

# 通过 Cross-Context Attention 和纹理增强去除面部高光

Hongsheng Zheng, Wenju Xu, Zhenyu Wang, Xiao Lu, and Chunxia Xiao\*, *Member, IEEE*

**摘要:** 面部高光去除旨在识别并去除面部图像中的镜面高光成分, 确保生成的图像具有一致的面部色调和高保真纹理细节。现有的方法很难同时去除高光并恢复受干扰区域的细节, 通常会导致镜面残留或扭曲的局部细节 (即纹理、照明和颜色)。为了解决这些问题, 这项工作提出了一种新颖的两阶段面部高光去除网络 (FHR-Net), 它主要由跨上下文注意力模块 (CCAM) 和纹理增强模块 (TEM) 组成。在第一阶段, 根据检测到的高光掩模, CCAM 显式地整合跨上下文信息以获得与周围面部上下文一致的粗略高光去除结果。在粗略结果的基础上, 第二阶段的 TEM 利用分片注意力来细化高光区域的纹理细节, 从而产生高保真面部图像。为了提高去除的高光区域和非高光区域之间的一致性, 这项工作引入了面部特征损失, 使处理后的高光干扰区域与周围的面部结构很好地对齐。此外, 为了解决研究界缺乏高质量数据集的问题并满足数据驱动的面部高光去除的培训需求, 本工作通过交叉极化构建了真实世界的配对面部镜面漫反射 (PFSD) 数据集。在 PFSD 等数据集上的实验结果表明, FHR-Net 可以有效去除面部高光并恢复原始颜色和纹理细节。

**索引术语**——面部高光去除、跨上下文注意力、纹理增强、数据集和面部特征丢失。

## 一、简介

UE to the special illumination environment and oily 由于人体皮肤的特性, 面部表面容易呈现镜面高光。不可忽略的镜面高光成分破坏了面部纹理细节的完整性和连续性, 导致肤色不一致、精细结构特征丧失、亮度变化不自然等, 从而导致面部图像的真实感和审美情趣受到破坏, 尤其是强度不足。照明条件或尝试捕捉时

这项工作得到了国家自然科学基金 (No. 61972298 和 No. 62372336) 的部分支持。郑洪生、王振宇和肖春霞来自武汉大学计算机学院, 中国武汉 (电子邮件: hszheng@whu.edu.cn, 2021282110202@whu.edu.cn, cxxiao@whu.edu.cn)。Wenju Xu 来自亚马逊, Palo Alto, CA 94301 USA (电子邮件: xuwenju@amazon.com)。Xiao Lu 来自湖南师范大学工程与设计学院, 长沙, 中国 (电子邮件: luxiao@hunnu.edu.cn)\*。肖春霞是通讯作者。版权所有 © 20xx IEEE。允许个人使用本材料。但是, 若要将本材料用于任何其他目的, 必须通过发送电子邮件至 pubs-permissions@ieee.org 获得 IEEE 的许可。



图 1. 我们的面部高光去除结果的可视化示例。(a) 输入带有镜面高光的面部图像。(b) 我们的高光去除方法有效地提高了面部图像的整体美学质量。

微妙的面部表情。在这种情况下, 高光去除对于提高面部图像的质量和美感至关重要, 如图 1 所示。此外, 面部高光去除是提高下游计算机视觉任务性能的常见预处理, 例如人脸识别 [1]、3D 人脸重建 [2] 和人物检测 [3]。解决高光干扰对于生成视觉上令人满意的面部图像至关重要, 这引起了人们的广泛关注。然而, 面部高光区域中有用的上下文信息的可用性有限, 使得去除高光并恢复一致的肤色和纹理具有挑战性。此外, 人脸固有的复杂轮廓和局部纹理变化使得面部高光去除比自然物体更加困难。为了解决这些挑战, 我们引入了一种新颖的面部高光去除方法, 该方法可以有效地去除面部图像中的高光损坏, 同时恢复高保真纹理细节和肤色。早期的作品依赖于手工制作的先验知识, 例如纹理细节的相似性 [4]、图像暗通道 [5]、一致的颜色 [6]、[7] 和纯色照明 [8]–[12], 用于高光去除。然而, 这些先行要求在实际应用中很难满足。近年来, 基于深度学习的方法 [13]–[19] 在去除自然物体的高光方面取得了令人满意的结果。然而, 由于面部皮肤的几何和纹理特性, 将这些方法直接应用于面部高光去除具有挑战性。Chen et al 提出的皮肤反射模型。[20] 还展示了与其他物体相比皮肤的独特机制。无法区分面部和物体高光会导致高光去除算法的性能下降, 从而导致恢复的皮肤区域中出现不切实际的语义信息。还有一些作品 [21]、[22] 专注于面部高光去除。然而,

由于缺乏高质量的面部高光去除训练数据集, 性能并不令人满意。为了解决上述问题, 我们首先使用交叉偏振摄影构建真实世界的高质量配对面部镜面漫反射 (PFSD) 数据集。与现有数据集相比, PFSD 表现出更高的强度和更多样化的面部镜面反射。这一特性增强了现实场景中复杂光照条件下面部高光去除的效果。、强力高光去除。直观上, 面部的高光区域在其相邻区域中具有相似的纹理细节和肤色。因此, 这些邻近区域可以作为恢复面部高光区域的理想指南。基于这一观察, 为了有效去除面部高光, 同时恢复连贯的面部细节, 我们引入了一种新颖的从粗到细的两阶段面部高光去除网络 (FHR-Net)。具体来说, 在粗略阶段, FHR-Net 首先识别高光区域并使用检测模块生成相应的高光掩模。利用该掩模, 它自动将面部图像分为非高光区域、高光区域和最近邻周围区域。然后, 我们提出了一个跨上下文注意模块 (CCAM) 来整合来自非高光区域和最近邻周围区域的信息, 以指导高光区域的照明和肤色恢复, 从而实现粗略的高光去除。在细化阶段, 我们提出了纹理增强模块 (TEM) 进一步恢复和增强纹理细节, 利用地面实况 (GT) 作为辅助信息。The TEM 结合了增强的编码器-解码器架构和补丁式注意力机制。该架构从相应的 GT 中提取高频纹理特征并将其整合到粗略结果中, 实现最终的高保真高光去除结果。综上所述, 主要贡献如下: · 提出了一个两阶段的 FHR-Net 网络通过有效的结构重建和纹理细节增强以由粗到细的方式完成面部高光去除任务。提出了跨上下文注意模块 (CCAM) 和纹理增强模块 (TEM) 来实现跨区域面部结构重建和相干纹理细节恢复。· 构建基于交叉偏振的真实世界配对面部镜面漫反射 (PFSD) 数据集, 这有助于更好地处理更强的面部高光。

以图像去噪的方式去除高光。金等人[5]通过暗通道和最大后验法得到近似的漫反射图像。谭等人。[8]使用输入图像及其无镜面图像的反射基函数系数。明石等人。[25]通过非负矩阵分解同时估计漫反射颜色和分离镜面反射分量。魏等[26]利用光场的偏振特性来消除镜面反射。他们还提出了一种基于二色反射模型的  $x$  均值聚类方法来恢复图像中的不饱和高光像素[27]。但是, 这些方法无法解决具有复杂纹理和照明的图像, 例如面部图像。最近, 深度学习在高光去除任务中的应用变得越来越普遍。林等人[28]利用多分类判别器来解决高光去除残留问题。宋等人。[29]提出了一种多级曲率引导反射去除网络来逐步重建背景层信息并细化纹理细节。李等人。[30]设计了一种卷积形式的特征融合金字塔模块, 该网络可以充分利用多尺度特征和上下文信息来去除自然物体的表面反射光。傅等人。[31]提出了一种能够同时检测和去除图像中的镜面高光的多任务网络。吴等人。[14]在 GAN 网络中引入了注意力机制来直接建模镜面反射分量和漫射分量之间的关系。傅等人。[32]提出了一种三阶段高光去除网络, 该网络在大规模合成数据集上进行训练, 可以解决视觉伪影、颜色失真和色调偏差等问题。吴等人。[33]通过结合 Unet 和 Transformer 架构来检测并消除单个图像中的亮点。钟等人。[34]介绍了第一个利用语言指导进行图像反射分离的网络。这项工作采用交叉注意力和对比学习来建立语言描述和图像传输/反射层之间的对应关系。然而, 此类具有详细语言描述的数据集是公开有限的, 这可能会限制该方法的发展。面部高光去除。Deepflash [35] 将闪光灯自拍照转变为工作室品质的肖像, 显着增强图像美感, 同时有效管理面部的刺眼阴影和镜面反射光等问题。朱等人。[36]构建了可控的人像图像采集系统, 并通过最优的数据处理策略得到训练好的网络。任等人。[37]提出了一种在图像空间而不是 UV 空间中学习面部反射图来指导 3D 面部重建的方法, 并提取面部镜面反射分量以进行重新照明。韩等人。[38]采用组合照明模型来有效地表示真实照明, 并采用可变形的面部反照率模型作为反射优先实验来区分面部的漫反射和镜面反射。然而, 这些方法主要集中于提取面部的镜面反射分量, 而不是专门用于它们的去除。李等人。[20]、[39]第一次踏足这个领域

## 二.相关工作

单图像高光去除。大多数传统的高光去除方法都是基于颜色直方图分布的分析。阿图西等人。[23]提出了一项关于反射去除的调查, 总结了各种去除镜面反射的方法, 并分析了现有方法的优点和缺点。谭等人。[9]基于色度分析迭代比较了输入图像与其无镜面反射图像之间强度的对数微分。杨等人。[24]使用边缘保留低通滤波器

并提出了一种人脸图像高光去除方法。[21]构建了一个名为FaceSDIM的高分辨率人脸数据集，并从内在分解的角度解决了GAN网络的高光去除问题。但在某些光照条件下（人脸阴影较大），性能较差。[40]提出了一种基于朗伯一致性的轻量级优化方法，用于去除多视图面部图像中的镜面高光反射。[41]设计了一个用于面部区域的高光掩模提取模块。但是，它可能会将其他白色物体误识别为高光，从而导致颜色偏差。朱特等人。[42]使用条件生成对抗网络[43]来生成无高光图像，并通过高光检测引入新的掩模损失。但它无法解决复杂高光区域的色彩饱和度问题，并且会产生色彩失真的结果。穆罕默德等人。[22]引入了一个新的数据集 Spec-Face，包括被镜面照明损坏的面部图像和相应的漫反射图像。然而，基于深度学习的方法是数据驱动的，获得鲁棒的训练模型很大程度上取决于数据集的多样性和复杂性。由于缺乏高质量的面部数据集，这些方法在真实场景中的面部高光去除任务中的性能并不令人满意。

