

文章编号:1007-2780(2024)08-1024-13

基于多特征融合的人像 HDR 图像处理算法

吴春林¹, 张永爱^{1,2}, 林志贤¹, 郭太良², 林鹏飞³, 林坚普^{1*}

- (1. 福州大学 先进制造学院, 福建 泉州 362200;
2. 福州大学 物理与信息工程学院, 福建 福州 350108;
3. 东京大学 信息科学技术学院, 日本 东京 113-8657)

摘要: 基于深度学习的高动态范围 (High Dynamic Range, HDR) 图像处理算法在处理含有人像的图片时存在皮肤偏色问题。针对此问题, 本文提出了一种基于多特征融合的人像 HDR 图像处理算法 U²HDRnet。该算法由皮肤特征提取模块、三边特征提取模块、色彩重建模块 3 部分构成。首先, 皮肤特征提取模块分离出皮肤区域的颜色和位置信息; 其次, 三边特征提取模块分别提取图片的局部特征、全局特征和语义特征, 并与皮肤特征融合; 最后, 色彩重建模块对网格做空间和颜色深度上的插值。此外, 本文引入改进的自注意力与卷积融合模块以提升 HDR 的处理效果; 同时本文还制作了人像 HDR 数据集 PortraitHDR, 填补了该领域内数据集的空白。实验结果显示, U²HDRnet 的 PSNR 达 31.42 dB, SSIM 达 0.985, 均优于目前常见的 HDR 算法, 在获得高质量人像 HDR 图像的同时避免了皮肤的失真。

关键词: 深度学习; 高动态范围; 皮肤特征提取; 注意力机制; 色彩重建

中图分类号: TP391 **文献标识码:** A **doi:** 10.37188/CJLCD.2023-0255

HDR image processing algorithm for portrait based on multi feature fusion

WU Chunlin¹, ZHANG Yongai^{1,2}, LIN Zhixian¹, GUO Tailiang², LIN Pengfei³, LIN Jianpu^{1*}

- (1. School of Advanced Manufacturing, Fuzhou University, Quanzhou 362200, China;
2. School of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China;
3. Graduate School of Information Science and Technology, University of Tokyo, Tokyo 113-8657, Japan)

Abstract: Deep learning based high dynamic range (HDR) image processing algorithms has the problem of skin color deviation when processing images containing human figures. In response to this issue, this article proposes a portrait HDR image processing algorithm based on multi feature fusion—U²HDRnet. This algorithm consists of three parts: skin feature extraction module, trilateral feature extraction module

收稿日期: 2023-07-19; **修订日期:** 2023-10-12.

基金项目: 国家重点研发计划 (No.2021YFB3600603); 福建省自然科学基金 (No.2020J01468); 福建省教育厅中青年
教师教育科研项目 (No.JAT210030)

Supported by National Key R&D Program of China (No.2016YFB0401503); Natural Science Foundation of Fujian Province (No.2020J01468); Educational Research Project for Young and Middle-aged Teachers of Fujian Provincial Department of Education (No.JAT210030)

*通信联系人, E-mail: ljp@fzu.edu.cn

and color reconstruction module. Firstly, the skin feature extraction module separates the color and position information of the skin region. Secondly, the trilateral feature extraction module extracts local features, global features and semantic features of the image, and fuses them with skin features. Finally, the color reconstruction module interpolates the grid in terms of space and color depth. In addition, this article adds an improved fusion module of self attention and convolution to improve the processing performance of HDR. At the same time, this article also produces the PortraitHDR dataset for portraits, filling the gap in the dataset in this field. The test results show that the PSNR of U²HDRnet reaches 31.42 dB, and the SSIM reaches 0.985, both of which are superior to the commonly used HDR algorithms. They obtain high-quality portrait HDR images while avoiding skin distortion.

Key words: deep learning; high dynamic range; skin feature extraction; attention mechanism; color reconstruction

1 引言

高动态范围图像(High Dynamic Range, HDR)比起低动态范围(Low Dynamic Range, LDR)图像更能表现出图像的细节,还原真实的场景信息,一直以来吸引了很多研究者的目光。随着深度学习的发展,基于深度学习的HDR图像处理算法大量涌现。这些算法可分为两类:一类是基于多帧融合的方法,另一类是基于单帧重建的方法。

多帧融合的HDR算法通过多张LDR图像序列,应用多种计算方法构建出HDR图像,其参数量大、对算力要求较高。而单帧HDR算法在取得了较好效果的同时,其计算量更小、处理速度更快,因此从单张LDR图像中重建出HDR图像更加经济。Wu等人^[1]提出了一种基于重映射和多尺度曝光融合的简单有效的HDR成像方法,只需要采集一组短、长曝光图像便可以还原出HDR图像。Lu等人^[2]提出了一种多尺度梯度色调映射算法,在压缩HDR图像的动态范围的同时,可以更好地保留边界、纹理等细节信息。不同于Lu等人的方法,Xia等人^[3]用风格转移的方式实现了HDR,但这种方法对图片暗处的细节处理较差,同时会产生不真实的结果。Lee等人^[4]提出了一种基于卷积神经网络(Convolutional Neural Network, CNN)从LDR图片中还原HDR图片的方法,称为HSVNet,该模型使用HSV色彩空间,使网络能够识别饱和区域并自适应地关注关键组件。Khan^[5]等人提出了一个名为FHDR的网络,通过反馈网络从LDR中重构HDR,通过

多次迭代进行学习完成HDR重建。但这些方法在处理存在过曝问题的图片时都会使过曝变得更加严重。Liu等人^[6]发现,之前的方法在处理欠饱和与过饱和像素时,应用相同的卷积滤波器在训练过程中会产生模糊现象。为了克服这一问题,他们提出了一种新的网络结构CNN-HDR,利用特征掩蔽机制减少饱和区域特征的贡献,取得了较好效果。这种应用掩膜来引导不同饱和度和区域进行图像重建的思想同样被用于Liba等人^[7]的研究中,其创建了一个精确的天空掩膜并使用它们来改善天空的外观,通过引入掩膜将天空进行了特别的处理和计算,方便对天空进行白平衡处理、亮度调整、对比度增强和降噪等处理。该算法在对天空场景的处理中取得了很好的效果。

在图像处理任务中,注意力机制^[8]可以在众多的输入信息中聚焦于对当前任务更为关键的信息,降低对其他信息的关注度,提高任务处理的效率和准确性。不同注意力机制的作用原理各具特色,如Channel Attention Module^[9]强调图像或视频中不同区域的重要性;Self Attention^[10]能够对序列中不同位置的元素进行加权;Dual Attention^[11]从两个正交的角度进行Self Attention;Neighborhood Attention^[12]将自注意力定位到最近的相邻像素,具有线性的时间和空间复杂度;ACmix^[13]既享受自注意和卷积的好处,又与纯卷积或自注意相比具有最小的计算开销。

