DOI:10.19651/j.cnki.emt.2416330

基于层次特征聚合的自动人像抠图*

汪小梅[1,2] 谭 棉[1,2] 罗太维[1,2] 冯夫健[1,2]
(1.贵州民族大学数据科学与信息工程学院 贵阳 550025; 2.贵州省模式识别与智能系统重点实验室贵州民族大学 贵阳 550025)

摘 要:针对抠图任务中人像毛发这类细微结构局部区域出现的误抠问题,其本质是此区域中混合信息所导致的图像像素点透明度遮罩回归预测不准确问题,提出一种端到端的层次特征聚合抠图网络模型。该模型通过一个共享编码器和两个独立解码器,利用通道和位置注意力机制,以层次特征聚合方式聚合低层次纹理线索和高层次语义信息,能够在没有额外输入的情况下从单个人像的精细边界和自适应语义中感知前景透明度遮罩。基于此,结合交叉熵损失、未知区域的透明度遮罩预测损失和结构性损失,以引导层次特征聚合抠图网络模型完善前景整体结构,恢复毛发纹理细节。为验证所设计模型的有效性,在自建的 MCP-1k 和公开的 P3M-500-NP 数据集上进行验证分析,实验结果表明所提模型在 MSE 和 SAD 指标上的误差分别为 0.007 6,25.59 与 0.007 2,25.52,与其他典型深度抠图模型相比在恢复人像细微毛发和完善人像语义结构方面有较大提升,解决人像毛发区域的误抠问题。

关键词:图像抠图;透明度遮罩;特征融合;层次特征聚合;自动人像抠图
中图分类号:TP391.41 **文献标识码**:A **国家标准学科分类代码**:510.4050

Hierarchical feature aggregation for automatic portrait matting

Wang Xiaomei[1,2] Tan Mian[1,2] Luo Taiwei[1,2] Feng Fujian[1,2]
(1. School of Data Science and Information Engineering, Guizhou Minzu University, Guiyang 550025, China;
2. Key Laboratory of Pattern Recognition and Intelligent System, Guizhou Minzu University, Guiyang 550025, China)

Abstract: Addressing the issue of erroneous extraction of fine structures such as human hair in image matting tasks, the problem essentially stemmed from inaccurate prediction of pixel alpha mattes due to mixed information within these regions. To address this problem, a novel end-to-end hierarchical feature aggregation matting network model is proposed. This model incorporates a shared encoder and two independent decoders, leveraging channel and positional attention mechanisms to aggregate low-level texture clues and high-level semantic information in a hierarchical manner. It enables perceiving foreground transparency masks from fine boundaries of individual portraits and adaptive semantics without additional inputs. To guide the hierarchical feature aggregation matting network model in refining the overall foreground structure and restoring hair texture details, cross-entropy loss, alpha matte prediction loss for unknown regions, and structural losses are integrated. To validate the effectiveness of the proposed model, experiments were conducted on the self-constructed MCP-1k dataset and the publicly available P3M-500-NP dataset. Experimental results demonstrated that the proposed model achieved errors of 0.0076 MSE and 25.59 SAD on MCP-1k dataset, and 0.0072 MSE and 25.52 SAD on P3M-500-NP dataset, respectively. Compared with other typical deep matting models, it showed significant improvements in restoring fine human hair and enhancing semantic structure in portraits, effectively addressing the issue of erroneous extraction in human hair regions.

Keywords: image matting; alpha matte; feature fusion; hierarchical feature aggregation; automatic portrait matting

0 引 言

自然图像抠图旨在从图像中精准提取出感兴趣的前景对象对应的透明度遮罩,其广泛应用于图像合成[1]、前景提取[2]及电影制作[3]等领域。自 Porter 等[4]引入 alpha 通道构造图像前景遮罩提取模型,自然图像抠图问题则定义为

收稿日期:2024-06-28
*基金项目:贵州省科技计划项目(黔科合基础-ZK[2022]一般 195,黔科合基础-ZK[2023]一般 143,黔科合平台人才-ZCKJ[2021]007)、贵州省教育厅自然科学研究项目(黔教技[2023]012 号,黔教技[2022]015 号,黔教技[2023]061 号)、贵州省模式识别与智能系统重点实验室开放课题(GZMUKL[2022]KF01,GZMUKL[2022]KF05)项目资助

病态问题。

为求解该问题,传统抠图方法通过引入三分图作为额外输入,并人为设计采样策略或对图像像素的分布进行统计假设以完成抠图任务。例如,Feng 等[5]提出基于微搜索的前景遮罩提取算法,通过对像素的有效决策子集的搜索代替对决策集的搜索,解决高分辨率图像抠图问题中的前景遮罩抠反问题。Yao 等[6]结合非局部搜索法和抠图拉普拉斯法的特点,提出基于多尺度空间融合的亲和抠图方法,该方法有效避免了由于采样策略不适用导致丢失最优像素对的问题。尽管传统方法在研究像素之间的相关性以及前背景像素对的采样规则方面取得了一定的研究基础,但其主要使用颜色和纹理底层特征,而忽略了具有语义信息的高级特征,很容易在复杂的人像抠图情况下失效。

