文章编号:1671-4598(2025)12-0237-09 DOI:10.16526/j.cnki.11-4762/tp.2025.12.031 中图分类号:TP391 文献标识码:A

基于形状感知语义对齐的双分支分割网络

庄祉珊[1,2]，吴静静[1,2]，赵迎龙[3]，魏 斌[3]
(1. 江南大学 智能制造学院，江苏 无锡 214122；
2. 江苏省食品先进制造装备技术重点实验室，江苏 无锡 214122；
3. 江苏省特种设备安全监督检验研究院无锡分院，江苏 无锡 214000)

摘要：针对复杂工况下缺陷干扰，脏污噪声和镜头模糊导致目标分割精度低的问题，提出了一种基于形状感知语义对齐的双分支分割网络；针对由深层到浅层的语义传播错误导致的低精度问题，设计了语义流对齐模块，学习特征图之间的偏移量辅助信息对齐；引入了注意力引导自选择融合模块，结合深层信息和浅层信息特性来指导更准确的分割；设计了形状感知损失函数，利用形状特征引导网络关注难以分割的边界区域解决噪声与目标粘连的问题，提高了分割性能；在自建芯片数据集上进行的综合实验证实，此方法提高了特征表示和分割性能，相比基线网络，*mIoU* 达到了94.4%（提升了2.1%），速度达到了48.86 FPS（提高21%），实现了精度与速度的平衡，可满足实际工业应用；在CamVid数据集上，相比基线网络，*mIoU* 为65.1%（提升了3.0%），同时参数量减少4.6%，证实了所提算法的普适性。

关键词：分割；双分支网络；语义流；注意力机制；距离映射

Attention-Guided Shape-Aware Bilateral Segmentation Network

ZHUANG Zhishan[1,2], WU Jingjing[1,2], ZHAO Yinglong[3], WEI Bin[3]
(1. School of intelligent manufacturing, Jiangnan University, Wuxi 214122, China;
2. Jiangsu Key Laboratory of Advanced Food Manufacturing Equipment and Technology, Wuxi 214122, China;
3. Wuxi Branch of Jiangsu Institute of Special Equipment Safety Supervision and Inspection, Wuxi 214000, China)

Abstract: To address the low target segmentation accuracy caused by defect interference, dirty noise, and blurred lenses in complex scenarios, a attention-guided shape-aware bilateral segmentation network was proposed. To solve the low precision caused by semantic propagation errors from deep to shallow layers, a semantic flow alignment module was designed to learn the offset between feature maps and assist information alignment. An attention-guided self-selective fusion module was introduced to guide more accurate segmentation by combining deep and shallow information characteristics. A shape-aware loss function was designed, the shape features were used to guide the network to pay more attention to difficult boundary regions to solve the noise and object adhesion, improving segmentation performance. Comprehensive experiments on the self-built chip dataset show that, this method is superior in the feature representation and segmentation performance to the baseline network, with the mIoU reaching 94.4% (an improvement of 2.1%), the speed reaching 48.86 FPS (an improvement of 21%), achieving a balance between accuracy and speed and meeting actual industrial applications. On the CamVid dataset, the mIoU is 65.1% (an improvement of 3.0%) while the number of parameters is reduced by 4.6%, proving the universality of the proposed algorithm.

Keywords: segmentation; double-branch network; semantic flow; attention mechanism; distance map

0 引言

芯片表面字符分割是利用机器视觉进行工业检测程序中的关键一环。然而，从图像中完全自动地进行文本分割，尤其是从工业场景下获取的图像中，一直是一个挑战性问题。在芯片表面字符印刷过程中，经常出现划

收稿日期:2024-11-14； **修回日期**:2024-12-21。
基金项目:国家自然科学基金项目(62072416,61873246);河南省科技研发计划联合基金重点项目(235200810022)。
作者简介:庄祉珊(1999-),女,硕士。
通讯作者:吴静静(1982-),女,博士,副教授。
引用格式:庄祉珊,吴静静,赵迎龙,等.基于形状感知语义对齐的双分支分割网络[J].计算机测量与控制,2025,33(12):237-245.

痕干扰、镜头模糊和生产线上环境污染物干扰等问题。这些因素给图像分割带来了重大挑战。此外，芯片产品上印刷的表面字符的样式多样性进一步增加了分割过程的复杂性。

在深度学习出现之前，诸如边缘检测[1]、轮廓寻找[2]和纹理分析[3]等方法依赖于明确定义的约束和标准来区分图像中的对象差异。然而，这些方法通常鲁棒性不佳。早期的语义分割网络通常采用编码器－解码器架构[4-5]。编码器通过卷积或池化逐步扩大其感受野，而解码器则使用反卷积或上采样从编码器的输出中恢复分辨率。然而，在编码器－解码器网络的下采样阶段保持足够的细节信息是具有挑战性的。为了增强性能，研究多集中在增强上下文细节的捕获[6-7]和增强强语义信息的特征表示[8-9]两个角度。类似 FPN 的模型[10-11]使用自上而下的方法融合特征图。PSPNets[9]引入了金字塔池化模块（PPM，pyramid pooling module）以有效捕获多尺度上下文信息，而 HRNet[12]使用多路径和双边连接来学习和整合不同尺度的表示。但是这些方法不仅消耗过多的计算能力，而且在准确性上的提升十分有限。文献［13－14］在网络的最后几个阶段使用空洞卷积[15]，以产生具有强语义表示的特征图，同时保持高分辨率。受此启发，DeepLab 系列[8,16-17]通过在网络中使用不同扩张率的空洞卷积[15]，实现了对先前工作的显著改进，此方法可以扩展视野而不降低空间分辨率。然而，需要注意的是，空洞卷积由于其不规则的内存访问模式，在硬件实现方面具有很大的挑战。此外，研究认为，特征不对齐问题是影响语义分割性能的关键因素，性能的瓶颈被怀疑源于不同分辨率特征图中信息的错位和融合无效[18]。经过多次计算，如上/下采样和残差连接，通过简单采样恢复信息并确保不同层次之间的完整信息交互变得困难。为了解决这个问题，AlignSeg[19]引入了特征聚合模块和对齐上下文建模模块，SFNet[18]应用转换偏移来对齐特征图。SFNet［20］提出了流对齐模块（FAM），以在同一阶段内对齐特征图，以实现更有效的融合。可变形卷积[20]在规则网格采样位置添加 2D 偏移，帮助高层 CNN 层在空间位置上进行语义编码。