本文的研究内容是进行人像场景下的单帧HDR图像重建。在单帧HDR中,人像场景下的HDR处理一直是一个难点,主要因为皮肤区

域容易受到其他颜色的干扰,会发生不同程度的偏色。以前的研究往往通过降低整张图片的饱和度、对比度等方式对偏色进行抑制,这种方法会使得整体画面质量损失较大。本文受Liu等人^[6]和Liba等人^[7]的启发,用加入掩膜来对特定区域进行处理的思想,提出了U²HDRnet算法。该算法包含3个模块:皮肤特征提取模块、三边特征提取模块、色彩重建模块。皮肤信息提取模块构建了一个皮肤掩膜,提取了皮肤特征,将皮肤和非皮肤区域分离;三边特征提取模块会提取图片的多种信息,并且融合皮肤特征,最终输出一个公共的特征集;在色彩重建模块中,会再次加入皮肤区域的权重参数,抑制了皮肤区域可能产生的发红、发黄等偏色问题,从而避免了皮肤区域的失真。为了提高最终的成像质量,本文通过加入注意力模块的方式来提升网络的特征提取能力,过滤掉无关内容,聚焦关键信息。此外,本文创建了一个人像HDR数据集PortraitHDR,该数据集包含5 197张人像图片,填补了该领域内人像数据集的空白。实验结果表明,在该数据集上,本文所提出算法的PSNR达31.42,SSIM达0.985,在客观指标和主观视觉效果上均优于现有主流算法。

2 基本原理

通过特征提取网络执行HDR图像处理任务的策略由Gharbi等人^[14]提出,其构建了一个用于HDR处理任务的双边图像增强网络,称之为HDRnet,通过提取特征信息引导对不同区域执行不同程度HDR增强的方案获得了较好的效果。

HDRnet能在全分辨率图像上实时执行HDR增强,同时还可以捕捉高频效果。其通过切片、仿射颜色变换、全分辨率损失3种方案在低分辨率下进行了大量处理。HDRnet用双边滤波的方式保留边缘和细节信息。双边滤波是一种图像滤波方法,它考虑了像素之间的空间距离和像素值之间的差异。在双边滤波中,像素的权重由空间距离和像素值差异共同决定。此外,Gharbi等人还提出了一个基于深度学习的两边学习网络,用于学习图像的局部和全局特征,并结合双边滤波原理进行图像增强。该网络由局部特征提取

器和全局特征提取器两个主要组件组成。局部特征提取器使用CNN来提取图像的局部特征。它通过学习滤波器权重来捕捉图像的低级细节和纹理信息。全局特征提取器使用全连接层或者其他适当的结构来学习图像的全局特征。该网络可以捕捉图像的整体和上下文信息,之后将全局信息和局部信息相结合再将其转换到双边网格之中。此外,通过对输入图像做仿射变换得到引导图,并将其用来引导前面的双边网格做空间和颜色深度上的插值,恢复到与原来图像一样大小。最后将得到的特征对原图像做仿射变换,得到输出图像。该网络能够很好地提取图片的局部特征和全局特征。

HDRnet提供的这种先提取特征,再融合特征,最后再进行图像重建的方法,为本文的基于多特征融合的人像HDR图像处理算法提供了理论支撑。

3 U²HDRnet 算法模型

U²HDRnet选取visuAAL^[15]作为皮肤分割模块的预训练数据集。采用迁移学习的策略,通过读取在visuAAL数据集训练好的U²net^[16]网络参数对皮肤信息进行提取。同时,三边特征提取模块对图片进行全局信息、局部信息和语义信息的提取,并融合皮肤信息,得到最终的特征集。最后通过融合后的特征信息来引导双边网格做空间和颜色深度上的插值,并在最后阶段再次输入皮肤部分的权重,从而确保皮肤区域色彩的稳定性。

3.1 U²HDRnet 结构

由图1可知U²HDRnet模型主要由3部分构成,即三边特征提取模块、皮肤特征提取模块和色彩重建模块。皮肤特征提取模块所提取到的颜色和位置信息是网络后续计算的重要基础,其处理结果将输入到另外两个模块中,确保皮肤部分不会发生偏色。三边特征提取模块负责各类特征的提取。色彩重建模块会执行最小但关键的工作量:它将学习一个灰度图,使用新切片节点将仿射系数的网格上采样回全分辨率。然后将这些像素的局部仿射变换应用于全分辨率的输入,同时在计算中充分考虑皮肤部分占据的权重,产生最终的输出。

