673	0.9616	0.9663	0.9751	0.9718	0.9777	0.9782	0.9747	0.9543	**0.9794**
UE	0.1344	0.1951	0.1577	0.1464	0.1287	0.1221	0.1305	0.1070	0.1061	0.1189	0.2071	**0.1043**
CO	0.3419	0.4400	0.4101	0.2017	0.2406	0.1787	0.3710	**0.3963**	0.3863	0.3731	0.3544	0.3639
Time(s)	0.0975	41.2560	**0.0120**	1.8780	0.9511	0.7662	0.0260	0.0855	0.0765	0.0510	0.4043	0.0920

注：粗体表示最优值。

5.4.4 泛化能力

超像素分割任务对算法的泛化性能提出了较高要求，尤其是在不同数据集和应用场景下的适应能力。为验证所提方法在医学图像领域中的适应能力，我们选用视网膜血管分割领域的标准医学数据集DRIVE进行实验对比。具体而言，我们在该数据集上与经典的SLIC方法以及近年来代表性的超像素分割方法CRTrees、VSS、ESNet和CDS进行了系统对比，涵盖了传统方法与深度学习方法两大类别，以全面评估所提方法在医学图像场景中的适应性。

如图14所示和表4所示，在医学图像数据集DRIVE上，我们的方法在BR、ASA和UE等关键指标上均优于现有最先进方法，展现出更强的鲁棒性

和泛化能力。这得益于长程依赖建模，使得模型能够更好地理解图像中细长复杂的血管区域，提高了其对血管的分割精度。其次，显式融合边界信息提升了模型在低对比度区域中的表现，能够更准确地分割模糊的血管结构。最后，基于边界概率的自适应损失函数能够根据医学图像中边缘的复杂性，自动优化边界区域的语义一致性，进一步提高了模型的泛化性能和分割精度。

5.4.5　可视化

图15展示了不同方法在超像素数量为300时的分割结果。相较于当前主流的CNN-based方法（如FGSLT、AINet、ESNet和CDS），我们的方法通过SSM建模长程依赖关系，在语义一致性方面表现更优，能够更精准地区分不同类别区域，如猫的眼睛、飞机的机翼等。此外，通过在编码过程中显式引入边缘信息，增强了模型对边界的感知能力，使得分割结果在边界复杂的区域（如树枝、地砖接缝等）更加准确。最后，基于边界概率的自适应损失函数进一步优化了超像素的形状规则性与边界贴合性。在非边界区域（如天空等大面积平坦区域），超像素更加均匀规则，而在边界区域则能更紧密贴合真实轮廓，实现了超像素规则性与分割精度的最佳平衡。

表4　当超像素数量为300时，不同方法在DRIVE数据集上的数值结果

	SLIC	CRTrees	VSS	ESNet	CDS	Ours
BR	0.4530	0.5569	0.6008	0.6111	0.5886	**0.6200**
ASA	0.9174	0.9255	0.9270	0.9245	0.9247	**0.9273**
UE	0.4722	0.4813	0.4509	0.4478	0.4606	**0.4470**
CO	**0.4047**	0.2836	0.3217	0.3912	0.3715	0.3382
Time (s)	0.1211	0.0132	1.9422	0.0935	**0.0870**	0.1152

注：粗体表示最优值。

图15　当超像素数量为300时，不同超像素分割算法的可视化对比

5.5　显著性检测

显著性检测任务通过模拟人类视觉的注意机制，自动识别图像中最具视觉吸引力或关注度的区域。显著性检测可作为前置模块，为图像分类、目标检测等下游任务提供显著性引导，引导模型关注关键区域，从而提升感知效率与准确性。为了验证所提方法在实际应用中的有效性，我们将超像素分割作为预处理步骤，并在显著性检测任务中进行测试。具体而言，用WSBB、CDS、VSS、ESNet以及我们的方法替换原始应用中的超像素分割方法（SLIC），在PASCAL-S数据集上开展了对比实验。该数据集提供了高质量的显著性区域标注，广泛用