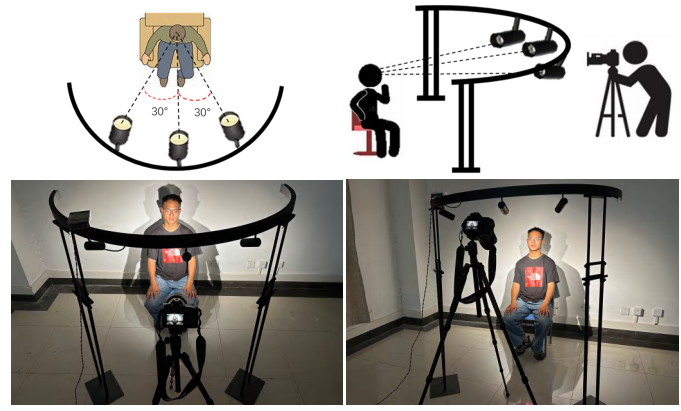


图 2. PFSD 数据集的构建概述。顶行示出了示意图，而底行描绘了相应的真实世界捕获的场景。

### B. 数据集构建

根据二向色反射模型[46]，面部图像可以分解为漫反射和镜面反射分量： $I(p) = I_d(p) + I_s(p)$  (1)

其中  $p$  表示像素， $I_d$  和  $I_s$  分别表示漫反射和镜面反射分量。交叉偏振有效地分离了漫反射和镜面反射分量，光学特性由马鲁斯定律描述（方程（2）） $I_s = I_{in} \cos^2(\theta)$ 。（2）

## 三. PFSD数据集

### A. 背景

高质量的数据集是数据驱动方法的基石。许多现有的自然物体高光去除方法利用交叉偏振技术来获取镜面反射-漫反射数据对[14]、[44]、[45]。然而，专门用于面部高光去除的数据集很少。唯一公开可用的真实世界面部数据集 FaceSDIM [21] 受到多种质量限制的阻碍，包括色调变化、照明不足和低强度高光。通过比较图 1 的顶行和底行可以看出，3 FaceSDIM 整体光照偏暗，高光成分不明显。高光区域是局部的和单一的，缺乏解决复杂的面部高光问题所需的丰富性和复杂性。地面实况不准确并且包含高亮残差，特别是在红框区域。此外，镜面高光图像与其相应的漫反射地面实况之间存在色调偏差。构建面部数据集比自然对象更具挑战性，主要是因为使用交叉偏振摄影获得具有和不具有镜面成分的配对面部图像通常需要调整光源或偏振角度。照明环境的这些变化通常会导致拍摄对象无意的面部运动或表情变化。因此，在实践中实现完美无抖动的配对数据是很困难的。

对于穿过两个正交偏光板的光，其中  $I_{in}$  为入射光的强度， $I_s$  为透过两个偏光板后镜面反射的强度， $\theta$  为两个偏光板之间的偏振方向角。当两个偏振片的偏振方向相互垂直（交叉偏振），即  $\theta=90^\circ$  时，镜面反射分量  $I_s=0$ ，即可得到无镜面反射的面部图像。为了满足高保真面部高光去除要求，我们基于该技术构建了一个新的配对面部镜面漫反射（PFSD）数据集。收集流程图如图2所示。我们的设置包括一台佳能6D Mark II相机和三个受控控制器色温为1500k的灯。中央灯作为主光源，直接照亮拍摄对象的面部，另外两盏作为环境光，照亮拍摄对象的周围。我们观察到环境光的使用有助于减少色调偏差。为了产生交叉偏振效果，偏振器安装在相机和所有三个光源的前面。一个特殊的旋转装置调节主光源前面的偏光器的角度，控制它与相机偏光器之间的角度  $\theta$ 。其他两个环境光源前面的偏振器始终垂直于相机偏振器，即  $\theta = 90^\circ$ 。这种配置确保只有主光源产生高光，从而保持捕获图像的不受干扰的质量。



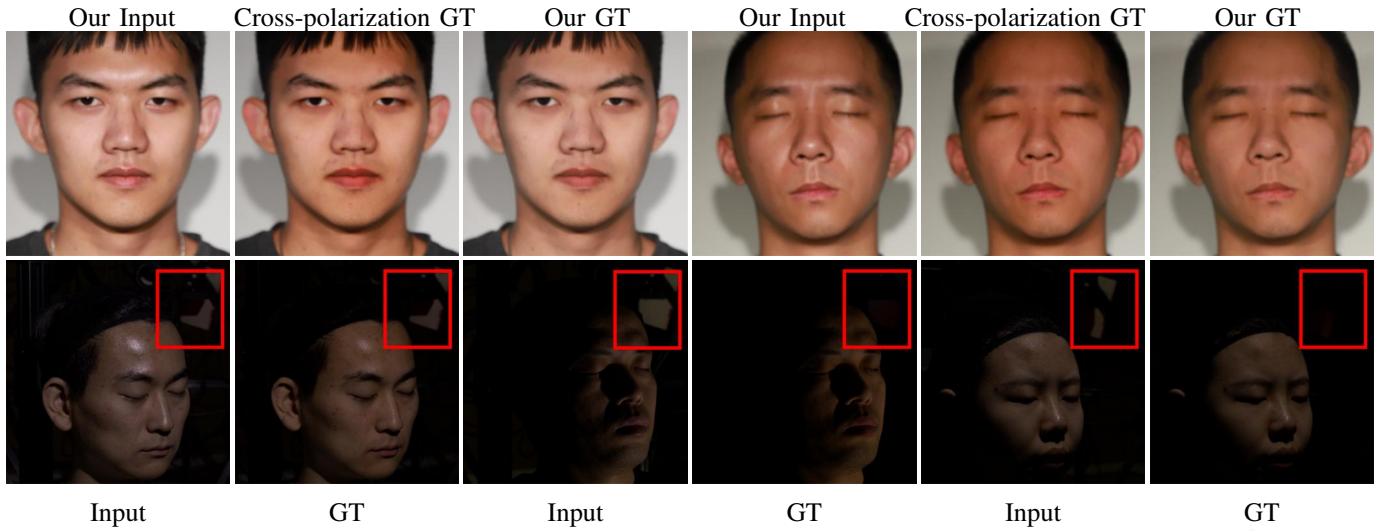


图 3. 在顶行中, 我们分别显示了带有高光的面部图像、通过交叉偏振技术收集的 GT 以及通过色调传输获得的 GT。底行显示来自 FaceSDIM [21] 的数据。

表 各数据集统计比较

Dataset	SHIQ	CLH	LIME	PSD	FaceSDIM	Ours
Counts	<b>16K</b>	1K	10K	13K	3K	4K
Object	✓	✓	✓	✓	×	×
Face	×	×	✓	×	✓	✓
Mean(HI)	0.138	0.109	<b>0.207</b>	0.104	0.113	0.167
Std(HI)	0.182	0.105	<b>0.216</b>	0.121	0.129	0.174

当主光偏振角与相机角度平行时 (即 $\theta=0^\circ$ ) , 得到高亮图像  $I_W$  为:

$$I_W = \frac{1}{2} I_D + I_S, \quad (3)$$

其中 $I_D$ 是所有光线的漫反射分量, 经过多次折射后有多不同方向的分量, 所以一半会被相机的偏光镜滤掉。  $I_S$  为主光源的镜面反射分量, 由于偏振角与相机角度平行, 因此不会被相机偏振片过滤。 · 当主光源偏振角与光源角度垂直时相机 (即  $\theta = 90^\circ$  ) , 得到的结果图像  $I_{WO}$  是没有高光的:

$$I_{WO} = \frac{1}{2} I_D, \quad (4)$$

其中  $I_S$  和  $I_D$  的一半将被相机的偏光镜过滤。根据该方法收集了 4000 多对图像。随后, 进行数据清洗, 进一步提高其质量。尽管图像是在受控照明环境中获取的, 但交叉偏振技术会滤除光源的镜面反射分量, 这可能会导致整体色调出现偏差。为了实现更真实和自然的外观, 我们应用颜色

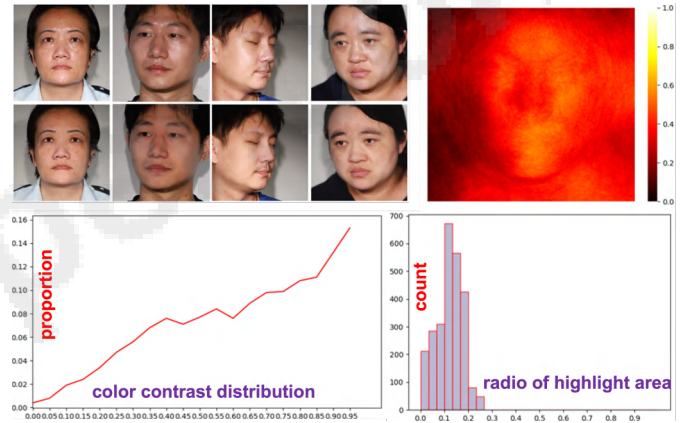


图 4. 左上: 从我们的 PFSD 数据集中选择的配对图像展示了带有突出显示的真实面孔 (第一张图像中的顶行) 及其 GT (底部)。右上: 突出显示区域比例以百分比形式说明了突出显示区域的大小整个图像区域。左下: 颜色对比度分布量化突出显示和非突出显示区域之间的颜色差异。右下: 高光位置分布表示图像中高光的空间分布。

转移[47]以将高亮图像的色调与其地面实况 (GT) 相匹配。该过程的示例如图 3 的顶行所示。此外, 由于捕获过程中不可避免的细微抖动, 我们还利用 SIFT [48] 来对齐漫反射镜面图像。

### C. 分析

在选项卡中。I, 我们将 PFSD 数据集与其他公开可用的数据集在样本数量、目标类别和突出强度 (HI) 方面进行比较。SHIQ [31]、CLH [31] 和 PSD [14] 都专注于自然物体的高光去除, 不涉及面部图像。相比之下, FaceSDIM [21] 专注于面部高光去除, 但高光强度 (HI) 低得多。LIME [49] 是一个合成数据集, 不适合高保真面部高光去除。我们的 PFSD 包含 4,070 对更高强度的

