

王才丽, 邹红梅, 黎旭, 等. 基于上下文增强的图像抠图方法研究[J]. 智能计算机与应用, 2026, 16(2): 29-34. DOI: 10.20169/j. issn. 2095-2163. 25092502

基于上下文增强的图像抠图方法研究

王才丽^{1,2}, 邹红梅^{1,2}, 黎旭^{1,2}, 吴瑞瑾^{1,2}, 谭棉^{1,2}

(1 贵州民族大学 数据科学与信息工程学院, 贵阳 550025;

2 贵州民族大学 贵州省模式识别与智能系统重点实验室, 贵阳 550025)

摘要: 本文针对当前图像抠图模型中存在的上下文信息捕获不足问题展开研究。现有抠图方法在处理复杂场景时, 往往难以有效建模长距离依赖关系, 导致对全局上下文信息的利用不充分。为了解决这一问题, 本文提出了跨层信息传递机制, 使模型能够在不同层级间有效共享和整合语义信息, 从而增强对全局上下文的理解能力, 提升模型对复杂场景的适应性和边缘细节的保持能力。在 Composition-1 k 测试集上的实验结果表明, 跨层信息传递机制增强了模型上下文信息的捕获, 从而提升抠图质量。与基线模型相比, 本文所提的方法在增强模型对全局上下文的理解上有较好的表现。

关键词: 自然图像抠图; 跨层信息传递; 门控记忆单元; 上下文信息

中图分类号: TP391

文献标志码: A

文章编号: 2095-2163(2026)02-0029-06

Research on image matting method based on context enhancement

WANG Caili^{1,2}, ZOU Hongmei^{1,2}, LI Xu^{1,2}, WU Ruijin^{1,2}, TAN Mian^{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 of Guizhou Province, Guizhou Minzu University, Guiyang 550025, China)

Abstract: This paper conducts research on the problem of insufficient context information capture existing in the current image matting model. Existing matting methods often struggle to effectively model long-distance dependencies when dealing with complex scenarios, resulting in insufficient utilization of global context information. To address this issue, this paper proposes a cross-layer information transmission mechanism, enabling the model to effectively share and integrate semantic information among different levels, thereby enhancing the understanding ability of the global context and improving the model's adaptability to complex scenarios and the ability to preserve edge details. The experimental results on the Composition-1 k test set show that the cross-layer information transfer mechanism enhances the capture of model context information, thereby improving the matting quality. Compared with the baseline model, the proposed method performs better in enhancing the model's understanding of the global context.

Key words: natural image matting; cross-layer information transmission; gated memory unit; context information

0 引言

自然图像抠图是指从给定的图像中精确提取出感兴趣的前景对象, 是计算机视觉领域中的重要任务之一, 广泛应用于图像编辑、图像合成^[1]、电影后期制作^[2]以及自动驾驶^[3]。图像抠图是通过计算每个像素的透明度来获得 alpha 遮罩进行前背景区分, 其中图像的前景包括人类、动物之类的实体, 另

外像头发发丝、灯泡、水之类对象拥有精细且复杂边缘, 很难得到 alpha 遮罩值。在数学上, 图像抠图任务被定义为前景图像和背景图像的线性组合, 即:

$$I_i = \alpha_i F_i + (1 - \alpha_i) B_i, \alpha \in [0, 1] \quad (1)$$

其中, α_i 表示像素 i 处的透明度值, 取值范围为 $[0, 1]$; I_i 表示输入图像中像素 i 处的 RGB 值; F_i 表示前景图像中像素 i 处的 RGB 值; B_i 表示背景图像中像素 i 处的 RGB 值。根据方程可知, 在这个线性

基金项目: 国家自然科学基金(62566012); 贵州省教育厅自然科学研究项目(黔教技[2023]061号); 贵州省科技厅科技计划项目(黔科合基础-ZK[2023]一般143); 贵州省研究生教育教学改革项目(2024YJSJGX041)。

作者简介: 王才丽(1999—), 女, 硕士研究生, 主要研究方向: 图像抠图; 邹红梅(2000—), 女, 硕士研究生, 主要研究方向: 迁移学习, 优化算法; 黎旭(2001—), 女, 硕士研究生, 主要研究方向: 虚拟数字人; 吴瑞瑾(2000—), 男, 硕士研究生, 主要研究方向: 图像抠图。

