doi:10.3969/j.issn.1673−9833.2025.05.009

TF-ME：多尺度特征增强的透明物体分割网络

郭　扬，邓晓军，肖世康，孙元昊

（湖南工业大学 计算机学院，湖南 株洲 412007）

摘　要：针对透明物体会继承来自背景的信息且传统卷积神经网络中感受野的限制等问题，提出了基于 Transformer 和多尺度特征增强的透明物体分割网络 TF-ME。模型采用 CNN 结合 Transformer 的混合结构，在特征提取阶段，设计了多尺度特征融合模块，有效整合全局与局部信息，提升了模型对不同尺寸透明物体的分割效果；此外，对前馈神经网络进行了重新设计，增强了 Transformer 编码器的上下文理解能力。为验证所提算法的有效性，在 Trans10K-v2 数据集上进行了对比实验。实验结果表明，所提方法在 11 种透明物体分割中的 ACC 和 MIoU 值分别达到了 94.68% 和 73.39%，相较于其他算法，该模型的性能明显提升。

关键词：透明物体；语义分割；Transformer；前馈神经网络；特征融合

中图分类号：TP391.4　　**文献标志码**：A　　**文章编号**：1673−9833(2025)05−0058−09

引文格式：郭　扬，邓晓军，肖世康，等 . TF-ME：多尺度特征增强的透明物体分割网络 [J]. 湖南工业大学学报，2025，39(5)：58−66.

TF-ME: Transparent Object Segmentation Network with Multi-Scale Feature Enhancement

GUO Yang，DENG Xiaojun，XIAO Shikang，SUN Yuanhao

（School of Computer，Hunan University of Technology，Zhuzhou Hunan 412007，China）

Abstract：In view of the transparent objects inheriting information from the background and the limitation of receptive fields in traditional convolutional neural networks, a transparent object segmentation network TF-ME has been proposed based on Transformer and multi-scale feature enhancement. The model adopts a hybrid structure of CNN and Transformer. In the feature extraction stage, a multi-scale feature fusion module is designed to effectively integrate global and local information, thus improving the segmentation effect of the model on transparent objects of different sizes. In addition, the feedforward neural network is redesigned for an enhancement of the context understanding ability of the Transformer encoder, followed by comparative experiments conducted on the Trans10K-v2 dataset for a verification of the effectiveness of the proposed algorithm. The experimental results show that the proposed method achieves 94.68% ACC and 73.39% MIoU in 11 types of transparent object segmentation, respectively. Compared with other algorithms, the performance of the proposed model has been significantly improved.

Keywords：transparent object；semantic segmentation；Transformer；feedforward neural network；feature fusion

收稿日期：2024−07−14

基金项目：湖南省自然科学基金资助项目（2024JJ7148）

作者简介：郭　扬，男，湖南工业大学硕士生，主要研究方向为计算机视觉，E-mail：2456573678@qq.com

通信作者：邓晓军，男，湖南工业大学教授，硕士生导师，主要研究方向为智能信息处理和图像处理，E-mail：little_army@hut.edu.cn

0　引言

透明物体分割是语义分割领域的一项重要任务。不同于常规物体分割，透明物体的边界模糊不清，常与复杂背景相互交叠，因此透明物体分割是语义分割领域的难点。透明物体分割应用得十分广泛，如在工业质检领域，对生产的玻璃制品进行缺陷检测；在智能机器人领域，家庭服务机器人需要抓取透明杯子、透明瓶子。精确分割透明物体是缺陷检测和抓取的前提，因此透明物体分割也是语义分割领域的重点。

国内外学者在精确分割透明物体方面进行了许多探索。A. Kalra 等 [1] 将透明物体分割问题转化为光偏振问题，利用偏振相机捕捉光波的旋转，从而确定透明物体的边缘。TOM-Net[2] 将透明物体抠图转化为折射光路预测问题，该方法在合成数据集上取得了较好的效果，但缺乏对真实数据集的评估。TransLab[3] 增加了对透明物体边界信息的关注，提升了对透明物体的分割精度。MFENet[4] 通过融合不同尺度的特征扩大了模型的感受野，但由于 CNN 架构的限制对透明物体分割性能提升有限。随着 Transformer[5] 在计算机视觉领域的成功应用，许多学者也将其应用到透明物体分割领域。Trans2seg[6] 使用 Transformer 作为分割模型的编码器，通过长距离的建模机制获取了细节信息，提供了全局感受野，进一步提升了对透明物体的分割精度。CTNet[7] 使用 CNN 提取透明物体的 RGB 图像和 TIR 图像特征，并使用 Transformer 将这两种特征融合，从而实现了多模态透明物体分割。这些方法在透明物体分割任务上的精度虽然有一定的提升，但并没有较好地关注透明物体的多尺度信息，导致针对复杂背景下的透明物体分割存在不足。

针对透明物体的特点，提出基于 Transformer 和多尺度特征增强的透明物体分割网络（Transformer and multi scale feature enhancement network，TF-ME）。主要工作包括重新设计了 Transformer 编码器中的前馈神经网络层，引入局部特征关注模块（local feature attention module，LFA）提高了编码器的上下文理解能力。此外，设计了多尺度特征融合模块（multi scale feature fusion module，MFF），该模块由全局特征融合模块（global feature fusion module，GFF）和全局特征分发模块（global feature distribution module，GFD）构成。GFF 模块能高效聚合 CNN 提取的不同尺度特征，GFD 模块提取全局特征中的关键信息，并将其注入局部特征，提高了模型对于不同大小透明物体的分割效果。