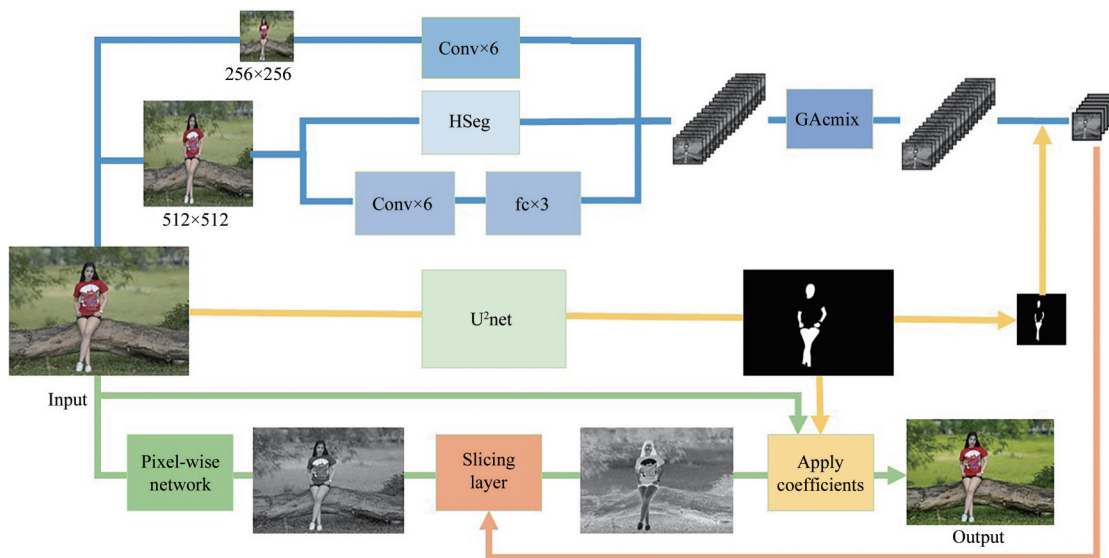


图1 U²HDRnet 网络结构图

Fig. 1 Network structure diagram of U²HDRnet

3.2 三边特征提取模块

为了对图片的多种特征信息进行充分提取,本文提出了一种三边特征提取网络。该网络由3条支路组成,其中局部特征提取支路在低分辨率下进行,语义信息提取支路和全局信息提取支路在高分辨率下进行,并在多种特征融合时通过添加注意力模块的方式提升特征提取效果。在三边特征提取模块中执行了大部分的计算。在HDR处理任务中,通常不仅依赖于局部图像特征,还依赖于全局图像特征和语义特征,如直方图、平均强度,甚至场景类别。因此,本模块的三边网络的特征提取部分进一步分为局部特征提取、全局特征提取和语义特征提取。局部特征提取在低分辨率流下工作,其输入的分辨率为 256×256 。低分辨率流的输入首先通过5次卷积计算降低图片的空间分辨率,计算结束后再次进行两次卷积计算,得出其最终的局部特征。高分辨率流的输入分辨率为 512×512 ,全局特征提取支路在6次卷积计算后,通过3次全连接计算得到最终的全局特征并输出结果。语义特征提取支路在HSeg模块中进行4次下采样之后,再进行一次上采样,提取深度语义信息,处理后的信息将进行两次卷积,之后输出最终的语义信息。各分辨率流计算完毕后,对输出的3部分特征和皮肤特征进行融合以产生最终的仿射系数。

表1展示了不同支路的通道数。可以发现,

在语义信息提取支路,其通道的深度达512,超过了另外两条支路。由此可见,该支路相对另外两条支路能更好地提取到语义信息。三边网络中的HSeg模块如图2所示。由图2可知,HSeg的结构在Segnet^[17]网络的基础上做出了改进,在进行一次上采样操作后,通过两次卷积操作,再输出网络所需要的语义信息。改进后结构的输出能够和另外两条特征提取支路的信息保持相同的维度,方便特征信息的融合。

表1 不同特征提取的支路通道数

Tab. 1 Number of branch channels for different feature extraction

支路	局部特征支路	全局特征支路	语义信息支路
Channels	64	256	512

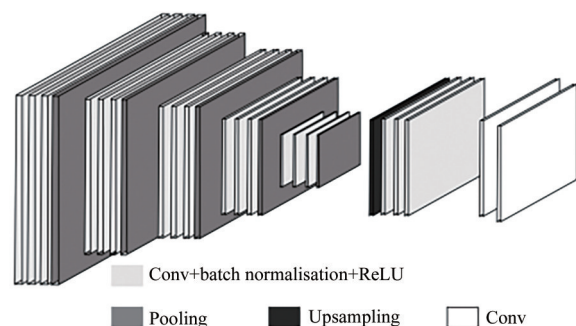


图2 HSeg网络结构示意图

Fig. 2 Schematic diagram of HSeg network structure

在得到三边网络输出的特征后,使用一个逐点的仿射变换去融合全局特征、局部特征和语义特征。使用 ReLU 函数进行激活,同时将 3 部分的仿射相加,之后将肤色区域的系数乘以 0.8 进行弱化,产生最终的仿射系数。最终得到了一个 $16 \times 16 \times 64$ 的特征矩阵,将其展开为 $16 \times 16 \times 8$ 的双边网格,每个格子有一个 3×4 的仿射颜色变换矩阵。

3.3 GAcmix 模块

为了使 ACmix 能够适应当前的 HDR 任务,提高其对于全局信息的感知能力,本文用全局自注意力 (Global-self-attention, GSA) 模块^[18]替代其中的 self-attention 模块。GSA 相比于 self-attention 更加关注序列中所有位置之间的全局关系,它允许每个位置与整个序列进行交互,从而更好地捕捉全局上的依赖关系,所以更加适合 HDR 任务。图 3 和图 4 分别展示了 GSA 模块和 GAcmix 的

结构,本文将在后续的实验证明此改进的有效性。

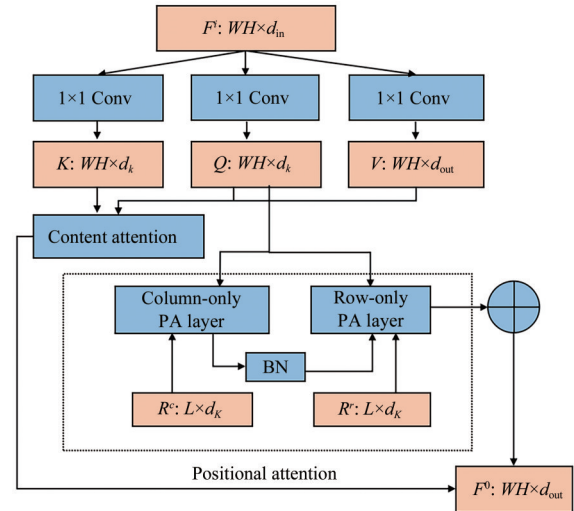


图3 GSA模块结构示意图

Fig. 3 Schematic diagram of GSA module structure

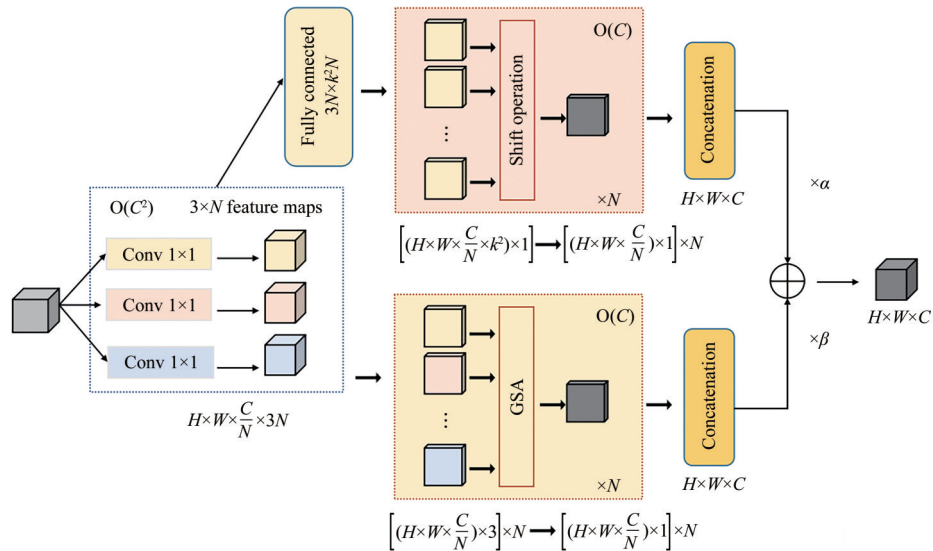


图4 GAcmix模块结构示意图

Fig. 4 Schematic diagram of GAcmix module structure

3.4 皮肤特征提取模块

本文采用 U^2net ^[16] 作为提取皮肤特征的网络。U型结构的图像分割网络是在 FCN 网络^[19] 的基础上构建,最早在 2015 年由 Ronneberger 等人^[20] 提出并命名为 U-Net。在此基础上,后续发展出了很多改进型的 U 型网络,如 UNET++^[21]、UNET 3+^[22] 等。但 U^2net 是其中的集大成者,其提出了一种名为 RSU 的新模块。每个 RSU 本