于显著性检测研究。在PASCAL-S数据集上，我们的模型能更稳定地突出显著区域，并更好地保留目标物体边界，如图16所示。为了客观评估不同超像素分割方法对显著性检测任务的影响，采用精度(Precision)和召回率(Recall)作为衡量指标。精度是被正确标记为显著的像素占提取区域总像素的比例，召回率则衡量被正确检测的显著像素占所有真实显著像素的比例。我们将各超像素方法生成的显著图进行归一化处理，并通过设置的一系列固定阈值对其进行二值化，生成预测的二值图。然后，将预测的二值图与真实显著性掩码进行比较，计算不同阈值下的精度与召回率(公式(35))，最终绘制 PR 曲线(图17)以评估模型的显著区域检测性能。

$$Precision=\frac{|S\cap G|}{|S|},\ Recall=\frac{|S\cap G|}{|G|} \tag{35}$$

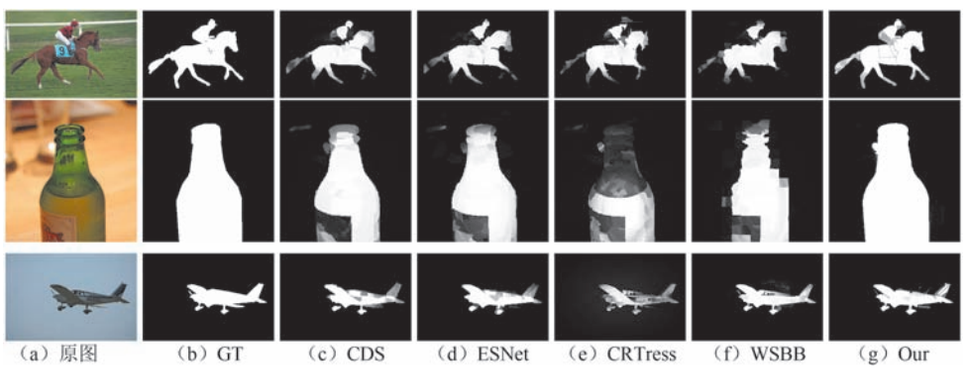


图16 使用不同的超像素分割算法进行预处理，并在PASCAL-S数据集上进行显著性检测的示例

图17显示，当召回率在0.3至0.7之间时，我们的精度优于其他算法。需要说明的是，PASCAL-S数据集同时具有高质量的分割与显著性标注，可用于超像素分割评估，也广泛应用于显著性检测任务。因此，适用于验证我们的EE-SSM模型在复杂场景下对显著区域的识别能力。而BSD、FASH和DRIVE数据集均不包含显著性区域的标注信息，因此无法在这些数据集上使用图17中的指标(如精度和召回率)进行评估。

5.6 消融实验

为验证我们方法的有效性，使用纯CCN编码器和解码器作为基线进行消融实验。首先，使用SSM构建编码器；然后，使用边缘强化框架来引入显式的边缘信息，这里边缘特征直接与编码器提取的特征进行相加。接下来，采用CMFM优化边缘特征与编码器特征的融合过程。然后，在此基础上，使用提出的损失函数对网络进行优化。最后，利用边缘结果 $Edge^F$ 进一步优化超像素分割的效果，实验结果见表5。从表5中可以看出，随着不同模块和方法的使用，BR、ASA和UE等指标显著提升，表明所提方法有效提高了超像素分割性能。CO指标的先降后升则反映了超像素准确性和规则性之间的矛盾：准确性提升时，规则性有所下降。然而，在使用我们提出的损失函数后，规则性和准确性均得到了提升，表明我们的方法在两者平衡方面具有良好的效果。最后，规则性的下降主要是由于在优化分割结果时，边界区域的超像素紧密贴合物体边界，从而降低了其规则性。

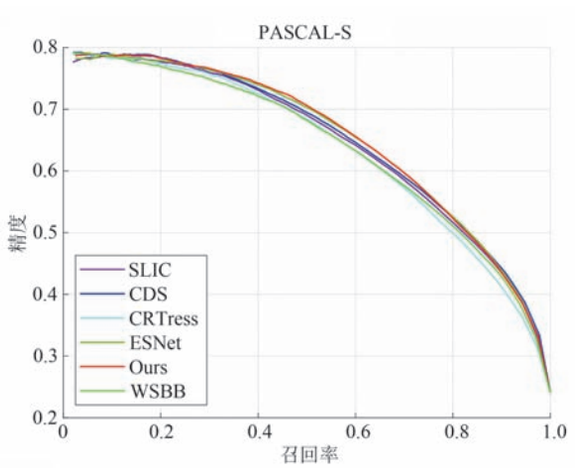


图17 PR曲线

表5 消融实验结果(BSD500数据集，超像素个数300)