为了解决工业检测领域高实时性能的需求，引入了各种高效和加速的 CNN 模型。为了简化复杂性并提高速度，一些模型通过调整输入尺寸[21]、裁剪[22]、剪枝网络通道或牺牲空间细节来修改输入维度。BiSeNet[23]的作者提出了一种双分支网络架构，该架构由两个不同深度的分支组成，一个用于上下文嵌入，另一个用于细节解析，并通过特征融合模块增强上下文和详细信息的融合。此模型既可以通过大感受野可以有效捕获上下文依赖性，同时可以捕获空间细节信息，有利于对于准确的边界划分和小尺度目标识别。然而，由于大多骨干网络是从最初为图像分类设计的网络中迁移过来的，因此它可能会影响分割精度。为了解决这个问题，文献［24］开发了一种短期密集连接（STDC，short-term dense concatenate）结构，以有效提取接收感受野和多尺度信息，解决了 BiSeNet 骨干网络的限制。同时，消除了多余的路径，并加入了细节引导模块，增强了架构对语义分割的适用性。文献［25］和［26］使用双分支网络结构解决了图像分割中的难题。一些后续工作在此基础上增强了其表示能力或减少了模型复杂性[27-29]。此外，DDRNet[30]引入了双边连接以促进上下文和详细分支之间的信息交换，极大提高了实时语义分割的精度。尽管这些网络在捕获语义信息和细节感知方面做出了巨大努力，但将语义和详细信息有效融合的挑战尚未完全解决。

由于生产环境下图片质量不理想（如对比度低，噪声高等）、带有干扰性背景及芯片字符样式方向不统一等原因，芯片文本分割仍然具有极大挑战，特别是噪声附着在对象边界上的情境下。以往的研究侧重于通过残差膨胀单元或额外的边缘约束来获取边缘信息，以补充丢失的细节并提高准确性[31-34]。PID-Net[35]使用 3 个分支来解析细节、上下文和边界信息，采用边界注意力来指导融合，在增强边界信息的提取方面表现出色，特别是在处理复杂边缘时。SMU-Net[36]提出了感知单元来约束分割边缘，有效提高了分割的准确性和边缘的平滑度。然而，很少研究关注其他的形态信息，例如分割目标的位置和形状，实际上，形态信息可以合作并指导分割性能的提高[37]。文献［38］提出了一种从距离图派生的创新损失惩罚项，文献［39］主张通过从真实标签生成的距离图对边界附近的分割不准确进行惩罚，这类方法在处理边界问题时显示出了独特的效果，能够在不增加模型复杂度的前提下提高分割精度。由此可见形状信息在提升分割模型的边界处理能力的重要作用。

为了解决上述问题，本文提出了一种改进 STDC 的实时分割网络架构（AS-BiseNet，attention-guided shape-aware bilateral segmentation network），用于工业场景中的文本分割。它使用了 STDC [24]提出的高效骨干网络，移除了耗时的空间路径以及原始架构中的注意力细化模块（Attention Refinement Module）和特征融合模块（Feature Fusion Module），引入了语义流对齐模块来对齐多级特征图，注意力引导自选择融合模块来平衡细节和上下文信息，并使用形状感知损失函数来提高网络对难以分割像素的关注。

本文的贡献总结如下：

1）提出了一个语义流对齐模块，用于学习两个相邻特征图的偏移量，帮助动态准确地对齐特征图。该模

块解决了由于深层到浅层的语义传播错误导致的低精度问题。

2）提出了一个注意力引导自选择融合模块，它利用通道注意力和空间注意力生成加权图，然后使用权重融合特征图。与原始模块相比，性能显著提高。

3）提出了一个从真实标签掩模派生的惩罚项，以创建形状感知损失函数。这个函数通过引导网络关注难以分割的边界区域，提高了分割性能。

4）提出了一个高效且通用的框架，名为 AS-BiseNet，该框架在工业场景文本分割任务中展现出卓越的性能。通过集成语义流对齐模块、注意力引导自选择融合模块以及形状感知损失函数，AS-BiseNet 能够在保证实时性的同时，显著提升分割精度和鲁棒性。

提出的框架在广泛的实验中展示了强大的性能以及普适性。实验结果表明，AS-BiseNet 在 CamVid 上表现良好。它在 CamVid 上实现了 65.1% 的 $mIoU$，比基线网络（STDC-Seg）提高了 3.0%，同时参数量减少了 4.6%。

1　AS-BiseNet 模型

本文提出了一种新颖高效的架构，使用 STDC 编码器，结合了两个改进的注意力机制模块，用于不同层次特征图的有效融合，以及提出一个新颖的损失惩罚项，以增强目标形状提取。

1.1　网络架构

基于语义流对齐模块（SFAM，semantic flow alignment module）、注意力引导自选择融合模块（ASFM，attention-guided self-selective fusion module）和形状感知损失函数（Shape-aware Loss），本文提出了 AS-BiseNet 用于工业场景中的文本分割（见图 1）。

在本文提出的方法中，使用了两个 SFAM 模块，通过自底而上层级叠加的方式融合相邻层级的高层特征图。随后，将 SFAM 模块融合后的最终结果特征图与最深层的特征信息进行叠加，得到高级语义特征。最后，通过 ASFM 模块将高级语义特征与 SFAM 的输出特征进行融合，同时保留了细节和语义信息，最终经过分割头，得到分割结果。