通信作者: 谭棉(1984—), 女, 高级实验师, 主要研究方向: 图像处理及智能算法。Email: tanmian@gzmu.edu.cn。

收稿日期: 2025-09-25

组合中, F_i 和 B_i 、 α_i 都是未知的,因此图像抠图是一个高度不适定问题。为了解决这个问题,许多算法引入了额外的辅助信息,如三分图^[4-5]、背景图^[6]等,有效缩小未知区域的解空间。三分图是将输入图片划分为3类,像素值为0的部分标注为背景区域,像素值为255的区域被标注为前景,像素值为128的区域被标注为未知区域,是传统和基于深度学习方法中最常见的选择。

传统的方法主要有基于采样的抠图方法^[7-9]和基于传播的抠图方法^[10-12]。Bayesian matting^[7]使用高斯模型对局部前景和背景颜色进行概率建模,并通过最大后验概率同时估计最优的不透明度、前景和背景颜色。He等学者^[8]提出了一种全局采样方法,利用图像中所有可用的样本,这样避免遗漏合格样本,并且为了解决样本选择过程中的迷糊性问题,定义了一种简单而且有效的代价函数。冯夫健等学者^[12]结合决策集分解策略和优化信息传递策略,提出了基于微搜索的前景遮罩提取算法,通过在有效决策子集中的搜索代替整个决策集的搜索,有效解决了高分辨率图像中部分前景背景信息相反的问题,提高高分辨率图像的抠图精度。尽管传统的图像抠图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基础性成果,但是传统的方法没有考虑前景对象或者背景场景的任何语义信息,因此对于复杂场景的(如具有相似颜色的背景、纹理复杂的背景等)抠图效果不佳。

近年来,随着深度神经网络的引入,这个问题得到了很大程度的改善。基于深度学习的抠图模型将高级语义信息融入到抠图网络中,具有非常强的表达能力,减少了传统机器学习方法中的需要手工设计特征的工作量。Xu等学者^[13]利用深度卷积编码器-解码器网络量,将图像和对应的三分图作为输

入,预测图像的透明度遮罩,这个方法使得图像抠图算法取得了新的进展,但是在一些复杂背景和半透明区域图像中产生模糊的纹理和边界。Li等学者^[14]利用周围的信息来估计局部透明度,引入了上下文注意力模块。在该模块中,使用图像的低层特征作为引导,并在该引导的基础上执行alpha特征传输。Sun等学者^[15]通过对三分图区域内的补丁结构分析,将传统的三分图扩展到语义三分图,同时学习了一个多类别判别器在语义层面对透明度遮罩进行正则化,学习了内容敏感权重来平衡不同的正则化损失,提高了复杂场景下的抠图精度。

尽管上述的抠图方法取得了一定的进展,但是现有的抠图方法在处理复杂场景时对上下文信息的捕获能力不足,高层网络无法直接利用低层已学习到的全局语义信息,只能重新估计,造成冗余计算与误差累积。为了实现输入的先验信息跨层流动与可控更新,同时避免记忆污染,本文提出跨层信息传递机制,实现先验信息的跨层流动与可控更新,精确重建高频细节,从而提升模型对细节的精细化处理。

1 网络模型结构

本文设计的跨层信息传递机制(Cross-layer Information Transmission Mechanism, CITM)框架如图1所示。本文的方法以matteformer^[16]为基础框架,整体架构基于编码器-解码器设计。通过引入跨层先验记忆更新模块(Cross-Layer Prior Memory),该模块作为编码器相邻层级间的信息桥梁,实现了前景、背景和未知区域先验知识的跨层传递(这里提到的先验信息是经过matteformer模型prior-tokens模块处理后的精细化先验信息),将原本只在层内循环的先验信息,实现了层间传播,允许网络在不同层级间共享语义理解,实现多尺度语义信息的融合。