近年来,基于深度学习的抠图模型将高级语义信息融入抠图任务中,利用其强大的表征能力学习鉴别图像特征,在很大程度上解决了传统抠图方法特征表示有限的问题。这类方法主要分为基于三分图的抠图模型和端到端的自动抠图模型。DIM[7]模型是最早将卷积神经网络应用于抠图的代表性模型,随后又有一系列有价值的工作推进了最先进的抠图性能。例如,Zhou 等[8]提出一种注意力转移网络,同时使用前景对象相关细节和高级语义特征来估计透明度遮罩,并降低计算复杂度。Liu 等[9]设计了一个深度图像抠图模型来增强图像的细粒度细节,解决了图像毛发等细微特征丢失问题。Li 等[10]模拟基于亲和方法的信息流,为图像抠图设计引导上下文注意模块以学习低级亲和力,直接在全局范围内传播高度不透明度信息,进一步提高抠图效果。虽然基于三分图的深度抠图模型在图像毛发和复杂背景等细微特征丢失方面提高了抠图精度,但所设计模型的训练和测试依赖于三分图,这限制了抠图的应用场景,且获得三分图的方式对于一般用户而言是困难的。为解决这类模型面临的困难,结合全局分割和局部抠图的自动抠图模型备受关注。例如 Chen 等[11]提出利用语义信息自动生成三分图,首次实现了通过神经网络预测三分图的抠图算法。此外,Zhang 等[12]研究了一种分割和抠图融合的两阶段网络用以隐式生成三分图,提高抠图精度。Liu 等[13]提出一种利用粗细标注数据混合训练来提高抠图性能的模型。Ke 等[14]通过显式约束同时优化一系列子目标,提出一种同时考虑低分辨率语义和高分辨率细节特征的轻量级网络,解决了自动和快速的人像抠图问题。Qiao 等[15]利用空间注意力和通道注意力来过滤高级语义和外观特征,设计了注意力机制引导的层级结构聚合图像抠图算法,在自动抠图领域中取得突破性的进展。

虽然上述抠图方法在自然图像抠图领域取得显著成就,相对而言,关于人像毛发这类细微结构深度模型的研究相对较少,其面对人像多毛等细微结构时,存在语义分割不准确及人像毛发细节难以恢复的现象,导致在人像毛发局部区域出现误抠问题。这限制了人像抠图领域的高质量发展,因此,在提升人像语义信息和外观细节线索方面需要进一步研究。

本文针对自动人像抠图方法在人像毛发局部区域出现的误抠问题,受目标分解思想的启发,将抠图任务分解为高层次语义信息和低层次纹理线索协同融合的子任务,即语义分割、细节预测和层次特征聚合。提出一种端到端的层次特征聚合抠图(hierarchical feature aggregation for matting,HFAM)网络模型,该模型由一个共享编码器和两个独立的解码器组成,以协同的方式完成抠图任务,然后进行语义信息与细节线索的有效融合。为使语义分割阶段得到准确的先验信息,设计多尺度特征提取模块(multi-scale feature extraction module,MFEM)学习骨干网络不同感受野下的特征,用于提升人像语义特征感知能力,同时利用形态学生成的三分图对全局语义解码器模块生成的全局图进行监督,保证前景目标结构性完整。针对人像毛发结构难以恢复的问题,在细节解码器前利用 BB 模块[16]融合未知区域的细节信息,同时设计位置注意力机制,使其对所提取的低级特征重新加权,避免冗余的背景信息和低级纹理特征对预测的影响。由于语义分割和细节抠图特征之间存在差异,将语义分割和细节预测分支的输出利用融合函数进行融合,得到精细化的透明度遮罩。

1 HFAM 网络模型结构

针对图像中细微结构复杂性和深层次高级特征难以准确捕捉,导致深度自动抠图模型在人像毛发等细微结构处出现误抠,分别设计语义分割分支和细节预测分支,并提出 HFAM 网络模型,其框架如图 1 所示。

由图 1 可知,所设计的 HFAM 网络模型结构遵循 GFM[17]模型框架,其基本视觉特征从共享编码器学习,从两个单独解码器中学习各自的特征。本文选择 resnet-101[18]作为共享编码器提取图像特征,编码器以单个 RGB 图像作为输入,并通过 5 个卷积块 $E_0 \sim E_4$ 进行处理,每个块的分辨率降低了一半。首先通过共享编码器提取图像多尺度特征,主干网络的特征层次表示为 $\{f_0, f_1, f_2, f_3, f_4\}$。接着将主干网络最后三层特征 $\{f_2, f_3, f_4\}$ 利用 MFEM 进行多尺度上下文信息的提取与融合,将融合之后的特征传送到语义解码器中,以恢复原始图像分辨率,该过程利用网络自动生成的三分图进行监督。同时以编码器第一、二层高分辨率的低级纹理特征 $\{f_0, f_1\}$ 作为 BB 模块的输入,利用语义解码器得到原始图像大小的语义图作为位置注意力图过滤冗余的低级纹理信息,将经过处理之后的低层次特征传送到细节解码器中得到细节预测图。最后将两个解码器的输出利用融合函数进行像素级拼接得到最终预测的前景透明度遮罩,该过程引入结构性损失判别预测结果与真实结果的差异是否显著。

