

人像抠图无监督语义精修算法

曾广荣, 程良伦, 卢增

广东工业大学, 广东 广州

Email: 994820121@qq.com

收稿日期: 2020年12月25日; 录用日期: 2021年1月19日; 发布日期: 2021年1月26日

摘要

针对当前不使用三分图作为先验知识的人像抠图算法在远景人像抠图任务中存在多余的干扰信息、人像边缘轮廓粗糙、人体携带物品易与背景混淆等问题, 提出了人像抠图无监督语义精修算法。该算法由人像边框感知模块与无监督语义精修模块组成。人像边框感知模块首先使用了行人检测模型识别出所有人像, 并结合边框感知算法来去除多余的干扰信息。无监督语义精修模块利用了无监督语义分割模型提取特征, 然后使用语义精修算法进行人像轮廓的修复。实验表明, 在自制的远景人像数据集中, 使用主流的人像抠图算法作为基线, 并加入人像抠图无监督语义精修算法后, 效果得到了明显的提高, 人体携带物品也能精准识别, 人像轮廓也更加清晰。同时在半身人像数据集中, 效果也有一定的提升, 表明了该算法也具有泛用性。

关键词

人像抠图, 深度学习, 语义精修

Unsupervised Semantic Human Matting Refinement

Guangrong Zeng, Lianglun Cheng, Zeng Lu

Guangdong University of Technology, Guangzhou Guangdong

Email: 994820121@qq.com

Received: Dec. 25th, 2020; accepted: Jan. 19th, 2021; published: Jan. 26th, 2021

Abstract

Aiming at the problems of human matting algorithm, which does not use trimap as a prior knowledge, such as redundant interference information, rough portrait edge contour, and easy confusion between objects carried by human body and the background, an unsupervised semantic matting refinement algorithm is proposed. The algorithm consists of a human body edge perception module and an unsupervised semantic matting refinement module. The human body edge perception module first uses a pedestrian detection model to identify all human bodies, and then combines the edge perception algorithm to remove redundant interference information. The unsupervised semantic matting refinement module uses an unsupervised semantic segmentation model to extract features, and then uses the semantic matting refinement algorithm to repair the human body contour. Experiments show that in the self-made long-distance human body data set, using the mainstream human body matting algorithm as a baseline, and adding the human body matting refinement algorithm, the effect is significantly improved, human body carried objects can be accurately identified, and the human body contour is also clearer. At the same time, in the half-body human body data set, the effect also has a certain improvement, indicating that the algorithm also has generalizability.

ting algorithm for human matting is proposed. The algorithm is composed of human border sensing module and unsupervised semantic refinement module. The portraits border sensing module firstly uses the pedestrian detection model to identify all the portraits, and combines the border sensing algorithm to remove the redundant interference information. Unsupervised semantic refinement module uses unsupervised semantic segmentation model to extract features, and then uses semantic refinement algorithm to repair the portrait contour. Experiments show that in the self-made long-term portrait data set, the mainstream portrait matting algorithm is used as the baseline, and the unsupervised semantic refinement algorithm for portrait matting is added. The effect is significantly improved, and the objects carried by the human body can also be accurately identified. The outline is also clearer. At the same time, in the portrait data set, the effect is also improved to some extent, indicating that the algorithm is also generalized.

Keywords

Human Matting, Deep Learning, Semantic Refinement

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着相机、手机等硬件设备的发展，需要后期处理的图像也越来越多。但由于图像具有多样性与复杂性，无论传统图像处理方法还是深度学习方法，都难以做到对任何输入完全鲁棒。所以，针对特定类别的图像抠图算法[1]更具有实际应用意义。

人像抠图是语义分割[2] [3]的一种，是从图像或者视频中提取出人像前景，使人像与背景准确分离的一种技术。该技术在现实中有着广泛的应用，比如虚拟现实、增强现实、电影制作和摄影合成，这些应用主要基于图像合成技术。在图像合成技术中，通过人像抠图技术可以得到所需要的 mask，mask 中每一个像素点的值表示了原始图像每个像素是否属于人像前景，其精度直接影响合成图像质量的优劣。