身其实就是一个缩小版的 U-Net,最后所有的 RSU 连接在一起,形成一个大 U 形的结构,命名为 U^2net 。 U^2net 较 U-Net 网络和其他 U-Net 改进型网络相比,由于通过其内部多个 RSU 结构多次进行特征的提取,因此在进行小目标的分割时,有着很好的效果。在人像场景中,由于皮肤区域所占比重较小,因此将 U^2net 作为皮肤特征提取模块最为合适。 U^2net 的网络结构如图 5 所示。

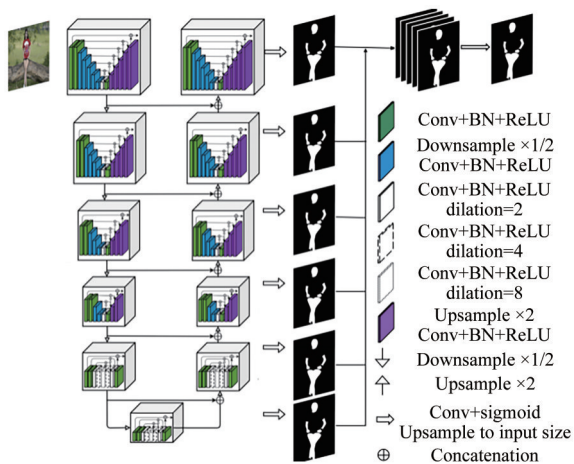


图5 U²net网络结构示意图

Fig. 5 Schematic diagram of U²net network structure

visuAA数据集包含了日常的绝大部分场景,并且对皮肤区域进行了手动标注。本文选择其作为皮肤信息提取模块的预训练网络。通过迁移学习的方式,先在 visuAA 数据集上对网络进行预训练,在之后的训练中,再读取其预训练的参数对图片进行处理,这样就生成了所需的皮肤掩膜。其处理效果如图6所示。

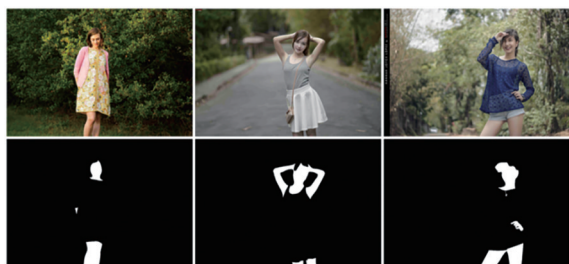


图6 皮肤分割效果

Fig. 6 Skin segmentation effect

由图6的分割效果可以发现,预训练后的U²net网络皮肤分割能力优秀,其生成了高质量的皮肤掩膜,在后续的处理中有着重要作用。

皮肤信息将在三边网络完成特征提取后,与多种特征信息进行融合,从而避免皮肤区域的偏色。同时在色彩重建网络中,通过皮肤掩膜的引导来弱化对皮肤区域的颜色增强,从而保证了HDR处理后皮肤颜色的稳定。

3.5 色彩重建模块

色彩重建模块在全分辨率下工作,执行的计算最少,但在捕获高频效果和保留边缘中具有关

键作用。本文沿用了Gharbi等人^[14]所使用的切片节点,该节点用基于学习的引导映射在仿射系数网格中执行数据的查找。通过用全分辨率切片网格获得高分辨率的仿射系数,同时对每个像素应用局部颜色变换以产生最终输出。

在将原图通过像素网络变为引导图的辅助网络时,将在全分辨率下做一个简单的点态非线性变换,其计算方法如公式(1)所示:

$$g[x, y] = b + \sum_{c=0}^2 \rho_c(M_c^T \cdot \odot_c[x, y] + b'_c), \quad (1)$$

其中: M_c^T 是 3×3 颜色变换矩阵的行; b 和 b'_c 是标量偏差; ρ_c 是分段线性传递函数,参数化为16个带阈值 $t_{c,i}$ 和斜率 $a_{c,i}$ 的缩放ReLU函数的和,其计算公式如式(2)所示:

$$\rho_c(x) = \sum_{i=0}^{15} a_{c,i} \max(x - t_{c,i}, 0). \quad (2)$$

参数 M 、 a 、 t 、 b 、 b' 与其他网络参数一起学习。 M 被初始化为恒等。 a 、 t 、 b 和 b' 被初始化,使得每个都是一个在 $[0, 1]$ 上的恒等映射。

3.6 损失函数

本文采用Charbonnier Loss^[23]和VGG Loss^[24]作为损失函数。Charbonnier Loss舍弃了 L_1 和 L_2 这两种损失函数,因为它们都会造成图像过于平滑,缺乏感官上的照片真实感。该函数具有极高的稳定性,其定义如式(3)所示:

$$L_{\text{charbonnier}} = \frac{1}{N} \sum_{i=1}^N \sqrt{(x_i - y_i)^2 + \epsilon^2}, \quad (3)$$

其中: x_i 是模型预测的第 i 个样本的输出值, y_i 是真实的标签或者期望的输出值, N 是样本的总数, ϵ 是平滑项常数。相对于 L_1 和 L_2 ,其损失曲线更平滑,在值接近零点时由于常数 ϵ 的存在,梯度也不会太小,避免了梯度消失;远离零点时值的梯度由于开方,梯度也不会太大,从而避免了梯度爆炸。具有Charbonnier损失函数的深度网络可以更好地处理异常值,与 L_2 损失函数相比,提高了SR性能。

VGG Loss的思想是使用VGG神经网络提取图像特征,比较生成的图像和风格图像在这些特征上的相似程度。具体而言,将生成的图像、风格图像和内容图像都输入到预训练的VGG网络中,提取出每一层的特征图,然后计算生成的图像和风格图像在每一层特征图上的差异,并将这些差异相加作为损失函数。由此可使得结果

更贴近主观评价的结果,公式如式(4)所示:

$$L_{\text{VGG}} = \frac{1}{N} \sum_{i=1}^N \frac{1}{M} \sum_{j=1}^M (\phi(x_i)_j - \phi(y_i)_j)^2, \quad (4)$$