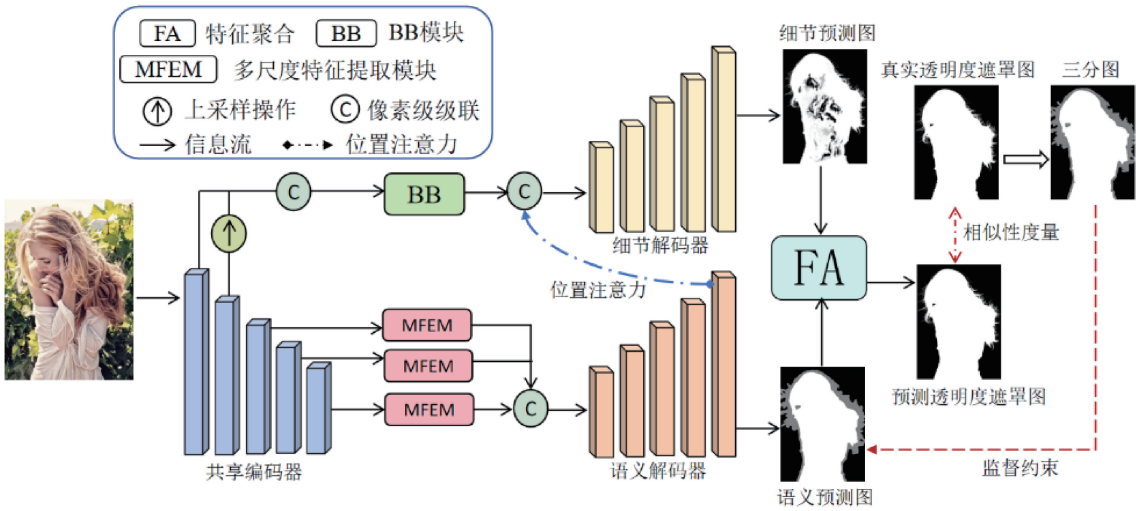

图 1　HFAM 网络模型结构图

Fig. 1　The structural diagram of the HFAM network model

1.1　语义分割分支

与图像分割算法[19-20]类似，HFAM 模型第一步是从输入图像中识别图像前景的主体结构，并将其他部分作为未知区域或背景，为此，语义解码器需要有一个很大的接受域来学习高级语义信息。如图 1 所示，语义解码器与共享编码器结构一致，对称地堆叠五个卷积块 $D_0^s \sim D_4^s$，每个块由 3 个顺序的 3×3 卷积层和一个上采样层组成。为准确捕捉图像深层次高级特征，在共享编码器与语义解码器之间增加 MFEM 提取全局上下文信息，然后通过元素级连接到每个语义解码器块 D_i^s。具体地，将共享编码器提取的最后三层特征 $\{f_2, f_3, f_4\}$ 经 MFEM 处理后进行融合，将经过处理的特征传输到语义解码器中得到带有未知区域的语义分割图，使用由真实透明度遮罩膨胀腐蚀得到的三分图监督语义分割图。MFEM 由 ASPP 模块[21]和 SK 注意力[22]构成，其结构如图 2 所示。

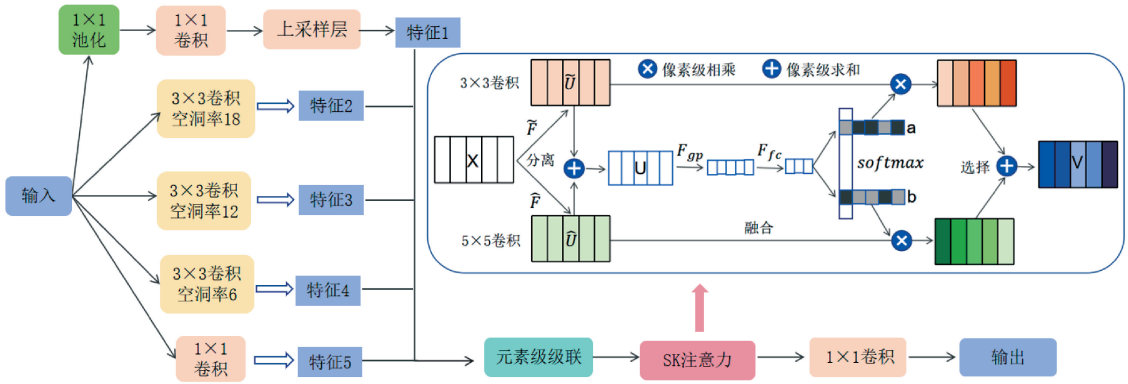

图 2　MFEM 结构图

Fig. 2　The structural diagram of the MFEM

标准 ASPP 模块利用空洞卷积进行多尺度特征提取，并将标准卷积用于多尺度特征融合，但提取的语义特征对前景结构回归的贡献不等。因此，对金字塔特征进行通道注意，即在 ASPP 模块中将 1×1 卷积、池化金字塔（不同空洞率的卷积组成）和池化层的输出特征图级联之后加入 SK 注意力，再进行 1×1 卷积输出。SK 注意力能够根据输入动态选择适合于图像抠图的高级语义特征，并保留前景轮廓和类别属性，提高模型对于不同尺度特征的感知能力和区分能力。该分支使用交叉熵损失（L_{ce}）进行监督训练，如式(1)所示。

$$L_{ce} = -\sum_{c=1}^{3} \alpha_g^c \log(\alpha_s^c),\ \alpha_g^c, \alpha_s^c \in [0,1] \tag{1}$$