在这个架构中，SFAM 模块有效增强了相邻层级特征图之间的上下文关联，而 ASFM 模块则协同整合了细节纹理信息与更广泛的上下文感知。此外，形状感知损失函数通过利用分割目标的形状信息，帮助网络进一步提升边缘分割的性能。

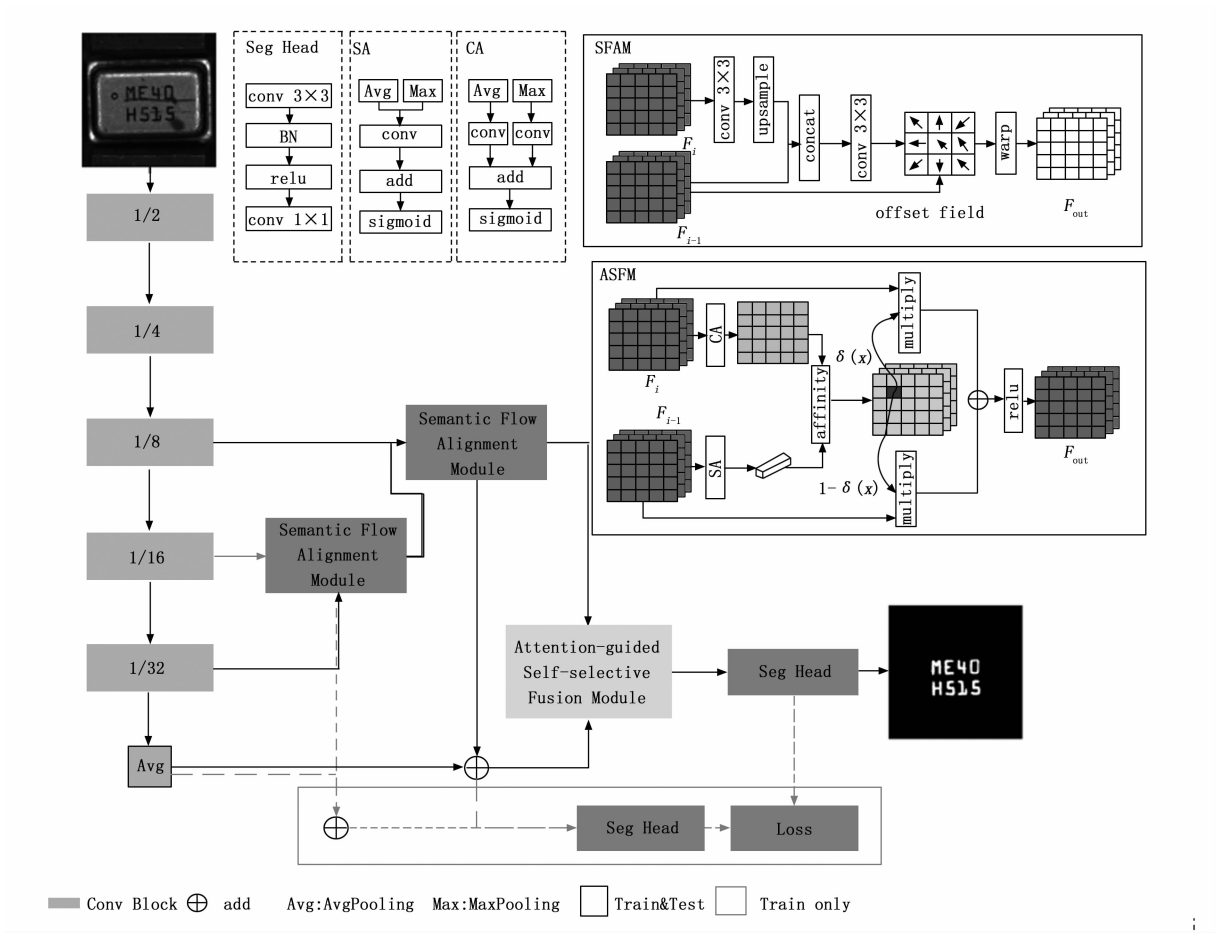


图 1　AS-Bisenet 整体结构图

1.2 语义流对齐模块

高层特征恢复到详细的高分辨率图像仍然是一个持续研究的难题。传统的卷积神经网络（CNN）在处理上下文信息时能力不足，并且由于残差连接、下采样和上采样等操作，使得特征图的对齐变得复杂，导致难以从原始图像中精确恢复细节。文献［8］在网络的最后几层使用空洞卷积，但这种方法显著增加了计算需求，并降低了推理速度。特征金字塔网络（FPN）[40]通过将浅层高分辨率图像的详细特征整合到深层特征图中，试图提升细节恢复的精度，但与高分辨率特征图相比，其精度提升仍然有限。本文提出了一个语义流对齐模块（SFAM），灵感来自可变形卷积网络（DCN，deformable convolution networks）[20]，设计提出语义流场来表示两个相邻特征图之间的由于复杂操作导致的特征偏移量，通过更好地将语义信息传递到高分辨率图像中，从而有效提升了特征图中丢失细节的恢复效果。

首先，生成语义流场。给定两个相邻的特征图，高层特征图为 F_i，低层特征图为 F_{i-1}，将 F_i 通过一个 1×1 卷积压缩到与 F_{i-1} 相同的通道深度，并通过一个具有 3×3 核大小的双线性插值层将 F_i 上采样到 F_{i-1} 相同的大小。然后将其在通道维度上拼接，使用一个 3×3 的卷积层提取融合特征，获得语义流场信息 $\{\Delta P_n \mid n=1,\cdots,N\}$，其中 $N=|R|$。上述步骤可以写成：

$$\Delta P_n = conv_p\{cat[con(F_i),u(F_{i-1})]\} \tag{1}$$

其中：$cat(\cdot)$ 表示拼接操作，$conv_p(\cdot)$ 是 3×3 卷积层，$conv$ 表示 1×1 卷积层，u 表示上采样操作。

其次，对低级特征图上的每个特征点进行 warp 生成。网格 R 定位为一个扩张率为 1 的 3×3 大小的核，$R=\{(-1,1),(-1,0),\cdots,(0,1)(1,1)\}$。使用了带有偏移量的网格 R 在低层特征图上进行采样，使用双线性插值得到偏移特征点。然后根据权重 w 对采样值加权求和，最终输出结果。对于输出特征图 F_{out} 上的每个位置 P_0，其值如公式（2）：

$$F_{\text{out}} = \sum_{P_n\in R} w(P_n)\cdot F_{i-1}(P_0+P_n+\Delta P_n) \tag{2}$$