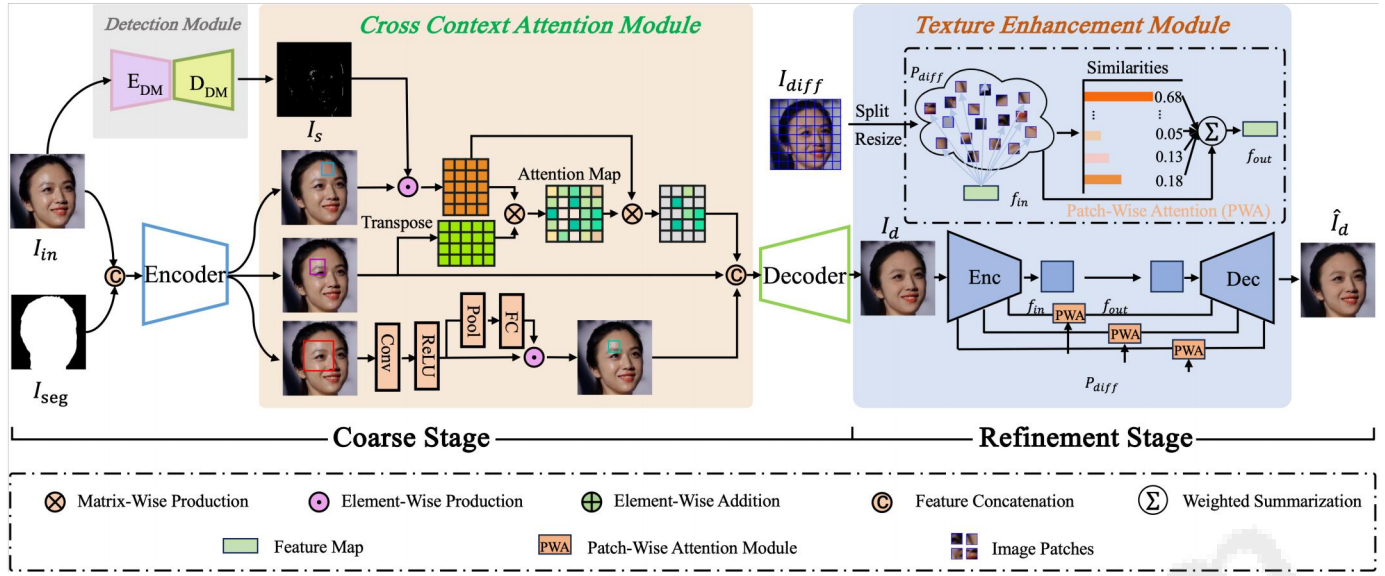


图 5. 所提出的 FHR-Net 概述。该网络设计有两个阶段。Coarse 阶段包含一个 Detection Module 和一个 Cross-Context Attention Module, 在这个阶段, 它获得检测结果  $I_s$  并提供具有颜色和照明恢复的粗漫漫漫图像  $I_d$ 。在细化阶段, 它通过纹理增强模块增强粗略结果的纹理细节。

以及更多样化的高光图像, 更好地解决了强高光去除面部图像的挑战。图 4 中的数据示例说明了 PFSD 的两个主要改进: 1) 色调一致性。我们观察到, 在捕捉过程中添加适量的环境光可以有效减少数据对色调偏差。2) 数据质量。我们的面部数据是从现实场景中捕获的, 提供了复杂多样的样本。照明方向和位置的变化导致高光区域的复杂分布。招募了大约 200 名不同年龄的志愿者来收集不同照明环境下的面部高光图像, 从而产生代表各种皮肤状况的不同数据。

#### B. 检测模块

我们首先应用检测模块来获取高光位置的掩模  $I$ 。该掩模代表了面部图像中高光成分的准确位置, 用于将面部分为高光区域和非高光区域。非高光区域指导高光区域的去除和恢复。检测模块采用 U-Net 架构, 其编码器和解码器通过残差模块进行跳跃连接[51]。残差网络包括门控卷积[52]、扩张卷积和池化层。全局池化操作利用 sigmoid 函数提供通道级关注。该模块最终通过将注意力激活与跳跃连接的特征相乘来输出高亮蒙版。

#### 四. 提议的方法

##### A. FHR-Net 概述

所提出的面部高光去除网络 (FHR-Net) 的框架如图 5 所示, 旨在通过两个阶段去除面部高光, 同时保持面部保真度: 用于建立有效结构的粗略阶段和用于增强纹理细节的细化阶段。在粗略阶段, FHR-Net 将具有镜面高光的面部图像  $I_{in}$  及其相应的分割图像  $I_{seg}$  作为输入,  $I_{seg}$  由 U2-Net [50] 获得。FHR-Net 重建与高光检测模块指示的区域相关的整体结构。它使用跨上下文注意力模块 (CCAM) 恢复整体颜色、照明和纹理信息。在细化阶段, 纹理增强模块 (TEM) 增强受干扰区域的细节, 确保高亮区域的纹理信息与实际皮肤纹理非常相似。

##### C. 跨上下文注意力模块

尽管面部图像中高光的分布通常是稀疏的, 但非高光的周围信息可以潜在地指导受干扰区域的恢复。基于此观察, 面部区域被分为三部分: 非高光区域、高光区域和最近的周围区域。如图 6 所示, 跨上下文注意力模块 (CCAM) 可以充分利用此上下文信息来去除面部高光。CCAM 由三个分支组成, 分别集成非亮点区域的上下文特征、亮点区域的局部上下文和最近周围区域的上下文特征[53]。目的是获得指导高亮区域恢复的上下文信息  $F_{ns}$ 。它首先从编码器的原始输入  $I_{in}$  和分段  $I_{seg}$  中提取非突出显示局部上下文  $F_{loc1}$  和突出显示局部上下文  $F_{loc2}$ 。将  $F_{loc1}$  与高光掩模  $I_s$  逐元素相乘, 得到特征  $F_{nloc1} \in \mathbb{R}^{W \times H \times C}$ 。

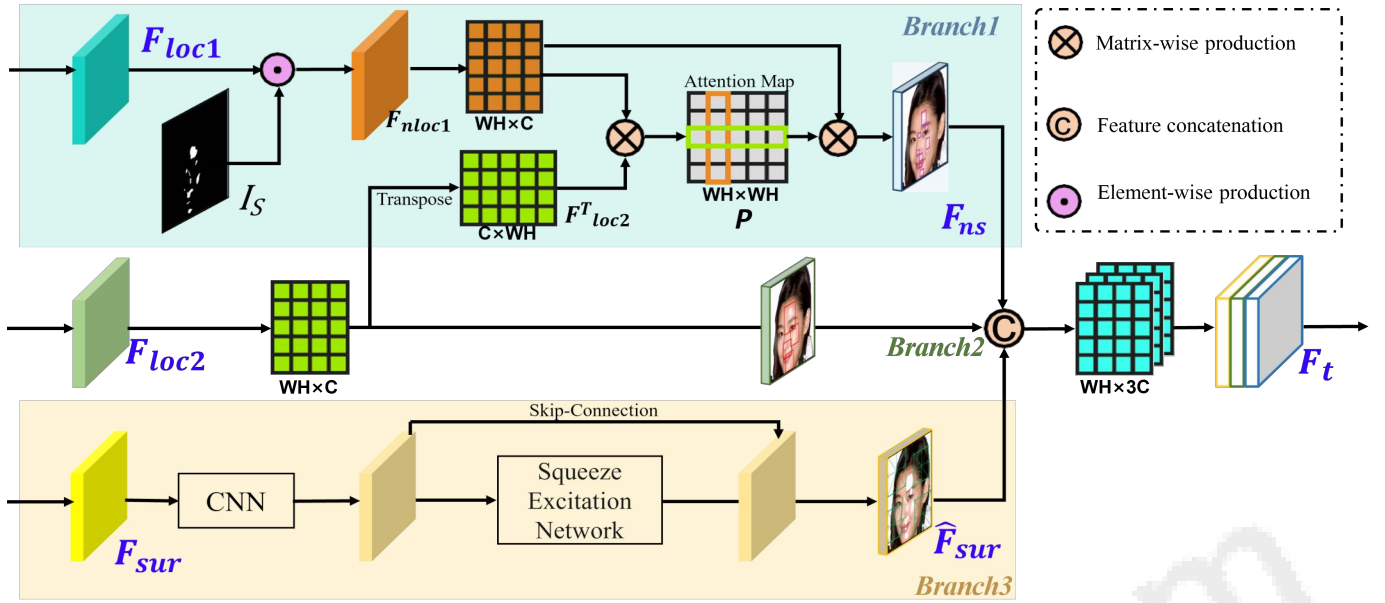


图 6. 跨上下文注意力模块。输入是高亮掩模  $I_S$ 、非高亮上下文特征  $F_{loc1}$ 、高亮上下文特征  $F_{loc2}$  和高亮区域的周围上下文特征  $F_{sur}$ 。输出  $F_t$  是不同分支中跨上下文特征的聚合信息。

包含非高亮区域上下文信息。然后将重构的  $F_{nloc1}$  乘以转置的  $F_{loc2} \in \mathbb{R}^{W \times H \times C}$  得到注意力图  $P \in \mathbb{R}^{W \times H \times W \times H}$ 。其公式为：

$$P = F_{nloc1} \times F_{loc2}^T. \quad (5)$$

$P$  矩阵中的值表示  $F_{nloc1}$  的每一列与  $F_{loc2}$  的每一行之间的余弦相似度[54]。然后我们使用 sigmoid 函数对  $P$  进行归一化，并将其与重构后的  $F_{nloc1}$  相乘。结果  $F_{ns}$  表明了非高亮区域中每个位置对于恢复高亮区域信息的重要性：

$$F_{ns} = \text{Sigmoid}(P) \times F_{nloc1}. \quad (6)$$

分支2。该分支有效地利用高亮区域  $F_{loc2}$  的局部上下文特征来恢复高亮区域中的原始信息。由于面部肤色在小的光亮区域往往是一致的。该分支通过skip连接恢复具有本地上下文特征的信息，以去除小突出区域中的突出显示。分支3。该分支有效地利用了最近的周围环境特征  $F_{sur}$ ，它是通过扩张卷积调制原始图像而获得的。与普通卷积相比，扩张卷积具有更大的感受野，因此  $F_{sur}$  可以捕获高亮像素点周围的上下文信息。基于挤压激励网络[55]，使用全局池化层和全连接层来获取权重向量，进一步将周围上下文特征  $F_{sur}$  细化为  $F_{sur}$ 。对于大的高亮区域，局部上下文特征是不够的，网络能够有效地学习恢复信息。在这种情况下，周围区域的上下文特征应该

用作高光去除的指导。CCAM模块的最终输出  $F_t$  可以表示为：

$$F_t = \text{Concat}(F_{ns}, F_{loc2}, F_{sur}). \quad (7)$$

#### D. 纹理增强模块

尽管粗略结果在突出显示区域中表现出良好恢复的颜色和照明，但它可能缺乏忠实的纹理细节。出现这种差异的原因是，虽然突出显示区域的颜色和照明与全局肖像一致，但纹理信息更多地取决于精确的局部上下文。为了获得更真实的去除结果，纹理增强模块（TEM）使用GT I diff 提供的结构特征进一步细化I d 的纹理。TEM 结合了门控卷积和扩张卷积，采用类似于[56]中的细化网络的设计。同时，我们还提出了一种结构特征注意机制，该机制根据 I diff f 计算结构相关性和聚合特征，以提高面部纹理的一致性和真实感。TEM 的实现如图 5 的细化阶段部分所示。模块首先将 I diff f 划分为  $N$  个  $32 \times 32$  大小的块，然后将它们重新整形为块 p diff f 以匹配编码特征  $f_{in}$  的分辨率。根据距离信息计算不同块 p diff f 和编码特征  $f_{in}$  之间的相似度，其公式为：

$$\hat{D}_{x,y} = \tanh\left(-\frac{D_{x,y} - \text{mean}}{\text{std}}\right), \quad (8)$$