对于图像 I ，人像抠图旨在找到人像前景 F 与背景 B 的最佳线性组合。对于任意的像素点 i ，可以满足以下公式：

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

其中 I_i 为图像 I 的像素点 i ， α_i 表示在像素点 i 的前景不透明度。对于 RGB 图像，有 7 个未知变量： $F_i^{(r)}$ ， $F_i^{(g)}$ ， $F_i^{(b)}$ ， $B_i^{(r)}$ ， $B_i^{(g)}$ ， $B_i^{(b)}$ ， α_i ，而已知变量仅有 3 个： $I_i^{(r)}$ ， $I_i^{(g)}$ ， $I_i^{(b)}$ 。由于已知变量远小于未知变量，因此需要用户提供额外的先验知识才能求解，所以多数图像抠图算法[4] [5] [6] [7]利用三分图 (trimap) 作为先验知识，再进行图像抠图任务。比如 Cho [8] 等人结合了 CF 抠图[9]与 KNN [5] 抠图两种方法，以两种方法的输出和 RGB 图像作为 CNN 网络的输入来学习图像与 mask 的映射关系。Xu 等人[10]用 RGB 图像与三分图来输入并集成了编码解码器结构和细化网络来预测 mask。Lutz [11] 等人则使用生成式对抗网络进行图像抠图任务。三分图是用户指定一部分已知的前景区域与已知的背景区域，然后求解未知的区域 α 。然而这个方法需要用户有一定水平的专业知识而且可能需要很长的时间才能得到满意的效果。因此近年来有学者尝试舍弃三分图作为先验知识并以端到端的方式学习图像与对应的 mask 之间的映射关系。比如 Chen [12] 等人从图像中学习隐性语义约束而不是用三分图或者涂鸦等先验知识来生成 mask。Liu 等人[13]利用粗糙注释数据与精细注释数据相结合的方法，在不使用三分图的情况下，实现端

到端的 mask 生成。但是这些方法在远景人像抠图中并不能达到满意的效果，原因有三个，首先是远景容易出现多余的干扰信息、其次是人体边缘轮廓的细化，最后是人体的携带物体容易与背景进行混淆。

因此，本文自制了远景人像数据集来进行实验。并提出了人像抠图无监督语义精修算法对 mask 进行语义精修。为了验证算法的效果，本文以 SHM 算法作为基线，并加上人像抠图无监督语义精修算法进行对比实验分析，表明了本文算法的优越性能。本文贡献如下：1) 为了去除人像背景噪音，结合了行人检测模型并提出了人像边框感知算法。2) 把无监督语义分割网络应用于人像抠图，并提出了人像精修模块来细化人体边缘轮廓和分离背景与人体携带物。3) 自制了远景人像抠图数据集，并使用充足的实验验证了算法的有效性。

2. 人像抠图无监督语义精修算法

由于使用当前主流的人像抠图模型对远景人像进行人像抠图得到的 mask 存在噪音以及语义缺失问题。针对这些问题，本文提出了人像抠图无监督语义精修算法，该算法可以对人像抠图后的 mask 进行图像去噪和语义精修。算法如图 1 所示，由人像边框感知模块与无监督语义精修模块组成。人像边框感知模块通过行人检测模型与人像边框感知算法对输入的 mask 进行去噪。无监督语义精修模块则通过无监督语义分割网络并结合人像精修算法对去噪后的 mask 进行语义精修，得到更为精细完整的人像抠图。

2.1. 人像边框感知模块

如图 1 中的人像边框感知模块所示，人像边框感知模块的输入为 RGB 图像和 mask，首先使用行人检测模型[14]对 RGB 图像进行行人检测，得到人像边框 $B = (x_1, x_2, y_1, y_2)$ ，其中 (x_1, y_1) 表示人像边框左上角的坐标， (x_2, y_2) 表示人像边框右下角的坐标。随后，使用人像边框感知算法对边框 B 进行修正并同时去噪。如图 2 所示，该算法主要分为两个部分：人像边框内的修正算法与人像边框外的噪音去除方法。人像边框内的修正算法用于边框 B 没有完全框住人像的场景，该算法可以对人像边框 B 进行修正以达到完全框住人像的目的。

