

张岚岚. 面向自然图像透明区域的特征感知抠图[J]. 智能计算机与应用, 2024, 14(12): 139–144. DOI: 10.20169/j.issn.2095-2163.24071102

面向自然图像透明区域的特征感知抠图

张岚岚^{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: 自然图像抠图作为计算机视觉的重要任务之一, 是一种通过预测精确的 alpha matte 来将前景物体从图像中分离出来的技术。由于透明区域前背景像素高度混合, 在提取透明区域遮罩时易出现误抠问题。为了解决这一问题, 本文提出透明特征感知抠图模型, 模型中带有专门针对透明区域遮罩误抠设计的透明特征感知模块。该模块能够多尺度提取上下文信息, 精确感知并分离前背景像素高度混合的透明区域, 同时提取透明的细致特征, 从而有效解决了自然图像中透明区域的遮罩误抠问题。在 Composition-1k 测试集上的实验结果表明, 透明特征感知抠图模型在解决透明区域遮罩误抠问题上具有一定的优越性。与经典的 DIM 抠图方法相比, 本文方法在 *MSE*、*SAD*、*Grad* 和 *Conn* 四个评价指标上分别提升了 45.71%、30.95%、47.74% 和 39.96%。

关键词: 图像抠图; 透明区域; 感知; 上下文信息

中图分类号: TP391

文献标志码: A

文章编号: 2095-2163(2024)12-0139-06

Feature-aware matting for transparent regions in natural images

ZHANG Lanlan^{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: Natural image matting, as a critical task in computer vision, involves separating foreground objects from images by predicting precise alpha mattes. Due to highly blended foreground-background pixels in transparent regions, errors in mask extraction are common. To address this issue, this paper proposes a transparent feature perception matting model equipped with a transparent feature perception module, which is specifically designed to mitigate mis-matting in transparent region masks. This module extracts contextual information across multiple scales, accurately perceives and separates highly blended foreground-background pixels in transparent regions, and extracts detailed transparent features, effectively resolving mis-matting issues in natural images. Experimental results on the Composition-1k dataset demonstrate the superiority of the transparent feature perception matting model while addressing mis-matting issues in transparent region masks. Compared to classical DIM matting methods, the proposed approach shows improvements of 45.71% in *MSE*, 30.95% in *SAD*, 47.74% in *Grad*, and 39.96% in *Conn* evaluation metrics.

Key words: image matting; transparent regions; perception; contextual information

0 引言

自然图像抠图作为计算机视觉的重要任务之一, 广泛应用于图像合成^[1]、虚拟现实^[2]等多个领域。是一种通过预测精确的 alpha matte 来将前景物体从图像中分离出来的技术。自然图像抠图把图

像 I 建模为前景 F 与背景 B 的凸组合^[3], 即在该模型中, 任意像素点 p 的颜色 I_p 可以由原始图像中的前景颜色 F_p 和背景颜色 B_p 线性表示, 数学表达式如下所示:

$$I_p = \alpha_p F_p + (1 - \alpha_p) B_p \quad (1)$$

其中, $\alpha_p \in [0, 1]$, 表示输入图像 I 中像素 p 处所对应前景物体的不透明程度, 即透明度遮罩

基金项目: 贵州省科技计划项目 (QKHJCZK2022YB195, QKHJCZK2023YB143, QKHPTRCZCKJ2021007); 贵州省教育厅自然科学研究项目 (QJJ2023061, QJJ2023012, QJJ2022015)。

作者简介: 张岚岚 (1998—), 女, 硕士研究生, 主要研究方向: 数字图像处理。Email: 1302947404@qq.com。

收稿日期: 2024-07-11

哈尔滨工业大学主办 ◆ 专题设计与应用

(alpha matte)。式(1)中只有 I_p 是已知, α_p 、 F_p 和 B_p 未知, 所以该方程是一个欠定方程。为了解决这一问题, 许多抠图方法利用三分图(Trimap)作为一种约束信息来缩减解空间。三分图由3个部分组成, 即: 白色、黑色和灰色区域, 分别表示前景、背景和未知区域。