1　算法介绍

算法的网络结构如图 1 所示，主要包括骨干网络模块、GFF 模块、GFD 模块、编码器模块和解码器模块。首先，通过骨干网络模块提取图像不同尺度的特征；其次，将不同尺度的特征通过 GFF 模块进行融合得到全局特征；随后，将全局特征输入 GFD 模块去冗余，提取全局特征中的关键信息；最后，将全局特征中的关键信息注入骨干网络提取出的第一层和第四层特征。图像特征在送入编码器之前需要展平为二维特征序列，并在二维序列添加位置嵌入，经过多个自注意力层和前馈神经网络层后，得到编码特征映射。然后，将一组可学习的类原型嵌入和特征编码映射送入解码器获取维度为 $(N, M, H/16, W/16)$ 的注意力特征图，其中 N 为类别数目，M 为多头注意力机制的头数，H 为图像的高度，W 为图像的宽度；最后，将注意力特征图进行 4 倍上采样与骨干网络提取出来的浅层特征进行融合，得到最终的分割结果。

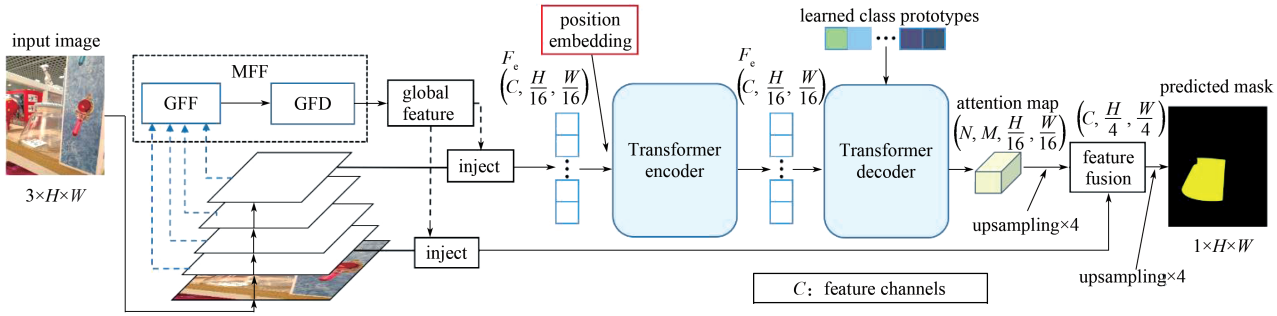


图 1　网络结构示意图

Fig. 1　Network structure diagram

1.1　骨干网络模块

由于透明物体自身纹理特征较弱，极易受到背景物体的干扰。因此，课题组选择具有强大特征提取能力的 ResNet 系列 [8] 作为骨干网络，网络的前三层分别进行 4 倍、8 倍、16 倍的下采样。在语义分割中，下采样倍率过大将难以还原图像，从而导致分割的结果不理想，所以第四层并没有进行下采样。

1.2 多尺度特征融合模块

在计算机视觉领域，特征融合能够有效提升模型性能。MCFA-UNet[9] 使用双重注意实现不同路径的多尺度特征信息融合，可减少编、解码路径之间的语义差异。JD 机制 [10] 通过不同层次的特征多次融合注入以增强模型的特征融合能力，但多次融合分发增加了模型的复杂度。文献 [11] 将不同尺度特征聚合，获取的特征更全面，细化了语义分割结果。受此启发，课题组设计了全局特征融合模块，其工作流程如图 2 所示。对于骨干网络模块输出的不同尺度的特征图，使用均值池化下采样法调整至下采样的 16 倍，随后将调整后的特征进行拼接得到全局特征，在获取全局特征的同时兼顾计算复杂度进行高效聚合。

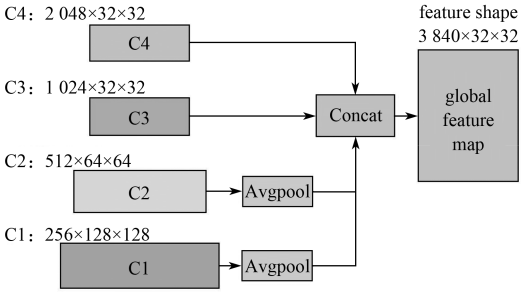
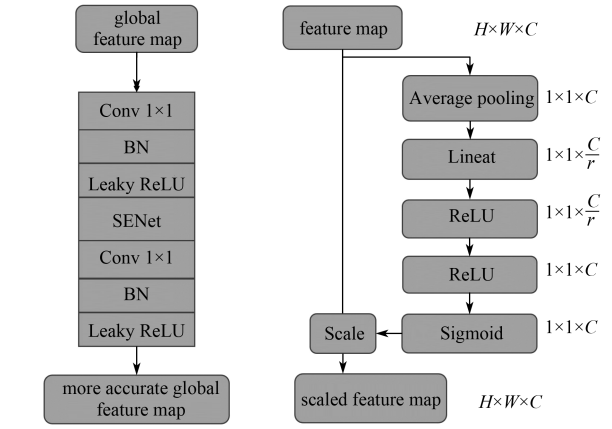