其中  $D_{x,y}$  是来自编码特征  $f_{in}$  的  $x$ -th patch 和来自 p diff f 的第  $y$  个块之间的欧几里得距离 [57]，mean 指平均值，而 std 是标准

所有  $D_{x,y}$  的偏差。然后  $D_{x,y}$  的注意力分数计算如下:  $S_{x,y} =$

$y=1$

由于斑块和面部局部结构之间具有很强的结构相似性, 因此在这项工作中采用了斑块式注意。根据人脸纹理局部强相关性的观察, 相邻人脸图像块的纹理信息是连续一致的。因此, 具有相似结构的图像块应该表现出相似的纹理。从结构特征获得注意力分数后, 增强特征可以通过以下方式计算:

$$\hat{F}_x^i = \sum_{y=1}^N S(x, y) F_y^i, \quad (10)$$

其中  $i \in \{2, 3\}$  代表编码器-解码器结构中编码特征的尺度级别。  $F_y^i$  是  $p_{diff}$  的第  $y$  个块, 其大小与编码器第  $i$  层中的特征相同。然后  $F_y^i$  将通过 skip 连接被发送到上采样层。总之, 所提出的面部高光去除方法 FHR-Net 的总体过程在算法 1 中详细阐述。请注意, CCAM 模块中的 Branch2 (图 6) 是一个跳跃连接, PWA 代表补丁式注意力。

#### Algorithm 1 Coarse-to-fine FHR-Net

输入: 输入图像  $I_{in}$ 、分割图像  $I_{seg}$ 、漫反射图像 (GT)  $I_{diff}$  输出: 无高亮图像  $\hat{I}_d$

COARSE STAGE  $I_s \leftarrow$  检测模块( $I_{in}$ )  $F_{loc1}$ ,  $F_{loc2}$ ,  $F_{sur} \leftarrow$  编码器( $\text{concat}(I_{in}, I_{seg})$ )  $F_{ns} \leftarrow$  CCAM Branch1( $I_s, F_{loc1}$ ,  $\text{Transpose}(F_{loc2})$ )  $F_{sur} \leftarrow$  CCAM Branch3( $F_{sur}$ )  $I_d \leftarrow$  Decoder( $\text{concat}(F_{ns}, F_{loc2}, F_{sur})$ ) 返回粗略去除结果  $I_d$

细化阶段  $p_{diff} \leftarrow$  分割和调整大小( $I_{diff}$ )  $f_{enc}, f_{in} \leftarrow p_{diff}$  编码器( $I_d$ )  $f_{out} \leftarrow$  PWA( $f_{in}, p_{diff}$ )  $\hat{I}_d \leftarrow$  解码器( $f_{enc}, f_{out}$ ) 返回无高亮显示图像  $\hat{I}_d$

#### E. 损失函数

所提出的模型通过最小化损失函数的组合来优化, 包括: 重建损失。这项工作计算了去除结果图像  $I_d$  与其对应的真实图像  $I_{gt}$  之间的像素损失, 并使用  $\ell_2$  损失来减少这种差异:

$$\ell_R(I_{gt}, I_d) = \|I_{gt} - I_d\|_2^2. \quad (11)$$

甘损失。本项采用鉴别器来提高粗级 CCAM 的性能。鉴别器学习对抗性地区分生成的图像是否

突出显示区域内的补丁是真实的还是虚假的, 在推理过程中无需引入额外的计算。此外, 铰链损失[58]用于约束 GAN 网络的训练:

$$\ell_{GAN}(I_{gt}, I_d) = BCE(\sigma(D(I_{gt}) - D(I_d)), g) + BCE(\sigma(D(I_{gt}) - D(I_d)), d), \quad (12)$$

其中  $g=1$  和  $d=0$  用于训练生成器, 而  $g=0$  和  $d=1$  用于判别器。  $\sigma$  是 sigmoid 激活函数, BCE 是二元交叉熵。检测损失。本项使用  $\ell_1$  损失来计算检测模块生成的  $I_s$  与其对应的真实掩模  $I_M$  之间的差异:

$$\ell_D(I_M, I_s) = \|I_M - I_s\|_1, \quad (13)$$

其中  $I_M$  是作为输入图像  $I_{in}$  和相应的 GT  $I_{gt}$  人脸特征损失之间的缩放差值而获得的。该术语引入了面部特征损失, 以确保整个面部结构的一致性。通过比较去除结果  $I_d$  和真实的漫反射地面实况  $I_{gt}$  之间的人脸特征差异。该术语约束网络尽可能多地去除受干扰的区域。损失表示为:

$$\ell_{Face}(I_{gt}, I_{out}) = \|\phi_f(I_{gt}) - \phi_f(I_{out})\|_1, \quad (14)$$

其中  $\phi_f$  表示从预训练模型 Facenet [59] 中提取的人脸特征。最终, 总损失函数表示为:

$$\ell_{All} = \lambda_1 \ell_D(I_s, I_M) + \lambda_2 \ell_R(I_d, I_{gt}) + \lambda_3 \ell_{Face}(I_d, I_{gt}) + \lambda_4 \ell_R(\hat{I}_d, I_{gt}) + \lambda_5 \ell_{Face}(\hat{I}_d, I_{gt}) + \lambda_6 \ell_{GAN}(I_d, I_{gt}), \quad (15)$$

where  $I_d$  is the result of Coarse stage and  $\hat{I}_d$  is the result of the Refinement stage. We set weights for various loss terms based on ablation study:  $\lambda_1 = \lambda_4 = \lambda_5 = 1.0$ ,  $\lambda_2 = \lambda_6 = 0.5$ , and  $\lambda_3 = 0.01$ . More details can be found in section V.C.

## 五、实验

### A. 实施细节

收集的 PFSD 数据集分为包含 3200 对图像的训练集和包含 870 对图像的测试集。FHR-Net 在 PFSD 训练集上使用 4 个 NVIDIA GeForce RTX 2080Ti GPU 进行 100 个 epoch 的训练, 大约需要 7 小时。采用 Adam 优化器,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , 遵循深度学习的一般范式。我们将初始学习率设置为  $10^{-4}$ , 每 5 个 epoch 以 0.4 的衰减系数下降, 直到达到  $10^{-5}$ 。第 20 个 epoch 之后, 执行纹理增强模块, 以确保细化阶段的训练具有足够精细的结构特征。从第 40 个 epoch 开始, 引入人脸特征损失以减轻粗糙结果对人脸特征提取的影响。

### B. 与 SOTA 方法的比较

这项工作对 FHR-Net 与其他最先进的高光去除模型进行了比较分析。具体来说, 对于面部图像, 我们将我们的方法与



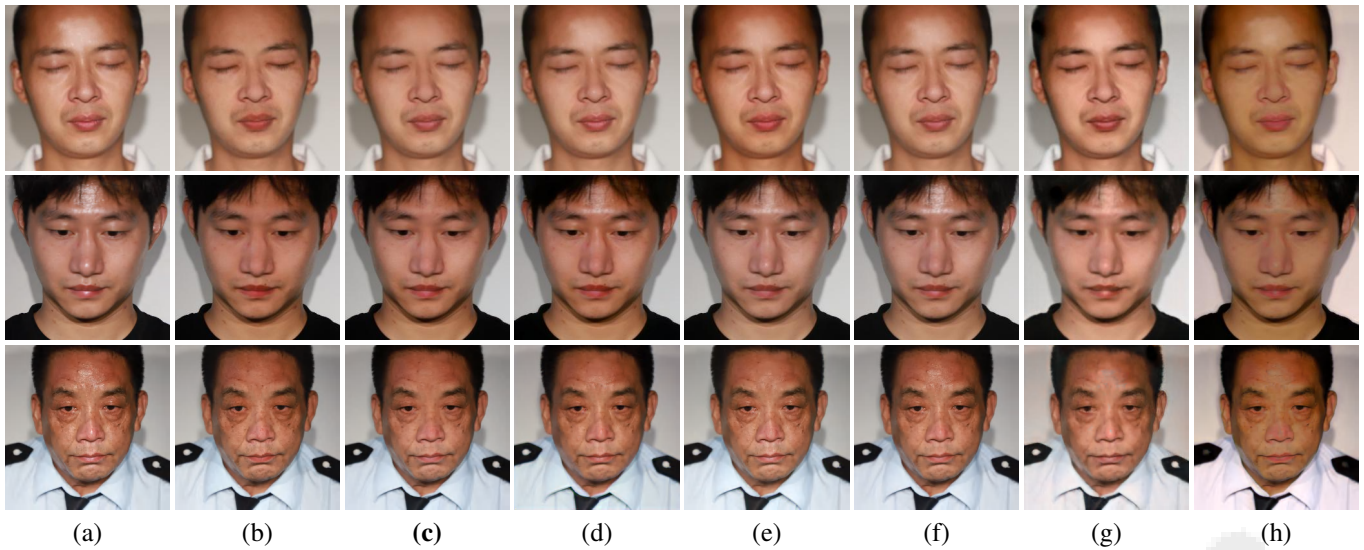


图 7. PFSD 数据集上高光去除的定性比较结果。(a) 输入; (b) GT; (c) 我们的; (d) 傅[32]; (e) 吴[14]; (f) 吴[33]; (g) 李[20]; (h) 梁[21];

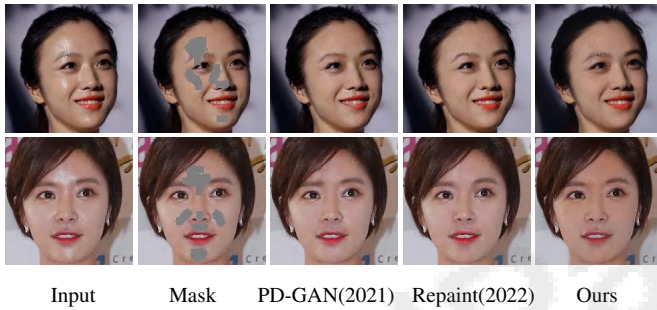


图 8. 与修复方法 PD-GAN [62] 和 Repaint [63] 的视觉比较。遮罩指示需要修复的位置, 通过遮挡输入上的突出显示区域获得。

梁等人。[21] 和李等人。[20], 而对于自然场景, 我们评估了三种方法: Fu 等人。[32] 和吴等人。[33][14]。为了定量评估这些方法的性能, 我们采用以下指标: LPIPS [60]、DreamSim [61]、SSIM 和 MS-SSIM。鉴于修复技术在某种程度上类似于我们的任务, 我们还将 FHR-Net 与最近的修复方法 PD-GAN [62] 和 Repaint [63] 进行比较。结果如图 8 和表 2 所示。二. 在定量指标和视觉表示方面, FHR-Net 的表现与基于 DDPM 的 Repaint 相当, 两种方法都取得了令人满意的结果。此外, FHR-Net 展示了优于 PD-GAN 的整体数值优势, 有效去除镜面反射分量而不留下残留高光。与我们的 PFSD 数据集进行比较。我们的 PFSD 数据集的视觉比较如图 7 所示, 数值比较详见表 1。三. 据观察, Fu 等人[32]去除脸上的大部分高光, 但仍保留一些残留物。此外, 由于许多数据集中缺乏内在的事实真相, 限制了可扩展性, 因此他们的训练具有挑战性。尽管如此, 该方法结合了色调校正, 导致色调与实际颜色紧密一致。