其中: N 表示样本数量, M 表示选择的VGG中间层的通道数量, x_i 表示生成图像中的第 i 个样本, y_i 表示目标图像中的第 i 个样本, $(x_i)_j$ 表示生成图像在VGG网络的第 j 个中间层输出的特征, $(y_i)_j$ 表示生成图像在VGG网络的第 j 个中间层输出的特征。

通过Charbonnier Loss和VGG Loss这两个损失函数的组合,U²HDRnet模型可以在训练过程中同时考虑图像重建误差和内容相似性,从而提高图像增强效果。

4 实验结果与分析

4.1 实验平台

本文使用Pytorch框架部署整个模型,实验环境为Ubuntu 18.04. 系统,显卡为4块NVIDIA GeForce RTX2080Ti,单卡显存为11 GB。

4.2 实验数据集

本文采用公开数据集MIT-Adobe FiveK作为日常场景的数据集。但由于人像HDR领域存在数据空白,所以本文制作了一个人像HDR的数据集并将其命名PortraitHDR。该数据集部分图片通过DXO取得。DXO是一家专业影像机构,其提供的HDR图片具有极高的质量。另外部分图片是经专业摄影师通过专业HDR相机拍摄,并通过后期处理而成。整理后的PortraitHDR数据集由5 497张LDR人像和5 497张HDR人像图片组成,包含黑人、白人、黄种人3种肤色人种,同时包含了常见的街道、树木、天空、室内等多种常见的背景。此外还选取了不同光照强度和天气条件下的图片,基本包含了日常生活中的各种场景。经过随机选取的20人的主观评价可知,PortraitHDR数据集中的HDR图片比LDR图片具有更广阔的动态范围,捕捉到了更多细节,包括高光和阴影部分。照片同时显示出明亮和细节丰富的高光区域以及暗部区域,其颜色鲜艳,对比度较高,符合HDR图像的标准。

本文选取LDR和HDR图片各5 197张作为训练数据集。同时选择了LDR和HDR图片各300张作为测试数据集,用以测试训练后模型的

HDR效果。在训练时,本文通过翻转、平移、裁切等不影响HDR处理效果的数据增强方式对数据集进行扩充。

4.3 评价指标

本文采用客观评价和主观评价两种评价方式,其中客观评价指标采用峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)和结构相似性(Structure Similarity Index Measure, SSIM)。同时通过对比肤色区域的直方图对肤色重构效果进行评判。PSNR是一种重要的图像质量评价标准,是最普遍、最广泛使用的评鉴图像的客观量测法,其值越大,表明图片质量越高。PSNR被定义为:

$$\text{PSNR} = 10 \cdot \lg \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right), \quad (5)$$

其中:MAX是图像像素值的最大可能取值,对于8位灰度图像其值为255;MSE是均方误差,表示重建图像和原始图像之间像素值差异的平方的平均值。

由于PSNR的分数无法与人眼看到的视觉品质完全一致,所以本文还采用了另外一个评价指标SSIM,SSIM在评价图像质量上更能符合人类的视觉特性。SSIM使用的两张图像中,一张为未经压缩的无失真图像,在本实验中其对应的是测试集中的理想HDR图片;另一张为处理后的图像,在本实验中其对应的是网络处理后的HDR图片。对于 x 和 y 两个图像的结构相似性,可按式(6)求出:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (6)$$

其中: μ_x 表示图像 x 的亮度均值; μ_y 表示图像 y 的亮度均值; σ_x 表示图像 x 的亮度方差; σ_y 表示图像 y 的亮度方差; σ_{xy} 表示图像 x 和图像 y 的亮度协方差; c_1 和 c_2 是两个用于稳定计算的常数,它们可以避免分母为0的情况。

SSIM值的范围为[0,1],值越大代表图像越相似。当两张图片完全一样时,SSIM值为1。在后续的实验中,通过计算测试集中的理想HDR图片与网络处理后的HDR图片的SSIM,可以判断是否达到了理想效果。

直方图是一种用于可视化数据分布的图形工具,可用于分析和比较数据的频率分布。在图像处理中,直方图通常用于表示图像中各个像素

值的分布情况。对于彩色图像,可以通过计算其不同色彩的直方图的方式简单且快速对两张图片进行相似性评估。通过将网络处理后皮肤区域的图片的直方图同理想的 ground truth 的直方图进行对比,可以对肤色重构效果进行评价。

4.4 消融实验

本文设计了多组消融实验对模型的性能进行充分的评估。为了验证本文所提出三边特征提取网络的性能,本文在 FiveK 数据集上设计了消融实验来对比不同特征提取网络之间的性能差异,其具体表现如表2所示。

表2 不同特征提取网络的效果对比

Tab. 2 Comparison of different feature extraction networks

策略	特征提取网络改进策略			PSNR
	语义特征	局部特征	全局特征 动物	
Gharbi		✓	✓	28.42
Ours(Bilateral)		✓	✓	29.23
Ours(Bilateral)	✓	✓		29.94
Ours(Bilateral)	✓		✓	29.98
Ours(Trilatera)	✓	✓	✓	30.73

通过表2可知,相较于 Gharbi 等人^[14]的双边特征提取网络,本文所提出的特征提取网络在仅调用全局和局部特征提取流时,其 PSNR 在 Gharbi 等人的基础上有 2.8% 的进步。当在特征提取网络中加入语义信息提取支路时,其 PSNR 取得了 8.1% 的进步。结合了语义信息的三边网络相较于 Gharbi 等人的双边网络 PSNR 数值高出 2.31。这一方面说明了本文所提出的三边网络性能更好,相较于双边网络可以提取到更多的图片信息;另一方面说明在 HDR 图像处理任务中,通过提升网络的特征提取能力来提升其 HDR 处理能力的策略是有效的。

在三边特征提取模块中,为了使其能聚焦于关键信息,忽略次要信息或噪声干扰,本文采取了添加注意力机制的策略。为了探究不同注意力机制对输出的影响,本文在三边网络的不同支路设计了消融实验,用 FiveK 数据集作为训练和测试数据集,比较了 Channel Attention^[9]、SA^[10]、Dual^[11]、ACmix^[13]、NA^[12]和 GACmix 分别作用于局部特征提取支路、全局特征提取支路、语义特征提取支路和特征融合支路的效果,结果见表3。

表3 不同注意力模块作用于不同支路的效果对比

Tab. 3 Comparison of effects of different attention modules on different brands

注意力模块	PSNR			
	局部 信息 支路	全局 信息 支路	语义 信息 支路	特征 融合 支路
GACmix	30.68	30.51	30.75	31.36
Dual attention	29.99	30.27	30.72	31.04
NA	29.98	30.42	29.62	31.26
Channel Attention	30.49	30.28	30.37	31.21
SA	30.08	30.25	30.52	30.77
ACmix	30.06	30.14	30.21	31.03