图 2 全局特征融合模块示意图

Fig. 2 Schematic diagram of the global feature fusion module

全局特征中包含浅层特征的细粒度信息和深层特征的语义信息，直接注入局部特征会产生大量的冗余特征，从而导致模型精度下降。图 3 为全局特征分发模块示意图。图 3 中，C 为 1 × 1 卷积调整后的通道数，r 为缩放因子。



a）全局特征分发模块结构 b）SE-Net 模块结构

图 3 全局特征分发模块示意图

Fig. 3 Schematic diagram of the global feature distribution module

全局特征分发模块结构如图 3a 所示，由两个 1 × 1 卷积和通道注意力模块 SE-Net[12] 构成，每个卷积层后有一个 BN 层和激活函数 Leaky ReLU[13]。SE-Net 采用通道压缩激励机制，对于通道之间的相关信息建模，能够有效强化重要特征，抑制无用、冗余特征。SE-Net 模块的结构如图 3b 所示。

GFD 模块工作流程分为 3 步骤：

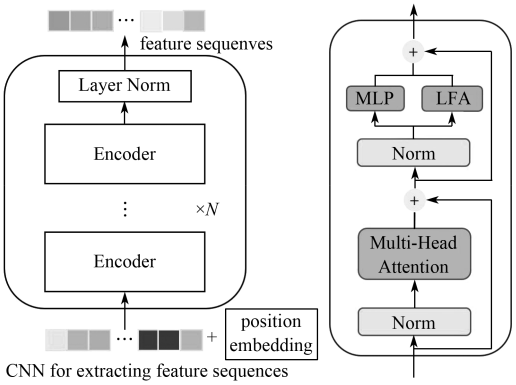
1）全局特征图由 1 × 1 卷积进行通道调整，以减少计算复杂度。

2）调整通道后输出的特征图输入 SE-Net 模块，SE-Net 采用通道压缩激励机制，达到去除冗余特征的目的。

3）将准确的全局特征图通过 1 × 1 卷积调整至适合注入局部特征图的通道数目。

1.3 编码器模块

编码器模块的结构如图 4a 所示，编码器模块由多个编码层堆叠而成。编码层的结构如图 4b 所示，编码层由多头自注意力模块、层归一化和前馈神经网络层组成。前馈神经网络层由多层线性感知器模块（multi-layer perceptron，MLP）和 LFA 模块构成。



a）编码器模块结构 b）编码层结构

图 4 编码器模块示意图

Fig. 4 Encoder module schematic diagram

编码器工作流程分为 3 个步骤：

1）将卷积神经网络提取出来的特征展平为二维序列，并在二维序列添加位置嵌入，补充空间信息的缺失。

2）特征序列输入多头注意力模块，再输入前馈神经网络层进行处理。

3）通过残差连接和归一化操作的输出被传递给下一个子层级或作为最终的编码器输出。

多头注意力机制能够并行地处理输入序列的不同部分，并且能够在不同位置之间建立关联。公式表示如下：

$$Attention\_out=Attention(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V})。\quad (1)$$

式中 $\boldsymbol{Q}$、$\boldsymbol{K}$ 和 $\boldsymbol{V}$ 分别为查询、键和值，由输入的特征序列与之对应的权值矩阵 $\boldsymbol{W}^{Q}$、$\boldsymbol{W}^{K}$、$\boldsymbol{W}^{V}$ 的乘积得到。

注意力计算过程：首先，计算查询和键之间的点积，然后将结果除以缩放因子；再经过 *SoftMax* 函数得到注意力权重，注意力权重与数值向量相乘得到每

个头部的注意力输出，注意力计算公式如下：

$$Attention(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = SoftMax\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V}。\quad（2）$$

式中：d_k 为键维度，除以$\sqrt{d_k}$是为了避免输入过大，避免 *SoftMax* 函数的值过大和过小，从而解决梯度消失问题，并实现归一化的效果。

Transformer 的优势在于能够建立图像不同区域之间的联系，从而给模型提供全局感受野。但标准编码器将特征展平为二维序列进行计算，导致忽略了空间维度上的局部特征。文献 [14] 中提出 Transformer 中标准的编码器结构利用本地上下文的能力有限；文献 [15] 在编码器中的前馈神经网络层引入深度可分卷积，减少了编码器的计算复杂度；文献 [16] 在多头自注意力机制旁并行多尺度卷积，使得捕捉到的信息更加全面。综合上述研究，本文对前馈神经网络层进行了重新设计，其结构如图 5 所示。

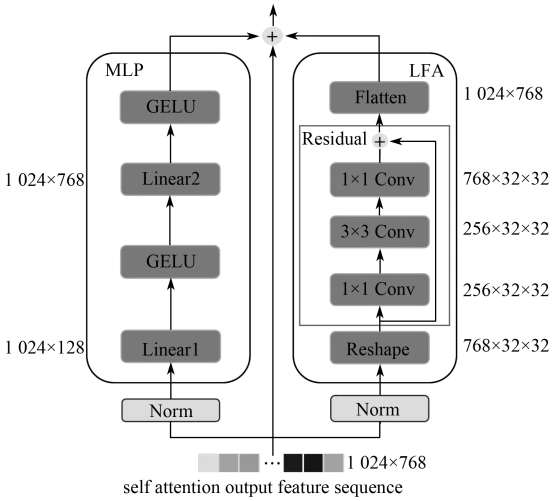


图 5　前馈神经网络层示意图

Fig. 5　Schematic diagram of the feedforward neural network layer

前馈神经网络层由 MLP 和 LFA 模块构成。LFA 模块由 Reshape 操作、Flatten 操作和残差模块构成。前馈神经网络层将 LFA 模块输出的带有空间局部信息的特征序列与 MLP 的输出特征序列相结合，提高了编码器上下文的理解能力。LFA 模块工作流程分为 3 个步骤：

