

基于边缘引导的自校正皮肤检测

郑顺源, 胡良校, 吕晓倩, 孙鑫, 张盛平

引用本文

郑顺源, 胡良校, 吕晓倩, 孙鑫, 张盛平. [基于边缘引导的自校正皮肤检测](#)[J]. 计算机科学, 2022, 49(11): 141-147.

ZHENG Shun-yuan, HU Liang-xiao, LYU Xiao-qian, SUN Xin, ZHANG Sheng-ping. [Edge Guided Self-correction Skin Detection](#)[J]. Computer Science, 2022, 49(11): 141-147.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于多路径特征提取的实时语义分割方法](#)

Real-time Semantic Segmentation Method Based on Multi-path Feature Extraction

计算机科学, 2022, 49(7): 120-126. <https://doi.org/10.11896/jsjcx.210500157>

[基于注意力机制和多任务学习的阿尔茨海默症分类](#)

Alzheimer's Disease Classification Method Based on Attention Mechanism and Multi-task Learning

计算机科学, 2022, 49(6A): 60-65. <https://doi.org/10.11896/jsjcx.201200072>

[共享浅层参数多任务学习的脑出血图像分割与分类](#)

Intracerebral Hemorrhage Image Segmentation and Classification Based on Multi-task Learning of Shared Shallow Parameters

计算机科学, 2022, 49(4): 203-208. <https://doi.org/10.11896/jsjcx.201000153>

[基于特征定位与融合的行人重识别算法](#)

Person Re-identification Based on Feature Location and Fusion

计算机科学, 2022, 49(3): 170-178. <https://doi.org/10.11896/jsjcx.210100132>

[面向出租车空载时间预测的多任务时空图卷积网络](#)

Multi-task Spatial-Temporal Graph Convolutional Network for Taxi Idle Time Prediction

计算机科学, 2021, 48(7): 112-117. <https://doi.org/10.11896/jsjcx.201000089>

基于边缘引导的自校正皮肤检测

郑顺源 胡良校 吕晓倩 孙鑫 张盛平

哈尔滨工业大学计算机科学与技术学院 山东 威海 264209

(sawyer0503@gmail.com)

摘要 皮肤检测作为计算机视觉领域中的研究热点多年来被广泛研究,且仍然是一项具有挑战性的任务。尽管目前的方法在许多常规场景下取得了成功,但仍然存在预测不完整和泛化能力差等问题。针对该问题,提出了一种基于边缘引导的神经网络,并且由大量经过自校正的皮肤检测数据驱动网络训练,实现鲁棒的皮肤检测。首先,提出一种基于多任务学习的网络,对皮肤检测和边缘检测两个任务进行联合优化。进一步,提出边缘注意力模块,将预测所得的边缘检测结果通过该模块重新融合到皮肤检测支流中。最后,提出一种自校正算法,通过借助人眼解析任务中的大量低质量数据以增强皮肤检测模型的泛化能力。通过自校正算法对带噪声标签的优化,逐步消除使用带噪声标签进行监督训练的副作用。实验结果表明,所提皮肤检测方法优于现有的其他方法。

关键词: 皮肤检测;边缘检测;多任务学习;自校正算法

中图法分类号 TP391

Edge Guided Self-correction Skin Detection

ZHENG Shun-yuan, HU Liang-xiao, LYU Xiao-qian, SUN Xin and ZHANG Sheng-ping

College of Computer Science and Technology, Harbin Institute of Technology, Weihai, Shandong 264209, China

Abstract Skin detection has been a widely studied computer vision topic for many years, whereas remains a challenging task. Previous methods celebrate their success in various ordinary scenarios but still suffer from fragmentary prediction and poor generalization. To address this issue, this paper proposes an edge guided network driven by a massive self-corrected skin detection dataset for robust skin detection. To be specific, a multi-task learning based network which conducts skin detection and edge detection jointly is proposed. The predicted edge map is further converged to the skin detection stream via an edge attention module. Meanwhile, to engage a large-scale of low-quality data from the human parsing task to strengthen the generalization of the network, a self-correction algorithm is adapted to prune the side effect of supervised by noisy labels with continuously polishing up those defects during the training process. Experimental results indicate that the proposed method outperforms the state-of-the-art in skin detection.