其中：P_n 枚举 R 中的位置。

1.3 注意力引导自选择融合模块

低层特征图保留了丰富的细节信息，而高层特征图则捕获了更高级的语义信息。直接进行特征图的乘法或拼接操作可能会导致细节信息的丢失，从而无法有效学习到关键特征。直接融合原始的细节信息和低频上下文信息可能会导致大量信息被淹没，影响模型的性能。因此，本文提出了一种注意力引导自选择融合模块（ASFM），旨在通过高层特征图与低层特征图的相互引导，提升特征融合的效果。

具体来说，ASFM 利用低层特征图中的细节和边缘信息来指导高层特征图更好地选择边界和其他关键细节信息。先对高层特征应用通道注意力[41]以提取重要通道，对低层特征应用空间注意力[42]以提取重要的空间信息。通过这两种注意力机制，我们可以更好地捕捉到特征图中的关键信息。然后，我们通过亲和力操作获得包含通道和空间信息的权重图，利用权重图指导高低层特征图融合。

首先注意力模块将高层特征图 F_i 和低层特征图 F_{i-1} 作为输入，并产生加权特征图 V_C 和 V_S。数学上，这些步骤可以写成：

$$V_C = Sigmoid\{conv[MaxPooling(F_i)]+conv[AvgPooling(F_i)]\} \tag{3}$$

$$V_S = Sigmoid\{conv\{concat[MaxPooling(F_{i-1}),AvgPooling(F_{i-1})]\}\} \tag{4}$$

在获得注意力加权特征后，通过亲和力操作生成权重图。Sigmoid 函数的输出可以写成公式（5）。最后，ASFM 利用权重图 δ 对高低层特征 F_i 和 F_{i-1} 进行逐元素相乘，而后相加输出融合特征。上述过程表示为：

$$\delta = \text{sigmoid}(V_S\cdot V_C) \tag{5}$$

$$F_{\text{out}} = \text{Relu}(\delta\cdot F_i+(1-\delta)\cdot F_{i-1}) \tag{6}$$

1.4 形状感知损失函数

以往的研究致力于通过残差扩张单元或附加的边缘约束来捕获边缘信息，以增强边缘分割。然而，当噪声干扰与分割目标在特征上难以分辨时，形状特征信息对提高分割性能就具有重要意义。因此，本文提出了一种改进的交叉熵损失函数，以距离图映射作为语义分割的惩罚参数。

首先，生成来自真实掩模的形状约束映射 $D(x,y)$。$E=\{e_{k,l}\}$ 表示对象边界的点集合，可以通过优化函数计算如下：

$$D(x,y) = \Gamma[\min_{e_{k,l}\in E} d(r_{x,y},e_{k,l})] \tag{7}$$

其中：$d(r_{x,y},e_{k,l})$ 表示所述图像中的每个像素点 $r_{x,y}$ 与所述边界像素 $e_{k,l}$ 之间的坐标距离，并定义 $\Gamma\{A\}$ 为 $max\{A\}-A$。由于图像有多个目标区域，需要对形状约束映射进行归一化，以确保分割区域之间的相对一致性。上述步骤表示为：

$$\begin{cases}\mu_D = \dfrac{1}{m}\sum_{i=1}^{m} D_i(x,y)\\ \sigma_D^2 = \dfrac{1}{m}\sum_{i=1}^{m}(D_i(x,y)-\mu_D)^2\\ D_{\text{nor}}(x,y) = \dfrac{D_i(x,y)-\mu_D}{\sqrt{\delta_D^2+\varepsilon}}\end{cases} \tag{8}$$

其中：m 表示该集合 D 的个数。正则化后的距离图 D_{nor} 用于惩罚训练过程中的预测误差。本文认为，不仅应该关注边缘区域的单个像素，而且应该对边缘区域的像素进行减弱的关注。因此，通过距离变换函

数对距离图 D_{nor} 进行数学变换，生成一个形状约束图，如图 2（c）所示。直观地说，如果像素点更接近边缘，像素所对应的值就会更大，从而绘制目标区域的形状轮廓，并在训练过程中为边缘区域的像素分配更大的损失。在训练过程中，目标是最小化公式（9）计算的损失值：

$$L_{\text{distance}} = \frac{1}{N} \cdot \sum_{i=1}^{N} (e^{D_{\text{nor}}} + 3^{D_{\text{nor}}}) \sum_{j=1}^{K} - y_j \log \hat{y}_j \quad (9)$$

式中，N，K 分别为样本数量和种类的数量，y 为真实值，$\hat{y}$ 为网络预测值，· 是点乘。

(a) 标签图　　(b) 距离图　　(c) 形状约束图

图 2　状感知损失函数计算过程可视化

本文对两个语义流对齐模块和注意力引导自选择融合模块的输出进行监督。提出的 AS-Bisenet 的总损失函数 L_{total} 可以表示为：

$$L_{\text{total}} = L_{\text{ohemce}} + L_{\text{distance}} \quad (10)$$

L_{ohemce} 是结合了在线难例挖掘（OHEM，online hard example mining）和交叉熵损失（Cross-Entropy Loss）的损失函数，阈值设为 0.7。L_{distance} 为本文提出的形状感知损失函数。

2　实验结果与分析

2.1　数据集

本文在自建芯片数据集 MESH-chip 和 CamVid 数据集[43]上进行了实验。

自建芯片数据集 MESH-chip 是在自然工业场景中捕获的芯片图像的集合。它包含 1 288 张精细注释的图像，分别分为训练、验证和测试集，分别包含 902、271 和 115 张图像。本文使用了两个类（对象和背景）。该数据集的分辨率为 640×640。包括多种工况的图片，如油墨污染、划痕和灰尘污渍，这对于缺陷干扰，脏污噪声和镜头模糊导致的分割不佳的问题至关重要。