其中，α_s^c 与 α_g^c 分别表示第 c 类的语义分割图预测概率和真实透明度遮罩概率。

1.2　细节预测分支

图像抠图要求精确的前景边界，而高级语义特征不能提供细微的纹理细节。因此，本文设计细节预测分支用来处理前景主体周围的未知区域，该分支的细节解码器与语

义解码器的基本结构一致,即对称堆叠的五个卷积块 $D_o^d \sim D_4^d$。为了提取到更丰富的低层结构特征,使用骨干网第一层提取的二倍下采样特征和第二层的四倍下采样特征作为该分支的输入。对四倍下采样特征上采样到第一层特征的尺寸大小,并将两层的特征进行连接,使用 BB 模块进行特征融合并降维,BB 模块由 3 个扩张的卷积层组成。

此外,在每个编码器块 E_i 和细节解码器块 D_i^d 之间添加一个跳跃链接,以保留前景的外观细节线索,这些外观线索可以描绘复杂的图像纹理,兼容透明度遮罩感知所需的边界精度。尽管外观细节线索显示出足够的图像纹理,但也包含了除前景内部或周围区域的其他冗余背景信息,为过滤低层纹理特征中存在的冗余背景信息,使用语义分割图作为位置注意力图,让其指导细节预测图中的信息选择,强调前景内部的外观线索,而去除属于背景的纹理和细节。为保证对未知区域的每个像素点预测的准确率,该分支用未知区域的 alpha 损失 (L_α^u) 表示细节预测的损失,如式(2)所示。

$$L_\alpha^u = \frac{\sum_i \sqrt{((\alpha^i - \alpha_d^i))^2 \times U_u^i + \varepsilon^2}}{\sum_i U_u^i}, U_u^i \in (0,1) \tag{2}$$

其中,α_d^i 表示细节预测图中像素点 i 的透明度遮罩值,U_u^i 表示未知区域内的像素集合,其作用是让 L_α^u 关注前景的未知区域,如果像素在未知区域内,则它的值为 1,否则为 0。

1.3 层次特征聚合分支

原始 RGB 图像经由两个目标不同的分支后,分别得到语义分割图 α_s 及细节预测图 α_d。受融合层次特征的启发[23],将两个分支最终的输出图像利用函数进行融合,使网络在学习的过程中兼顾图像前景的高层次语义信息和低层次细节信息。

优先信任从语义分支预测得到的分割图,因此,语义分割图中预测的绝对前景和背景区域将会保持不变,对于未知区域,用细节预测图中对应的像素来替换,以此得到最终的透明度遮罩图。融合函数用 f 表示为:

$$f(x_i) \rightarrow \alpha_p = \begin{cases} \alpha_s, & \alpha_s \in \{0,1\} \\ \alpha_d, & \alpha_s \in (0,1) \end{cases} \tag{3}$$

其中,i 表示图像的像素索引,α_s、α_d 和 α_p 分别表示语义分割图、细节预测图和网络最终预测的透明度遮罩图。

为保持抠图前景目标结构的完整性和预测的准确性,对于经过融合函数 f 处理后的模型总输出,引用结构性损失 (L_{ssim}) 和 alpha 损失 (L_α) 进行监督,如式(4)~(6)所示。

$$L_f = L_\alpha + L_{ssim} \tag{4}$$

$$L_\alpha = \| \alpha_p - \alpha_g \|_1 \tag{5}$$

$$L_{ssim} = 1 - \frac{(2\mu_p\mu_g + c_1)(2\sigma_{pg} + c_2))}{(\mu_p^2 + \mu_g^2 + c_1)(\sigma_p^2 + \sigma_g^2 + c_2)} \tag{6}$$

其中,μ_p、μ_g 和 μ_p^2、μ_g^2 分别为网络模型最终预测的透明度遮罩和真实透明度遮罩之间的平均值及标准差,σ_{pg} 为网络模型最终预测的透明度遮罩和真实透明度遮罩之间的协方差,c_1、c_2 为两个不同的常数。

整个 HFAM 模型通过 L_{ce}、L_α^u 和 L_f 的总和进行端到端训练,总的训练损失表示为:

$$L = \lambda_s L_{ce} + \lambda_d L_\alpha^u + \lambda_f L_f \tag{7}$$

其中,λ_s、λ_d 和 λ_f 是平衡这 3 个损失的超参数,设置为:$\lambda_s = \lambda_f = 1$ 和 $\lambda_d = 10$。