Keywords Skin detection, Edge detection, Multi-task learning, Self-correction algorithm

1 引言

皮肤检测或皮肤分割任务的目的是为图像每个像素分配一个二元标签以表示皮肤像素和非皮肤像素。该任务是一项传统计算机视觉任务,亦为许多其他计算机视觉任务的重要预处理过程,如人像美化^[1]、人脸检测^[2]、手势检测^[3]、姿态估计^[4]等。许多现有方法在复杂场景中仍然存在明显缺陷,如光照特殊场景、带有阴影的场景、背景与皮肤相似度较高的场景、人像尺度变化较大的场景、具有密集人群的场景等。

皮肤检测最直接的方法是在特定的颜色空间(如 RGB, YCbCr 和 HSV)中定义边界或超平面,这类方法可称为基于

阈值的方法^[5-7],在阈值范围内的像素被定义为皮肤区域。这类方法对光照和人种敏感,同时与皮肤颜色相似的饰品或着装易造成算法误判。基于机器学习的方法^[8-12]在一定程度上缓解了该问题,它们首先基于局部特征得到描述符,然后设计分类器逐像素预测其是否属于皮肤像素。但此类方法受制于分类器的学习能力。近年来,深度学习技术取得了巨大的成功,同样为皮肤检测任务带来巨大提升^[13-17]。尽管基于深度学习的方法已经使皮肤检测任务中的许多问题得到明显改善,但在性能上仍有较大改进空间。在前文所述的复杂场景中,皮肤区域预测往往会碎裂。此外,较高的标注成本限制了皮肤检测数据集的规模。

到稿日期:2022-06-02 返修日期:2022-07-22

基金项目:国家自然科学基金(61872112);山东省泰山学者人才计划(tsqn201812106)

This work was supported by the National Natural Science Foundation of China(61872112) and Taishan Scholars Program of Shandong Province(tsqn201812106).

通信作者:吕晓倩(xiaoqian_hit@gmail.com)

本文提出了一个基于边缘引导的多任务学习网络,结合一种针对噪声标签的自校正算法,同时从网络结构和训练数据两个方面解决上述问题。在多任务学习网络的实现中,增加一个神经网络支路用于回归皮肤区域的边缘。额外的边缘信息有助于共用的编码器更关注皮肤边缘。然后,预测的边缘结果通过边缘注意力模块(Edge Attention Module, EAM)反馈到皮肤检测支路。由于 EAM 的设计是基于一种轻量级注意力模块^[18],因此它只会额外引入有限的计算复杂度。本文在 3 个不同的特征图尺度上设置多个 EAM 模块,在每个特征图尺度的金字塔池化(Atrous Spatial Pyramid Pooling, ASPP)模块^[19]之前和之后分别添加。此外,为缓解训练数据不足的问题,本文提出一种标签自校正技术^[20]。鉴于人体解析任务中有大量公开的标签完备数据^[21-27],利用这些数据解决皮肤检测这一与之有许多共同点的任务是一种自然的想法。但是,这些可用的数据同样存在问题,如图 1 所示,人体解析任务标签重组为皮肤检测标签后可能会出现标注错误的问题,因此这些标签被视为带噪声标签或伪标签。在带噪声标签上直接训练可能会导致预测质量较低,本文提出的自校正算法通过在训练过程中对标签进行优化以降低噪声带来的负面影响。当网络通过训练收敛到最优的同时,标签中的错误会得到纠正,如图 1 所示。大量实验证明本文提出的方法具有良好的皮肤检测性能。

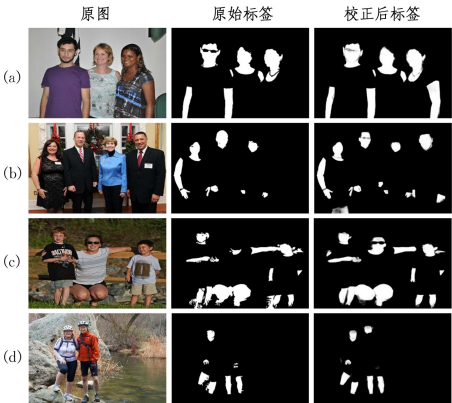


图1 皮肤检测原始标签及自校正结果

Fig. 1 Original skin detection labels and refinement results by self-correction strategy

2 相关工作

2.1 皮肤检测

皮肤检测方法可以分为 3 类:基于阈值的方法^[5-7]、基于机器学习的方法^[8-12]以及基于深度学习的方法^[13-17]。基于阈值的方法在特定色彩空间中定义边界或超平面。Gomez 等^[5]提出了一种构造性归纳算法,在规范化 RGB 空间上使用受限覆盖算法。在文献[6]中,Cheddad 等设计了一个色彩空间,该空间包含了在区分灰度图和其非编码红色灰度图中得到的错误信号。文献[7]提出了一种动态阈值方法以区分皮肤像素。Jones 等^[8]认为直方图模型在准确性和计算成本方面优于混合模型。Phung 等^[9]的研究同样证实了该观点。Tan 等^[10]将平滑的 2D 直方图和高斯模型相结合,用于自动人体皮肤检测,该研究引入眼睛检测器以优化皮肤检测模型。

Hwang 等^[11]基于图像的多层图表示,提出了一种多种子传播方法。近年来,基于深度学习的方法极大地促进了该领域研究的发展。Zuo 等^[13]将递归神经网络层融合到全卷积神经网络中以强调语义上下文依赖性。Kim 等^[14]提出分别基于 VGG 网络和 NiN 网络的两种卷积神经网络。为弥补昂贵的标注成本造成的数据短缺,He 等^[15]提出了一种半监督方法,该方法以人体分割为辅助任务。Tarasiewicz 等^[16]基于 U-Net 提出轻量级皮肤检测模型 Skinny,通过加入多尺度模块来增强多尺度皮肤检测性能。除彩色图像外,灰度图像^[12]和近红外图像的皮肤检测^[17]同样吸引了学者们的广泛关注。