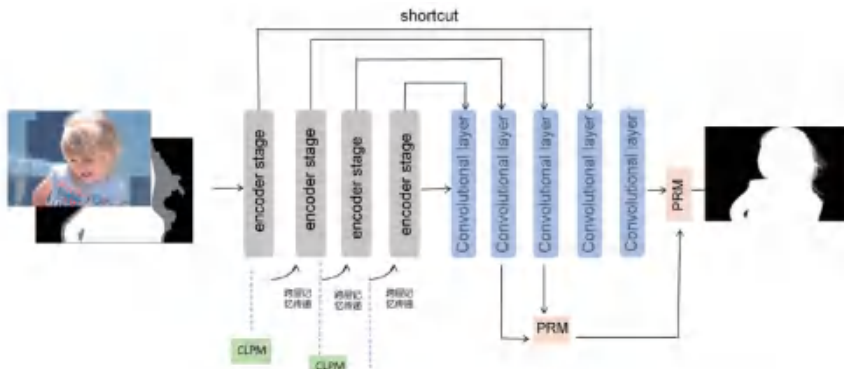


图1 网络框架图

Fig. 1 Network framework

1.1 跨层先验记忆更新模块

本文深入分析了模型中信息传递的规律,多数方法未能充分挖掘先验信息在多层次特征融合和跨尺度语义传递中的潜力。基于此,研究采用 mattemformer 的编码器结构,在保持原有 block 间信息传递的基础上实现层间信息传递。原有三分图信息经过 prior-tokens 模块的处理具有不同的语义特性。传统方法将这些信息隔离在单一层级内处理,导致深层网络无法利用浅层提取的细节特征,而浅层网络也缺乏来自深层的语义指导。受循环神经网络^[17]在序列建模中成功管理长效记忆的启发,设计了一个轻量级的跨层记忆更新模块。该模块作为一个轻量级的门控记忆单元,作用于 2 个相邻的编码层级之间,对来自不同层的信息进行更新,以便在深层网络利用过去的信息。同时,模块会接受来自第 L 层最后一个块的输出先验 P_{new} ,以及第 L 层之前维护的历史记忆状态 P_{old} ,对此可以表示为:

$$P_{new} = \{P_{new}^{fg}, P_{new}^{bg}, P_{new}^{uk}\} \quad (2)$$

其中, $P \in R^{B \times C}$, 对于前景、背景和未知区域这 3 种先验,研究为其分别配备了独立的门控循环单元,以实现更具针对性的控制。记忆更新过程对于

每种先验类型 $s \in \{fg, bg, uk\}$ 是独立且并行的,其数学表达如下:

$$P_{out}^s = GRU^s(P_{new}^s, P_{old}^s) \quad (3)$$

其中,GRU 的内部门控机制由更新门 z_t 和重置门 r_t 控制,推得的公式为:

$$z_t^s = \sigma(W_z^s \cdot [P_{new}^s, P_{old}^s]) \quad (4)$$

$$r_t^s = \sigma(W_r^s [P_{new}^s, P_{old}^s]) \quad (5)$$

$$\tilde{h}_t = \tanh(W_h^s \cdot [p_{new}, r_t \odot p_{old}]) \quad (6)$$

$$p_{out} = (1 - z_t) \odot p_{old} + z_t \odot \tilde{h}_t \quad (7)$$

更新门控制了历史记忆中有多少信息应被保留至未来,而重置门决定了如何将新的输入与历史记忆相结合以产生候选激活。

在每一个训练迭代或处理每一张输入图像时,研究中首先将所有先验记忆的内部状态初始化为空,以确保不同训练样本间的计算独立性。在每个阶段的末尾,该阶段最终输出的先验信息被送入跨层先验记忆更新模块(CLPM),与上一阶段的记忆状态(如果是第一个阶段,则为空)进行门控融合。融合后的结果被存储下来,作为新的记忆状态传递给下一个阶段。跨层先验记忆更新模块如图 2 所示。