CamVid（Cambridge-driving Labeled Video Database）[43]是一个超过 10 分钟的高质量视频的集合，分辨率为 720×960。它从电影序列中提取 701 幅图像，并精细地标记 11 个类别，用于语义分割任务。其中 367 张图片进行训练，101 张用于验证，233 张用于测试。

2.2　实验环境与参数设置

本文采用带有动量为 0.9 和权重衰减为 5e-4 的小批量随机梯度下降（SGD）[44]进行训练。类似于[23]中的方法，采用“poly”学习率策略，初始学习率逐渐减少。初始学习率设定为 0.001，这是为了在确保训练稳定性的同时，尽可能地加快收敛速度。衰减因子为 0.9，在训练后期，随着模型逐渐接近最优解，减小学习率有助于模型在细粒度上进行更精细的调整，从而避免在最优解附近产生过大的震荡。训练批次大小在 CamVid 和自建芯片图像数据集上均设置为 8，这一设置是基于内存限制和计算效率的考虑。两者在前 1000 次迭代中均应用线性预热方法[45]。参照文献［18］和［24］，本文还使用了在线难例挖掘（OHEM）[46]。在所有实验中，使用 PyTorch 1.11.0 [47]，在 CUDA 11.3 下，通过单卡 NVIDIA GeForce RTX 3090 GPU 进行。

本文采用颜色抖动、随机水平翻转、随机裁剪和随机缩放进行数据增强。在 CamVid 和 MESH-chip 芯片图像数据集上的裁剪分辨率分别为 720×960 和 640×640。缩放比例包括{0.125，0.25，0.375，0.5，0.625，0.75，0.875，1.0，1.125，1.25，1.375，1.5}。

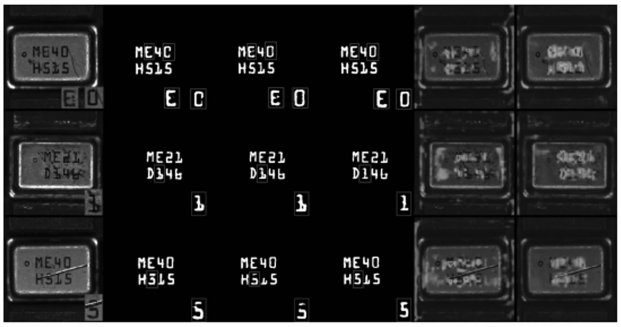
2.3　消融实验

本文提出了一个语义流对齐模块（SFAM），旨在解决不同层级特征图之间传播过程中深层语义信息错位的问题。基于可变形卷积，本文通过结合相邻层级特征的信息来计算卷积偏移量。通过实验验证，模块在 *mIoU*、*Dice* 和 *Recall* 指标上的表现优于双线性插值、最近邻插值等传统插值方法、DCN 和基准网络的特征融合模块（如表 1）。在 *Boundary _ IoU* 指标上，SFAM 高于插值算法 2.9%，在 *Recall* 指标上，SFAM 高于插值算法 6.6%，显示其将细节信息完整恢复到高分辨图像上的能力。SFAM 模块的预测结果的可视化表明，SFAM 特征比 DCN 具有更结构化的表现，并且在物体边界的准确性上更优（如图 3），分割文本更加完整，减轻伪影对分割的干扰。热力图表示在加入 SFAM 之后，网络对粘连边缘的关注度更高。

表 1　语义流对齐模块的消融实验

Method	*mIoU*	*Boundary_IoU*	*Dice*	*Recall*	*FLOPS*/G	*Params*/M
bilinear	0.922	0.844	0.880	0.797	**35.75**	**12.60**
nearest	0.924	0.823	0.854	0.753	**35.75**	**12.60**
deconvolution	0.929	0.851	0.891	0.818	48.48	14.44
baseline	0.923	0.865	0.864	0.773	36.75	14.24
SFAM	**0.935**	**0.873**	**0.914**	**0.863**	35.96	13.58

注意力自选择模块利用来自不同特征图层级的信息，高层特征图和低层特征图之间相互引导。该模块通过激活函数自适应地保留详细信息和语义信息。如图 4 所示，Grad-CAM 的可视化结果表明，低层特征图上网络对边缘信息的关注度更高，高层特征图上则更关注语



(a) input (b) baseline (c) DCN (d) SFAM (e) before SFAM (f) after SFAM

图 3 语义流对齐模块可视化

义部分，经过该模块的修正，融合特征图中亮暗区域的对比更加显著。网络更准确地关注到正确的像素，并生成更细致的分割边界。其消融实验结果如表 2 所示。

(a) input (b) F_i (c) F_{i-1} (d) output of ASFM

图 4 注意引导自选择融合模块的 Grad-CAM 可视化

表 2 MESH-chip 数据集上消融实验

	STDC	SFAM	ASFM	Shape-aware Loss	*mIoU*	*FLOPs*/G	*Para ms*/M	*FPS*
Test 1	√				0.923	36.75	14.24	62.11
Test 2	√	√			0.934	39.26	14.13	47.13
Test 3	√		√		0.935	33.49	13.70	61.38
Test 4	√			√	0.932	36.75	14.24	64.83
Test 5	√	√	√	√	**0.944**	35.96	13.58	**48.86**

距离变换函数旨在为物体边缘的外围像素分配较高的值，并随着像素逐渐远离边缘时逐步降低分配的值。为实现这一目标，本文对多种指数函数和幂函数及其组合进行了系统的评估，并通过不同组合函数的实验来验证其有效性。实验结果如表 3 所示，所有实验均采用 STDC 作为骨干网络，移除了基线网络中的细节损失，并保持相同的实验设置。基于 *mIoU* 指标，最终选择了一对指数函数的组合作为距离变换算法。