1）将多头自注意力模块输出的特征序列进行归一化，并使用 Reshape 操作增加特征维度。

2）将增加维度后的特征送入残差模块，残差模块中 1 × 1 卷积负责调整通道数，3 × 3 卷积捕获空间维度的局部信息。

3）展平收缩通道，以匹配输入通道尺寸。

1.4　解码器模块

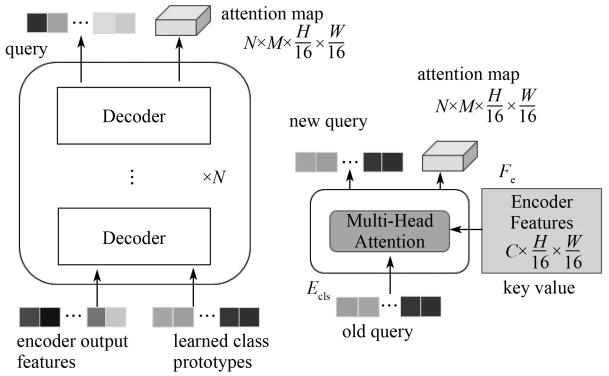
解码器模块的结构如图 6a 所示，解码器模块由多个解码层堆叠而成。解码层由多头注意力模块、层归一化和多层线性感知机模块构成。解码层的输入和输出信息如图 6b 所示，将一组可学习的类原型嵌入（E_{cls}）作为查询、编码器输出的编码特征（F_{e}）作为键和值输入解码层，解码层输出新的 E_{cls} 和注意力特征图。E_{cls} 通过编码器输出的特征和自身进行迭代学习，E^{n}_{cls}更新规则如下：

$$E^{n+1}_{\mathrm{cls}} = SoftMax(E^{n}_{\mathrm{cls}}, F_{\mathrm{e}})\cdot F_{\mathrm{e}}。\quad（3）$$

式中 n 为迭代次数。

每经过一个解码层就需要进行一次迭代，提供新的可学习的类原型嵌入。

解码器模块输出的最后一层注意力特征图上采样到（N, M, H/4, W/4）与骨干网络模块提取的 4 倍下采样的特征图融合为（N, M+C, H/4, W/4），使用卷积等操作将特征图转化为（N, H/4, W/4）。最后，对特征图的像素点逐一进行 Argmax 操作获得最终的分割结果。



a）解码器模块结构　　b）解码层输入和输出信息

图 6　解码器模块结构示意图

Fig. 6　Decoder module schematic diagram

2　实验

2.1　实验环境

本文实验平台的操作系统为 Ubuntu，硬件环境为 CPU 42 核，内存 82 GB，8 块显存 24 GB 的 GPU，软件环境为 Python 3.8、Pytorch 1.7.1、OpenCV-4.7.0。

2.2　数据集与预处理

本文选择的数据集是 Trans10K-v2[6]，共有 11 个类别的透明物体，包含 10 428 幅图像，训练集、验证集和测试集分别为 5 000, 1 000, 4 428 幅。数据集中的图像具有丰富的遮挡、空间尺度、透视畸变等特点。图 7 展示 Trans10K-v2 数据集中部分图像。Trans10K-v2 数据集中图像的分辨率范围为 [850, 1 200] × [1 100,1 500]，为了便于模型的特征提取和训练中并行化计算，在训练和推理过程中使用随机缩放和裁剪等方式将图像统一调整至 512 × 512。

图 7 Trans10K-v2 部分图像

Fig. 7 Trans10K-v2 partial images

2.3 实验评估标准

评价指标选取了在语义分割领域中广泛使用的 3 个指标，即像素精度（pixel accuracy，ACC）、平均交并比（mean intersection over union，MIoU）和类别交并比（category intersection over union，category IoU），以对透明物体分割的模型性能进行基准测试。

2.4 模型训练和参数设置

课题组选择 ResNet-50 作为特征提取网络，为了提高训练效率，使用在 MS-COCO[17]（Microsoft Common Objects in Context）数据集上的训练权重初始化网络参数。每块 GPU 批处理大小设置为 4，优化器采用 Adam，动量设置成 0.9，权重衰减设为 0.000 1。采用多项式学习率衰减策略，训练初期以较大的学习率使模型快速收敛，逐步减小的学习率有助于提高训练后期模型的稳定性。初始学习率为 10^{-4}，多项式的幂为 0.9，总共迭代 50 轮。

Transformer 模型的尺寸主要由嵌入维度（embedding dim）、编解码器层数（layer depth）和多层线性感知机比率（MLP ratio）3 个超参数决定。为了探讨模型的尺寸对于分割性能的影响，本文设计了 3 种尺寸的模型。模型尺寸对于分割性能影响的实验结果如表 1 所示，随着 embedding dim、layer depth 和 MLP ratio 的增加，模型参数量逐步增大，MIoU 先增大后减小，对于透明物体分割并不是增加模型的尺寸就能提升算法的精度。多头自注意力机制头部数目也是十分重要的超参数，自注意力机制通过并行处理多个独立的注意力头，能够捕捉更加丰富的信息。针对头部数的实验结果如表 2 所示，模型的 embedding dim 设置为 256，layer depth 设置为 4，MLP ratio 设置为 3，随着头部数增多，GFlops 逐渐增大，MIoU 随之先增大后减小，并不是说多头自注意力机制中头部数目越多，模型表达能力越强、分割精度越高。因此，本文在实验中多头注意力机制的头数设置为 8，嵌入维度设置为 256，编解码器层数均设置为 4，多层线性感知机的比率设置为 3。