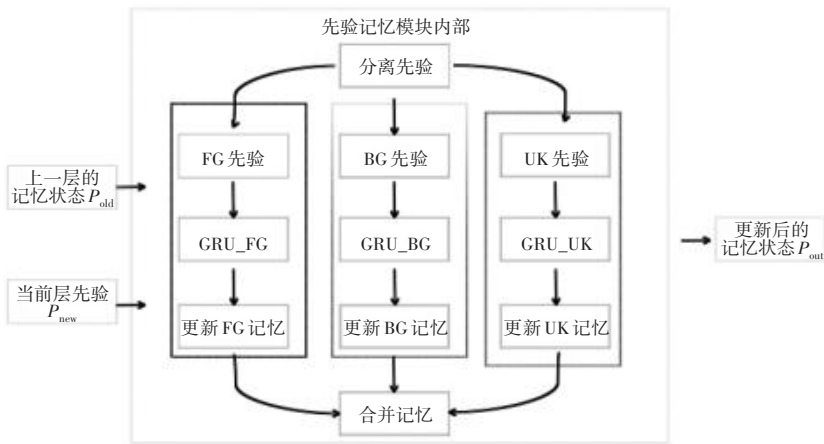


图 2 跨层先验记忆模块

Fig. 2 Cross-layer prior memory module

跨层先验记忆模块允许先验信息随着网络深度的增加而不断被精炼。模型可以自适应地选择信任历史记忆或新的观测,从而逐步提炼出更可靠、更准确的全局先验。这就为关键先验信息提供了一条受保护的“高速公路”,使其能够绕过主干网络中的复杂变换路径,有效地缓解了信息在深层传递过程中的衰减与噪声累积问题。

1.2 渐进细化模块

在解码过程中,采用 Yu 等学者^[18]提出的渐进

细化模块(Progressive Refinement Module, PRM),以粗到细的方式预测精确的透明度遮罩。PRM 模块是一个多阶段的网络结构,在每一个特征级别都有一个预测输出,通过选择性地融合上一层和当前层的预测输出进行逐步细化。具体来说,当前层输出的自引导掩码是由上一层预测的 α 值生成的。计算公式如下:

$$g_l = \begin{cases} 1, & 0 < \alpha_{l-1} < 1 \\ 0, & \text{其他} \end{cases} \quad (8)$$

其中, g_l 表示当前 l 层的掩码, α_{l-1} 表示 $l-1$ 层的抠图输出。定义的 $0 < \alpha < 1$ 是未知的, 对 α_{l-1} 进行上采样以匹配当前层的原始抠图输出 α'_l , 利用自引导掩码 g_l 计算更新后的 α_l , 公式如下:

$$\alpha_l = \alpha'_l g_l + \alpha_{l-1} (1 - g_l) \quad (9)$$

其中, α_l 表示当前层 l 计算更新后的预测输出, 用当前层的原始输出 α'_l 替换 α_{l-1} 的未知区域。

1.3 损失函数

本文采用的损失函数分别为 L_1 损失, 合成损失 L_{comp} 和拉普拉斯损失 L_{lap} ^[19], 总损失函数 L 是 L_1 , L_{comp} 和 L_{lap} 的加权和, 计算公式如下:

$$L = L_1 + L_{\text{comp}} + L_{\text{lap}} \quad (10)$$

2 实验结果与分析

2.1 实验数据

本文使用的数据集为 Xu 等学者^[13]提出的 Composition-1 k 数据集, 其中 Composition-1 k 测试集包含 50 个独特的前景图像。其中, 每个都与来自 VOC 2012 数据集的 20 个背景图像通过计算加以合成, 以构建具有 1 000 个不同合成图像的测试集。合成图像的公式为:

$$I = \alpha F + (1 - \alpha) B \quad (11)$$

Composition-1 k 训练集具有 431 个独特的前景图像和对应的 alpha 遮罩, 与 COCO 数据集构建包含 43 100 个不同的合成图像。

2.2 实验设置

实验的配置环境为深度学习框架 Python-3.8+ Cuda11.8 版本, 编译软件为 Jupyterlab Python, 实验均由 NVIDIA A100 GPU 及 15 G 内存的服务器运行的。在训练阶段, 批次大小设置为 20, 使用 Adam 优化器, 学习率初始化为 4×10^{-4} 。

为了评估本文的方法, 文中使用 4 个常用指标: 绝对误差之和 (SAD)、均方误差 (MSE)、梯度误差 (Grad) 和连通误差 (Conn)^[20]。这些评价指标的值越低, 表示透明度遮罩的质量越高。为了便于阅读, 将 MSE 的值缩放为 10^{-3} 。

2.3 对比实验

为了验证提出模型的有效性, 将本文的方法与 Closed-Form、KNN Matting、DIM、IndexNet、GCA、MGM_{trimap}、LCDM 方法进行实验对比。本文的方法在 Composition-1 k 数据集上进行训练和测试。实验结果见表 1。