如图 2(a)所示，行人的乐器在行人检测模型没有检测出来的情况下，人像边框内的修正算法可以很好地进行边框修正。具体实现方法如算法 1 所示。在得到完整的人像边框后，本文使用人像边框外噪音去除方法进行去噪。如图 2(b)所示，该方法只保留面积最大的人像边框内的前景信息，并把该边框之外的前景信息设为背景信息，以去除人像背景的噪音。

算法 1. 人像边框内的修正算法

Input: $B = (x_1, x_2, y_1, y_2)$ 表示人像边框， $mask$ 表示透明度遮罩。

Output: B' 表示修正后的人像边框。

- 1) initialize set $B' = 0$
- 2) for $i \leftarrow x_1$ to x_2 do
- 3) while $mask_{i,y_1} > 0$
- 4) do $y_1 \leftarrow y_1 + 1$ end
- 5) while $mask_{i,y_2} > 0$
- 6) do $y_2 \leftarrow y_2 - 1$ end
- 7) end for
- 8) for $i \leftarrow y_2$ to y_1 do
- 9) while $mask_{x_1,i} > 0$
- 10) do $x_1 \leftarrow x_1 - 1$ end

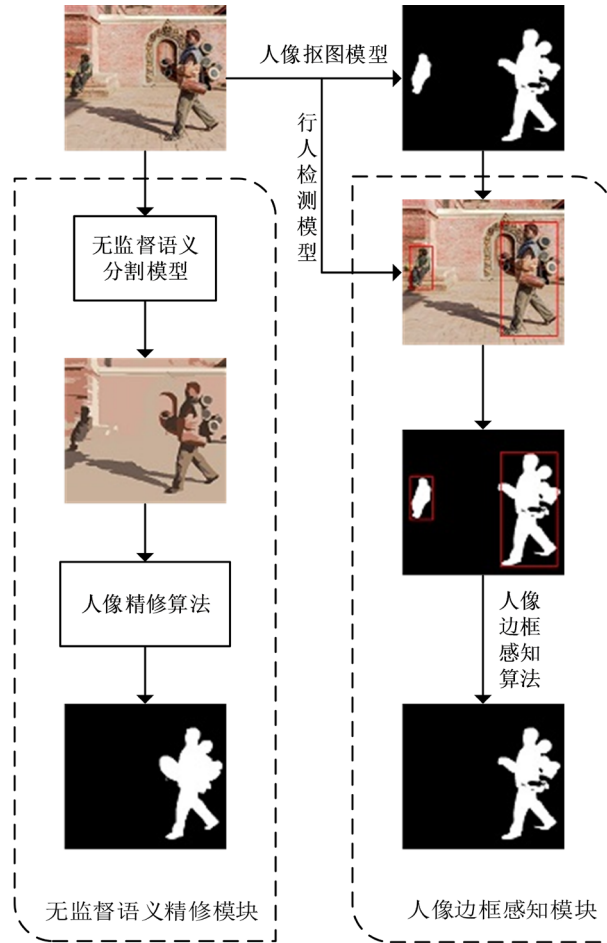


Figure 1. The flow chart
图 1. 整体流程图

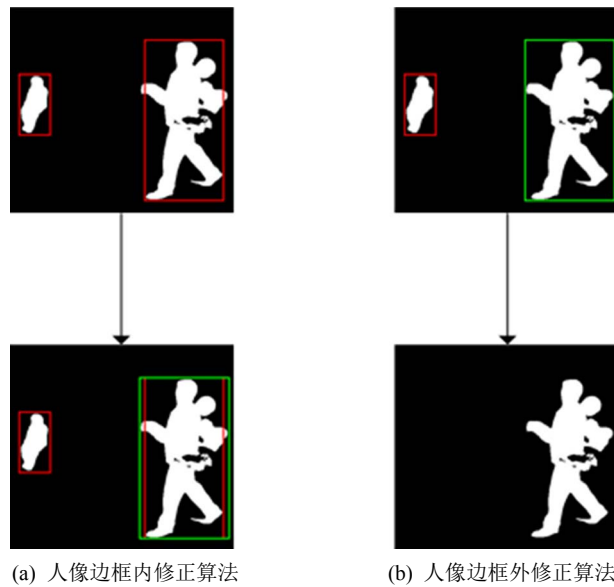


Figure 2. Portrait border sensing algorithm flow chart
图 2. 人像边框感知算法流程图

- 11) while $mask_{x_2,i} > 0$
- 12) do $x_2 \leftarrow x_2 + 1$ end
- 13) end for