表 1 不同尺寸模型的实验结果

Table 1 Experiments results on models of different sizes

ID	embedding	depth	MLP ratio	MParames	MIoU/%
1	128	2	2	38.70	69.23
2	256	4	3	57.75	73.39
3	768	8	4	228.36	72.51

表 2 多头自注意力机制头部数实验结果

Table 2 Multiple-head self-attention mechanism head count experiment results

ID	num heads	model-scale	GFlops	MIoU/%
1	4	256-4-3	49.60	72.94
2	8	256-4-3	50.81	73.39
3	16	256-4-3	53.24	73.12

2.5 对比的语义分割算法

为了评估本文算法性能，将在 Trans10K-v2 数据集上与下列算法进行对比实验。FCN[18] 是现代语义分割算法的基础；U-Net[19] 采用编码器－解码器结构和跳跃连接有效地处理医学图像分割；TransLab 针对透明物体增加了边界关注；DeepLabv3+[20] 不仅结合了空洞卷积和多尺度融合等技术提升语义分割的精度，还采用了深度可分卷积提高计算效率；Trans2Seg 是基于 Transformer 的透明物体分割算法，由于增加了全局感受野，因此获得了更高质量的细节。

2.6 对比实验结果

实验结果如表 3 所示。

表 3 Trans10K-v2 数据集上本文算法对比其他分割算法实验结果

Table 3 Experimental results with the proposed algorithm and other segmentation algorithms compared on the Trans10K-v2 dataset

method	ACC/%	MIoU/%	category IOU/%											
			bg	shelf	jar	freezer	window	door	eyeglass	cup	wall	bowl	bottle	box
FCN[18]	91.65	62.75	93.62	38.84	56.05	58.76	46.91	50.74	82.56	78.71	68.78	57.87	73.66	46.54
U-Net[19]	81.90	29.23	86.34	8.76	15.18	19.02	27.13	24.73	17.26	53.40	47.36	11.97	37.79	1.77
Translab[3]	92.67	69.00	93.90	54.36	64.48	65.14	54.58	57.72	79.85	81.61	72.82	69.63	77.5	56.43
DeepLabv3+[20]	92.75	68.87	93.82	51.29	64.65	65.71	55.26	57.19	77.06	81.89	72.64	70.81	77.44	58.63
Trans2Seg[6]	94.14	72.15	95.35	53.43	67.82	64.20	59.64	60.56	88.52	86.67	75.99	73.98	82.43	57.17
ours	94.68	73.39	96.05	50.98	69.28	67.36	61.51	63.39	88.86	88.58	77.64	74.98	82.04	59.92

由表 3 可知，本文算法在 Trans10K-v2 数据集的实验结果最好，其 ACC 和 MIoU 分别达到了 94.68% 和 73.39%，相较于性能较好的 Trans2seg 算法，ACC 和 MIoU 分别提升了 0.54% 和 1.24%。

同时，本文算法对比 Trans2seg 整体的分割效果都有较好的性能提升。对于小件透明物体，如杯子、碗和罐子，分别提升了 1.91%, 1%, 1.46%；对于大件透明物体，如冰柜、窗户、门、墙和盒子，分别提升了 3.16%, 1.87%, 2.83%, 1.65%, 2.75%。

2.7　消融实验

本文的算法是以 Trans2Seg 为基础，增加了 MFF 模块，使输入编解码器的特征更加准确。并且改进了编码器中前馈神经网络层，增加了 LFA 模块，提高编码器上下文的理解能力。为了验证这些模块的性能，进行了全面的消融实验。具体包括以下两个方面：第一，前馈神经网络中增加局部特征关注模块的对比实验；第二，骨干特征提取网络引入多尺度特征模块的对比试验。表 4 为消融实验结果，图 8 为消融实验可视化训练过程。

表 4　本文算法消融实验结果

Table 4　Experimental results of the proposed algorithm ablation

ID	LFA	MFF	SE-Net	ACC/%	MIoU/%
1	×	×		94.16	72.11
2	√	×		94.23	72.65
3	×	√	×	93.66	69.45
4	×	√	√	94.56	73.15
5	√	√	√	94.68	73.39

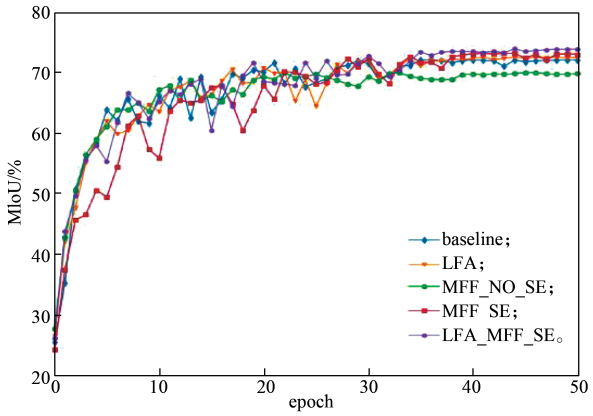


图 8　消融实验可视化训练过程

Fig. 8　Visualization training process for ablation experiments

第一组实验中验证了 Transformer 编码器中有无 LFA 模块对实验结果的影响，实验结果如表 4 所示。本文使用标准的编码器和图 4b 增加了 LFA 模块的编码器进行对比，结果显示有 LFA 模块的编码器结构拥有更加高的分割精度。