	BR	ASA	UE	CO
CNN,CNN	0.7961	0.9424	0.2468	**0.4050**
Mamba ,CNN	0.8414	0.9503	0.2267	0.3646
Mamba ,CNN,F^n_{edge}	0.8723	0.9545	0.2136	0.3548
Mamba ,CNN,F^n_{edge},CMFM	0.8942	0.9566	0.1992	0.3363
Mamba ,CNN,F^n_{edge},CMFM,(22)	0.9124	0.9622	0.1833	0.3464
完整方法	**0.9215**	**0.9635**	**0.1748**	0.3300

注：粗体表示最优值。

此外，为了验证我们提出的即插即用边缘强化框架对当前超像素分割方法的贡献，我们分别对传统方法和深度学习方法进行了不同的优化处理：对于传统方法，直接使用边缘结果对其输出进行优化；而对于深度学习方法，则将边缘强化框架集成到网络中。实验结果如表6所示，证明了我们的框架能为传统和深度学习方法中带来明显的性能提升。

表6　即插即用边缘强化框架的有效性（BSD500数据集，超像素个数300）

	BR	ASA	UE	CO
VSS	0.8613	0.9298	0.2900	**0.1701**
VSS+$Edge^F$	**0.9117**	**0.9407**	**0.2366**	0.1642
WSBB	0.7302	0.9338	0.2998	**0.3677**
WSBB+$Edge^F$	**0.9024**	**0.9445**	**0.2166**	0.3118
ESNet	0.8800	0.9605	0.1877	**0.3889**
ESNet+$Edge^F$+F^n_{Edge}	**0.9110**	0.9616	**0.1804**	0.3363
AINet	0.8700	0.9584	0.1983	**0.3556**
AINet+$Edge^F$+F^n_{Edge}	**0.9119**	**0.9605**	**0.1835**	0.3106

注：粗体表示最优值。

5.7　阈值T的选择

在边界概率的计算过程中，引入阈值T可以有效过滤伪边缘，从而提高边界计算的准确性[25]。即，当像素的边界概率$t<T$时，将其视为非边界像素，其对应的边界概率设为0，以减少伪边缘带来的干扰。为评估阈值T对模型性能的影响，我们在BSD500验证集上进行了参数敏感性分析。为了公平评估阈值T对模型分割性能的影响，在实验中省略了最终的后处理优化步骤（第4.3节中的“分割结果优化”），以排除其他因素干扰。

表7展示了不同阈值设置下模型在分割准确性（BR，ASA和UE）与超像素规则性（CO）方面的表现。当阈值$T=0$时，引入了大量的伪边缘，影响了分割精度。此外，错误的边界信息也使得非边界区域出现形状不规则的超像素，导致规则性下降。随着T值逐渐增大，伪边缘得到有效抑制，边界概率的计算更加准确，模型的分割性能（BR，ASA和UE）和超像素规则性指标（CO）均呈上升趋势。当$T>5$时，虽然规则性指标取得了最优，但部分边缘强度较弱的边缘被误去除，导致边界信息不完整，从而使分割精度下降。综合来看，当$T=5$时，模型在准确性指标上表现最优，同时保持了较高的规则性。因此，本文采用$T=5$作为默认设置。

表7　不同阈值T下模型的性能分析（BS500验证集，像素个数300）

T	BR	ASA	UE	CO
0	0.8953	0.9570	0.1983	0.3360
1	0.8963	0.9573	0.1976	0.3365
2	0.8983	0.9578	0.1962	0.3372
3	0.9013	0.9586	0.1937	0.3386
4	0.9058	0.9600	0.1895	0.3412
5	**0.9124**	**0.9622**	**0.1833**	0.3464
6	0.9123	0.9620	0.1836	0.3578
7	0.9115	0.9616	0.1844	**0.3679**

注：粗体表示最优值。

6　结　论

本研究提出了一种用于超像素分割的边缘强化状态空间模型。我们利用SSM构建编码器，实现对图像全局结构的有效感知。更重要的是，设计了一种轻量级的边缘强化框架，在推理过程中显式引入边界信息，并结合交叉注意力与选择性扫描机制，实现边缘特征与编码特征的深度融合，使模型在处理复杂轮廓与低对比度区域时表现更优。我们还提出了一种基于边界概率的自适应损失函数。该函数使模型在边界区域优先关注语义一致性，而在非边界区域则更侧重于超像素的空间紧致性，从而在分割精度和形状规则性之间实现了更优的平衡。实验结果表明，我们的模型在多个关键指标上超越了现有最先进技术，并在不同类型数据集上展现出良好的泛化能力。