表 II 与修复方法的定量比较结果,

并标记出最佳结果

	Ours	Repaint [63]	PD-GAN [62]
LPIPS↓	<b>0.097</b>	0.101	0.124
DreamSim↓	0.109	<b>0.107</b>	0.141
SSIM↑	<b>0.943</b>	0.940	0.935
MS-SSIM↑	<b>0.951</b>	0.949	0.940

表 III 我们的 PFSD 数据集的定性比较以及最佳数据集

结果已标记

Method	LPIPS↓	DreamSim↓	SSIM↑	MS-SSIM↑
Liang <i>et al.</i> [21]	0.454	0.534	0.859	0.874
Li <i>et al.</i> [20]	0.392	0.416	0.893	0.905
Wu <i>et al.</i> [33]	0.211	0.222	0.907	0.918
Wu <i>et al.</i> [14]	0.187	0.158	0.928	0.933
Fu <i>et al.</i> [32]	0.131	0.145	0.934	0.941
<b>Ours</b>	<b>0.097</b>	<b>0.109</b>	<b>0.945</b>	<b>0.953</b>

基本事实。吴等人。[33] 和吴等人。[14] 在去除高光部分方面表现出有限的能力, 通常会产生明显的视觉伪影。这些方法还与面部的纹理细节和色调恢复作斗争, 产生的结果并不令人满意。李等人。[20] 根据人类皮肤和面部的物理统计特性进行高光去除。然而, 他们的方法在我们的 PFSD 数据集上表现出相对较弱的泛化能力, 导致不干净的高光去除和整体色调偏差。梁等人提出的网络结构。[21] 需要反照率、光线和法线信息, 当仅使用高光和漫反射图像进行训练时, 导致性能不理想。相比之下, 我们的 FHR-Net 产生视觉上令人满意的效果, 并且



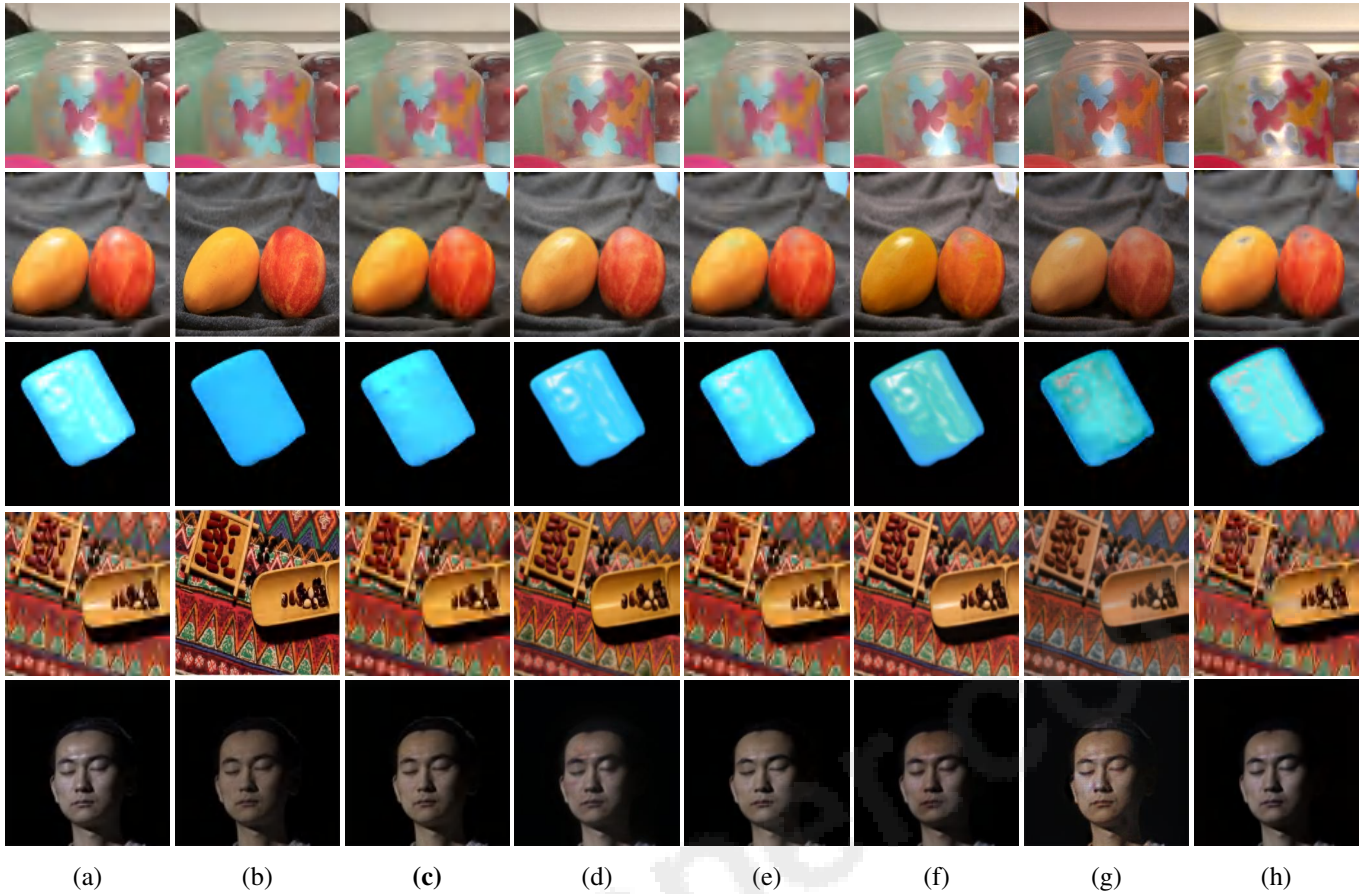


图 9. 更多公共数据集的定性比较结果: SHIQ [31]、CLH [31]、LIME [49]、PSD [14] 和 FaceSDIM [21]。 (a) 输入; (b) GT; (c) 我们的; (d) 傅[32]; (e) 吴[14]; (f) 吴[33]; (g) 李[20]; (h) 梁[21];

表4 在更多公开数据集上的定量比较, 并标注了最好的结果

Dataset	SHIQ				CLH				LIME			
Methods	LPIPS↓	DreamSim↓	SSIM↑	MS-SSIM↑	LPIPS↓	DreamSim↓	SSIM↑	MS-SSIM↑	LPIPS↓	DreamSim↓	SSIM↑	MS-SSIM↑
Liang <i>et al.</i> [21]	0.487	0.505	0.874	0.881	0.409	0.414	0.862	0.872	0.464	0.480	0.873	0.886
Li <i>et al.</i> [20]	0.510	0.493	0.855	0.863	0.434	0.451	0.843	0.855	0.487	0.525	0.856	0.866
Wu <i>et al.</i> [33]	0.291	0.311	0.902	0.907	0.327	0.342	0.904	0.911	0.389	0.403	0.887	0.895
Wu <i>et al.</i> [14]	0.153	0.182	0.937	0.944	0.209	0.236	0.915	0.921	0.317	0.357	0.901	0.911
Fu <i>et al.</i> [32]	0.221	0.253	0.933	0.939	0.183	0.170	0.917	0.924	0.261	0.216	0.909	0.917
<b>Ours</b>	<b>0.113</b>	<b>0.130</b>	<b>0.940</b>	<b>0.952</b>	<b>0.106</b>	<b>0.137</b>	<b>0.932</b>	<b>0.939</b>	<b>0.204</b>	<b>0.165</b>	<b>0.920</b>	<b>0.931</b>

高质量的结果, 在数值指标方面优于其他方法。我们的 FHR-Net 通过集成跨上下文注意力和纹理增强, 在高光去除和纹理细节恢复方面表现出色。此外, 它保持了面部图像的色调一致性, 如图7的前两行所示。与其他公共数据集的比较。为了验证 FHR-Net 的通用性, 我们在几个公共数据集上测试了它的性能。结果如图9和表1所示。四. 我们观察到, 与其他方法相比, FHR-Net 表现出较高的通用性, 并取得了更好的定性和定量结果。定性地讲, FHR-Net 有效地去除了各种物体的强烈高光, 在不引入纹理失真

的情况下恢复局部细节。从数量上讲, 它获得了最低的 LPIPS 和 DreamSim 值, 以及最高的 SSIM 和 MS-SSIM 值, 表明生成的图像与地面实况 (GT) 之间的感知差异最小。[32] 和吴等人。[33]展示了不完整的高光去除结果, 导致残留和明显的纹理失真, 在 FaceSDIM 中尤其明显。而吴等人。[14]表现相对较好, 但在消除强亮点方面仍遇到困难。利特等人。[20]介绍了一种基于皮肤黑色素和血红蛋白的高光约束模型。然而, 这种方法很难应用于自然物体, 导致高光去除效果不佳, 色差和纹理失真更加严重。梁等人。[21]引入许多异常