HFAM 模型通过语义分割分支和细节预测分支从单个图像中分别学习高层次语义信息和低层次纹理线索,最后以层次特征聚合方式聚合两部分信息,该模型解决了人像毛发这类细微结构局部区域易出现的误抠问题。下面将通过自建 MCP-1k 和公开 P3M-500-NP[24] 抠图数据集对上述所设计模型进行验证分析。

2 实验与结果分析

为验证 HFAM 模型的有效性,在 MCP-1k 和 P3M-500-NP 抠图数据集上进行综合评估,与相关深度抠图方法进行视觉效果和定量结果对比。

2.1 实验环境与参数设置

所有实验均在 PyCharm 平台下,使用深度学习框架 Python-3.8.17 + cuda1.8 版本进行,编译软件为 Jupyterlab-pytorch,使用单一的 NVIDIA GeForce RTX A100 GPU 及 15GB 内存的服务器运行。

为加快训练过程,防止过拟合,实验以 ImageNet 数据集[18] 上预训练的 resnet-101 作为特征提取网络,而其他层则从高斯分布中随机初始化。为了进行训练,所有被输入的图像都被随机裁剪到[640 × 640]、[960 × 960]和[1 280×1 280],将裁剪后的图像大小调整为 512×512 的分辨率,并以 0.5 的概率随机翻转,批量大小为 8,Adam 作为优化器,动量为 0.9,权重衰减为 0.000 5,学习率为 1×10^{-5},其他参数按照文献[17]设置。

2.2 数据集与评价指标

本文人像抠图数据是收集现有公开数据集中的人像图像混合而成的,具体包括 Adobe Matting[7],PPM-100[14],Distinctions-646[15] 和 Real-world Portrait-636[25],共收集 1 323 张人像前景。依照文献[7]的合成路线,将收集数据按 9∶1 的比例随机划分为 1 190 张训练集和 133 张测试集,然后将前景图像随机合成到 BG-20k[17] 背景数据集,其中每张训练与测试前景分别随机选择 20 张背景与 10 张背景。这样创建了一个训练数据集有 1 190 个独特的前景对象和 23 800 张图像,命名为 MCP-20k,而测试数据集有 133 个独特的对象和 1 330 张图像,命名为 MCP-1k。

为对比 HFAM 模型与其他抠图模型的性能,采用绝对差之和(SAD)、均方误差(MSE)、梯度(Grad)和连通性(Conn)作为相应的评估指标[3],对 HFAM 模型的性能进

行评价，计算公式如下：

$$SAD = \sum_i | \alpha_p^i - \alpha_g^i | \tag{8}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\alpha_p^i - \alpha_g^i)^2 \tag{9}$$

$$Grad = \sum_i (\nabla\alpha_p^i - \nabla\alpha_g^i)^2 \tag{10}$$

$$Conn = \sum_i (\varphi(\alpha_p^i, \Lambda) - \varphi(\alpha_g^i, \Lambda)) \tag{11}$$

其中，α_p^i，α_g^i 分别为像素 i 处预测的透明度遮罩值和真实透明度遮罩值，N 为透明度遮罩图像素点的个数。$\nabla\alpha_p^i$，$\nabla\alpha_g^i$ 分别为预测的透明度遮罩与真实透明度在像素 i 处的标准化梯度幅值。Λ 表示预测的透明度遮罩与真实透明度遮罩均为完全不透明的面积最大的连通区域，函数 φ 描述像素 i 与 Λ 的连通程度。

2.3　对比实验结果分析

为验证 HFAM 模型在人像样本前景遮罩提取问题上的性能，在 MCP-1k 和 P3M-500-NP 数据集上与深度抠图模型进行比较，具体包括 LFM[12]，MODNet[14]，HAtt[15]，Vitae[24]，MGM[25]，FGI[26]，AIM[27]，和 P3MNet[28]，结果如表 1、图 3 所示。

表 1　在 MCP-1k 上不同抠图方法的误差对比结果

Table 1　Error comparison results of different matting methods on MCP-1k dataset

Method	MSE	SAD	Grad	Conn
MODNet[14]	0.012 9	33.53	48.72	48.72
Vitae[24]	0.010 8	26.15	54.39	**25.65**
FGI[26]	0.011 3	32.62	45.17	33.69
AIM[27]	0.024 0	49.82	66.60	49.27
P3MNet[28]	0.019 0	39.12	56.00	37.62
BASE	0.009 5	28.31	45.13	29.09
HFAM	**0.007 6**	**25.59**	**43.41**	27.64