致　谢　本研究工作得到了国家自然科学基金联合基金项目的资助（批准号：U22A2033、U24A20219），这为本文研究的顺利开展提供了重要保障和有力支撑。在此，作者表达衷心感谢。

参　考　文　献

[1] Kwak S, Hong S, Han B. Weakly supervised semantic segmentation using superpixel pooling network//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, UAS, 2017: 4111-4117

[2] Yuan Y, Zhu Z L, Yu H, et al. Watershed-based superpixels with global and local boundary marching. IEEE Transactions on Image Processing, 2020, 29(8): 7375-7388

[3] Ren X F, Malik J. Learning a classification model for segmentation//Proceedings of the 9th IEEE International

Conference on Computer Vision. Nice, France, 2003: 10-17

[4] Li Z G, Wu X M, Chang S F. Segmentation using superpixels: a bipartite graph partitioning approach//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012: 789-796

[5] Wang H O, Lu H, Guo Q, et al. Design of superpixel U-Net network for medical image segmentation. Journal of Computer-Aided Design & Computer Graphics, 2019, 31(6): 1007-1017 (in Chinese)
(王海鸥, 刘慧, 郭强, 等. 面向医学图像分割的超像素U-Net网络设计. 计算机辅助设计与图形学学报, 2019, 31(6): 1007-1017)

[6] Yang F, Lu H, Yang M H. Robust superpixel tracking. IEEE Transactions on Image Processing, 2014, 23(4): 1639-1651

[7] Yuan Y, Fang J, Wang Q. Robust superpixel tracking via depth fusion. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 24(1): 15-26

[8] Li Y l, Zhou Z, Wu W. Scene parsing based on a two-level conditional random field. Chinese Journal of Computers, 2013, 36(9) :1898-1907 (in Chinese)
(李艳丽, 周忠, 吴威. 一种双层条件随机场的场景解析方法. 计算机学报, 2013, 36(9):1898-1907)

[9] Mori G, Ren X F, Efros A A, et al. Recovering human body configurations: combining segmentation and recognition// Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, USA, 2004: II-II

[10] Cour T, Shi J. Recognizing objects by piecing together the segmentation puzzle//Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, USA, 2007: 1-8

[11] Wang Z, Feng J S, Yan S C, et al. Image classification via object-aware holistic superpixel selection. IEEE Transactions on Image Processing, 2013, 22(11): 4341-4352

[12] Cheng J, Liu J, Xu Y W, et al. Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. IEEE Transactions on Medical Imaging, 2013, 32(6): 1019-1032

[13] Liu Z, Zhang X, Luo S H, et al. Superpixel-based spatiotemporal saliency detection. IEEE Transactions on Circuits and Systems for Video Technology, 2014, 24(9): 1522-1540

[14] Chen B C, Tao X, Chen H, et al. Saliency detection via fusion of boundary connectivity and local contrast. Chinese Journal of Computers, 2020, 43(1):16-28 (in Chinese)
(陈炳才, 陶鑫, 陈慧, 等. 融合边界连通性与局部对比性的图像显著性检测. 计算机学报, 2020, 43(1):16-28)

[15] Jampani V, Sun D Q, Liu M Y, et al. Superpixel sampling networks//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany, 2018: 352-368

[16] Yang F T, Sun Q, Jin H L, et al. Superpixel segmentation with fully convolutional networks//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 13961-13970

[17] Wang Y X, Wei Y C, Qian X M, et al. AINet: association implantation for superpixel segmentation//Proceedings of the 18th IEEE/CVF International Conference on Computer Vision. Montreal, Canada, 2021: 7058-7067

[18] Xu S, Wei S K, Ruan T, et al. Learning invariant inter-pixel correlations for superpixel generation//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada, 2024: 6351-6359

[19] Xu S, Wei S K, Ruan T, et al. ESNet: an efficient framework for superpixel segmentation. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 34(7): 5389-5399