进一步地，本文设计了一系列实验，旨在评估不同损失函数在网络中的插入位置对其性能的影响。本文通过在不同位置（如 ASFM、SFAM1 和 SFAM2）添加损失函数，进行了 *mIoU* 指标的实验，以探索损失函数的最优插入点。所有实验均采用 STDC 作为主干网络，且未引入细节损失，确保实验设置的一致性。如表 4 所示，最佳的实验结果出现在将损失函数添加在 ASFM 模块、SFAM1 模块以及 SFAM2 模块之后。这些结果表明，在特定的融合模块之后插入损失函数，能够更有效地引导网络学习关键特征。

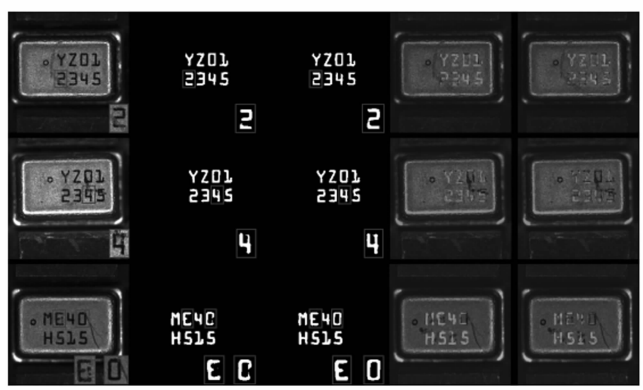
表 3 MESH-chip 数据集上距离变换函数对比实验

变换函数	*mIoU*
$y=x0.2$	0.931
$y=3x$	0.930
$y=ex$	0.931
$y=ex+3x$	0.932
$y=2ex$	0.928

表 4 MESH-chip 数据集上损失函数位置对比实验

After ASFM	After SFAM1 (1/16)	After SFAM2 (1/8)	Stage2 (1/4)	Stage3 (1/8)	*mIoU*
√	√	√			**0.932**
		√	√	√	0.930
√	√	√	√	√	0.927

与 STDC 中的细节损失相比，形状感知损失能够更清晰地区分噪声像素与目标像素，尤其是在边界分割方面表现出显著的提升。如图 5 所示，从热力图中可以观察到，在噪声与分割目标相近的情况下，形状感知损失显著提高了对目标区域的关注度，使分割结果更加完整和连贯。



(a) input (b) output without loss (c) output with loss (d) grad-cam without loss (e) grad-cam with loss

图 5 形状感知损失函数 Grad-CAM 可视化

部分定量结果如表 4 所示。在移除了基线网络的特征增强模块和边缘分支后，改进模型的参数量减少 0.66 M，每秒帧数（FPS）提高 13.25。

2.4 与其他方法的对比

如表 5 所示，AS-BiseNet 在自建芯片数据集上取

得了显著的性能提升，达到了 94.4% 的 *mIoU*，相比 STDC-1-Seg 提高了 2.1%。与其他方法相比，AS-BiseNet 在分割精度上表现出显著的改进。尽管其计算负载（FLOPs）略微超过了 Bisenet 和 BiSenetv2，但仍低于 STDC 和 ICnet，在模型性能和计算成本之间取得了良好的平衡。这充分证明了本文所提出方法的卓越能力。如图 6 所示，脏污、划痕、模糊以及文本边界粘连等因素对文本分割任务带来了显著挑战。与基线网络相比，本文模型在处理这些干扰因素时表现出更强的鲁棒性，能够更好地完成不同干扰下的文本分割任务。从热力图可以看出，本文模型对难以分割的边界区域给予了更高的关注度，从而提升了分割的准确性。此外，通过特征图的对比可以发现，本文模型能够更有效地恢复关键的边界信息，进一步增强了分割效果。

表 5　MESH-chip 数据集对比实验。No 表示没有主干网络

Model	Resolution	Backbone	*mIoU* /val	*mIoU* (test)	*FLOPs* /G	*Params* /M
ICnet	640×640	Resnet50	0.919	0.911	57.80	26.25
Bisenet	640×640	NO	0.922	0.921	23.83	13.42
Bisenetv2	640×640	NO	0.916	0.920	27.79	5.19
STDC	640×640	NO	0.933	0.923	36.75	14.24
Ours	640×640	NO	0.940	0.944	35.96	13.58

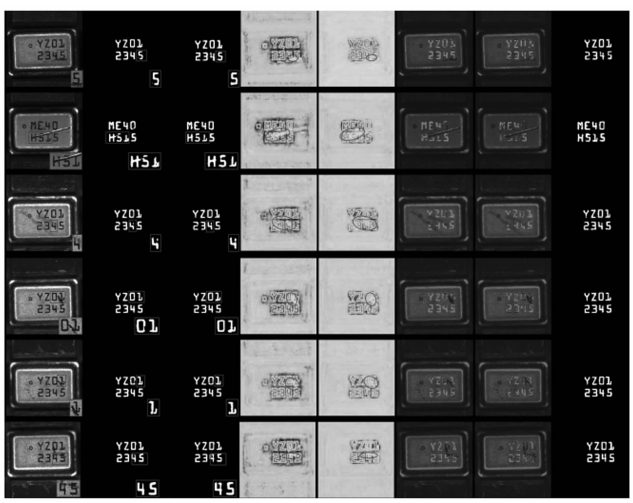


图 6　AS-BiseNet 和 STDC-1-Seg 在 MESH-chip 测试数据集上的对比可视化

在 CamVid 数据集上，我们采用 STDC 骨干网络构建的 AS-BiseNet 实现了 65.1% 的 *mIoU*，相较于 STDC-1-Seg 有 3%的显著提升。表 6 详细展示了所提出的网络架构的消融实验结果，这些实验结果有力地证明了本文引入的各项模块和损失函数的有效性。如图 7 所示，与基线网络相比，本文的网络在预测图中揭示了更多的结构细节，特别是在目标物体的边缘和细小结构上表现得更为出色。这种改进使得目标的分割结果更加清晰，有效地避免了目标被背景所淹没的情况。此外，表 7 还将本文的方法与其他分割算法进行了全面比较。结果显示，AS-BiseNet 不仅在分割精度上具有竞争力，而且在计算效率和模型大小等方面也表现出色。这些对比充分展示了 AS-BiseNet 在语义分割任务中的优越性能和潜力。