在第二组对比实验中验证有无 MFF 模块之间的实验差异，并验证了 MFF 模块中引入注意力机制对冗余信息的抑制效果，实验结果如表 4 所示。特征提取网络如图 1 左侧所示的网络结构，首先验证了有无多尺度特征融合模块对实验结果的影响，实验结果显示，无 MFF 模块的 MIoU 为 72.11%，增加 MFF 模块后 MIoU 降至 69.45%。在 MFD 模块中引入通道注意力机制 SE-Net 后 MIoU 为 73.15%。实验结果表明，多尺度特征融合有利于提升图像分割的精度，因本文特征融合采用局部特征拼接的方式，其中包含大量的信息冗余，直接注入导致了模型的分割精度下降，所以需要引入通道注意力机制 SE-Net 对冗余特征进行抑制。

2.8　可视化实验结果展示和分析

Trans10K-v2 数据集总共有 11 类透明物体，按体积大小可以分为两大类。本文将对这两类物体的分割结果进行可视化效果展示。为了更加直观地评估本文算法的性能，本文将与其他算法进行可视化结果对比分析。

图 9 所示为采用本文算法对小件透明物体的分割结果展示图。

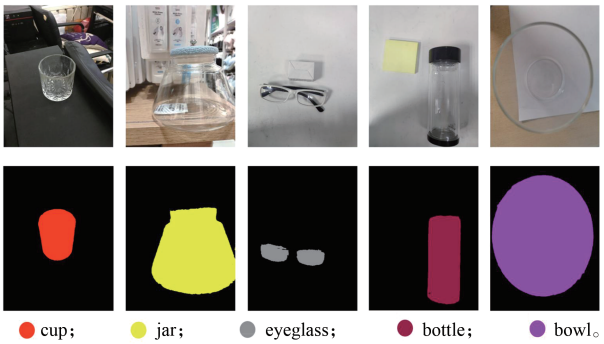


图 9　小件透明物体分割结果展示图

Fig. 9　Display of segmentation results for small transparent objects

这类透明物体的分割应用十分广泛。例如，智能家庭机器人在工作中可能涉及抓取透明物体；工业生产中对杯子、瓶子等透明物体进行缺陷检测。从图 9 所示结果来看，本文算法能够有效地分割这类透明物体，并且物体的边缘十分精细，这对于实际应用非常有意义。小件的透明物体在生活中通常都是多个摆放在一起，这给本文算法带来了巨大的挑战。

图 10 为本文算法对于多个透明物体分割结果展示图。如图 10 所示，对于多个同一类别的物品，分割结果较为精准。但在不同类别透明物体重叠部分的轮廓判断和在光线昏暗的情况下的分割存在不足，待后续改进。

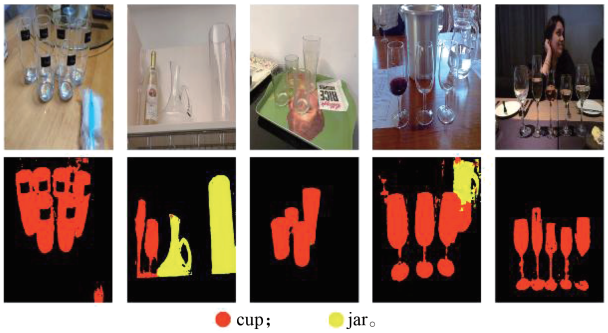


图 10 多个透明物体分割结果展示图

Fig. 10 Display of segmentation results for multiple transparent objects

图 11 为本文算法对大件透明物体的分割结果展示图。这类物品的共同特点就是它们的背景信息特别丰富，因此分割具有一定的难度。智能服务机器人分割出玻璃门、柜子和窗户，才能在工作中避开这类透明物体，精确分割这类物体具有一定的实际意义。

图 11 大件透明物体分割结果展示图

Fig. 11 Display of segmentation results for large transparent objects

本文算法相较于 Trans2seg 整体的分割效果有较好的性能提升。尤其是对杯子、冰柜、门和盒子，附图 1 展示了这几类物品分割对比图。从图中可以看出本文算法在透明物体分割的边缘更加精细。本文提出的多尺度特征增强方法通过全局特征注入局部的方式有效地提升了模型的理解能力，相较于 Trans2seg 对冰柜、门和盒子这类体积较大并且拥有复杂背景的透明物体，分割结果更加精确。

附图 2 是本文方法与其他方法的可视化实验结果对比图。在附图 2a 中，基于 CNN 和基于 Transformer 的分割方法都能有效地分割单一的小件透明物体，但基于 Transformer 方法分割结果的边缘更加精细。一旦涉及多个或者背景复杂的透明物体，基于 Transformer 方法的优势就体现出来了。如附图 2b 和图 2c，多个透明物体摆放在一起时，基于 CNN 的方法将酒杯错误地归类成罐子，基于 Transformer 的分割算法并没有出现这一问题。附图 2b 中 Trans2seg 虽然分隔出杯子和瓶子的轮廓，但将重叠区域归类为罐子，而本文算法在编码器中引入了 LFA 模块，增加了空间局部信息的关注，并没有出现这一问题。附图 2d 和附图 2e 中透明物体的背景信息十分丰富，基于 CNN 分割方法受到的干扰较大，而本文算法引入 MFF 模块，可全面、准确地将全局信息和局部信息融合，分割结果较其他算法更准确。通过上述结果可以发现，在透明物体分割领域，基于 Transformer 的算法比基于 CNN 的算法更适用，分割结果的边缘更加精细，像素类别判定得更加准确。本文方法在 Transformer 模型中引入的局部特征关注模块和多尺度特征增强方法，能更好地理解透明物体与其背景之间的关系，使模型在复杂背景下对透明物体的分割结果更加精准。