[20] Fu D Y, Dao T, Saab K K, et al. Hungry hungry hippos: towards language modeling with state space models. arXiv preprint arXiv: 2212.14052, 2022

[21] Smith J T H, Warrington A, Linderman S W. Simplified state space layers for sequence modeling. arXiv preprint arXiv: 2208.04933, 2022

[22] Gu A, Dao T. Mamba: linear-time sequence modeling with selective state spaces. arXiv preprint arXiv: 2312.00752, 2023

[23] Qian X, Li X M, Zhang C M. Weighted superpixel segmentation. The Visual Computer, 2019, 35(6): 985-996

[24] Yuan Y, Zhang W, Yu H, et al. Superpixels with content-adaptive criteria. IEEE Transactions on Image Processing, 2021, 30: 7702-7716

[25] Xu Y Y, Gao X F, Zhang C M, et al. High quality superpixel generation through regional decomposition. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 33(4): 1802-1815

[26] Wang W G, Shen J B, Shao L, et al. Correspondence driven saliency transfer. IEEE Transactions on Image Processing, 2016, 25(11): 5025-5034

[27] Grady L. Random walks for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(11): 1768-1783

[28] Liu M Y, Tuzel O, Ramalingam S, et al. Entropy rate superpixel segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Colorado, USA, 2011: 2097-2104

[29] Achanta R, Shaji A, Smith K, et al. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2274-2281

[30] Li Z Q, Chen J S. Superpixel segmentation using linear spectral clustering//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 1356-1363

[31] Van den Bergh M, Boix X, Roig G, et al. Seeds: superpixels extracted via energy-driven sampling//Proceedings of the 12th European Conference on Computer Vision. Florence, Italy, 2012: 13-26

[32] Zhang Z L, Li A H, Li C W. Superpixel segmentation algorithm based on clustering by finding density peaks. Chinese Journal of Computers, 2020, 43 (1):1-15 (in Chinese)
(张志龙, 李爱华, 李楚为. 基于密度峰值搜索聚类的超像素分割算法. 计算机学报, 2020, 43(1):1-15)

[33] Wang Y X, Wei Y C, Qian X M, et al. Generating superpixels for high-resolution images with decoupled patch calibration. Chinese Journal of Computers, 2024, 47(11): 2664-2677 (in Chinese)
(王亚雄，魏云超，钱学明，等. 基于解耦区域校准的高分辨率超像素生成算法. 计算机学报，2024，47(11):2664-2677)

[34] Pan X, Zhou Y F, Zhang Y F, et al. Fast generation of superpixels with lattice topology. IEEE Transactions on Image Processing, 2022, 31: 4828-4841

[35] Gu A, Goel K, Ré C. Efficiently modeling long sequences with structured state spaces. arXiv preprint arXiv: 2111.00396, 2021

[36] Liu Y, Tian Y, Zhao Y, et al. Vision mamba: efficient visual representation learning with bidirectional state space model// Proceedings of the 41st International Conference on Machine Learning. Vienna, Austria, 2024, 235:62429-62442

[37] Liu Y, Tian Y, Zhao Y Z, et al. Vmamba: visual state space model. arXiv preprint arXiv:2401.10166

[38] Wang Z, Ma C. Weak-mamba-unet: visual mamba makes cnn and vit work better for scribble-based medical image segmentation. arXiv preprint arXiv:2402.10887, 2024

[39] Ma J, Li F, Wang B. U-mamba: enhancing long-range dependency for biomedical image segmentation. arXiv preprint arXiv: 2401.04722, 2024

[40] Guo H, Li J M, Dai T, et al. MambaIR: a simple baseline for image restoration with state-space model//Proceedings of the 18th European Conference on Computer Vision. Milan, Italy, 2024: 222-241

[41] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, 2020

[42] Wan Z, Zhang P, Wang Y, et al. Sigma: siamese mamba network for multi-modal semantic segmentation//Proceedings of the 2025 IEEE/CVF Winter Conference on Applications of Computer Vision. Tucson, USA, 2025: 1734-1744

[43] Liu Y, Cheng M M, Hu X W, et al. Richer convolutional features for edge detection//Proceedings of the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017: 5872-5881

[44] Martin D, Fowlkes C, Tal D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics// Proceedings of the 8th IEEE International Conference on Computer Vision. Vancouver, Canada, 2001: 416-423