由表 1 可知, 在 Composition-1 k 测试集上, 本

文提出抠图的方法在 MSE, SAD, Grad, Conn 这 4 项评价指标上均获得了最优的结果, 比其他 8 种典型的抠图方法提取透明度遮罩的效果较好。相对于经典的深度学习 DIM 抠图方法, 本文的方法在 MSE, SAD, Grad, Conn 上分别提升了 62%, 42%, 63%, 42%, 在这 4 项评价指标上均有了显著提升。与 MGM 方法相比, 本文的方法在 MSE, SAD, Grad, Conn 上分别提升了 24.0%, 9.6%, 26.0%, 8.6%, 说明本文的方法提取前景遮罩有较好的性能。

表 1 在 Composition-1 k 测试集上的定量结果

Table 1 Quantitative results on the Composition-1 k test set

Method	MSE	SAD	Grad	Conn
Closed-Form	91.0	168.1	126.9	167.9
KNN Matting	103.0	175.4	124.1	176.4
DIM	14.0	50.4	31.0	50.8
IndexNet	13.0	45.8	25.9	43.7
LCDM	21.0	62.1	25.7	43.4
GCA	9.1	35.2	16.9	32.5
MGM _{trimap}	7.0	32.1	14.0	27.9
本文方法	5.3	29.0	10.3	25.2

2.4 消融实验

为了验证本文所提跨层先验记忆更新模块的有效性, 在 Composition-1 k 数据集上进行实验, 与 Baseline^[16] 进行对比分析。Baseline 在本文的配置环境中进行训练。实验结果见表 2。跨层先验记忆模块允许先验信息随着网络深度的增加而不断被精炼。模型可以自适应地选择信任历史记忆或新的观测, 即可逐步提炼出更可靠、更准确的全局先验, 从而引导模型进行前、背景分离。

表 2 在 Composition-1 k 数据集上的消融实验

Table 2 Ablation experiments on the Composition-1 k dataset

Method	MSE	SAD	Grad	Conn
Baseline	5.6	29.3	11.2	25.4
Baseline+CLPM	5.3	29.0	10.3	25.2

与基线模型相比, 本文所提的方法在 MSE, SAD, Grad, Conn 上分别提升了 5.0%, 1.0%, 8.0%, 0.8%, 虽然在 SAD, Conn 指标上提升不大, 但整体提取前景透明度遮罩的效果较好, 也说明了本文模型的有效性。

本文所提出的方法与其他经典的抠图方法在 Composition-1 k 数据集进行可视化比较, 可视化结果如图 3 所示。图 3 中, (a) 是原图, (b) ~ (e) 分别

是 GCA、MGM_{trimap}、Matteformer、本文方法预测的结果,(f)是真实透明度遮罩。从图3第3列中可以看出,本文的方法能很好地将蕾丝花纹比较清晰地抠出来,没有伪影,本文得到的结果比较接近真实透明度遮罩,能够更好地分离前景和背景。