3 结语

TF-ME 基于 Transformer 和多尺度特征增强的透明物体分割方法，主要对图像特征提取和编解码器进行了优化。设计了多尺度特征融合模块，充分融合全局特征和局部特征，增加模型对图像全局的理解，提升了不同尺寸的透明物体分割结果。在 Transformer 编码器中引入了局部特征关注模块，克服了标准的前馈神经网络利用本地上下文能力不足的问题。实验结果表明，本文算法能更好地理解透明物体与背景之间的关系，有效提升了透明物体分割的准确率和鲁棒性。本文算法对于复杂背景下的透明物体能够精确地分割，具有较好的应用前景。但在光线昏暗下和不同类别的透明物体重叠区域分割仍有不足。未来可以考虑引入透明物体边界信息增加约束条件，提升算法性能来改善不同类别透明物体重叠情况下的分割，以满足不同场景下对于透明物体分割的需求。

参考文献：

[1] KALRA A，TAAMAZYAN V，RAO S K，et al. Deep Polarization Cues for Transparent Object Segmentation [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle：IEEE，2020：8599−8608.

[2] CHEN G Y，HAN K，WONG K Y K. TOM-Net：Learning Transparent Object Matting from a Single Image[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City：IEEE，2018：9233−9241.

[3] XIE E Z，WANG W J，WANG W H，et al. Segmenting Transparent Objects in the Wild[C]//Computer Vision-ECCV 2020. Cham：Springer International Publishing，2020：696−711.

[4] 张博翔 . 基于深度学习的图像分割若干关键问题研究 [D]. 长春：吉林大学，2023.
ZHANG Boxiang. Research on Some Key Problems of Image Segmentation Based on Deep Learning[D]. Changchun：Jilin University，2023.

[5] VASWANI A，SHAZEER N，PARMAR N，et al. Attention is All You Need[J]. Advances in Neural Information Processing Systems，2017，30：5998−6008.

[6] XIE E Z，WANG W J，WANG W H，et al. Segmenting Transparent Objects in the Wild with Transformer[C]// Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence. California：International Joint Conferences on Artificial Intelligence Organization，2021：1194−1200.

[7] 潘惟兰，张荣芬，刘宇红，等 . 结合 CNN-Transformer 的跨模态透明物体分割 [J]. 计算机工程与应用，2025，61(4)：222−229.
PAN Weilan，ZHANG Rongfen，LIU Yuhong，et al. Cross-Modal Transparent Object Segmentation Combining CNN-Transformer[J]. Computer Engineering and Applications，2025，61(4)：222−229.

[8] HE K M，ZHANG X Y，REN S Q，et al. Deep Residual Learning for Image Recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas：IEEE，2016：770−778.

[9] ZHOU Y R，KONG Q Q，ZHU Y，et al. MCFA-UNet：Multiscale Cascaded Feature Attention U-Net for Liver Segmentation[J]. IRBM，2023，44(4)：100789.

[10] WANG C，HE W，NIE Y，et al. Gold-YOLO：Efficient Object Detector via Gather-and-Distribute Mechanism[J]. Advances in Neural Information Processing Systems，2023，36：51094−51112.

[11] 徐兆忠，彭　力，戴菲菲 . 多尺度特征对齐聚合的语义分割方法 [J]. 激光与光电子学进展，2023，60(2)：0215004.
XU Zhaozhong，PENG Li，DAI Feifei. Semantic Segmentation Method Based on Multiscale Feature Alignment and Aggregation[J]. Laser & Optoelectronics Progress，2023，60(2)：0215004.

[12] HU J，SHEN L，ALBANIE S，et al. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2020，42(8)：2011−2023.

[13] HE K M，ZHANG X，REN S，et al. Delving Deep into Rectifiers：Surpassing Human-Level Performance on Imagenet Classification[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago：IEEE，2015：1026−1034.

[14] YUAN K，GUO S P，LIU Z W，et al. Incorporating Convolution Designs into Visual Transformers[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal：IEEE，2021：559−568.

[15] WANG Z D，CUN X D，BAO J M，et al. Uformer：a General U-Shaped Transformer for Image Restoration[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans：IEEE，2022：17662−17672.

[16] 朱松豪，孙冬轩，宋　杰 . 基于 Transformer 的透明物体图像语义分割 [J]. 南京邮电大学学报 (自然科学版)，2023，43(4)：83−92.
ZHU Songhao，SUN Dongxuan，SONG Jie. Semantic Segmentation of Transparent Objects via Transformer[J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition)，2023，43(4)：83−92.