- 14) $B' \leftarrow (x_1, x_2, y_1, y_2)$

2.2. 无监督语义精修模块

如图 1 中的无监督语义精修模块所示, 该模块包括无监督人像语义分割网络与人像精修算法, 它的输入为人像边框感知模块的输出和 RGB 图像, 输出为精修后的 mask。

无监督语义分割网络[15]如图 3 所示, 输入 RGB 三通道图像 P , 表示为 $P = \{x_n \in \mathbb{R}^3\}_{n=1}^N$, 其中 N 为像素点的个数。每个像素点先正则化到 $[0, 1]$ 区间内, 再使用无监督聚类算法 f 对图像 P 的像素点进行语义预分类, 分成 k 类, 即 $y_n = f(x_n)$, y_n 是一个 $1 \times k$ 的矩阵, 该矩阵元素为 1 的下标即为该像素点 n 的类别。给每个像素分配标签后, 卷积神经网络可以根据该标签实现无监督学习。该网络包括特征提取部分与特征分类部分。特征提取部分由 M 个特征提取块组成, 特征提取块包括卷积层, 批归一化层(BN)和线性激活层(ReLU), 其中在卷积层上使用的是 3×3 的卷积核。

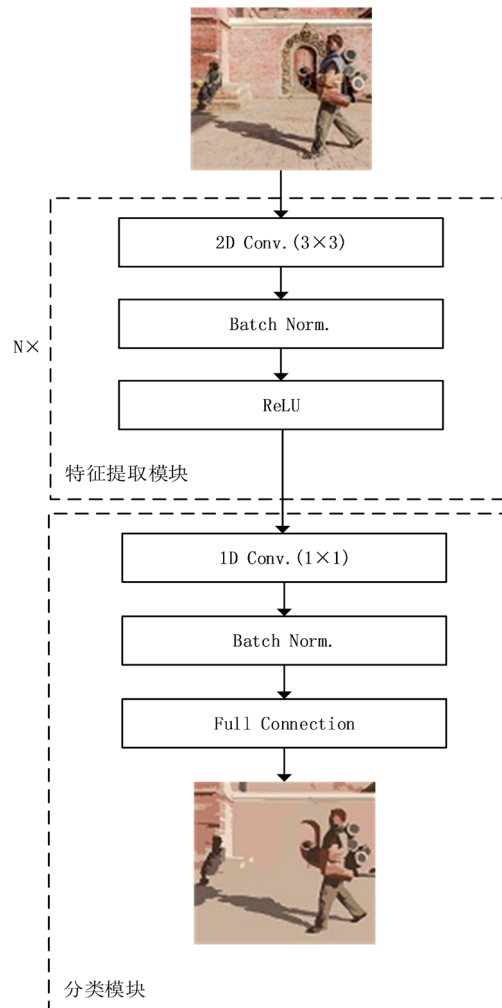


Figure 3. Unsupervised portrait semantic segmentation network
图 3. 无监督人像语义分割网络

$$h_i = \text{ReLU}(\text{BN}(\text{conv}_{3 \times 3}(h_{i-1}))) \quad (2)$$

其中 h_i 为第 i 个特征提取块的输出。批归一化层可以使输入到 ReLU 的图片接近正态分布。特征分类部分由使用 1×1 卷积核的卷积层, BN 层和全连接层组成。

$$\mathbf{G} = \text{classification}(\text{BN}(\text{conv}_{1 \times 1}(h_M))) \quad (3)$$