注：加粗位置表示最优值

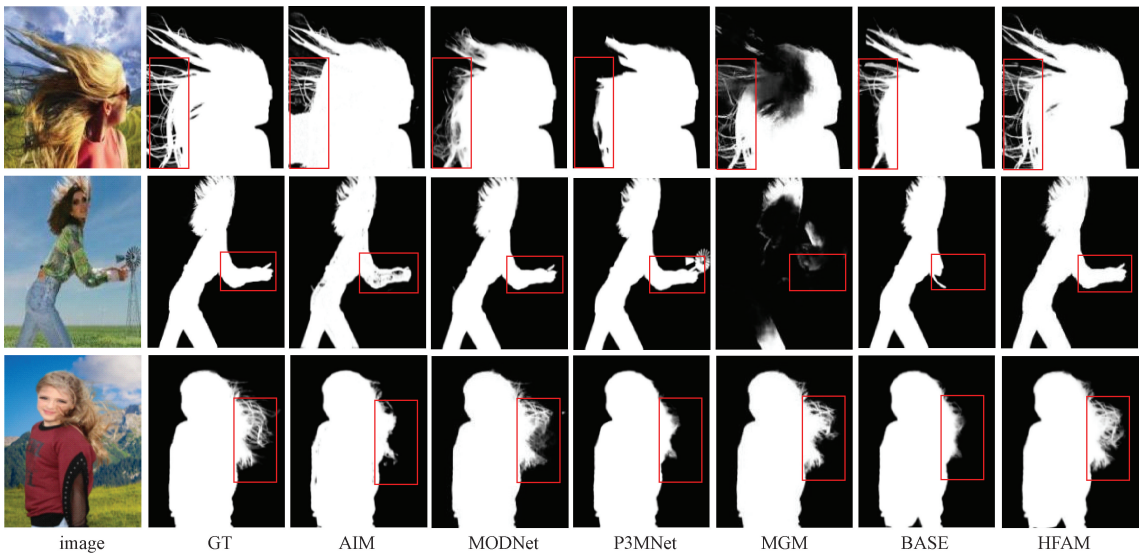

图 3　在 MCP-1k 上不同抠图方法视觉对比结果

Fig. 3　Visual comparison results of different matting methods on the MCP-1k dataset

表 1 展示了 HFAM 模型与其他抠图模型在 MCP-1k 上的误差结果。与所有抠图模型相比，在 MSE、SAD 和 Grad 指标中都取得了最好的成绩，即 0.007 6，25.59，43.41；但在 Conn 指标上的误差略高于 Vitae 模型，即 27.64 VS 25.65。对于 HFAM 模型在 Conn 指标上的提取精度略低于 Vitae 模型的主要原因在于 HFAM 模型两个分支之间的特征未提供相互指导，导致在前景边界附近产生断裂或错误的分割，从而最终 Conn 的值略高于 Vitae 模型。虽然 HFAM 模型在 Conn 指标上的误差略高于 Vitae 模型，但 HFAM 模型在其余四个指标中的表现是最好的。同时，HFAM 模型在所有指标上都远远优于 BASE 模型（改进的基线模型），说明本文所做的改进对人像细微毛发抠图是有效的。图 3 展示了各个抠图模型在 MCP-1k 数据集上的视觉效果。由图 3 看出，HFAM 模型有更完整的前景语义结构和细微的毛发纹理，整体上比其他抠图模型有更好的视觉效果。同时，比 BASE 模型有更强的语义分割和细节恢复能力。

此外，将 HFAM 模型与其他抠图模型在 P3M-500-NP 数据集上进行评估，验证抠图的泛化能力，结果如表 2，图 4 所示。

可以看出 HFAM 模型在 MSE 和 SAD 指标上远远优于基于三分图的抠图模型和自动抠图模型，其 MSE 值为 0.007 2，SAD 值为 25.52，这说明 HFAM 模型泛化能力更强。图 4 展示了各个抠图模型在 P3M-500-NP 数据集上的视觉效果。从图像来看，AIM 模型整体效果最差，例如在图 4 的第 1 幅图像中，将右手臂与毛发混合区域全部识别

表 2 在 P3M-500-NP 上不同抠图方法的误差对比结果

Table 2 Error comparison results of different matting methods on P3M-500-NP dataset

Method	MSE	SAD
LFM[12]	0.013 1	32.59
MODNet[14]	0.007 4	25.77
HAtt[15]	0.007 2	30.53
MGM * [25]	0.025 1	64.94
AIM[27]	0.011 8	30.35
BASE	0.009 3	27.35
HFAM	**0.007 2**	**25.52**

注:* 号表示基于三分图的抠图方法,加粗位置表示最优值

为前景,并没有区分出背景部分;在第 2 副图像中的头发部分产生明显的伪影,且头部上方结构不完整。此外,基于三分图的 MGM 模型在图 4 中的两幅图像上的视觉效果不佳,在毛发复杂的局部区域存在严重的误抠现象。相比之下,HFAM 模型总体上在主体结构和毛发细节上优于其他深度抠图模型,在毛发复杂的局部区域提取效果更精细。这说明本文所设计的模型在 P3M-500-NP 数据集上的泛化能力更强,所做的改进有效。

2.4 消融实验结果分析

为证明 HFAM 模型各变体的有效性,在 MCP-1k 数据上对其进行消融实验。MFEM,PA 和 L-BB 分别表示多尺度特征提取模块,位置注意力机制和以浅层次特征为输入的 BB 模块,实验结果如表 3 所示。