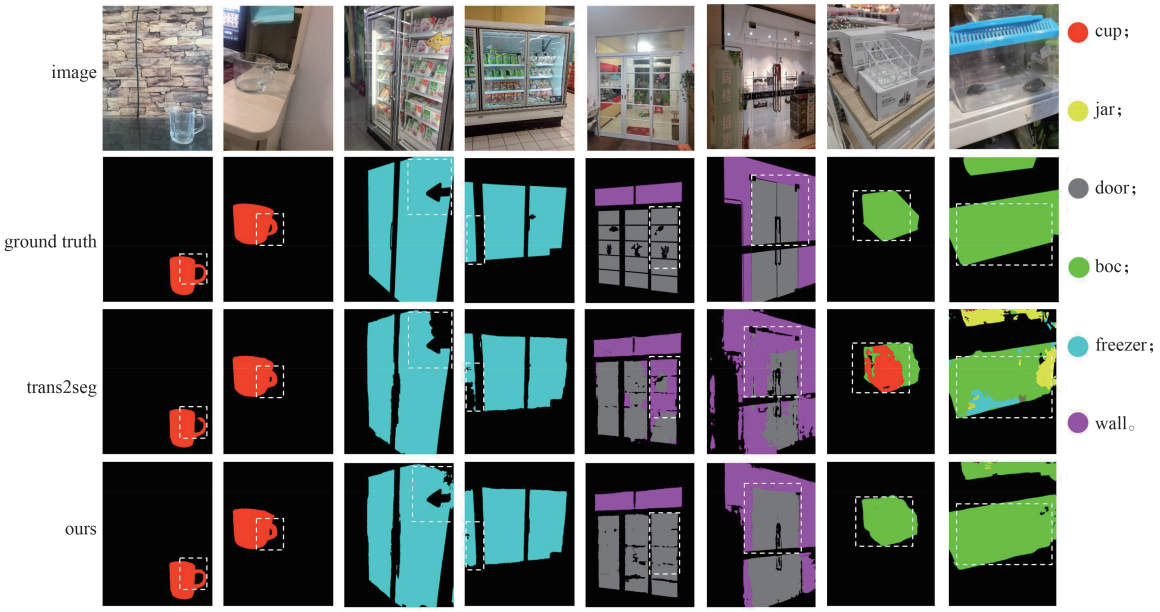
[17] LIN T Y，MAIRE M，BELONGIE S，et al. Microsoft COCO：Common Objects in Context[C]//Compater Vision-ECCV 2014. Cham：Springer International Publishing，2014：740−755.

[18] SHELHAMER E，LONG J，DARRELL T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2017，39(4)：640−651.

[19] RONNEBERGER O，FISCHER P，BROX T. U-Net：Convolutional Networks for Biomedical Image Segmentation[C]//Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Cham：Springer International Publishing，2015：234−241.

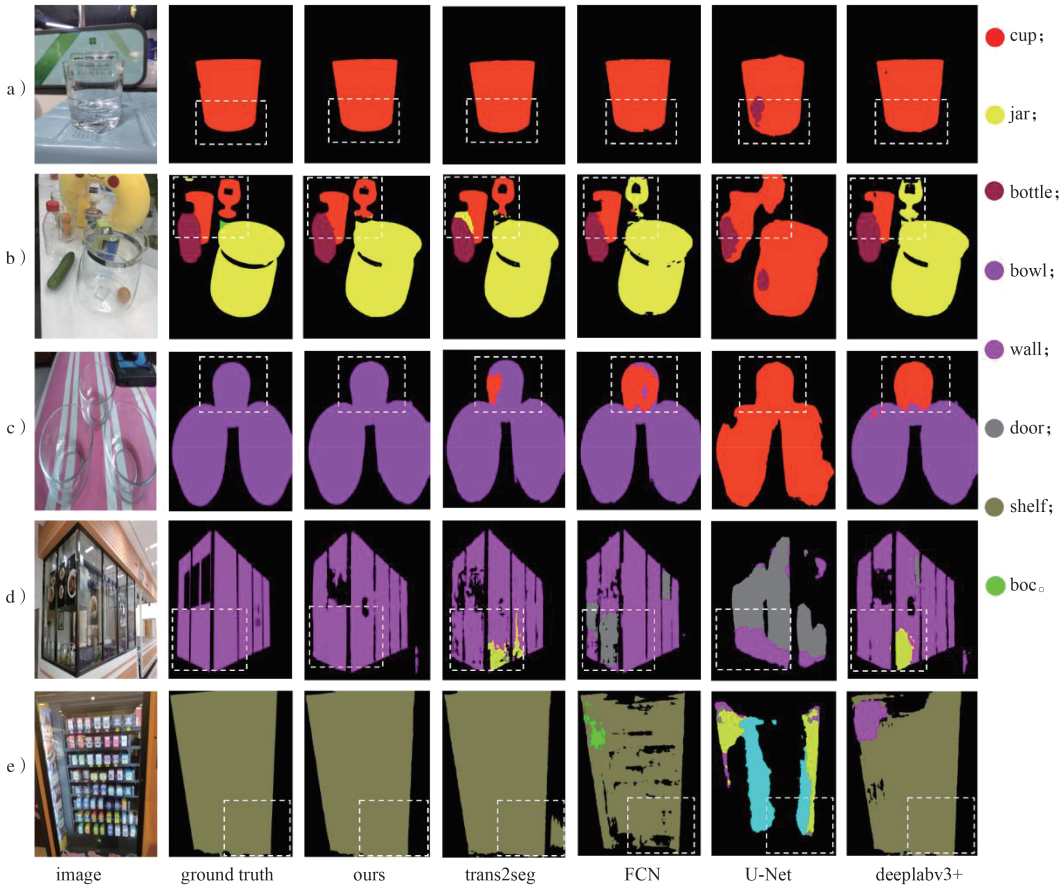
[20] CHEN L C，ZHU Y，PAPANDREOU G，et al. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation[C]//Proceedings of the European Conference on Computer vision (ECCV). Munich：IEEE，2018：801−818.

(责任编辑：申　剑)



附图 1 本文算法与 Trans2seg 分割结果对比图

Fig. 1 Comparison of segmentation results between the proposed algorithm and Trans2seg



附图 2 本文算法与其他算法分割结果对比图

Fig. 2 Comparison of segmentation results between the proposed algorithm and other algorithms