通过表3可知,不同注意力机制在作用于不同支路时的效果迥异。对比后可以发现,当 GACmix 模块作用于不同特征提取支路上时效果都较为理想,当其作用于特征融合支路时对网络的提升最大,其 PSNR 提升幅度为 2.2%,因此本文选择加入该注意力模块增强网络的特征提取能力。通过对比 Acmix 和 GACmix 的效果可知,改进后的 GACmix 效果更好,不论作用于何支路时相较于 Acmix 效果皆有一定的提升,平均提升幅度约为 1.5%,说明本文对 Acmix 做出的改进是有效的。

为了验证在 U²HDRnet 模型中加入皮肤掩膜后对人像处理能力是否有提升,本文对比了分别采用 FCN^[19]、Unet^[20]、Unet3+^[22]、U²net^[16]作为分割网络生成皮肤掩膜时,在 PortraitHDR 数据集上的性能差异,其结果如表4所示。

表4 加入不同皮肤分割模块前后的效果对比

Tab. 4 Comparison of effects before and after adding different skin segmentation module

策略	SSIM
No mask	0.915
+FCN	0.967
+Unet	0.964
+Unet3+	0.968
+U ² net	0.985

通过表4可知,是否加入皮肤掩膜对于 SSIM 数值有影响。SSIM 数值的高低能反映出处理后的图片与理想的 HDR 图片相比是否发生了偏色

或者处理过度。在加入皮肤掩膜后,SSIM 的数值获得了最高 7% 的提升,说明本文添加皮肤掩膜的策略是正确的。同时,通过表 4 可知,在加入不同的皮肤分割网络时其 SSIM 数值有所差异,在选取 U²net 作为皮肤提取网络时,结果最好,因此选择 U²net 作为本网络的皮肤提取模块。

图 7 呈现了加入 U²net 生成的皮肤掩膜前后的变化。可以发现,在没有加入皮肤掩膜处理时,其输出图片的皮肤颜色明显偏红。在加入皮肤掩膜后,肤色更为自然,观感更好,同时也更贴近原本的肤色。



图 7 加入皮肤掩膜前后的效果对比。(a) 输入图像;(b) 加入皮肤掩膜前;(c) 加入皮肤掩膜后。
Fig. 7 Comparison of effects before and after adding skin masks. (a) Input images; (b) Before adding skin mask; (c) After adding skin mask.

图 8 呈现了融合皮肤特征前后,其他区域具体的细节信息。可以发现,肤色信息的加入只对皮肤的偏色进行了抑制,其他区域的色彩并未发生变化,说明皮肤特征提取模块的加入既纠正了皮肤区域的偏色,又不会对非肤色区域产生负面影响。

图 9(a)~(c) 分别呈现了 ground truth、加入皮肤特征时、不加入皮肤特征时皮肤区域的直方图。通过观察直方图可以发现,当不加入皮肤特征时,其像素值在 0~40,与 ground truth 明显不同,说明此时颜色发生了偏色;在加入了皮肤特征之后,其直方图和 ground truth 表现一致,像素值并未发生偏离,说明皮肤特征的加入使皮肤区域不再失真,得到了理想的处理效果。直方图的表现和 SSIM 数值的表现一致,也说明了 SSIM

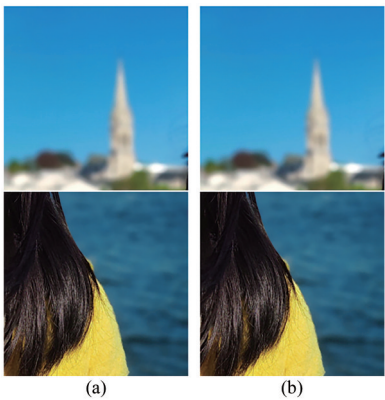


图 8 加入皮肤掩膜前后的细节对比。(a) 加入皮肤掩膜前;(b) 加入皮肤掩膜后。
Fig. 8 Comparison of details before and after adding skin mask. (a) Before adding skin mask; (b) After adding skin mask.

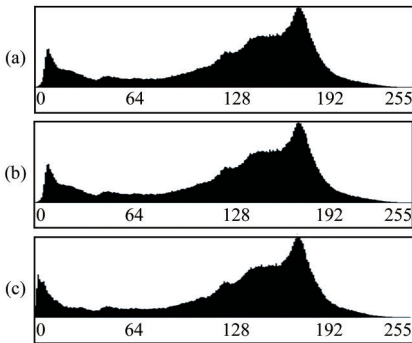


图 9 直方图对比。(a) 真值;(b) 加入皮肤特征;(c) 不加入皮肤特征。
Fig. 9 Histogram comparison. (a) Ground truth; (b) Adding the skin feature; (c) No adding the skin feature.

的数值可以呈现肤色是否发生偏色,其数值越高,说明整张图同 ground truth 越接近,肤色重构效果越一致。

为了探究皮肤特征与不同特征进行融合时,对输出的影响,本文比较了加入 U²net 作为皮肤提取模块前后,网络在 PortraitHDR 数据集上性能的差异。其结果如表 5 所示。

表 5 不同融合策略的效果对比

Tab. 5 Comparison of the effects of different fusion strategies					
策略	+三边特征 提取模块	+注意力 模块	+皮肤 分割模块	PSNR	SSIM
方案一	✓			30.73	0.909
方案二	✓	✓		31.46	0.915
方案三	✓	✓	✓	31.42	0.985

通过表5可知,不同特征分别与肤色特征进行融合时,对于SSIM的提升是巨大的,SSIM数值越大说明与ground truth越接近。表5的结果证明皮肤特征与局部特征、全局特征、语义特征进行融合时,都对网络的性能有提升。融合后的PSNR会有0.01的下降,这是因为为了防止皮肤区域发生偏色,在处理皮肤区域时会抑制过亮、过鲜艳,所以会带来峰值信噪比0.01的损失,但却带来了画面相似度7.3%的提升。此外,通过对比表2和表5的结果可以发现,采用特征

融合的策略比只提取单一特征的策略其性能更好。

依次使用不同组合策略在PortraitHDR数据集上进行训练,实验结果如表6所示。通过表6可知,本文所搭建的三边特征提取网的PSNR数值均高于30,通过加入通道注意力模块的方式可以进一步提升PSNR的数值。同时,皮肤提取模块的加入将SSIM提升了8.4%。这说明通过本文所采取的策略,在提升整体HDR处理效果的同时,可以减小人像场景下HDR处理的失真。

表6 不同组合策略的效果对比

Tab. 6 Comparison of the effects of different combination strategies

策略	局部特征	全局特征	语义特征	PSNR	PSNR(+U ² net)	SSIM	SSIM(+U ² net)
方案一	✓			27.32	27.29	0.866	0.916
方案二		✓		28.01	28.01	0.852	0.922
方案三			✓	27.98	27.96	0.867	0.928
方案四	✓	✓	✓	30.73	30.72	0.906	0.978