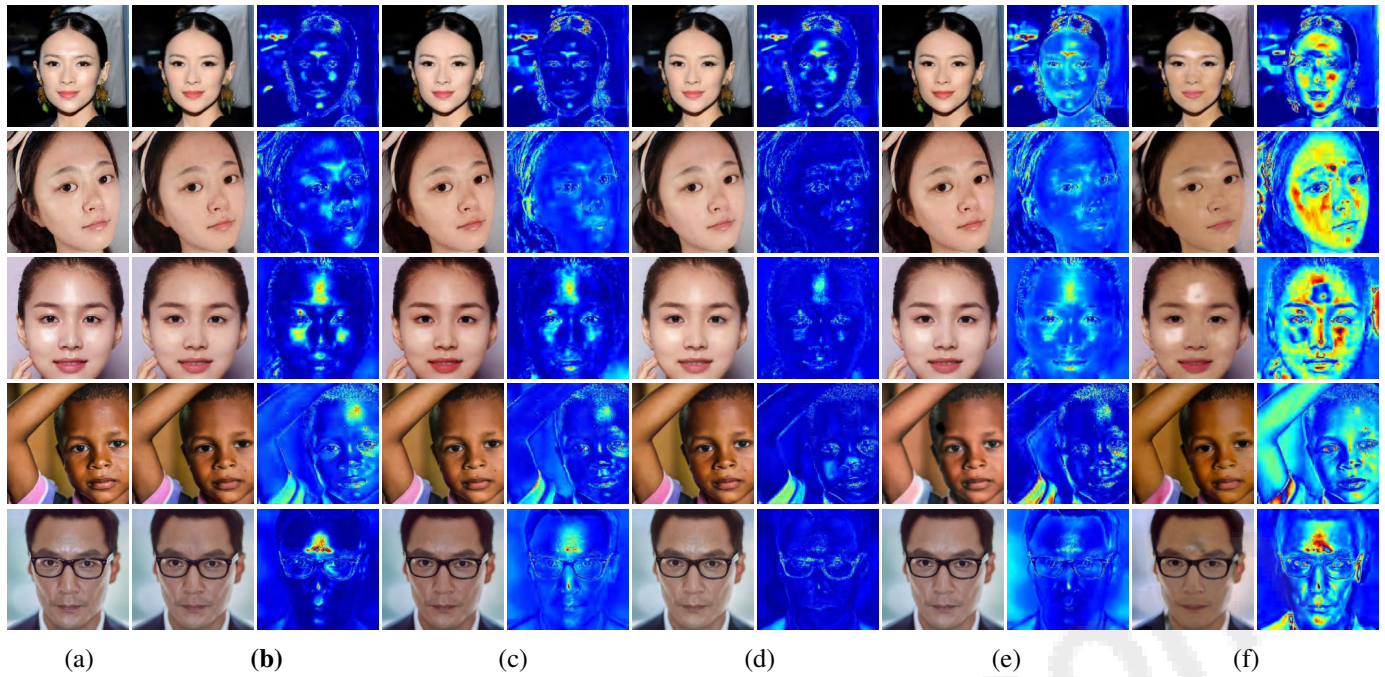


图 10. 野外图像高光去除的更多定性比较。第一列显示输入图像。接下来的每两列是高光去除结果以及输入和去除结果之间差异的可视化。(a) 输入; (b) 我们的; (c) 傅[32]; (d) 吴[14]; (e) 吴[33]; (f) 梁[21]。

黑色色块。相比之下，我们的方法有效地利用了非高光区域的有用信息，以获得更好的纹理和颜色恢复性能。野外图像的比较。我们还比较了这些方法在真实场景图像上的性能。为了更好地理解，我们将输入和高光去除结果之间的差异可视化，可以表示为  $I_{vis} = I_{in} - I_d$ 。结果如图10所示。可以看出，Fu等人。[32]有一些残留，特别是在强高光条件下，处理后的图像整体色调也发生了变化。吴等人。[14]和[33]不能有效去除野外的面部高光。Liang等人。[21]在实际数据中表现不佳，并且不能准确检测高光区域。结果存在较大的色块，与原始图像存在较大的颜色偏差，使得生成的图像缺乏真实感。相比之下，FHR-Net可以有效去除自然场景中的面部高光，并且效果比其他方法要好得多。这些结果也验证了我们的FHR-Net具有良好的泛化性，可以应用于收集的数据集和野外的一些复杂图像。对于输入和高光去除结果之间的差异，图10显示了更多细节。傅等人。[32]可以在一定程度上检测高光，但不能完全去除高光，并且纹理恢复不理想。乌特等人。[14]和[33]在弱高光区域上表现良好，在强高光区域上效果有限，并且出现色调不一致的情况。梁等人。[21]修改整个面部图像而不区分高光区域，导致在野生面部图像上的表现不佳。相比之下，FHR-Net 表现出强大的泛化能力，可以恢复高光区域内的结构和纹理，同时

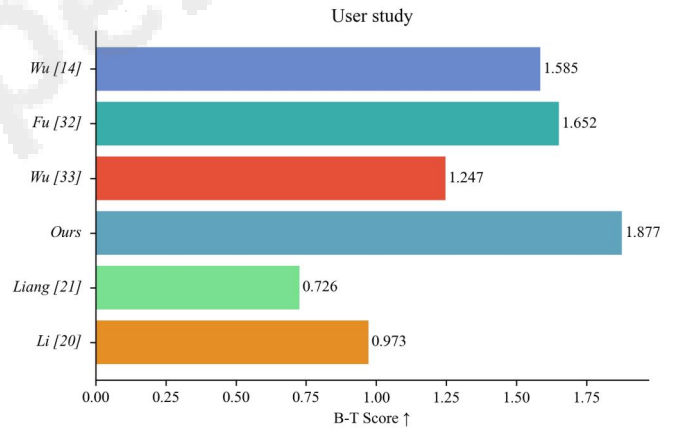


图 11. 使用 PFSD 数据集上的 B-T 分数进行的用户研究。

保护其他地区。我们的方法在具有强反射高光的自然面部图像上取得了更好的结果，生成了高保真度的恢复面部图像。对我们的 PFSD 数据集的用户研究。为了进一步比较我们的FHR-Net与其他五种方法（Fu等人[32]、Wu等人[14]、Wu等人[33]、Li等人[20]和Liang等人[21]），我们使用 PFSD 数据集进行了主观评估。具体来说，我们从测试集中选择了 20 张图像用于此实验。继[64]、[65]之后，为每个输入生成了六个相应的高光去除结果。然后从这六个图像中选择两个图像来创建图像对，每个输入产生  $62 = 15$  个图像对。对于这 20 张测试图像，我们获得了 300 个图像对。然后，我们邀请 50 名参与者主观评估这些图像，而不告知他们任务涉及



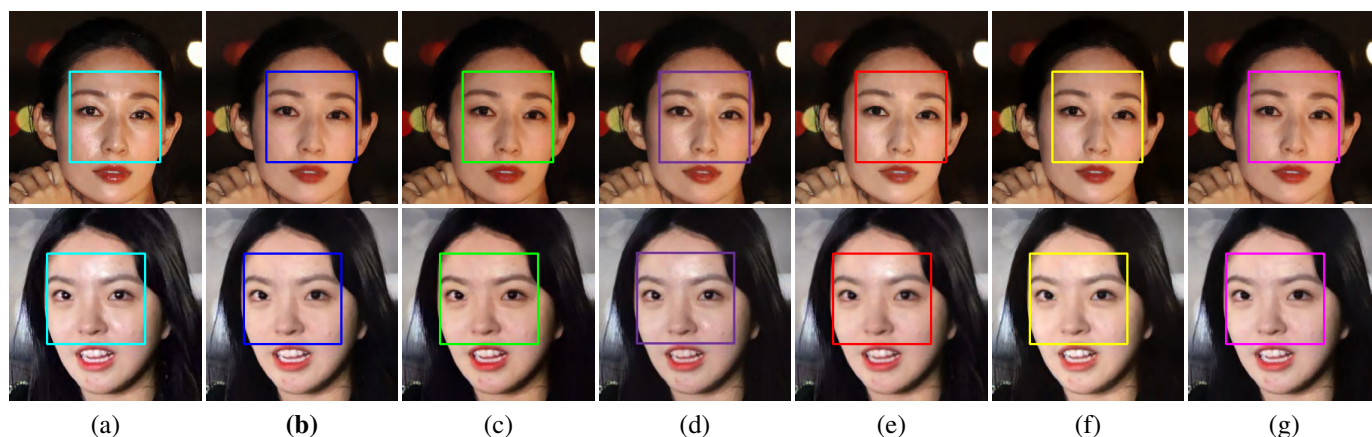


图 12. 消融研究结果的可视化。(a) 输入图像; (b) 我们的 FHR-Net; (c) 不带色调转移; (d) 不带检测模块; (e) 不带 CCAM; (f) 不带 TEM; (g) 没有面部特征损失。

突出显示删除。他们被要求在每对图像中选择他们认为最真实、最自然的图像。这些参与者大多数是大学生，男女比例为 2:3。在做出回应之前，他们有大约 5 秒的时间查看每一对，并可以根据自己的喜好自由调整观看距离。实验使用 iPhone 13 Pro 作为显示设备进行。我们最终从参与者那里收集了总共 15,000 份主观比较。最后，我们应用 Bradley-Terry 模型 (B-T 模型) [66] 来计算每种方法的全局排名分数。如图 11 所示，所提出的 FHR-Net 获得了最高的 B-T 分数，产生了让参与者更满意的高光去除结果。

### C. 消融研究

整体网络消融。进一步验证 FHR-Net 中构建模块的有效性，包括数据集预处理中的色调传输、检测模块、跨上下文注意模块 (CCAM)、纹理增强模块 (TEM) 和人脸特征损失，我们设计了多种变体。消融结果如图 12 和表 1 所示。五、

1) 色调转移: 实验表明，在没有色调转移的情况下，面部图像表现出明显的颜色不和谐。相反，使用未进行色调转移的数据集训练网络可以显著增强颜色保真度。这一改进使背景和面部区域的色彩更加一致，使删除的图像更接近现实世界中的自然颜色。请注意，虽然色调传输会改变图像的总色调，但它不会修改各个区域内的特定高光成分。2) 检测模块: 该模块识别高光的强度和位置，这是高光去除的先决步骤。通过比较图 12 中的图像 (b) 和 (d)，我们观察到在消融该模块后，模型很难去除面部高光，并保留了大量的高光残留。此外，数值量化指标也会恶化。这种下降主要是由于缺乏有关亮点区域及其背景的精确信息，

表 1 定量消融结果，最好的结果是

	MARKED			
	LPIPS↓	DreamSim↓	SSIM↑	MS-SSIM↑
<b>FHR-Net</b>	<b>0.097</b>	<b>0.109</b>	<b>0.945</b>	<b>0.953</b>
w/o Hue Transfer	0.107	0.121	0.928	0.934
w/o Detection Module	0.118	0.130	0.922	0.927
w/o CCAM	0.115	0.126	0.929	0.931
w/o TEM	0.104	0.119	0.932	0.937
w/o Face Feature Loss	0.099	0.115	0.943	0.950

这使得模型执行干净和准确的消除具有挑战性。3) 跨上下文注意模块: 我们的观察表明，CCAM 组件的缺失会导致视觉上令人不满意的结果，其特点是色调不一致和残留高光丰富。定性和定量结果都强调了 CCAM 在增强我们的模型中的关键作用。CCAM 明确聚合非高光区域和高光最近周围区域的上下文特征，指导高光去除。这种跨区域特征处理能力使 FHR-Net 能够有效去除面部高光。4) 纹理增强模块: 无需 TEM 组件然而，高光去除结果显得不太真实并且包含视觉伪影。TEM 利用结构特征注意力基于漫反射图像  $I_{diff}$  来处理局部上下文，进一步增强纹理信息并保留人脸的真实细节。5) 人脸特征损失: 这种损失减少了高光去除结果与高光去除结果之间的结构差异。GT 同时保持面部内容的一致性。这与我们专注于去除面部高光而不过度改变面部区域以外的背景的主要目标相一致。总之，色调转移是一项至关重要的预处理操作。使用有缺陷的数据集训练的网络在测试样本中表现出整体色调变化，导致处理前后的图像之间存在显著差异。高光去除的目的是确保非