2.2 边缘检测

边缘检测既可作为独立计算机视觉任务^[28-29],亦可作为语义分割^[30-31]或其相关任务的补充信息,如人体解析和显著性检测等。虽然边缘信息对语义分割等任务而言十分重要,但完美地勾勒出分割图的边缘仍是一项具有挑战性的任务。文献[30]对实例分割任务中降采样导致的边缘过于平滑问题进行了优化。Takikawa 等^[31]利用语义分割和边缘检测两个任务间的相关性提出了一种新的损失函数。在人体解析任务中,文献[25]使用实例级边缘检测将语义部件准确地归类到人体实例中。Zhao 等^[32]提出一个联合语义和边缘感知的网络架构。文献[20]和文献[33]共享类似的网络结构,其中包含一种简单而有效的语义上下文嵌入模块,用于对单个人体中各个部分的边缘进行感知。在显著性检测任务中,文献[34]提出一种预测-细化的网络结构以及混合损失,用于边缘可感知的显著性检测。文献[35]提出一种边缘引导网络,用于检测显著性物体。

2.3 皮肤检测数据集

皮肤检测是一项传统计算机视觉任务,常用的数据集^[8-9]出现时间较早。COMPAQ 数据集^[8]是第一个用于皮肤检测的大规模数据集,包含 13 640 张来自互联网的图像。ECU 数据集^[9]由 4 000 幅图像组成,其中约 1% 的图像由数码相机拍摄,其余图像是从互联网上人工采集得到,确保了背景场景、照明条件和皮肤类型的多样性。COMPAQ 和 ECU 数据集图像质量和标注质量均较低。为满足神经网络训练需要,研究人员构建了 Mutual Guidance 数据集^[15],该数据集包含 5 000 幅图像,图像质量和标注质量均较高,但是大部分图像为单人图像。

3 基于边缘引导的自校正皮肤检测方法

3.1 基于边缘引导的多任务学习网络

3.1.1 网络结构

基于深度学习的方法极大地提升了光照复杂和背景杂乱等挑战性场景中的皮肤检测性能,但仍不能令人满意。当面对该类场景时,预测结果趋于碎裂,即在皮肤区域中有若干区域被误判为非皮肤像素,导致预测的皮肤不完整。有缺陷的预测会对下游任务带来潜在干扰,如在人像美化任务中,皮肤美白算法的执行会导致误判区域的颜色与周围区域存在显著差别。针对该问题,本节提出一种多任务学习方法,引入边缘检测进行联合学习;然后设计一种边缘注意力机制,增强高级语义信息和低级语义信息间的融合。

本文构建的基于边缘引导的深度全卷积神经网络如图2所示。鉴于 DeepLab-v3+^[19]在许多语义分割任务中的优秀性能,本文选择以该结构作为基准网络模型进行皮肤检测,并以 ResNet101^[36]为主干网络。在 DeepLab-v3+的基础上增加一个额外的网络分支用于回归皮肤区域边缘,以此构成多任务学习框架。皮肤检测和边缘检测两个任务共享编码器。边缘检测解码器分支与原始 DeepLab-

v3+设置保持一致。皮肤检测解码器分支的设计中,经过 ASPP 后的高级语义特征首先进行 $2\times$ 上采样,然后与低级特征 F_2 融合。融合后再次进行 $2\times$ 上采样,最后与低级特征 F_1 融合。在双分支网络结构基础上,将边缘检测结果通过边缘注意力模块 EAM 反馈到皮肤检测分支中。EAM 模块添加在 3 个特征图尺度上,每个尺度在 ASPP 模块前后分别添加。