4.5 对比实验

为了验证U²HDRnet模型对自然场景的处理能力,同时对比相较于主流HDR算法,本文所提算法是否优秀,本文将肤色信息的引导因子置为

0,在FiveK数据集上对比了U²HDRnet同style-transfer^[3]、HSVNet^[4]、FHDR^[5]、CNN-HDR^[6]、HDRnet^[14]算法在自然场景下的表现效果,其结果见表7。

表7 不同算法在FiveK数据集上的效果

Tab. 7 Effects of different algorithms on the FiveK dataset

算法	PSNR					
	风景	建筑	静物	车辆	动物	食品
HSVNet	26.52	23.35	25.22	25.41	22.18	26.37
HDRnet	28.96	29.65	28.22	27.47	25.14	28.01
Style-transfer	26.74	27.51	26.94	28.33	27.29	27.25
CNN-HDR	23.42	23.15	23.86	22.02	21.23	22.51
FHDR	29.93	30.52	30.39	30.85	27.58	29.86
Ours (no mask)	31.26	31.38	30.39	30.67	31.84	31.45

通过实验发现,在未加入皮肤提取模块时,U²HDRnet在多个场景中的表现都优于常见算法。在处理车辆场景时,其性能对比FHDR算法有落后,通过查看数据集发现,FiveK数据集中车辆场景的色彩比较单一,大多是纯黑色或纯白色车辆,U²HDRnet所擅长的是深度特征提取,对于色彩过于单一的场景,其性能对比FHDR这种通过不断迭代提高饱和度的算法,其PSNR有0.18的落后,但是落后幅度只有0.5%。综合所有场景

来看,U²HDRnet比其他单帧HDR算法在不同场景下的PSNR值波动最小,且PSNR值都高于30,表明本文所构建的HDR图像处理网络具有很强的稳定性。

为了验证U²HDRnet在人像场景下的HDR处理能力,同时比较不同算法在人像场景下处理能力的优劣,本文在PortraitHDR数据集上对不同的算法进行了对比实验,结果如表8所示。

从表8可以看出,U²HDRnet在人像场景下,

表 8 不同 HDR 算法在人像场景下的效果
Tab. 8 Effect of different HDR algorithms in portrait scenes

算法	PSNR	SSIM	训练时间/h	推理时间/s
FHDR	28.32	0.829	20.6	0.24
HDRnet	28.38	0.904	16.8	0.10
Style-transfer	27.41	0.905	22.6	0.15
CNN-HDR	22.15	0.826	12.3	0.09
HSVNet	30.32	0.932	26.8	0.32
U ² HDRnet	31.42	0.985	17.2	0.10

其 PSNR 有着最高的得分,同时在 SSIM 上的表现好于其他网络。SSIM 所呈现的是网络处理后的图片与理想 HDR 图片的相似度。由于本文的研究重点是在人像场景下获取高质量 HDR 图像的同时抑制皮肤的偏色,所以 SSIM 指标更能体现是否达到了抑制皮肤偏色的效果。对比不同网络的训练和推理时间可以发现,CNN-HDR 的

训练和推理时间是最短的,但是其处理效果也是最差的。U²HDRnet 算法相较于采用双边特征网络的 HDRnet 算法,其训练时间和推理时间增加了 0.4 h,但是 PSNR 提升了 10.7%,同时 SSIM 提升了 8.9%。这是因为该算法中的 U²net 模块采取了迁移学习的策略,缩减了训练所花费的时长。因此,U²HDRnet 融合的特征信息虽然更多,但是其训练时间只增加了 2%。此外由于本文采取了添加注意力机制的方案过滤了部分无关参数,所以其推理速度在对比的算法中排名第二。结合各项客观指标来看,采用多特征融合的方式进行 HDR 图像处理的策略是成功的,U²HDRnet 算法的综合性能是各个算法中最好的。

除了客观评价指标,主观评判更能说明算法对于人像 HDR 处理效果的优劣。图 10 展示了 U²HDRnet 算法同其他数值表现相对较好算法的效果对比。

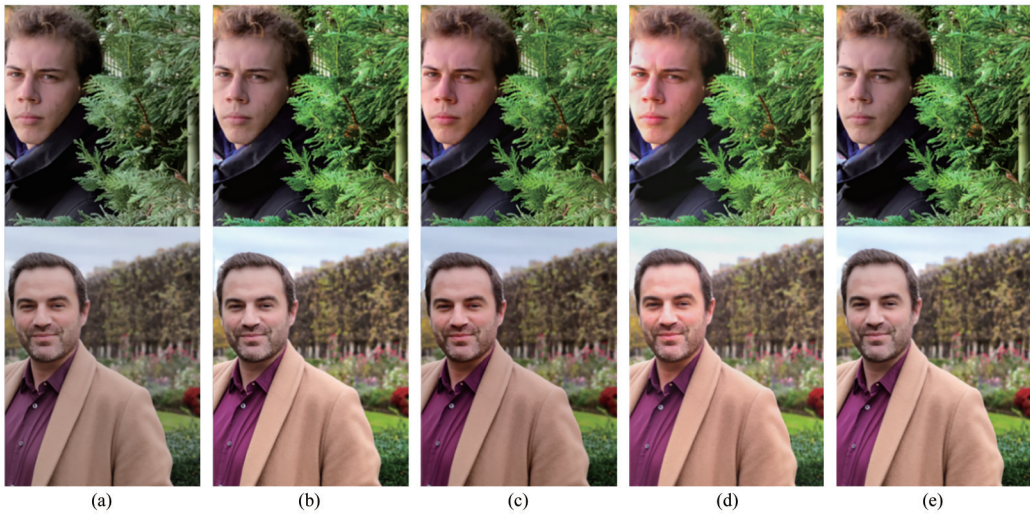


图 10 不同算法在 PortraitHDR 数据集上的效果对比。(a)输入图片;(b)理想输出图片;(c)FHSVNet 算法输出图片;(d)HDRnet 算法输出图片;(e)U²HDRnet 算法输出图片。

Fig. 10 Comparison of different algorithms on the PortraitHDR dataset. (a) Input images; (b) Ideal output images; (c) FHSVNet algorithm outputs images; (d) HDRnet algorithm outputs images; (e) U²HDRnet algorithm outputs images.