现有的抠图方法分为基于传统的抠图方法和基于深度学习的抠图方法。

传统的抠图方法分为采样和传播两种方式。Karacan 等学者^[4]将抠图采样过程建模为对已知区域像素进行稀疏子集选择的问题, 利用 KL 散度(Kullback-Leibler divergence)对样本集的稀疏度进行定量评价, 提出基于 KL 散度采样的抠图算法。该算法不仅实现全局采样, 提高了已知区域像素样本集的多样性, 而且控制了样本集合大小。Levin 等学者^[5]假设图像的前景和背景在局部区域内是平滑的, 提出了一种基于闭式解法的图像抠图方法。该方法通过构建一个基于拉普拉斯矩阵的线性系统, 来求解每个像素的 alpha 值, 从而实现精确的图像抠图。Chen 等学者^[6]提出一种基于 K 近邻的图像抠图方法。该方法利用 K 近邻算法在高维特征空间中搜索与待抠图像素相似的像素, 并基于这些相似像素的 alpha 值进行插值计算, 从而得到目标像素的 alpha 值。Zhu 等学者^[7]提出基于自适应透明度的传播抠图算法, 该算法从一个新的角度考虑图像抠图。根据图像中前景对象的透明度, 将其分为3类, 即高透明、强透明和弱透明, 可以根据输入图像的透明度类型进行相应的调整。传统抠图方法在非透明区域效果很好, 但由于透明区域的透明特性, 前背景像素高度混合, 传统方法没有考虑前景对象或背景场景的任何语义信息, 因此提取透明区域遮罩时容易失败。

由于深度卷积神经网络的成功研发, 基于学习的方法在自然图像抠图中占据主导地位。Xu 等学者^[8]提出了一种基于卷积神经网络的深度图像抠图方法, 该方法使用端到端的训练方式, 结合图像和三分图, 来预测每个像素的 alpha 值。Hou 等学者^[9]提出了一种基于上下文感知的自然图像抠图方法。该方法使用抠图编码器来学习局部特征, 使用上下文编码器来获得更多全局上下文信息, 将这2个编码器的输出连接在一起, 并将其送入解码器网络, 以同时估计前景和 alpha matte。Li 等学者^[10]提出了一种新的端到端自然图像抠图方法, 使用引导式上下文注意模块, 通过低级亲和度全局传

播高级不透明度信息, 实现精确抠图。虽然以上基于学习的方法利用神经网络从注释良好的数据集中学习颜色信息和自然结构, 其性能优于传统方法。但是在处理透明区域产生模糊的结构或纹理方面, 不能有效反映前景的特殊结构。

综上所述, 尽管基于深度学习的抠图方法已取得成果, 然而, 目前的深度模型较少关注透明复杂区域, 在处理透明区域时会出现模糊, 易导致前景遮罩误抠问题(即前景遮罩抠错)。因此, 对自然图像中透明区域的误抠问题仍需进一步的研究。

针对上述问题, 本文提出了一个透明特征感知模块。该模块通过结合 Inception 模块^[11]、扩张卷积^[12]、卷积块注意力(Convolutional Block Attention Module, CBAM)^[13] 和多维协作注意力(Multidimensional Collaborative Attention, MCA)^[14], 能够精确感知并分离前景和背景像素高度混合的透明区域, 同时提取前景的细致特征。透明特征感知模块在多尺度特征提取方面表现出色, Inception 模块通过并行的多尺度卷积核, 实现对不同尺寸的透明物体进行有效处理, 这种多尺度处理方法使得模块能够在不同尺度的透明物体上表现出优越的适应性, 确保了处理结果的精确性和一致性。模块中的扩张卷积进一步增强了对透明区域的感知能力。扩张卷积能够在不增加计算量的情况下扩展卷积核的感受野, 从而捕捉到更大范围内的上下文信息。这对于处理复杂背景和前景的混合区域尤为重要, 因其能够提供更多的背景信息, 帮助区分前景和背景。CBAM 通过通道注意力和空间注意力机制, 能够有效地筛选出最重要的特征, 抑制噪声, 提高透明区域的特征提取精度。MCA 能够在多个尺度上提取图像特征, 捕捉不同层次的信息, 整合了丰富的上下文信息, 能够精确捕捉并分离这些透明区域的混合像素, 提取出前景的细致特征。