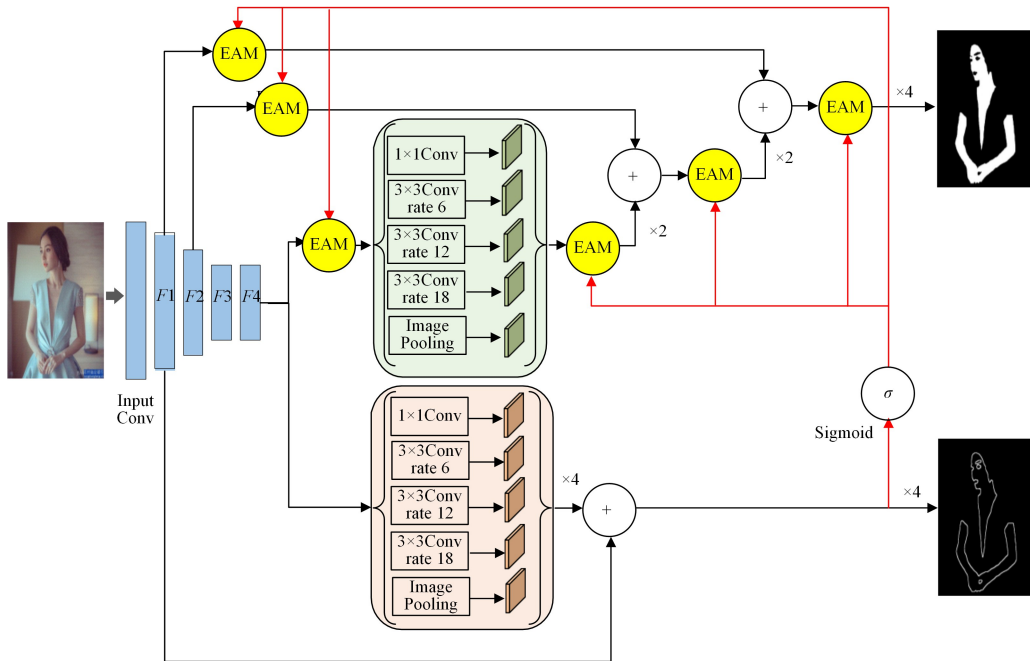


图2 融合边缘注意力模块的多任务学习网络结构

Fig. 2 Structure of our multi-task learning network with edge attention mechanism

3.1.2 边缘注意力模块

边缘注意力模块的设计如图3所示。图中 Edge 表示边缘检测分支的输出结果,首先将其缩放放到与输入特征图 X 相同大小,然后再与特征图 X 逐像素相乘后进行通道叠加。由于此时特征图通道数为原始特征图通道数的两倍,故需要选择更有价值的通道,去除冗余通道,因此选择级联 SE 模块^[18],该模块被视为有效的通道注意力机制。通过 Sigmoid 层规范化后的权重可视为每个通道的重要性。通过与特征图相乘,有价值的通道将被强调,反之将被抑制。由于输入该模块的特征图与输出特征图在特征图尺寸和通道数量上完全一致,故该模块可嵌入网络的任意位置。由于 SE 模块为轻量级注意力机制,因此 EAM 模块同样仅额外增加有限的计算复杂度。

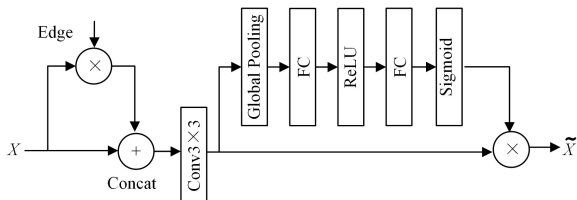


图3 边缘注意力模块网络结构

Fig. 3 Structure of edge attention module

3.1.3 损失函数

皮肤检测使用二元交叉熵损失函数,而边缘检测任务

正负样本不平衡,落在边缘区域的像素只占极小比例,故需要对损失函数进行加权以平衡正负样本的贡献。设 j 为输入图像 X 中的任意像素点,设计边缘检测部分损失函数为:

$$L_{\text{edge}} = -\beta \sum_{j \in Y_+} \log Pr(y_j = 1 | X) - \gamma \sum_{j \in Y_-} \log Pr(y_j = 0 | X) \quad (1)$$

其中, $\beta = |Y_-|/|Y|$, $\gamma = 1.1 * |Y_+|/|Y|$, $|Y_+|$ 和 $|Y_-|$ 分别表示真实标记中边缘像素集以及非边缘像素集。 $Pr(y_j = 1 | X)$ 为二值预测,表示每个点为边缘像素的置信度。网络整体的损失函数为:

$$L = L_{\text{skin}} + \mu L_{\text{edge}} \quad (2)$$

其中, L_{skin} 表示皮肤检测任务中二元交叉熵损失, L_{edge} 表示边缘检测任务中加权二元交叉熵损失, μ 为边缘损失权重系数。

3.2 自校正算法

3.2.1 数据集准备

皮肤检测可看作人体解析各个类别的子集,因此借用人体的解析数据训练皮肤检测网络是一种自然的想法。如图4所示,通过组合面部、颈部、手臂、手、腿、脚和躯干皮肤这几个人体解析标签中定义类别可获取皮肤检测标签。但是,并不是所有人体解析数据均可满足需求,表1为现有的公开人体解析数据集简析。Fashionista^[21], ATR^[23] 和 LIP^[24] 中的图像均为单人图像。PASCAL Person Part^[22] 标签将人体分为7类,许多类别被混杂为一类导致无法从原始人体解析标签中重新组合出皮肤标签,如“头发”和“脸”共享同一种类别标签。

CIHP^[25], MHP-v1^[26] 和 MHP-v2^[27] 对扩充数据量有较高的价值,平均每张图像中包含超过 3 个人体实例。MHP-v2 是 MHP-v1 的扩充,除部分服装类别定义更细致外,二者其他特点相同,因此下文统称 MHP 数据集。CIHP 数据集图像质量低于 MHP 数据集,且 CIHP 数据集包含过多小尺寸图像。因此, MHP 数据集被视为扩充数据的最优选择。