其中 \mathbf{G} 为 $k \times N$ 维的矩阵。最后网络的训练需要计算矩阵 \mathbf{G} 与标签 \mathbf{Y} 的损失, 本文使用 Softmax Loss 作为损失函数。

$$\text{Loss} = \frac{1}{N} \sum_{i=1}^N \left(- \sum_{j=1}^k y_{i,j} \log \frac{e^{G_{i,j}}}{\sum_{l=1}^k e^{G_{i,l}}} \right) \quad (4)$$

RGB 三通道图像经过无监督人像语义分割网络后每个像素点被分为 q 类。相同类别且连续的像素点构成区域 P_k , 其中 $k \in \{1, 2, \dots, q\}$ 。人像精修算法的核心在于统计 mask 在每个区域 P_k 中前景与背景的个数, 本文把像素点为黑色设为背景, 像素点为白色设为前景。如果 mask 在区域 P_k 中前景的数量与背景的数量比值大于 θ , 则把 mask 在区域 P_k 中的值都设置为前景的值, 如果前景的数量与背景的数量比值小于 $(1-\theta)$, 则把 mask 在区域 P_k 中的值都设置为背景的值。其中 θ 为超参数。该算法可以修改 mask 中的属于背景信息的前景信息, 同时可以细化人像及携带物品的边缘轮廓。

3. 实验设计及结果分析

3.1. 实验准备

由于目前人像抠图数据集大多为半身人像图片, 因此本文自制了一个远景人像数据集来进行实验设计与分析。该数据集的图像均采集于互联网, 而且大多选择人像前景与背景变化较小的并且有携带物体的图片。数据集共有 2000 张图片, 分辨率为 1920 像素 \times 988 像素, 标注内容包括 2703 个行人目标框, 2000 张 mask 图。本文将该数据集按 8:1:1 分为训练集、验证机、测试集。因此随机抽取 1600 张图像所为训练集, 200 张图像作为验证集, 200 张图像作为测试集。

本文的实验环境: ubuntu16.04 操作系统, Intel i7 中央处理器, 1 TB 固态硬盘, 32 GB 内存, NVIDIA GTX2080Ti GPU \times 2, 并使用 Pytorch 深度学习框架进行训练。

3.2. 实验结果与分析

由于 SHM 算法是在人像抠图任务中率先舍弃用三分图作为先验知识的算法, 并且是广泛使用的人像抠图算法, 因此本文选择把 SHM 算法当作基线进行对比实验分析。

本文采用 2 个广泛使用的评价指标进行评测, 分别为均方误差 MSE (mean-square error) 和绝对误差 SAD (sum of absolute differences)。这 2 个指标的值越低表示效果越好。公式如下:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2 \quad (5)$$

$$\text{SAD} = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (6)$$

其中 x_i 为实验获得的图像; y_i 是 ground true 图像。

在人像边框感知模块的实验中, 本文使用 YOLOv3 算法[14]在自制数据集上训练行人检测模型。对输入图片进行行人检测得到人像边框后, 使用人像边框感知算法对标签图片进行去噪。

在无监督语义精修模块的实验中, 迭代次数为 500, 学习率设为 0.0001, 并选用 Adam 算法[16]作为

优化器进行优化。对于不同的超参数 θ ，在表 1 中展示了使用 Felzenszwalb 算法[17]和 SLIC 算法[18]作为无监督聚类算法的实验效果，评价指标为使用语义精修算法后与使用之前降低均方误差。

Table 1. The results of two pre-classification algorithms in different hyperparameters θ
表 1. 两种预分类算法在不同超参数 θ 的实验结果