1 透明特征感知抠图模型

本文采用文献[10]中的基线模型, 并在此基础上引入透明特征感知模块, 形成透明特征感知抠图模型, 如图1所示。

图1中, 蓝色块表示基线模型的残差块, 橙色块是本文提出的透明特征感知模块, 该模块位于编码器末端。网络输入为 RGB 图像与三分图拼接形成的6通道张量, 通过编码器-解码器结构最终输出预测的前景遮罩。

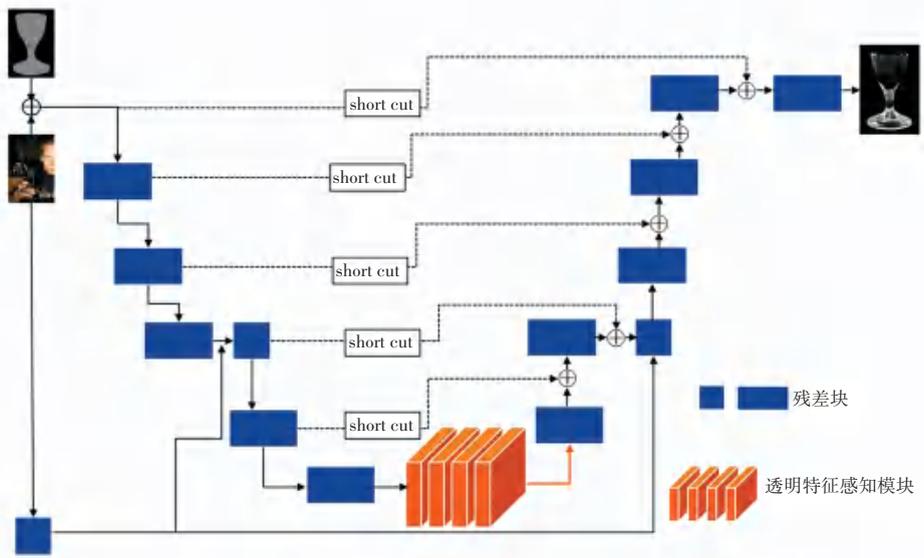


图 1 模型结构图

Fig. 1 Diagram of model structure

2 透明特征感知模块

CFP (Channel-wise Feature Pyramid) 能够捕捉不同尺度的上下文信息,但由于透明区域的特征往往细微且复杂,CFP 在处理透明区域的遮罩提取时仍存在一些不足。具体而言,CFP 在区分这些混合区域时的能力不足,未能充分利用注意力机制来动态调整特征权重,从而在复杂场景下提取透明区域的细致特征方面存在局限性。为了克服上述不足,受文献[15]启发,本文提出了透明特征感知模块。该模块中结合 Inception 模块、扩张卷积、CBAM 和 MCA 优势,能多尺度提取特征图、上下文信息并捕捉细致透明特征。该模块包含 K 个具有不同扩张率的 FP (Feature Pyramid) 通道,扩张率 $\{r_1, r_2, \dots, r_k\}$ 表示不同通道的扩张参数,这些扩张参数用于

调整卷积核的感受野,从而捕捉不同尺度的信息。模块首先使用 1×1 卷积将高维特征图投影到低维。然后,应用 CBAM 模块,更有效地关注透明物体前景的特征信息。CBAM 通过通道和空间注意力机制,自适应地调整特征权重,使模型能够更好地区分透明物体与背景。再将多个 FP 通道设置为具有不同扩张率的并行结构。将所有特征图连接到输入的维度,并使用另一个 1×1 卷积来激活。最后,经过 MCA 输出,MCA 模块通过多维协作注意力机制,在多个尺度上精确捕捉透明区域的细节,动态调整特征权重,增强了特征提取的精度和效率。

本文所提的透明特征感知模块有助模型能更有效捕获图像中透明区域的轮廓、边界和纹理信息等一些细微结构,有助于区分与背景不同的重要特征。透明特征感知模块如图 2 所示。