图 4 从人体解析标签重组皮肤检测标签效果

Fig. 4 Reassembled skin annotations from original human parsing annotations

表 1 人体解析公开数据集简析

Table 1 Overview of publicly available datasets for human parsing

数据集	实例数	图片数	类别数
Fashionista ^[21]	1	685	56
PASCAL Person Part ^[22]	2.2	3 533	7
ATR ^[23]	1	16 700	18
LIP ^[24]	1	40 462	20
CIHP ^[25]	3.4	32 940	20
MHP-v1 ^[26]	3	4 980	19
MHP-v2 ^[27]	3	20 403	59

虽然 MHP 数据集图像质量较高,但其标注水平未达预期。如图 5 所示,标签主要受两类错误影响,一类是分割边缘粗糙,另一类是类别标注错误。前者是由标注工具所致,该标注工具通过生成图像的多尺度超像素以加快标注速度。后者是由标注人员的失误所致,标注者错误地将两个类别标签混淆,如将 T 恤标注为躯干皮肤。因此,组合得到的原始标签仅可视为带噪声标签或伪标签,直接使用这些标签会导致错误的模式被网络学习。

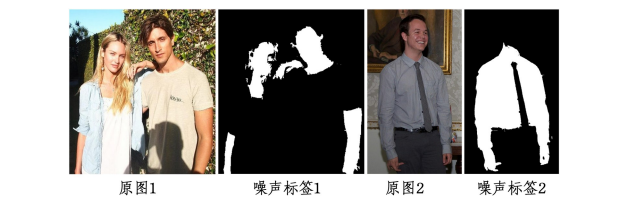


图 5 MHP 数据集两类标注错误示例

Fig. 5 Two types of annotation errors in MHP dataset

3.2.2 算法描述

本节提出一种自校正算法,在提升模型性能的同时对标签进行迭代优化,算法流程如图 6 所示。首先,使用带噪声标签进行初始化训练,该过程与常规的网络训练完全相同,共迭代 T_0 个周期。初始化步骤至关重要,因为后续的优化很大程度上依赖于该阶段模型的性能。即使只有伪标签可用,此阶段同样应力求得到一个相对可靠的模型,若非如此,后续的优化过程表现将随之劣化,从而对整体算法产生影响。初始化后的参数将保存在一个非训练的模型中,该模型仅用于生成新的伪标签。

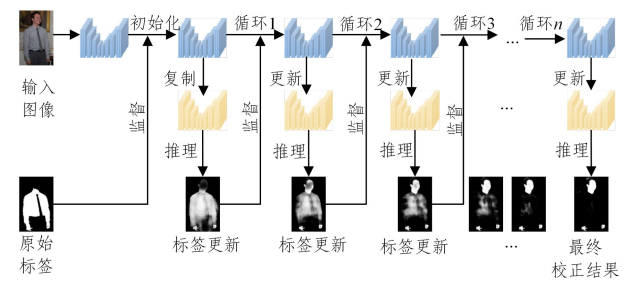


图 6 自校正算法流程图

Fig. 6 Illustration of self-correction pipeline

自校正过程以循环的方式进行,共包含 n 个循环,每个循环迭代 T 个周期。每个循环结束后,网络收敛到局部最优。由于局部最优值间存在差异,所设计的标签生成模型将整合多个局部最优值,因此使用滑动平均的方式更新参数。

$$\omega_n = \frac{n}{n+1}\omega_{n-1} + \frac{1}{n+1}\omega \tag{3}$$

其中, ω 为训练模型当前参数, n 为循环次数, ω_n 为标签生成模型在 n 次循环后的参数, ω_0 为初始化阶段后的模型参数。根据式(3)更新参数后,标签生成模型对所有训练数据进行一次前向计算,根据原始标签和推理结果生成新的伪标签。

$$y_n = \frac{n}{n+1}y + \frac{1}{n+1}y_0 \tag{4}$$

其中, y 为标签生成模型推理结果, y_0 为原始标签, y_n 为第 n 次循环后更新的伪标签,用于监督下一个循环的训练。神经网络具有较强的学习能力,可以很好地对噪声模式进行学习,故直接使用该算法无法达到预期效果。另一方面,神经网络尺度不变性较差,同一幅图像在不同尺度下的预测可能有较大差异,利用该特点可以改进伪标签生成过程。在进行前向计算时,为每一幅图像额外增加 3 个图像尺度,分别为原图的 0.75 倍、1.5 倍和 2 倍,在原尺度上计算其平均值作为最终标签。该操作可避免上文提到的缺陷,使生成的伪标签更平滑、准确。逐渐优化的标签同样会增强网络的鲁棒性。