通过图 10 可以发现,对比其他算法,U²HDRnet 在处理人像场景时优势明显,在抑制皮肤偏色上有着很好的效果。相比于其他网络,其在纠正了肤色的加深和偏色的同时,保留了皮肤的细节,同理想的 target 图片效果高度一致,没有发生失真。

通过客观评价指标结合主观评判可以发现,U²HDRnet 在获得高质量 HDR 处理效果的同时

保留了肤色信息和皮肤细节,在进行人像场景下的 HDR 处理时有着更好的视觉效果。

5 结 论

本文提出了一种基于多特征融合的人像 HDR 图像处理算法 U²HDRnet,旨在解决现存 HDR 算

法在处理人像图片时容易造成皮肤偏色的问题。该算法采用U²net作为皮肤特征提取的模块,将提取完成的皮肤特征输入到三边特征提取模块和色彩重建模块中。同时,本文所提出的三边特征提取模块表现优异,该模块的局部特征提取支路先将图像做下采样以降低图片分辨率,通过多次卷积提取局部信息;该模块的全局特征提取支路在对图片进行多次卷积计算后,再通过多次全连接提取全局特征;该模块的语义信息提取支路在HSeg模块中进行提取深层语义信息提取。最终对输出的3部分特征进行融合,在经改进后的

Acmmix模块计算后融合皮肤特征,产生最终的仿射变换系数。在色彩重建模块中,在其切片节点采用基于学习的引导映射。通过用全分辨率切片网格获得高分辨率的仿射系数,同时对每个像素应用局部颜色变换以产生最终输出,在最终的HDR重建中再次融合皮肤特征的权重,尽可能地保留了皮肤信息。实验数据表明,本文所提出算法的PSNR达31.42 dB,SSIM达0.985。结合主观评测结果可以发现,U²HDRnet算法在人像场景下的HDR处理能力优于目前常见的HDR算法。

参 考 文 献:

- [1] 吴玲凤,李娜,胡骏保. 基于重映射和曝光融合的HDR成像[J]. 液晶与显示,2021,36(12):1712-1719.
WU L F, LI N, HU J B. HDR imaging using remapping and exposure fusion [J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(12): 1712-1719. (in Chinese)
- [2] 芦碧波,皇甫珍珍,郭凯,等. 基于双边滤波的多尺度分层色调映射算法[J]. 液晶与显示,2018,33(9):816-822.
LU B B, HUANGFU Z Z, GUO K, *et al.* Bilateral filter based multiscale layer tone mapping algorithm [J]. *Chinese Journal of Liquid Crystals and Displays*, 2018, 33(9): 816-822. (in Chinese)
- [3] XIA X D, ZHANG M, XUE T F, *et al.* Joint bilateral learning for real-time universal photorealistic style transfer [C]. 16th European Conference on Computer Vision. Glasgow: Springer, 2020: 327-342.
- [4] LEE M J, RHEE C H, LEE C H. HSVNet: reconstructing HDR image from a single exposure LDR image with CNN [J]. *Applied Sciences*, 2022, 12(5): 2370.
- [5] KHAN Z, KHANNA M, RAMAN S. FHDR: HDR image reconstruction from a single LDR image using feedback network [C]. IEEE Global Conference on Signal and Information Processing (GlobalSIP). Ottawa: IEEE, 2019: 1-5.
- [6] LIU Y L, LAI W S, CHEN Y S, *et al.* Single-image HDR reconstruction by learning to reverse the camera pipelin [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020: 1648-1657.
- [7] LIBA O, MOVSHOVITZ-ATTIAS Y, CAI L Q, *et al.* Sky optimization: semantically aware image processing of skies in low-light photography [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Seattle: IEEE, 2020: 2230-2238.
- [8] MNIH V, HEES N, GRAVES A, *et al.* Recurrent models of visual attention [C]//*Proceedings of the 27th International Conference on Neural Information Processing Systems*. Montreal: MIT Press, 2014: 2204-2212.
- [9] ZHANG Y L, LI K P, LI K, *et al.* Image super-resolution using very deep residual channel attention networks [C]// *Proceedings of the 15th European Conference on Computer Vision (ECCV)*. Munich: Springer, 2018: 294-310.
- [10] ZHAO H S, JIA J Y, KOLTUN V. Exploring self-attention for image recognition [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020: 10073-10082.
- [11] FU J, LIU J, TIAN H J, *et al.* Dual attention network for scene segmentation [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019: 3141-3149.
- [12] HASSANI A, WALTON S, LI J C, *et al.* Neighborhood attention transformer [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver: IEEE, 2023: 6185-6194.
- [13] PAN X R, GE C J, LU R, *et al.* On the integration of self-attention and convolution [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans: IEEE, 2022: 805-815.
- [14] GHARBI M, CHEN J W, BARRON J T, *et al.* Deep bilateral learning for real-time image enhancement [J].

- ACM Transactions on Graphics*, 2017, 36(4): 118.
- [15] HASHEMIFARD K, FLOREZ-REVUELTA F. From garment to skin: the visuAAL skin segmentation dataset [C]. ICIAP 2022—Workshops on Image Analysis and Processing. Lecce: Springer, 2022: 59-70.
- [16] QIN X B, ZHANG Z C, HUANG C Y, *et al.* U²-Net: going deeper with nested U-structure for salient object detection [J]. *Pattern Recognition*, 2020, 106: 107404.
- [17] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481-2495.
- [18] SHEN Z R, BELLO I, VEMULAPALLI R, *et al.* Global self-attention networks for image recognition [J/OL]. *arXiv*, 2020: 2010.03019.
- [19] LONG J, SHEHMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston: IEEE, 2015: 3431-3440.
- [20] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]. 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich: Springer, 2015: 234-241.
- [21] ZHOU Z W, SIDDIQUEE M M R, TAJBAKHS N, *et al.* UNet++: a nested U-Net architecture for medical image segmentation [C]. 4th International Workshop on Deep Learning in Medical Image Analysis and 8th International Workshop on Multimodal Learning for Clinical Decision Support. Granada: Springer, 2018: 3-11.
- [22] HUANG H M, LIN L F, TONG R F, *et al.* UNet 3+: a full-scale connected UNet for medical image segmentation [C]. 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona: IEEE, 2020: 1055-1059.
- [23] LAI W S, HUANG J B, AHUJA N, *et al.* Fast and accurate image super-resolution with deep laplacian pyramid networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 41(11): 2599-2613.
- [24] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C]//*Proceedings of the 3rd International Conference on Learning Representations (ICLR)*. San Diego, Curran Associates Inc., 2015: 1-14.

作者简介:



吴春林,男,硕士研究生,2020年于西南交通大学希望学院获得学士学位,主要从事机器视觉及自动控制方面的研究。E-mail:218527109@fzu.edu.cn



林坚普,男,博士,讲师,2019年于福州大学获得博士学位,主要从事新型显示技术、图像处理技术、电子纸驱动与集成等方面的研究。E-mail:ljp@fzu.edu.cn