表 6　CamVidval 数据集的消融实验

	STDC	ASFM	SFAM	Shape-aware Loss	*mIoU*	*FPS*
Test 1	√				0.620	56.41
Test 2	√	√			0.621	56.88
Test 3	√	√	√		0.627	40.68
Test 4	√	√	√	√	**0.651**	**40.26**

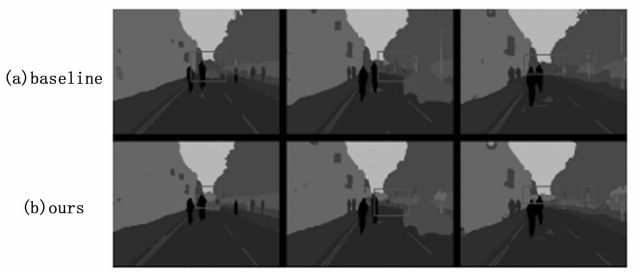


图 7　AS-BiseNet 和 STDC-1-Seg 在 CamVidval 数据集上的对比可视化

表 7　CamVid 数据集对比实验。No 表示没有主干网络

Model	Resolution	Backbone	*mIoU* /val	*FLOPs* /G	*Params* /M
Enet	720×960	No	0.513	**7.88**	**0.36**
Bisenet	720×960	No	0.634	23.83	13.42
Bisenetv2	720×960	No	0.593	27.79	5.19
STDC	720×960	No	0.621	36.75	14.24
LMDNet	720×960	Vgg16	0.635	—	—
DeepLab	720×960	Vgg16	0.616	—	65.10
SegNet	720×960	Vgg16	0.464	201.42	53.56
Ours	720×960	No	**0.651**	35.96	13.58

3　结束语

本文提出了 AS-BiseNet，这是一种用于工业场景文本的实时分割网络。该网络通过重新设计两个新的特征融合模块，显著提升了特征信息之间的交互效率，并更好地利用了低层和高层特征映射。语义流对齐模块（SFAM）：通过融合相邻特征图的信息，生成语义流，从而更有效地将高层特征传播到低层特征。与传统的空洞卷积和可变形卷积相比，SFAM 在减少计算量的同时提高了分割精度。注意力引导自选择融合模块（AS-FM）：通过引入注意力机制生成的权重，增强了特征表示的针对性，进一步提升了网络对复杂场景的适应能力。此外，本文还引入了一种形状感知损失函数，专门用于惩罚物体边界处的分割错误，从而有效提升边界分

割的准确性。实验结果表明，AS-BiseNet 在复杂工业场景下的分割准确性和实时性上均优于原始网络。在自建芯片数据集上，AS-BiseNet 达到了 94.4%的 $mIoU$，参数量减少了 4.6%，帧率提高了 21%。在 CamVid 数据集上，AS-BiseNet 实现了 65.1%的 $mIoU$，比原始网络[24]提高了 3.0%。未来工作将集中在进一步优化模型的实时性和降低计算成本，同时探索弱监督或无监督学习方法，以减少对标注数据的依赖。

参考文献：

[1] CANNY J. A computational approach to edge detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，1986：679－698.

[2] KASS M，WITKIN A，TERZOPOULOS D. Snakes：active contour models [J]. International Journal of Computer Vision，1988：321－331.

[3] MANJUNATH B S，CHELLAPPA R. Unsupervised texture segmentation using Markov random field models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，1991：478－482.

[4] 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，2017：2481－95.

[5] LONG，JONATHAN，SHELHAMER，et al. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence，2017：3431－40.

[6] RONNEBERGER O，FISCHER P，BROX T. U-Net：Convolutional networks for biomedical image segmentation [C] //Lecture Notes in Computer Science，Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015，2015：234－241.

[7] 杨瑞君，陈丽叶，程　燕. 基于多尺度边缘感知和增强的息肉图像分割 [J]. 计算机工程与应用，2025，61 (1)：272－281.

[8] CHEN L C，PAPANDREOU G，SCHROFF F，et al. Rethinking atrous convolution for semantic image segmentation [J]. ArXiv：Computer Vision and Pattern Recognition，2017.

[9] ZHAO H，SHI J，QI X，et al. Pyramid scene parsing network [C] //2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)，Honolulu，HI. 2017：1－6.

[10] ZHU Y，SAPRA K，REDA F A，et al. Improving semantic segmentation via video propagation and label relaxation [C] //2019 IEEE/CVF Conference on Computer Vision and Pattern Rec-ognition (CVPR)，Long Beach，CA，USA. 2019.

[11] LIN X，SUN S，HUANG W，et al. EAPT：Efficient attention pyramid transformer for image processing [J]. IEEE Transactions on Multimedia，2021，25：50－61.

[12] WANG J，SUN K，CHENG T，et al. Deep high-resolution representation learning for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2020，43 (10)：3349－64.

[13] FU J，LIU J，TIAN H，et al. Dual attention network for scene segmentation [C] //2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)，Long Beach，CA，USA. 2019.

[14] NAZIR A，CHEEMA M N，SHENG B，et al. OFF-eNET：An optimally fused fully end-to-end network for automatic dense volumetric 3D intracranial blood vessels segmentation [J]. IEEE Transactions on Image Processing，2020，29：7192－202.

[15] YU F，KOLTUN V. Multi-scale context aggregation by dilated convolutions [J]. Arxiv Preprint Arxiv：151107-122，2015.

[16] CHEN LC，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)：834－48.

[17] CHEN L C，ZHU Y，PAPANDREOU G，et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [M/OL] //Computer Vision-ECCV 2018，Lecture Notes in Computer Science，2018：833－851.