4 实验

4.1 实现细节

自校正训练过程使用 MHP 数据集^[26-27] 共 25 383 张图像。基准训练过程使用 Mutual Guidance 数据集^[15], 包含 4 530 张训练图像和 470 张测试图像。所有定量测试均在 Mutual Guidance 的测试集上完成。将边缘检测损失权重 μ 设为 5。自校正算法中, T_0 , T 和 n 分别设为 50, 20 和 5。初始化阶段和优化循环中使用 cosine 学习率下降策略,初始学习率分别设为 1×10^{-2} 和 5×10^{-3} 。使用文献^[35]中的方法根据皮肤检测标签生成边缘检测标签。将图像的较短边缩放为 640 像素,然后随机裁剪 512^2 图像块作为输入。所有实验均使用 Pytorch 完成,在一张具有 11 GB 显存的 NVIDIA RTX 2080Ti 训练和测试。

4.2 与其他方法对比

本节将本文方法与目前最优的方法进行对比,包括 DeepLab-v3+^[19]、He 等^[15]提出的方法和 Skinny^[16]。为了公平比较,我们在 MHP 数据集上对 DeepLab-v3+ 和 Skinny 进行预训练,并在 Mutual Guidance 训练集上进行微调。定量评估

结果如表 2 所列,本文提出的方法的 IoU 指标比 He 等的方法和 Skinny 分别高 5.76% 和 3.08%,比本文的基准模型 DeepLab-v3+高 3.89%。

表 2 Mutual Guidance 数据集定量评估结果
Table 2 Evaluation on Mutual Guidance dataset
(单位:%)

方法	IoU
DeepLab-v3+ ^[19]	83.05
He 等 ^[15]	81.18
Skinny ^[16]	83.86
本文	86.94

如图 7 所示,定性比较同样能体现本文方法的优势。图 7(a)和图 7(b)中,现有方法分割边缘结果不能尽如人意,部分由 MHP 数据集标注噪声导致,部分由皮肤区域干扰因素导致,如阴影、照明、发型等。本文提出的方法在图 7(c)和图 7(d)中所示的多人场景或背景复杂的场景中表现更优,DeepLab-v3+和 Skinny 将背景中许多区域误判为皮肤像素。图 7(e)和图 7(f)背景中出现许多与皮肤颜色类似的物体,图 7(g)和图 7(h)在饰品、头发等人体区域内出现,图 7(i)和图 7(j)中服装颜色与肤色类似,所提出的 EAM 模块使这两类场景中的误判大幅减少。

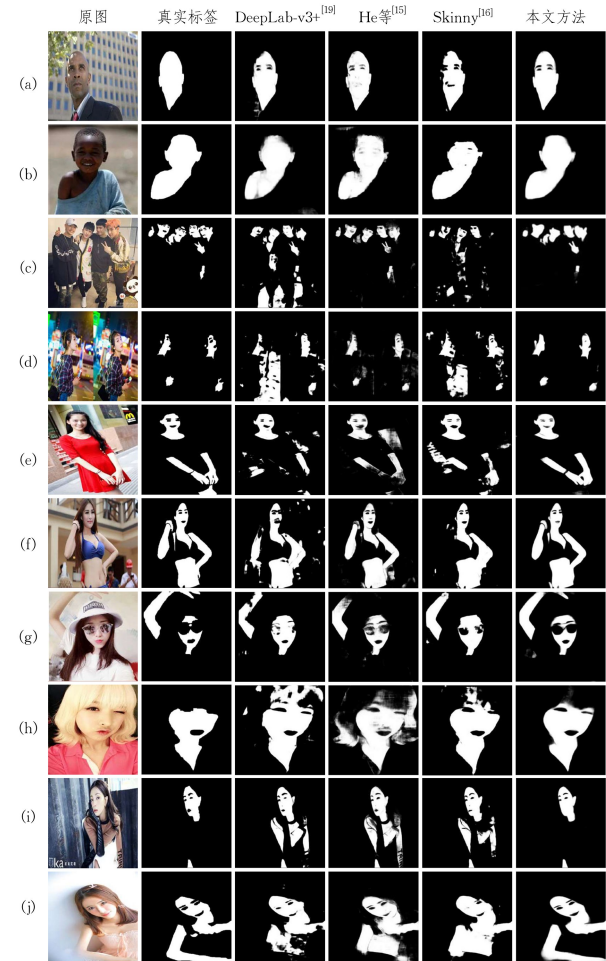


图 7 本文方法与其他皮肤检测方法的定性结果对比

Fig. 7 Qualitative results comparison of our method and other skin detection methods

4.3 消融分析

本节通过消融实验证明了本文方法中每个部分给结果

带来的影响,结果如表 3 所列。表中 M. T. L. 和 S. C. 分别表示多任务学习和自校正算法。图 2 中的 6 个 EAM 模块按图像尺度分为 3 组,设置在特征图 F1, F2 和 F4 尺度的 EAM 模块分别称为 EAM_{F1}, EAM_{F2} 和 EAM_{F4}。第一行列出的基准网络表示原始 DeepLab-v3+网络。从表 3 中可发现,仅增加一个辅助任务进行多任务学习时性能提升幅度仅为 0.14%。EAM_{F1}可使 IoU 提升 0.92%,添加 EAM_{F2}后可进一步增加到 1.17%,当 3 组 EAM 全部添加时,IoU 可提升 2.03%。以上结果表明,在网络多个阶段加入边缘注意引导,性能可稳步提升。自校正算法通过增强网络的泛化能力可以使 IoU 提升 2.5%。整体上,本文方法在基准网络基础上 IoU 提升了 6.53%。