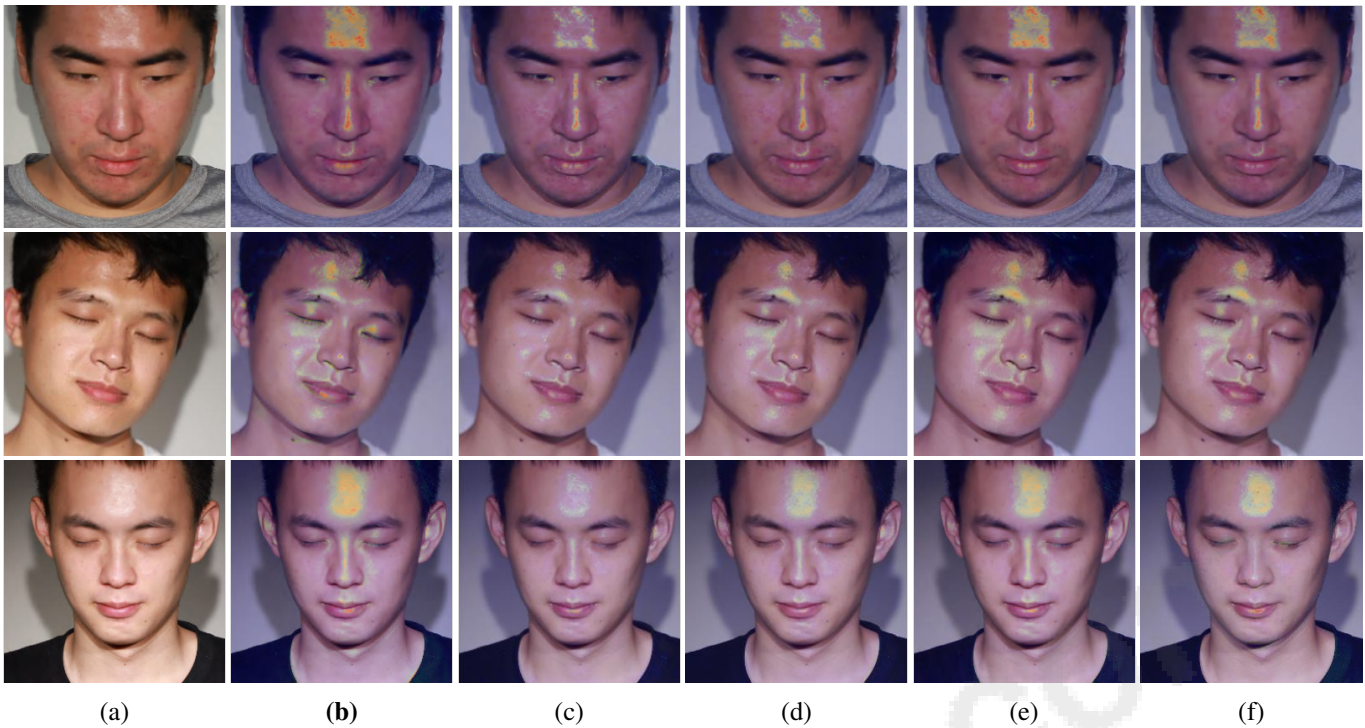


图 13. 我们的 CCAM 和不同分支的消融研究。我们通过使用不同的 CCAM 变体来可视化参与的亮点区域。(a) 输入图像；(b) 我们的 FHR-Net；(c) 不带 CCAM；(d) 不设分支机构 1；(e) 不设分支机构 2；(f) 不带分支 3。

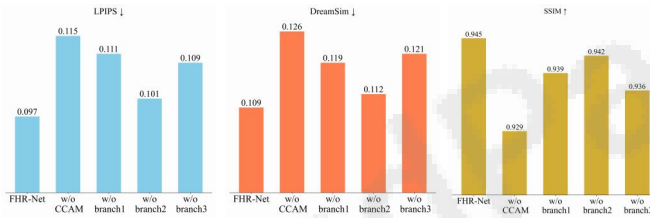


图 14. CCAM 不同分支的消融研究结果。

保留高亮区域，并准确恢复高亮区域下的信息。CCAM协同不同区域的多样化特征，有效去除面部图像中的高光成分，显着提升模型性能。根据粗略结果，TEM 以高保真度增强面部高光区域的纹理。CCAM 中的不同分支。在此消融中，每个分支都被删除以评估其有效性和对 CCAM 模块的影响。如图13和图14所示，Branch3 提供了最近的周围上下文特征，对于颜色和纹理恢复至关重要。在 Branch3 中，最近的周围上下文特征  $F_{sur}$  是通过扩张卷积获得的。与通过普通卷积获得的局部上下文特征相比，由于扩张卷积具有更大的感受野， $F_{sur}$  可以捕获每个像素周围更多的上下文信息。Branch1 生成注意力图，识别图像中可用于恢复突出显示区域下的原始信息的区域，表现出最高的有效性。Branch2 仅起到跳跃连接的作用，以保留原始特征，并且有效性最低。

表 PFSD 损失参数消融的直观结果

数据集，并标记了最佳结果

Setting of the Loss Hyper-Parameters		LPIPS↓	SSIM↑
$\lambda_1 = 0.5$	$\lambda_2 = 0.5, \lambda_3 = 0.5, \lambda_4 = 0.1, \lambda_5 = 0.1, \lambda_6 = 0$	0.113	0.923
$\lambda_1 = 1.0$		0.109	0.929
$\lambda_1 = 2.0$		0.110	0.927
$\lambda_2 = 1.0$	$\lambda_1 = 1.0, \lambda_3 = 0.5, \lambda_4 = 0.1, \lambda_5 = 0.1, \lambda_6 = 0$	0.105	0.933
$\lambda_2 = 2.0$		0.108	0.931
$\lambda_3 = 1.0$	$\lambda_1 = 1.0, \lambda_2 = 1.0, \lambda_4 = 0.1, \lambda_5 = 0.1, \lambda_6 = 0$	0.102	0.938
$\lambda_3 = 2.0$		0.104	0.937
$\lambda_4 = 0.5$	$\lambda_1 = 1.0, \lambda_2 = 1.0, \lambda_3 = 1.0, \lambda_5 = 0.1, \lambda_6 = 0$	0.100	0.941
$\lambda_4 = 1.0$		0.105	0.939
$\lambda_5 = 0.5$	$\lambda_1 = 1.0, \lambda_2 = 1.0, \lambda_3 = 1.0, \lambda_4 = 0.5, \lambda_6 = 0$	0.098	0.943
$\lambda_5 = 1.0$		0.098	0.941
$\lambda_6 = 0.01$	$\lambda_1 = 1.0, \lambda_2 = 1.0, \lambda_3 = 1.0, \lambda_4 = 0.5, \lambda_5 = 0.5$	<b>0.097</b>	<b>0.945</b>
$\lambda_6 = 0.05$		0.099	0.944

损耗参数的设置。拟议的 FHR-Net 包含多个损失项。我们通过消融实验确定最佳超参数。鉴于损失函数超参数有大量组合，将它们全部呈现出来是不切实际的。因此，我们根据每个网络组件的重要性根据经验初始化其损失超参数。然后，我们单独改变这些值以进行敏感性测试。标签。VI展示了PFSD数据集上不同损失超参数的LPIPS和SSIM分数。这里， $\lambda_1$  用于DetectionModule， $\lambda_2$  和  $\lambda_3$  代表粗略阶段的CCAM， $\lambda_4$  和  $\lambda_5$  对应细化阶段的TEM， $\lambda_6$  作为CCAM辅助训练的GAN损失。

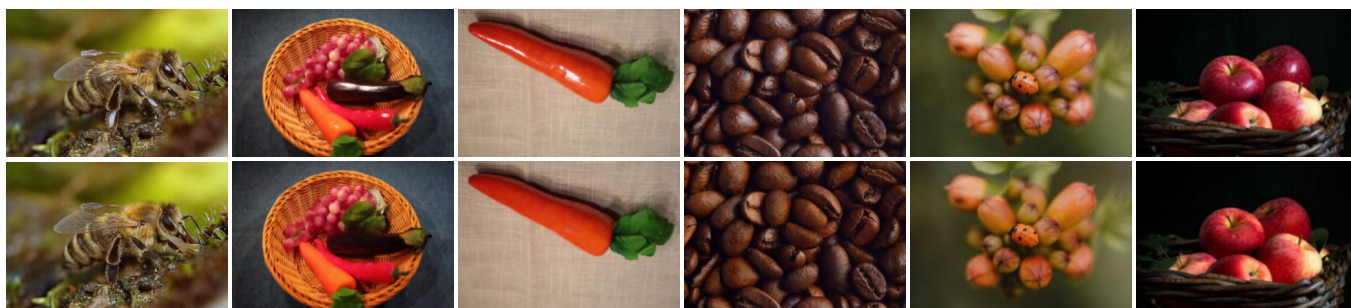


图 15. 单个自然物体上高光去除的结果。第一行显示具有镜面高光的对象，第二行显示来自 FHR-Net 的高光去除结果。

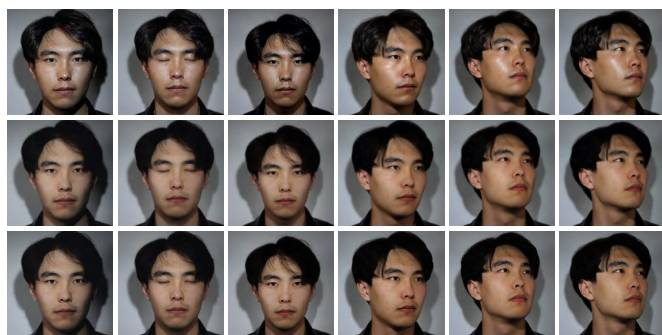


图 16. 不同照明条件下高光去除的结果。第一行显示带有面部高光的输入，第二行显示我们的 FHR-Net 的去除结果，第三行显示相应的地面实况 (GT)。

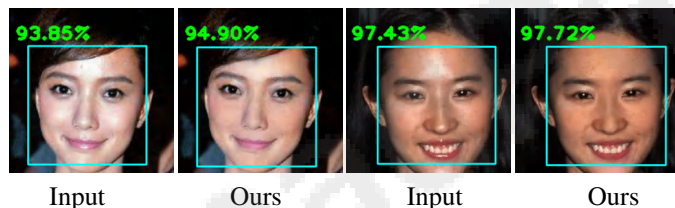


图17.对人脸识别准确率的影响。每对图像由带有高光的原始图像（左）和高光去除后的结果（右）组成。

检测损失获得人脸图像中高光的准确位置，是高光去除的前提，当 $\lambda_1 = 1$ 时达到最佳性能。此外，我们的实验表明，CCAM可以有效地恢复高光区域下的颜色和照明信息，以产生相对令人满意的初始结果。然而，纹理细节略有缺陷，随后在 TEM 中进行了微调。因此，我们为CCAM分配较高的权重，为TEM分配较低的权重。GAN损失仅作为增强CCAM性能的辅助，因此具有较低的权重。