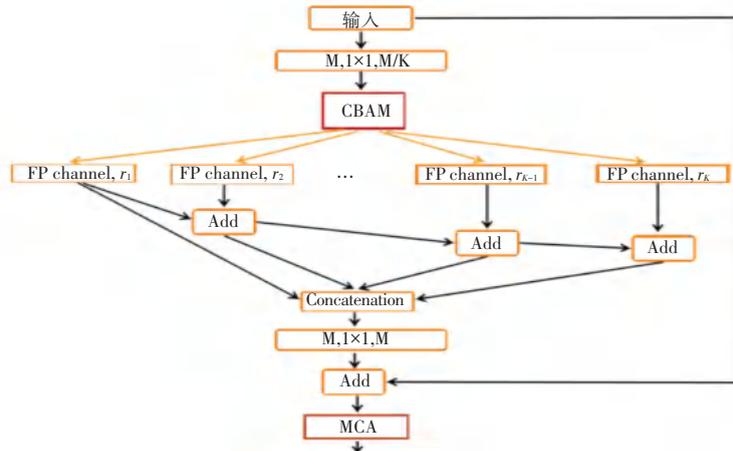


图 2 透明特征感知模块

Fig. 2 Transparent feature awareness module

3 损失函数

本文采用回归损失、合成损失^[8]和拉普拉斯损失^[9],并将其分别表示为 L_1 、 L_{comp} 、 L_{lap} 。这里将给出探讨分述如下。

(1) L_1 。具体公式如下:

$$L_1 = \frac{1}{U} \sum_{i \in U} |\hat{\alpha}_i - \alpha_i| \quad (2)$$

其中, U 表示三分图中标注的未知区域, α_i 和 $\hat{\alpha}_i$ 分别表示像素 i 处的预测透明度遮罩值和真实透明度遮罩值。

(2) L_{comp} 。具体公式如下:

$$L_{comp} = \sqrt{(c_p - c_g)^2 + \varepsilon^2} \quad (3)$$

其中, c_p 表示由预测透明度遮罩合成的图像; c_g 表示由真实透明度遮罩合成的图像; ε 表示一个小数,以避免合成损失为零。

(3) L_{lap} 。具体公式如下:

$$L_{lap} = \sum_{i=1}^5 2^{i-1} \|L^i(\hat{\alpha}) - L^i(\alpha)\|_1 \quad (4)$$

其中, L^i 表示alpha图的拉普拉斯金字塔的第 i 层, α 和 $\hat{\alpha}$ 分别表示预测透明度遮罩和真实透明度遮罩。

(4)总损失函数 L 。是 L_1 、 L_{comp} 、 L_{lap} 的加权和,即:

$$L = L_1 + L_{comp} + L_{lap} \quad (5)$$

4 实验结果与分析

本文在Composition-1k数据集^[8]上对透明特征感知抠图模型进行了200 000次迭代训练,该数据集包含431组前景图像和相应的alpha matte用于训练,50组用于测试。模型训练使用Adam优化器^[16],Adam优化器的参数设置为 $\beta_1 = 0.5$ 、 $\beta_2 = 0.999$ 。初始学习率为 4×10^{-4} ,批量大小为10。所有抠图模型的实现均在PyCharm平台下进行,使用深度学习框架PyTorch(版本为Python-3.8.17+cuda11.8)编写,编译软件为JupyterLab。所有实验均在NVIDIA GeForce RTX A100 GPU(30G内存)服务器上运行,整个训练过程耗时约2天。

为验证本文所提出方法的有效性,本文在Composition-1k测试集上评估其性能,选用4个评估指标来评估模型性能:均方误差(Mean Squared Error, MSE),绝对误差和(Sum of Absolution Difference, SAD),梯度误差(Gradient, $Grad$)和连

通性误差(Connectivity, $Conn$)^[3]。

本文方法与KNN^[6]、Closed Form^[5]、和DIM^[8]、IndexNet^[17]、GCA^[10]、CAM^[9]、A²U^[18]、PIAMatting^[19]和LCDM^[20]经典抠图方法进行定量比较,结果见表1,最优结果加粗表示。