表 3 本文方法各部分有效性分析
Table 3 Validity of each component of the proposed method
(单位:%)

M. T. L.	EAM _{F1}	EAM _{F2}	EAM _{F4}	S. C.	IoU
					80.41
✓					80.55
✓	✓				81.47
✓	✓	✓			81.72
✓	✓	✓	✓		82.58
				✓	82.91
✓	✓	✓	✓	✓	86.94

图 8 中的定性结果展示了所提出的边缘注意力模块对结果产生的影响。在图 8(a)—图 8(c)中,图像背景复杂,没有 EAM 模块的引导会导致背景中与肤色相似的颜色区域被误判为皮肤。图 8(d)—图 8(f)中,输入图像受光照影响严重,EAM 模块弥补了皮肤区域预测不完整的缺陷。从视觉上能直观地感受到,基于边缘引导的网络能较好地保持分割区域的完整性。

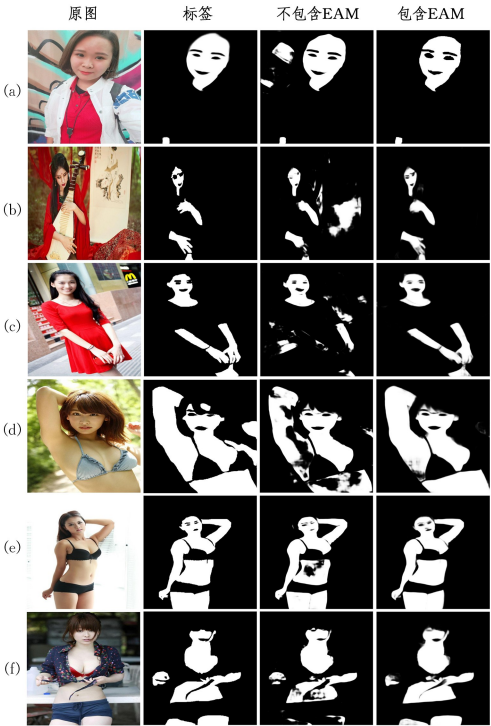


图 8 边缘注意力模块消融分析结果图

Fig. 8 Ablation analysis of edge attention module

结束语 本文提出了一种基于边缘引导的皮肤检测方法以及一种自校正算法。在实现中,增加一个额外分支用于边缘检测,并通过边缘注意力机制将预测结果进一步融合到皮肤检测分支中。自校正算法可以使网络基于带噪声标签开启训练过程,然后使带噪声标签和网络同时相互优化。此外,本文利用神经网络尺度不变性差的特点来改进使用带噪声标签进行监督训练带来的缺陷。边缘引导和自校正算法使网络在挑战性场景中鲁棒性增强。实验表明,本文方法在皮肤检测任务中优于现有的其他方法。