预分类算法	θ 取值	MSE ($\times 10^{-3}$)
SLIC	0.6	3.1
SLIC	0.7	5.6
SLIC	0.8	5.4
Felzenszwalb	0.6	4.2
Felzenszwalb	0.7	7.5
Felzenszwalb	0.8	6.8

该数据显示，使用 Felzenszwalb 算法在 θ 为 0.7 的时候，实验效果最好，MSE 平均降低 0.0075。因为 Felzenszwalb 算法相比于 SLIC 算法，对于变化幅度小的照片，能很好地处理细节；对于变化幅度大的照片，也能很好地忽略一些细节。同时发现当 θ 取值 0.6 时，MSE 降低最少，因为在区域 P_k 中前景与背景像素点数量的比值相差较小的话，改变前景或者背景的值可能会得到错误的结果。而当 θ 取值较大时，很有可能会忽略一些本应该改变的前景(背景)像素点。

为了验证人像抠图无监督语义精修算法(SHMR)在远景人像数据集的性能效果，在表 2 展示了具体的实验数据，该实验是预分类算法使用 Felzenszwalb 算法， θ 取 0.7 时，在有无人像抠图无监督语义精修算法(SHMR)的情况下，用 SHM 算法的实验效果，可以看出在加上了 SHMR 算法后效果得到了明显的提升。

Table 2. Comparison of experimental results in the perspective portrait dataset
表 2. 在远景人像数据集中的对比实验结果

方法	MSE	SAD
SHM	0.0142	0.0203
SHM + SHMR	0.0067	0.0114

为了验证在远景人像数据集中人像边框感知模块以及无监督语义精修模块的作用，本文进行了消融实验。如表 3 所示，在只使用人像边框感知模块的情况下，效果也有一定的提升。但是只使用无监督语义精修模块，效果反而变差，原因是在没有完全去除环境噪音的情况下使用无监督语义精修模块，会让噪音放大，使实验效果变差。

Table 3. Results of two module ablation experiments
表 3. 两个模块消融实验的实验结果

方法	MSE	SAD
只使用人像边框感知模块	0.0106	0.0179
只使用无监督语义精修模块	0.0230	0.0338
都使用	0.0067	0.0114

图 4 为在远景人像数据集中,使用 SHM 算法进行人像抠图以及 SHM 加上 SHMR 人像精修算法的实验效果图。可以看出,在有行人的环境中,SHM 算法容易产生额外噪音,而本文提出算法的人像边缘感知模块可以很好地去除噪音。在人像有携带物的情况下,人像整体也难以识别出来,在引入无监督语义精修模块后,人体携带物体也能准确识别,人像边缘轮廓也更加清晰。



Figure 4. Experimental renderings in the perspective portrait dataset
图 4. 在远景人像数据集的实验效果图

为了验证 SHMR 人像精修算法的泛用性能,本文同时在全身人像抠图数据集进行实验分析。使用的数据集为爱分割(aisegment.com)高质量标注并开源的人像抠图数据集。该数据集包含 34,427 张图像和对应的 matting 结果图,分辨率为 600 像素 × 800 像素,超参数的设置与远景人像数据集相同。表 4 展示了在全身人像数据集中的对比实验结果。

Table 4. Comparison of experimental results in the data set of half-length portraits
表 4. 在全身人像数据集中的对比实验结果

方法	MSE	SAD
SHM	0.0086	0.0132
SHM + SHMR	0.0071	0.0119

由表 4 可以看出, 在半身人像数据集中, 加入 SHMR 算法后效果同样得到提升, 但是 MSE 的降低并没在远景人像数据集中那么多, 原因是: 在远景人像数据集中存在很多环境噪音, 由于 SHMR 有人像边框感知模块, 环境噪音可以得到去除, 最后再进行人像边缘的精修。然而在半身人像数据集中, 环境噪音非常少, 主要侧重于人像边缘的精修, 所以 MSE 的降低并没有在远景人像数据集中那么多。