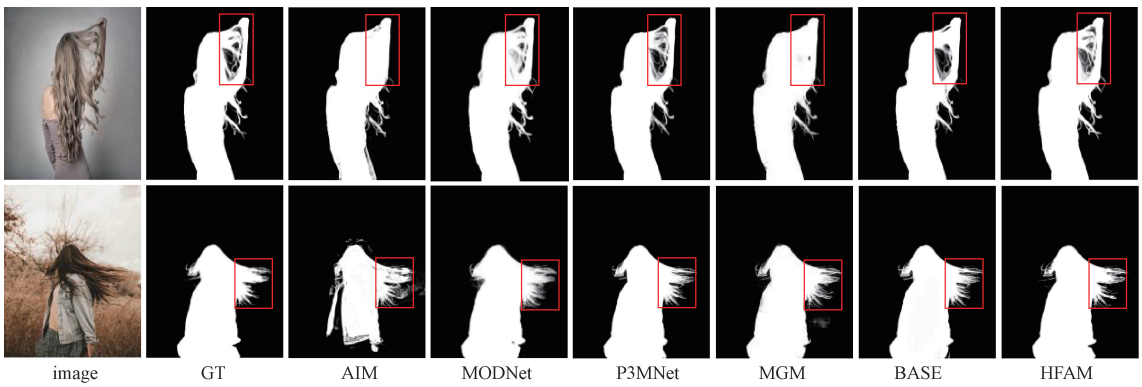

图 4 在 P3M-500-NP 上不同抠图方法视觉对比结果

Fig. 4 Visual comparison results of different matting methods on the P3M-500-NP dataset

表 3 在 MCP-1k 上进行消融试验的误差结果

Table 3 Error results of ablation experiments on MCP-1k dataset

MFEM	L-BB	PA	MSE	SAD	Grad	Conn
			0.009 5	28.31	45.13	29.09
Y			0.009 2	28.02	44.97	28.85
3Y			0.008 5	27.46	44.61	28.39
3Y		Y	0.008 7	27.75	44.42	28.06
3Y	Y		0.007 9	26.04	43.67	**27.43**
3Y	Y	Y	**0.007 6**	**25.59**	**43.41**	27.64

注:3Y 指编码器最后三层特征均通过 MFEM 处理,加粗位置表示最优值

由表 3 可知,在基线中融入 MFEM 和 3MFEM,五个评价指标误差均有显著下降,说明在解码器之前融合多尺度高级特征对提高抠图质量是有效的。此外,发现 3MFEM+PA 组合和 3MFEM+L-BB 组合对抠图任务均有效,且 3MFEM+L-BB 组合比 3MFEM+PA 组合减少误差效果更显著,说明以浅层特征作为细节预测分支的输入对恢复图像细节结构是有效的。值得注意的是,同时加入三个模块时,MSE,SAD 和 Grad 三个指标误差均减少,但 Conn 指标从 27.43 上升到 27.64,增加了 0.21。Conn 指标误差增加的主要原因在于位置注意力机制对低级特征重新加权时提取了不重要的冗余信息,导致图像局部区域不连通。虽然 HFAM 模型在 Conn 指标上结果不是最优的,但是 3MFEM+L-BB+PA 组合获得的透明度遮罩视觉效果更好,有更完整的语义结构和清晰的毛发细节。

3 结 论

本文提出了一种层次特征聚合抠图网络模型,该模型采用多尺度特征提取模块提取与抠图目标自适应的高层次语义信息,使用低层次浅层特征以恢复图像未知区域的细节信息,并执行位置注意力来过滤掉冗余的外观线索和背景信息,以层次特征聚合方式聚合低层次纹理线索和高层次语义信息,能够在没有额外输入的情况下从单个人像的精细边界和自适应语义中感知前景透明度遮罩。此外,本文利用合成技术构建了一个大规模的真实前景图像抠图数据集,以适用于端到端抠图模型的训练。实验结果表明,所提 HFAM 网络模型在处理人像细微毛发和捕获前景语义结构方面优于典型的深度抠图方法,有效解决人像

毛发局部区域的误抠问题。

在未来，将运用像素之间本身存在的信息，探索多实例人像抠图领域中实例细边界结构复杂区域及多毛发细微区域前景遮罩提取问题，以进一步扩展自然人像抠图的应用范围。

参考文献

[1] DING H H, ZHANG H, LIU C, et al. Deep interactive image matting with feature propagation[J]. IEEE Transactions on Image Processing, 2022, 31: 2421-2432.

[2] 冯夫健，黄翰，吴秋霞，等. 基于群体协同优化的高清图像前景遮罩提取算法[J]. 中国科学：信息科学，2020，50：424-437.
FENG F J, HUANG H, WU Q X, et al. An alpha matting based on collaborative swarm optimization for high-resolution images [J]. SCIENTIA SINICA Informationis, 2020, 50: 424-437.

[3] 梁椅辉，黄翰，蔡昭权，等. 自然图像抠图技术综述[J]. 计算机应用研究，2021，38(5)：1294-1301.
LIANG Y H, HUANG H, CAI ZH Q, et al. Survey of natural image matting[J]. Application Research of Computers, 2021, 38(5): 1294-1301.