参 考 文 献

- [1] VELUSAM S, PARIHAR R, KINI R, et al. FabSoftNet: Face beautification via dynamic skin smoothing, guided feathering, and texture restoration [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020:530-531.
- [2] CHEN W, WANG K, JIANG H, et al. Skin color modeling for face detection and segmentation: a review and a new approach [J]. Multimedia Tools and Applications, 2016, 75(2): 839-862.
- [3] RAUTARAY S S, AGRAWAL A. Vision based hand gesture recognition for human computer interaction: a survey [J]. Artificial Intelligence Review, 2015, 43(1): 1-54.
- [4] QIN X, GUO H, HE C, et al. Lightweight human pose estimation: CVC-net [J]. Multimedia Tools and Applications, 2022, 81(13): 17615-17637.
- [5] GOMEZ G, MORALES E. Automatic feature construction and a simple rule induction algorithm for skin detection [C] // Proceedings of the ICML workshop on Machine Learning in Computer Vision. 2002: 31-38.
- [6] CHEDDAD A, CONDELL J, CURRAN K, et al. A new colour space for skin tone detection [C] // Proceedings of the IEEE International Conference on Image Processing (ICIP). IEEE, 2009: 497-500.
- [7] YOGARAJAH P, CONDELL J, CURRAN K, et al. A dynamic threshold approach for skin segmentation in color images [C] // Proceedings of the IEEE International Conference on Image Processing. IEEE, 2010: 2225-2228.
- [8] JONES M J, REHG J M. Statistical color models with application to skin detection [J]. International Journal of Computer Vision, 2002, 46(1): 81-96.
- [9] PHUNG S L, BOUZERDOUM A, CHAI D. Skin segmentation using color pixel classification: analysis and comparison [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(1): 148-154.
- [10] TAN W R, CHAN C S, YOGARAJAH P, et al. A fusion approach for efficient human skin detection [J]. IEEE Transactions on Industrial Informatics, 2011, 8(1): 138-147.
- [11] HWANG I, KIM Y, CHO N I. Skin detection based on multi-seed propagation in a multi-layer graph for regional and color consistency [C] // Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2017: 1273-1277.
- [12] PARACCHINI M B M, MARCON M, VILLA F, et al. Fast Skin Segmentation on Low Resolution Grayscale Images for Remote PhotoPlethysmoGraphy [J]. IEEE MultiMedia, 2022, 29(1): 28-35.
- [13] ZUO H, FAN H, BLASCH E, et al. Combining convolutional and recurrent neural networks for human skin detection [J]. IEEE Signal Processing Letters, 2017, 24(3): 289-293.
- [14] KIM Y, HWANG I, CHO N I. Convolutional neural networks and training strategies for skin detection [C] // Proceedings of the IEEE International Conference on Image Processing (ICIP). IEEE, 2017: 3919-3923.
- [15] HE Y, SHI J, WANG C, et al. Semi-supervised skin detection by network with mutual guidance [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 2111-2120.
- [16] TARASIEWICZ T, NALEPA J, KAWULOK M. Skinny: A lightweight U-net for skin detection and segmentation [C] // Proceedings of the IEEE International Conference on Image Processing (ICIP). IEEE, 2020: 2386-2390.
- [17] PANDEY P, TYAGI A K, AMBEKAR S, et al. Unsupervised domain adaptation for semantic segmentation of NIR images through generative latent search [C] // Proceedings of the European Conference on Computer Vision. Cham: Springer, 2020: 413-429.
- [18] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7132-7141.
- [19] 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). 2018: 801-818.
- [20] LI P, XU Y, WEI Y, et al. Self-correction for human parsing [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(6): 3260-3271.
- [21] YAMAGUCHI K, KIAPOUR M H, ORTIZ L E, et al. Parsing clothing in fashion photographs [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012: 3570-3577.
- [22] CHEN X, MOTTAGHI R, LIU X, et al. Detect what you can: Detecting and representing objects using holistic models and body parts [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 1971-1978.
- [23] LIANG X, LIU S, SHEN X, et al. Deep human parsing with active template regression [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(12): 2402-2414.
- [24] GONG K, LIANG X, ZHANG D, et al. Look into person: Self-supervised structure-sensitive learning and a new benchmark for human parsing [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 932-940.
- [25] GONG K, LIANG X, LI Y, et al. Instance-level human parsing via part grouping network [C] // Proceedings of the European Conference on Computer Vision (ECCV). 2018: 770-785.
- [26] LI J, ZHAO J, WEI Y, et al. Multiple-human parsing in the wild [J]. arXiv:1705.07206, 2017.
- [27] ZHAO J, LI J, CHENG Y, et al. Understanding humans in

crowded scenes; Deep nested adversarial learning and a new benchmark for multi-human parsing[C] // Proceedings of the 26th ACM International Conference on Multimedia. 2018;792-800.

[28] POMA X S,RIBA E,SAPPA A. Dense extreme inception network; Towards a robust cnn model for edge detection[C]// Proceedings of the IEEE/ CVF winter Conference on Applications of Computer Vision. 2020;1923-1932.

[29] SU Z,LIU W,YU Z,et al. Pixel difference networks for efficient edge detection [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021;5117-5127.

[30] ZHANG G, LU X, TAN J, et al. Refinemask; Towards high-quality instance segmentation with fine-grained features[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021;6861-6869.

[31] TAKIKAWA T, ACUNA D,JAMPANI V, et al. Gated-scnn: Gated shape cnns for semantic segmentation[C]// Proceedings of the IEEE/ CVF International Conference on Computer Vision. 2019;5229-5238.

[32] ZHAO Y, LI J,ZHANG Y,et al. Multi-class part parsing with joint boundary-semantic awareness [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019; 9177-9186.

[33] RUAN T,LIU T,HUANG Z,et al. Devil in the details; Towards accurate single and multiple human parsing[C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2019;4814-4821.

[34] QIN X,ZHANG Z,HUANG C,et al. Basnet; Boundary-aware salient object detection [C] // Proceedings of the IEEE/CVF Conference on computer Vision and Pattern Recognition. 2019; 7479-7489.

[35] ZHAO J X,LIU J J,FAN D P,et al. EGNNet;Edge guidance network for salient object detection[C]// Proceedings of the IEEE/ CVF International Conference on Computer Vision. 2019;8779-8788.

[36] HE K,ZHANG X,REN S,et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016;770-778.



ZHENG Shun-yuan, born in 1997, postgraduate. His main research interests include computer vision and deep learning.



LYU Xiao-qian, born in 1995, postgraduate. Her main research interests include object detection, image enhancement and video analysis.

(责任编辑:何杨)