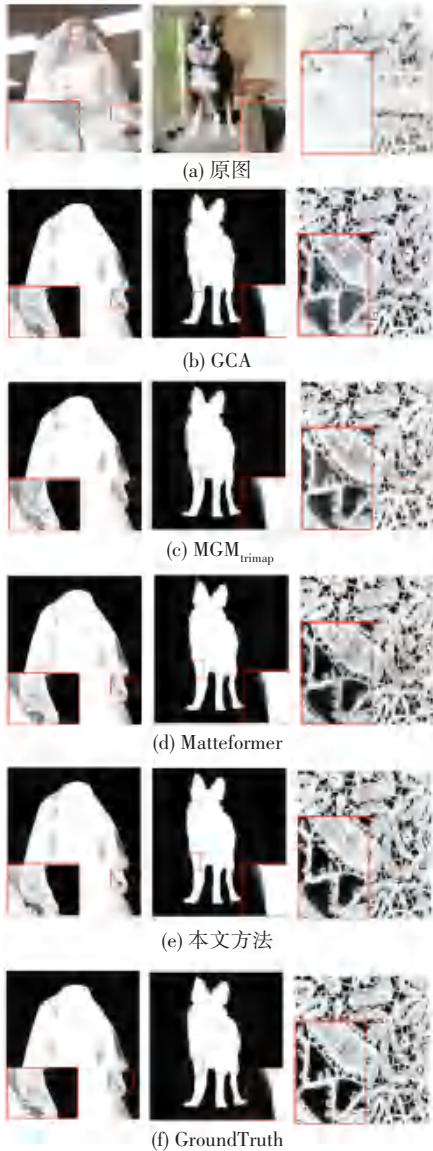


图3 本文与其他方法在 Composition-1 k 测试集上的可视化结果

Fig. 3 Visualization results of this paper and other methods on the Composition-1 k test set

3 结束语

本文围绕图像抠图任务中的上下文信息捕获不足问题进行了深入研究。通过分析现有方法的局限性,提出了一种跨层信息传递机制,设计了一个轻量级的跨层先验记忆更新模块(CLPM)。该模块通过引入门控循环单元(GRU),为前景、背景和未知区域三类先验信息分别建立了独立的跨层记忆传递通

路。所提出的跨层门控记忆机制能够有效解决 Transformer 传统层级中的先验信息隔离问题。研究结果表明,充分的上下文信息捕获对于提升抠图质量,特别是边缘细节的保持至关重要。该机制显著提升了模型对全局语义关系的理解能力。本文提出的方法简单有效,为相关研究提供了有益的借鉴。

参考文献

- [1] CHEN Qifeng, LI Dingzeyu, TANG Chi-Keung. KNN matting [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(9): 2175-2188.
- [2] 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.
- [3] RUAN Jiageng, CUI Hanghang, HUANG Yuhan, et al. A review of occluded objects detection in real complex scenarios for autonomous driving [J]. Green Energy and Intelligent Transportation, 2023, 2(3): 100092.
- [4] LU Hao, DAI Yutong, SHEN Chunhua, et al. Indices matter: Learning to index for deep image matting[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE, 2019: 3266-3275.
- [5] CAI Huanqia, XUE Fanglei, XU Lele, et al. Transmatting: Enhancing transparent objects matting with transformers [C]//Proceedings of the European Conference on Computer Vision. Cham: Springer, 2022: 253-269.
- [6] LIN Shanchuan, RYABTSEV A, SENGUPTA S, et al. Real-time high-resolution background matting[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 8762-8771.
- [7] CHUANG Y Y, CURLESS B, SALESIN D H, et al. A bayesian approach to digital matting [C]//Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001). Piscataway, NJ: IEEE, 2001, 2: II.
- [8] HE Kaiming, RHEMANN C, ROTHER C, et al. A global sampling method for alpha mattin[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2011). Piscataway, NJ: IEEE, 2011: 2049-2056.
- [9] SHAHRIAN E, RAJAN D, PRICE B, et al. Improving image matting using comprehensive sampling sets [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2013: 636-643.
- [10] SUN Jian, JIA Jiaya, TANG C K, et al. Poisson matting[C]//ACM SIGGRAPH 2004 Papers. New York: ACM, 2004: 315-321.
- [11] LEVIN A, LISCHINSKI D, WEISS Y. A closed-form solution to natural image matting [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 30(2): 228-242.
- [12] 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.
- [13] XU Ning, PRICE B, COHEN S, et al. Deep image matting [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 2970-2979.
- [14] LI Yaoyi, LU Hongtao. Natural image matting via guided

- contextual attention [C] // Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2020, 34 (7): 11234–11241.
- [15] SUN Yanan, TANG C K, TAI Y W. Semantic image matting [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 11120–11129.
- [16] PARK G T, SON S J, YOO J Y, et al. Matteformer: Transformer-based image matting via prior-tokens [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 11696–11706.
- [17] CHO K, VAN MERRIËNBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [J]. arXiv preprint arXiv, 1406.1078, 2014.
- [18] YU Qihang, ZHANG Jianming, ZHANG He, et al. Mask guided matting via progressive refinement network [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 1154–1163.
- [19] HOU Qiqi, LIU Feng. Context-aware image matting for simultaneous foreground and alpha estimation [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE, 2019: 4130–4139.
- [20] PORTER T, DUFF T. Compositing digital images [C] // Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM, 1984: 253–259.