[45] Li Y, Hou X D, Koch C, et al. The secrets of salient object segmentation//Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014: 280-287

[46] Yamaguchi K, Kiapour M H, Ortiz L E, et al. Parsing clothing in fashion photographs//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012: 3570-3577

[47] Staal J, Abràmoff M D, Niemeijer M, et al. Ridge-based vessel segmentation in color images of the retina. IEEE Transactions on Medical Imaging, 2004, 23(4): 501-509

[48] Russakovsky O, Deng J, Su H, et al. Imagenet large scale visual recognition challenge. International Journal of Computer Vision, 2015, 115(3): 211-252

[49] Rand W M. Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association, 1971, 66(336): 846-850

[50] Zhang Y X, Li X M, Gao X F, et al. A simple algorithm of superpixel segmentation with boundary constraint. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 27(7): 1502-1514

[51] Sun L M, Ma D Y, Pan X, et al. Weak-boundary sensitive superpixel segmentation based on local adaptive distance. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 33(5): 2302-2316

[52] Yan T M, Huang X L, Zhao Q F. Hierarchical superpixel segmentation by parallel CRTrees labeling. IEEE Transactions on Image Processing, 2022, 31: 4719-4732

[53] Zhou P, Kang X J, Ming A L. Vine spread for superpixel segmentation. IEEE Transactions on Image Processing, 2023, 32: 878-891

[54] Zhu L, She Q, Zhang B, et al. Learning the superpixel in a non-iterative and lifelong manner//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Online Meeting, USA, 2021: 1225-1234

XU Yun-Yang, Ph. D. candidate. His research interests are image super-resolution, and image segmentation.

FANG Le-Xin, Ph. D. candidate. Her research interest is medical image segmentation.

ZHANG Cai-Ming, Ph. D., professor, His research interests include digital image processing and time series analysis.

LI Xue-Mei, Ph. D., professor. Her research interests include digital image processing and time series analysis.

Background

Superpixel segmentation is a widely adopted technique in image processing and computer vision. Its fundamental idea lies in partitioning an image into a collection of small, coherent regions—referred to as superpixels. These regions exhibit high internal similarity in terms of color, texture, and brightness. Compared to processing at the individual pixel level, employing superpixels as the basic unit of computation offers enhanced semantic representation and substantially reduces the computational burden of subsequent processing.

Superpixel segmentation is often used in various visual tasks such as image segmentation, object detection, image compression, etc.

Limited by artificially designed features, traditional superpixel segmentation algorithms lack accuracy and generalization in complex scenes or cross-domain images. In recent years, CNN-based superpixel segmentation methods have gained increasing attention. These methods can automatically learn rich and hierarchical image representations from large-scale data, significantly improving both segmentation accuracy and cross-domain adaptability. However, several critical challenges remain:

(1) Due to the inherently local nature of CNN, current models struggle to capture long-range spatial dependencies within images. This limitation weakens their ability to integrate global semantic information, often resulting in the incorrect assignment of semantically dissimilar yet spatially adjacent regions to the same superpixel.

(2) Most existing methods rely on implicit cues such as color and texture to infer object boundaries, lacking explicit modeling of the relationship between boundary information and semantic understanding. Consequently, they are prone to segmentation errors, particularly in regions with complex boundaries or low contrast.

(3) Many deep learning-based approaches employ a dual-loss framework with fixed weights to jointly optimize representational consistency and spatial compactness. However, the use of fixed weights fails to adapt to the varying characteristics of different image regions. As a result, boundary regions are often segmented inaccurately, while non-boundary regions may lose regularity, ultimately harming the overall quality of the generated superpixels.

In this paper, we propose an Edge-Enhanced State Space Model (EE-SSM) to improve both the accuracy and regularity of superpixel segmentation through three key perspectives: global context modeling, boundary-aware semantic enhancement, and dynamic loss optimization. Specifically, EE-SSM adopts a state space encoder-convolutional decoder architecture, which strikes a balance between long-range dependency modeling and efficient inference. In particular, a plug-and-play lightweight edge enhancement framework for superpixel segmentation is designed to provide explicit edge information to the model. Finally, the model is trained using a newly proposed adaptive loss function based on boundary probability to solve the optimization problem of balancing superpixel segmentation accuracy and regularity.

This work was supported by the Joint Fund of the National Natural Science Foundation of China under Grant Nos. U22A2033, U24A20219.