[18] LI X，YOU A，ZHU Z，et al. Semantic flow for fast and accurate scene parsing [C] //Proceedings of the Computer Vision-ECCV 2020：16th European Conference，Glasgow，UK，Proceedings，Part I 16，F，2020.

[19] HUANG Z，WEI Y，WANG X，et al. AlignSeg：feature-aligned segmentation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2021：1－6.

[20] DAI J，QI H，XIONG Y，et al. Deformable convolutional networks [C] //2017 IEEE International Conference on Computer Vision (ICCV)，Venice，2017：1－6.

[21] ZHAO H，QI X，SHEN X，et al. ICNet for real-time semantic segmentation on high-resolution images [M]. Computer Vision-ECCV 2018，Lecture Notes in Computer Science，2018：418－34.

[22] PASZKE A，CHAURASIA A，KIM S，et al. ENet：a deep neural network architecture for real-time semantic segmentation [J]. ArXiv Preprint ArXiv：1606.02147，

2016.

[23] YU C，WANG J，PENG C，et al. BiSeNet：bilateral segmentation network for real-time semantic segmentation [C] //Computer Vision-ECCV 2018，Lecture Notes in Computer Science. 2018：334－349.

[24] FAN M，LAI S，HUANG J，et al. Rethinking BiSeNet for real-time semantic segmentation [C] //2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)，Nashville，TN，USA. 2021.

[25] 刘　波，王婷婷，刘　杰. 基于改进 BiSeNet V2 的手机盖板缺陷检测方法 [J]. 光学学报，2024，44 (16)：1－12.

[26] 张　凡，侯惠芳，张自豪，等. 医学图像分割中的双分支特征提取器及高效特征融合方法 [J/OL]. 计算机工程与应用，1－14 [2024－12－19]. http://kns.cnki.net/kcms/detail/11.2127.TP.20240819.1201.020.html.

[27] POUDEL R K，BONDE U，LIWICKI S，et al. ContextNet：exploring context and detail for semantic segmentation in real-time [J]. Arxiv：Computer Vision and Pattern Recognition，arXiv：Computer Vision and Pattern Recognition，2018.

[28] POUDEL R P，LIWICKI S，CIPOLLA R. Fast-scnn：fast semantic segmentation network [J]. Arxiv Preprint Arxiv：190204502，2019.

[29] YU C，GAO C，WANG J，et al. Bisenet v2：Bilateral network with guided aggregation for real-time semantic segmentation [J]. International Journal of Computer Vision，2021，129：3051－68.

[30] HONG Y，PAN H，SUN W，et al. Deep dual-resolution networks for real-time and accurate semantic segmentation of road scenes [J]. Arxiv Preprint ArXiv：210-106085，2021.

[31] WANG Y，SONG J，WANG S，et al. Surface-Framework structure：a neural network structure for weakening gridding effect in PCB mark-point semantic segmentation [J]. Plos One，2023，18 (7)：e0283809.

[32] XIAO P，YAN S，LONG J，et al. An adaptive coarse-to-fine framework for automatic first article inspection of flexographic printing labels [J]. Expert Systems with Applications，2023，227：120241.

[33] WANG H，LI X，HUO L，et al. Global and Edge Enhanced Transformer for Semantic Segmentation of Remote Sensing [J]. Applied Intelligence，2024：1－16.

[34] CAO H，LIU H，SONG E，et al. Dual-branch residual network for lung nodule segmentation [J]. Cornell University-arXiv，Cornell University-arXiv，2019.

[35] XU J，XIONG Z，BHATTACHARYYA S P. PIDNet：a real-time semantic segmentation network inspired by PID controllers [C] //Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2023.

[36] NING Z，ZHONG S，FENG Q，et al. SMU-Net：saliency-guided morphology-aware U-net for breast lesion segmentation in ultrasound image [J]. IEEE Transactions on Medical Imaging，2022：476－490.

[37] 吴英豪，杜晓刚，雷　涛，等. 基于形状特征引导的息肉分割网络 [J]. 计算机技术与发展，2024，34 (5)：52－59.

[38] CALIVA F，IRIONDO C，MARTINEZ A，et al. Distance map loss penalty term for semantic segmentation [J]. CoRR，2019，abs/1908.03679.

[39] KERVADEC H，BOUCHTIBA J，DESROSIERS C，et al. Boundary loss for highly unbalanced segmentation [J]. Medical Image Analysis，2021：101851.

[40] LIN T Y，DOLLAR P，GIRSHICK R，et al. Feature pyramid networks for object detection [C] //2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)，Honolulu，H-I. 2017.

[41] HU J，SH-EN L，ALBANIE S，et al. Squeeze-and-Excitation Networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2020：2011－23.

[42] PARK J，WOO S，LEE J Y，et al. BAM：Bottleneck Attention Module [J]. Arxiv：Computer Vision and Pattern Recognition，Arxiv：Computer Vision and Pattern Recognition，2018.

[43] BROSTOW G J，SHOTTON J，FAUQUEUR J，et al. Segmentation and recognition using structure from motion point clouds [M] //Lecture Notes in Computer Science，Computer Vision-ECCV 2008，2008：44－57.

[44] KRIZHEVSKY A，SUTSKEVER I，HINTON G E. Imagenet classification with deep convolutional neural networks [J]. Advances in Neural Information Processing Systems，2012：25.

[45] GOYAL P，DOLLáR P，GIRSHICK R，et al. Accurate，large minibatch sgd：Training imagenet in 1 hour [J]. Arxiv Preprint Arxiv：170602677，2017.

[46] SHRIVASTAVA A，GUPTA A，GIRSHICK R. Training region-based object detectors with online hard example mining [C] //2016 IEEE Conference on Computer Vision and Pattern Rec-ognition (CVPR)，Las Vegas，NV，USA. 2016.

[47] PASZKE A，GROSS S，MASSA F，et al. Pytorch：an imperative style，high-performance deep learning library [J]. Advances in neural Information Processing Systems，2019：32.