表1 Composition-1k测试集定量结果

Table 1 Quantitative results on Composition-1k test set				
方法	$MSE/10^{-3}$	SAD	$Grad$	$Conn$
KNN	103.0	175.4	124.1	176.4
Closed Form	91.0	168.1	126.9	167.9
DIM	14.0	50.4	31.0	50.8
IndexNet	13.0	45.8	25.9	43.7
CAM	8.0	35.8	17.3	33.2
GCA	9.1	35.3	16.9	32.5
A ² U	8.2	32.2	16.4	29.3
PIAMatting	9.0	36.4	16.9	31.5
LCDM	21.0	62.1	25.7	43.4
our	7.6	34.8	16.2	30.5

由表1可以知道,在Composition-1k测试集上,与其他9种抠图方法相比,本文提出的方法在MSE和Grad两个指标上均获得了最优的结果,表明其在处理图像抠图问题上具有显著优势。尽管在SAD和Conn指标上略低于A²U模型,但在图像细节保留和边缘清晰度的处理等方面表现突出,特别是在复杂背景下对透明物体的精准抠取和前景细节的保留上,取得了明显的提升。相比于经典DIM模型,MSE、SAD、Grad和Conn四个指标分别提升45.71%、30.95%、47.74%和39.96%,相比于基线GCA模型,MSE、SAD、Grad和Conn四个指标分别提升16.48%、1.42%、4.14%和6.15%。

本文所提出的方法与其他抠图方法在Composition-1k数据集进行可视化比较,其可视化结果如图3所示。图3中,(a)、(b)分别是Image和Trimap,(c)~(f)分别是IndexNet、A²U、GCA、本文方法预测的结果,(g)是真实遮罩(Ground Truth, GT)。从图3的可视化效果可以看出,本文的方法在处理透明区域时,相较于其他方法,生成的遮罩图效果更接近于真实遮罩图,误抠现象明显减少。在各方法可视化效果图对比中,本文所提方法处理的透明区域清晰度更高,能够更好地分离透明区域的前景和背景,提取出更精细的透明特征,且没有误抠到透明区域的背景,而其他方法明显将背景特征误抠出来了。尤需指出的是,第1列展示了在高透明度的玻璃物体上的表现,本文方法能够准确捕捉到高透明度物体的细节。第4列展示了提取网的遮罩可视化效果图,其中IndexNet、A²U和GCA方法抠取到的网遮罩是断开