图 5 为在半身人像数据集中, 使用 SHM 算法进行人像抠图以及 SHM 加上 SHMR 人像精修算法的实验效果图。对比远景人像数据集中的实验可以看出, 半身人像数据集中环境噪音明显大幅度减少, 并且主要侧重于人像轮廓的修复。而加入 SHMR 人像精修算法后人像轮廓修复得更加完整清晰, 这也验证了该算法的泛用性。

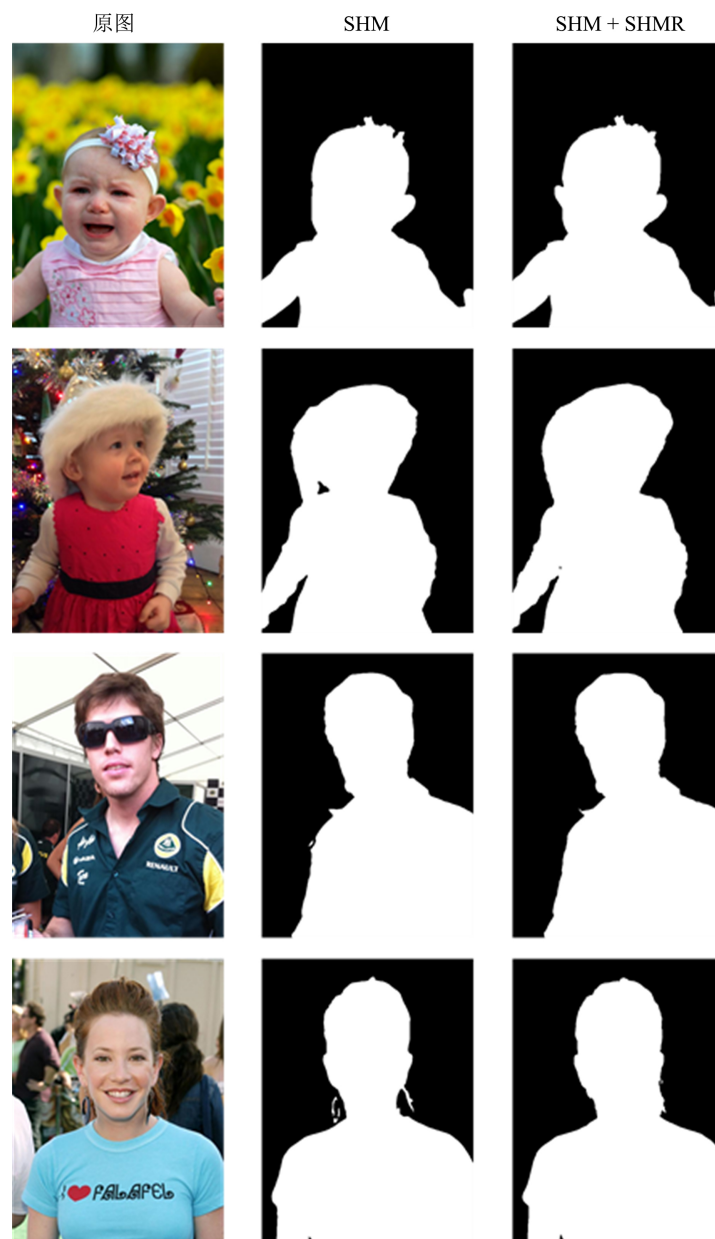


Figure 5. Experimental renderings of the bust portrait data set

图 5. 在半身人像数据集中的实验效果图

4. 结束语

本文针对远景人像抠图存在噪音、人体边缘轮廓粗糙、人体携带物易与背景混淆等问题,提出了人像抠图无监督语义精修算法。通过人像边框感知模块进行去噪并使用无监督语义精修模块优化细节,以达到人像前景与背景完整分割的目标。实验表明,在远景人像数据集中,在主流的人像抠图算法中加入人像抠图无监督语义精修算法后,效果有明显的提高,并且在半身人像数据集中,效果也有一定的提高。但由于本文将去噪模块与精修模块分割开来,也会导致误差的传递,接下来将针对端到端的人像抠图精修算法展开研究。