[4] PORTER T, DUFF T. Compositing digital images[C]. The 11th Annual Conference On Computer Graphics And Interactive Techniques-SIGGRAPH, 1984: 253-259.

[5] 冯夫健，杨圆，谭棉，等. 基于微搜索的高分辨率图像前景遮罩提取算法[J]. 模式识别与人工智能，2023，36(6)：530-543.
FENG F J, YANG Y, TAN M, et al. An alpha matting algorithm based on micro-scale searching for high-resolution images[J]. Pattern Recognition and Artificial Intelligence, 2023, 36(6): 530-543.

[6] YAO G L, JIANG D G, SUN J L. An affinity based matting method based on multi-scale space fusion[C]. The 33th Chinese Control and Decision Conference, 2021: 1572-1577.

[7] XU N, PRICE B, COHEN S, et al. Deep image matting[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 311-320.

[8] ZHOU F F, TIAN Y J, QI Z Q. Attention transfer network for nature image matting [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(6): 2192-2205.

[9] LIU C, DING H H, JIANG X D. Towards enhancing fine-grained details for image matting [C]. IEEE Winter Conference on Applications of Computer Vision, 2021: 385-393.

[10] LI Y Y, LU H T. Natural image matting via guided contextual attention [J]. AAAI, 2022, 34: 11450-11457.

[11] CHEN Q, GE T Z, XU Y Y, et al. Semantic human matting[C]. Proceedings of the 26th ACM International Conference on Multimedia, 2018: 618-626.

[12] ZHANG Y K, GONG L X, FAN L B, et al. A late fusion CNN for digital matting[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2019: 7469-7478.

[13] LIU J L, YAO Y, HOU W D, et al. Boosting semantic human matting with coarse annotations[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 8560-8569.

[14] KE Z H, SUN J Y, LI K C, et al. MODNet: Real-time trimap-free portrait matting via objective decomposition [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(1): 1140-1147.

[15] QIAO Y, LIU Y H, YANG X, et al. Attention-guided hierarchical structure aggregation for image matting [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 13673-13682.

[16] QIN X B, ZHANG Z C, HUANG C Y, et al. Basnet: Boundary-aware salient object detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2019: 7479-7489.

[17] LI J, ZHANG J, STEPHEN J, et al. Bridging composite and real: Towards end-to-end deep image matting[J]. International Journal of Computer Vision, 2022: 130(2): 246-266.

[18] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[19] 赵恩玄，何云勇，沈宽，等. 基于深度学习的铸件 CT 图像分割算法[J]. 仪器仪表学报，2023，44(11)：176-184.
ZHAO EN X, HE Y Y, SHEN K, et al. Casting CT image segmentation algorithm based on deep learning[J]. Chinese Journal of Scientific Instrument, 2023, 44(11): 176-184.

[20] 徐晓龙，俞晓春，何晓佳，等. 基于改进 U-Net 的街景图像语义分割方法[J]. 电子测量技术，2023，46(9)：117-123.
XUE X L, YU X C, HE X J, et al. Semantic segmentation method of street view image based on

improved U-Net [J]. Electronic Measurement Technology, 2023, 46(9): 117-123.

[21] CHEN L C, PAPANDREOU G, KOKKINOS L, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834-848.

[22] LI X, WANG W, HU X, et al. Selective kernel networks [C]. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, 510-519.

[23] 程德强, 陈杰, 寇旗旗, 等. 融合层次特征和注意力机制的轻量化矿井图像超分辨率重建方法[J]. 仪器仪表学报, 2022, 43(8): 73-84.
CHENG D Q, CHEN J, KOU Q Q, et al. Lightweight super-resolution reconstruction method based on hierarchical features fusion and attention mechanism for mine image[J]. Chinese Journal of Scientific Instrument, 2022, 43(8): 73-84.

[24] SIHAN M, Li J, ZHANGJ, et al. Rethinking portrait matting with privacy preserving [J]. International Journal of Computer Vision, 2023: 1-26.

[25] YU Q H, ZHANG J M, ZHANG H, et al. Mask guided matting via progressive refinement network [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 11.

[26] CHENG H, XU S G, JIANG X F, et al. Deep image matting with flexible guidance input [J]. ArXiv preprint arXiv: 2110.10898, 2021.

[27] LI J, ZHANG J, TAO D C. et al. Deep automatic natural image matting [C]. Thirtieth International Joint Conference on Artificial Intelligence, 2021: 800-806.

[28] LI J, SIHAN M, ZHANG J, et al. Privacy-preserving portrait matting [C]. 29th ACM International Conference on Multimedia, 2021: 3501-3509.

作者简介

汪小梅,硕士,主要研究方向为自然图像抠图。

E-mail:3146088016@qq.com

谭棉(通信作者),高级实验师,主要研究方向为图像处理及智能算法。

E-mail:tanmian@gzmz.edu.cn

罗太维,硕士,主要研究方向为工业瑕疵检测。

E-mail:2574573478@qq.com

冯夫健,高级实验师,主要研究方向为优化算法及机器视觉。

E-mail:fujian_feng@gzmz.edu.cn