的,而本文方法抠出来的网的遮罩是连接的。综上所述可知,本文方法在提取透明区域遮罩方面表现出明显优势。

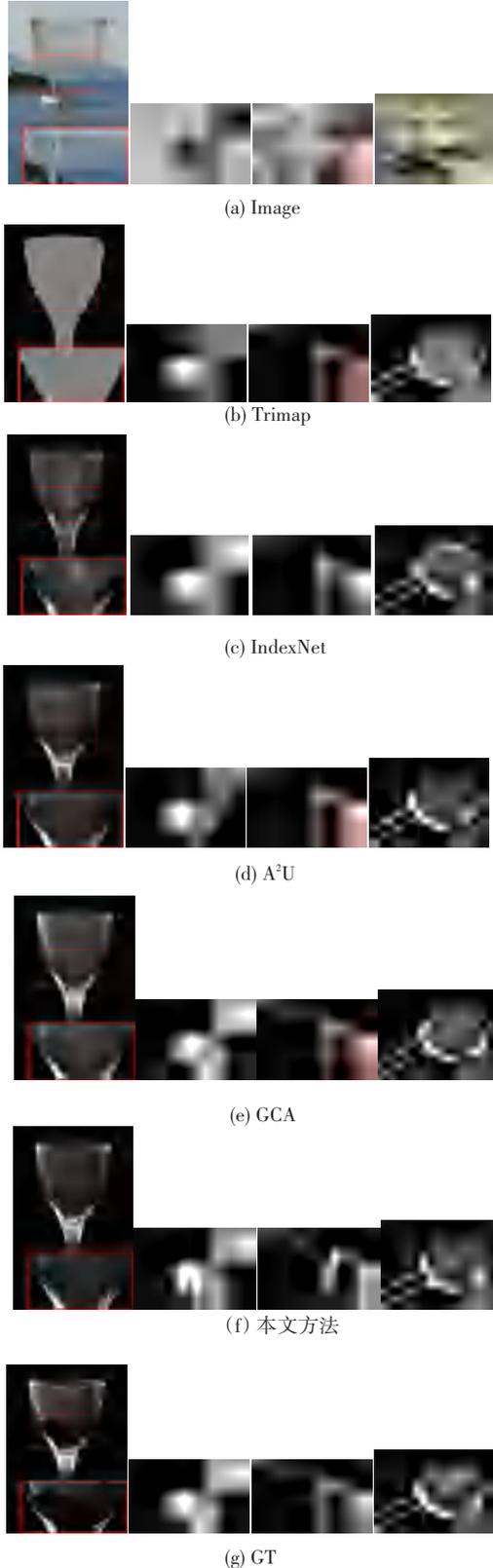


图 3 Composition-1k 测试集上的可视化

Fig. 3 Visualization on Composition-1k test set

5 结束语

本文提出透明特征感知抠图模型来解决在提取透明区域遮罩时易出现的误抠问题。为此,本文引入了透明特征感知模块,通过多尺度提取上下文信息,精确感知并分离前景和背景像素高度混合的透明区域,同时提取前景的细致特征,从而有效解决自然图像中透明区域的前景遮罩误抠问题。在 Composition-1k 测试集上的实验结果表明了该方法的优越性。

参考文献

- [1] MAQSOOD S, JAVED U, RIAZ M M, et al. Multiscale image matting based multi-focus image fusion technique [J]. Electronics, 2020, 9(3): 472.
- [2] CHEN Tao, CHENG Mingming, TAN Ping, et al. Sketch2photo: Internet image montage [J]. ACM Transactions on Graphics, 2009, 28(5): 1-10.
- [3] PORTER T, DUFF T. Compositing digital images [C]// Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques. Minneapolis, USA: dblp, 1984: 253-259.
- [4] KARACAN L, ERDEM A, ERDEM E. Image matting with KL-divergence based sparse sampling [C]// Proceedings of the IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2015: 424-432.
- [5] 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.
- [6] CHEN Qifeng, LI Dingzeyu, TANG C K. KNN matting [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(9): 2175-2188.
- [7] ZHU X, WANG P, HUANG Z. Adaptive propagation matting based on transparency of image [J]. Multimedia Tools and Applications, 2018, 77: 19089-19112.
- [8] 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.
- [9] 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.
- [10] LI Yaoyi, LU Hongtao. Natural image matting via guided contextual attention [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 11450-11457.
- [11] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 1-9.
- [12] HOLSCHIEDER M, KRONLAND-MARTINET R, MORLET J, et al. A real-time algorithm for signal analysis with the help of the wavelet transform [M]// COMBES J M, GROSSMANN A,

- TCHAMITCHIAN P. Wavelets. Inverse Problems and Theoretical Imaging. Cham: Springer, 1990: 286-297.
- [13] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module [M]//FERRARI V, HEBERT M, SMINCHISESCU C. Computer Vision - ECCV 2018. Lecture Notes in Computer Science. Cham:Springer, 2018,11211: 3-19.
- [14] YU Yang, ZHANG Yi, CHENG Zeyu, et al. MCA: Multidimensional collaborative attention in deep convolutional neural networks for image recognition [J]. Engineering Applications of Artificial Intelligence, 2023, 126: 107079.
- [15] LOU A, LOEW M. Cfpnet: Channel-wise feature pyramid for real-time semantic segmentation [C]//2021 IEEE International Conference on Image Processing (ICIP). Piscataway, NJ: IEEE, 2021: 1894-1898.
- [16] DIEDERIK P K. Adam: A method for stochastic optimization [J]. arXiv preprint arXiv,1412.6980, 2014.
- [17] 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.
- [18] DAI Yutong, LU Hao, SHEN Chunhua. Learning affinity-aware upsampling for deep image matting [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ:IEEE, 2021: 6841-6850.
- [19] LIU Yuyao, XIE Jiake, QIAO Yu, et al. Prior - induced information alignment for image matting [J]. IEEE Transactions on Multimedia, 2021, 24: 2727-2738.
- [20] FENG Fujian, HUANG Han, LIU Di, et al. Local complexity difference matting based on weight map and alpha mattes [J]. Multimedia Tools and Applications, 2022, 81 (30): 43357 - 43372.