参考文献

- [1] Wu, X., Fang, X.N., Chen, T., *et al.* (2020) JMNet: A Joint Matting Network for Automatic Human Matting. *Computation Visual Media*, **6**, 215-224. <https://doi.org/10.1007/s41095-020-0168-6>
- [2] Badrinarayanan, V., Kendall, A. and Cipolla, R. (2019) SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 2481-2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
- [3] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. *Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 3431-3440. <https://doi.org/10.1109/CVPR.2015.7298965>
- [4] Aksoy, Y., Aydin, T.O. and Pollefeys, M. (2017) Designing Effective Inter-Pixel Information Flow for Natural Image Matting. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 29-37. <https://doi.org/10.1109/CVPR.2017.32>
- [5] Chen, Q.F., *et al.* (2013) KNN Matting. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **35**, 2175-2188. <https://doi.org/10.1109/TPAMI.2013.18>
- [6] Cai, S., Zhang, X., Fan, H., *et al.* (2019) Disentangled Image Matting. *The IEEE International Conference on Computer Vision*, Seoul, 27-28 October 2019, 8818-8827. <https://doi.org/10.1109/ICCV.2019.00891>
- [7] Shen, X., Tao, X., Gao, H., *et al.* (2016) Deep Automatic Portrait Matting. *European Conference on Computer Vision*, Amsterdam, 8-16 October 2016, 92-107. https://doi.org/10.1007/978-3-319-46448-0_6
- [8] Cho, D., Tai, Y.W. and Kweon, I. (2016) Natural Image Matting Using Deep Convolutional Neural Networks. *The European Conference on Computer Vision*, Amsterdam, 8-16 October 2016, 626-643. https://doi.org/10.1007/978-3-319-46475-6_39
- [9] Levin, A. (2006) A Closed Form Solution to Natural Image Matting. *Proceedings of 2006 IEEE Conference on Computer Vision and Pattern Recognition*, New York, 17-22 June 2006, 228-242. <https://doi.org/10.1109/TPAMI.2007.1177>
- [10] Ning, X., Price, B., Cohen, S. and Huang, T. (2017) Deep Image Matting. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 2970-2979.
- [11] Lutz, S., Amplianitis, K. and Smolic, A. (2018) AlphaGAN: Generative Adversarial Networks for Natural Image Matting. *British Machine Vision Conference*, Newcastle, 3-6 September 2018, 259.
- [12] Chen, Q., Ge, T., Xu, Y., *et al.* (2018) Semantic Human Matting. *2018 ACM Multimedia Conference*, Seoul, 22-26 October 2018, 618-626. <https://doi.org/10.1145/3240508.3240610>
- [13] Liu, J., Yao, Y., Hou, W., *et al.* (2020) Boosting Semantic Human Matting with Coarse Annotations. *Proceedings of 2020 IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, 14-19 June 2020, 8560-8569. <https://doi.org/10.1109/CVPR42600.2020.00859>
- [14] Redmon, J. and Farhadi, A. (2018) YOLOv3: An Incremental Improvement.
- [15] Kanezaki, A. (2018) Unsupervised Image Segmentation by Backpropagation. *IEEE International Conference on Acoustics*, Calgary, 15-20 April 2018, 1543-1547. <https://doi.org/10.1109/ICASSP.2018.8462533>
- [16] Kingma, D.P. and Jimmy, B. (2014) Adam: A Method for Stochastic Optimization.
- [17] Felzenszwalb, P.F. and Huttenlocher, D.P. (2004) Efficient Graph-Based Image Segmentation. *International Journal of Computer Vision*, **59**, 167-181. <https://doi.org/10.1023/B:VISI.0000022288.19776.77>
- [18] Achanta, R., Shaji, A., Smith, K., *et al.* (2012) SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **34**, 2274-2282. <https://doi.org/10.1109/TPAMI.2012.120>