D.扩展实验各种光照条件下的高光去除。面部高光的强度、色调和形状不是恒定的，而是随着光照条件而变化。该实验演示了所提出的 FHR-Net 如何处理

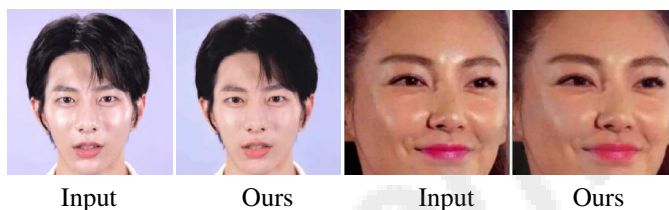


图 18. 我们网络的局限性包括高光去除后面部图像中存在高光残留，特别是当高光过强并分布在整个面部时。

照明变化。通过保持相机位置并逐渐调整主光源的色温从冷色调调到暖色调，我们获得了一系列输入及其相应的地面实况 (GT)。高光去除结果如图 2 的第二行所示。16. 可以看出，虽然面部高光随着光照条件的变化而变化，但所提出的方法仍然可以有效地去除高光，并得到满意的结果。自然物体上的高光去除。FHR-Net 在 PFSD 数据集上进行训练，还在其他几个专注于自然物体突出显示的公开数据集上进行了评估，包括 SHIQ、LIME 和 CLH。如图所示如图 15 所示，结果表明 FHR-Net 有效地去除了对象的高光，产生没有残留伪影的结果。值得注意的是，即使没有对公共对象高光数据集进行微调，FHR-Net 也能熟练地去除对象的高光，展示其在高光去除任务中强大的泛化能力。在下游任务上进行更多实验。高光去除是其他面部视觉算法的重要预处理。为了证明高光去除的增强效果，这项工作在人脸识别和3D人脸重建等下游任务上进行了实验。我们分别评估了在AdaFace上进行的识别面部高光图像和相应的非高光图像的性能[67]。结果如图17所示。高光可能会干扰面部识别的性能，特别是当它存在于面部识别的关键点附近时。我们的实验证明，使用所提出的 FHR-Net 预处理面部图像提高了识别的平均准确率。局限性。如图18所示，当镜面反射



高光成分过于强烈并且遍布整个脸部, 我们的 FHR-Net 无法完全去除它们而不留下高光残留物。出现这种困难的原因是高强度镜面高光导致几何和纹理信息的丢失, 而这些信息指导着数据不足的受干扰区域的恢复。因此, 纹理细节无法完全恢复以匹配真实图像。

六. 结论在这项工作中, 我们首先构建了一个高质量的配对面部镜面漫反射 (PFSD) 数据集, 用于面部高光去除。然后, 我们引入了一个名为 FHR-Net 的两阶段网络, 以有效去除面部图像中的高光成分。凭借其从粗到精的架构, FHR-Net 以高真实度和保真度恢复高光区域。在 PFSD 数据集和其他数据集上的实验结果表明, FHR-Net 优于现有的面部高光去除方法。未来, 我们将把我们的方法扩展到实时视频面部高光去除, 并探索针对现实场景的半监督或无监督方法。此外, 进一步细化该领域的数据集是值得努力的。

## REFERENCES

[1] M. Wang 和 W. Deng, “深度人脸识别: 一项调查”, 神经计算, 卷. 429, pp. 215–244, 2021. 1[2] Y. Feng, F. Wu, X. Shao, Y. Wang, and X. Zhou, “联合 3d 人脸重建和位置图回归网络的密集对齐,” 在 Proc. 欧元。会议。计算。维斯。(ECCV), 2018 年, 第 534–551 页。1[3] S. Sambolek 和 M. Ivasic-Kos, “使用深度 cnn 检测器在搜索和救援操作中自动进行人员检测”, IEEE Access, 卷. 第 9 页. 37 905–37 922, 2021. 1[4] P. Tan, L. Quan 和 S. Lin, “纹理表面上高光反射的分离”, Proc. 37 905–37 922, 2021. IEEE 会议计算。维斯。模式识别 (CVPR), 2006 年, 第 1855–1860 页。1[5] H. Kim, H. Jin, S. Hadap 和 I. Kweon, “使用暗通道先验的镜面反射分离”, Proc. IEEE 会议计算。维斯。模式识别。(CVPR), 2013 年, 第 1460–1467 页。1, 2[6] G. J. Klinker, S. A. Shafer 和 T. Kanade, “彩色图像中亮点的测量”, 国际计算机视觉杂志, 卷. 2, 没有. 1, 第 7–32 页, 1988 年。1[7] R. Bajcsy, S. W. Lee 和 A. Leonardis, “通过彩色图像分割检测漫反射和镜面界面反射和相互反射”, 国际计算机视觉杂志, 卷. 17, 没有. 第 3 页. 241–272, 1996. 1[8] R. T. Tan 和 K. Ikeuchi, “使用线性基函数对纹理表面进行反射分量分解”, Proc. 241–272, 1996. IEEE 计算会议。维斯。模式识别。(CVPR), 2005 年, 第 125–131 页。1, 2[9] K. Ikeuchi, D. Miyazaki, R. T. Tan 和 K. Ikeuchi, “使用单个图像分离纹理表面的反射分量”, 《数字归档文物》, 2008 年, 第 353–384 页。1, 2[10] Q. Yang, S. Wang 和 N. Ahuja, “使用双边过滤进行实时镜面高光去除”, Proc. 欧元。会议。计算。维斯。(ECCV)。施普林格, 2010 年, 第 87–100 页。1[11] J. Yang, L. Liu 和 S. Li, “在 hsi 颜色空间中分离镜面反射和漫反射分量”, 载于 IEEE 国际计算机视觉研讨会会议记录, 2013 年, 第 891 页 898. 1[12] B. Song, J. Zhou, X. Chen 和 S. Zhang, “使用原始 RGB 图像对进行真实场景反射去除”, IEEE Trans. 电路系统。视频技术, 卷. 33, 没有. 8, 第 3759–3773 页, 2023 年。1[13] G. Fu, Q. Zhang, C. Song, Q. Lin 和 C. Xiao, “真实世界图像的镜面高光去除”, 计算机图形学论坛, 卷. 38, 没有. 7. Wiley 在线图书馆, 2019 年, 第 253–263 页。1[14] Z. Wu, C. Zhuang, J. Shi, J. Guo, J. Xiao, X. Zhang 和 D.-M. Yan, “通过真实数据集构建实现单图像镜面高光去除”, IEEE 多媒体汇刊, 卷. 24, 第 3782–3793 页, 2022 年。1、2、3、4、8、9、10

[15] Y. Takimoto, R. Harakawa 和 M. Iwahashi, “利用镜面反射测量餐具三维表面的基于色调的灰色编码方法”, IEEE Trans. 电路系统。视频技术, 卷. PP, 不. 99, pp. 1–1, 2020. 1[16] R. Wan, B. Shi, H. Li, Y. Hong, L.-Y. Duan 和 A. C. Kot, “单图像反射消除算法基准测试”, IEEE Trans. 模式识别。马赫。智力, 卷. 45, 没有. 2, 第 1424–1441 页, 2023 年。1[17] K. Wei, J. Yang, Y. Fu, D. Wipf 和 H. Huang, “利用未对齐的训练数据和网络增强进行单图像反射去除”, Proc. IEEE 会议计算。维斯。模式识别。(CVPR), 2019 年, 第 8178–8187 页。1[18] G. Fu, Q. Zhang, L. Zhu, Q. Lin, Y. Wang, S. Fan, and C. Xiao, “走向高分辨率镜面高光检测”, 国际计算机视觉杂志, 卷. 132, 没有. 1, 第 95–117 页, 2024 年。1[19] K. Hu, Z. Huang 和 X. Wang, “基于改进的二色反射模型的高光去除网络”, 发表于 ICASSP 2024 - 2024 IEEE 国际声学会议、演讲和信号处理 (ICASSP), 2024 年, 第 2645–2649 页。1[20] C. Li, S. Lin, K. Zhou 和 K. Ikeuchi, “面部图像中的镜面高光去除”, Proc. IEEE 会议计算。维斯。模式识别 (CVPR), 2017 年, 第 3107–3116 页。1, 2, 8, 9, 10[21] 梁斌, 翁鼎, 屠志, 罗丽, 郝建, “基于偏振特性的人脸镜面去除及本征分解研究”, 《光学快报》, 卷. 29, 没有. 20, 第 32 256–32 270 页, 2021. 1, 3, 4, 8, 9, 10[22] S. Muhammad, M. N. Dailey, M. Farooq, M. F. Majeed 和 M. Ekpa-nyapong, “Spec-net 和 spec-cgan: 用于消除镜面反射的深度神经网络面孔”, 图像与视觉计算, 卷. 93, p. 103823, 2020. 1, 3[23] A. Artusi, F. Banterle 和 D. Chetverikov, “镜面反射消除方法的调查”, Comput. 图形。论坛, 卷. 30, 没有. 8, 第 2208–2230 页, 2011 年。2[24] Q. Yang, J. Tang 和 N. Ahuja, “高效且稳健的镜面高光去除”, IEEE Trans. 模式识别。马赫。智力, 卷. 37, 没有. 6, 第 1304–1311 页, 2014 年。2[25] Y. Akashi 和 T. Okatani, “通过稀疏非负矩阵分解分离反射分量”, 亚洲计算机视觉会议。施普林格, 2014 年, 第 611–625 页。2[26] 冯文, 程旭, 孙俊, 熊正, 翟正, “基于光场偏振特性的镜面高光去除及深度估计”, 光学通讯, 2016, 537, p. 129467, 2023. 2[27] W. Feng, X. Cheng, X. Li, Q. Liu, and Z. Zhai, “基于二色反射模型和基于优先级的光场相机自适应方向的镜面高光去除,” 工程中的光学和激光, 卷. 172, p. 107856, 2024. 2[28] J. Lin, M. E. Amine Seddik, M. Tamaazousti, Y. Tamaazousti 和 A. Bartoli, “深度多类对抗性镜面反射消除”, 斯堪的纳维亚图像分析会议, 2019 年, 第 3–15 页。2[29] B. Song, J. Zhou 和 H. Wu, “用于渐进式单图像反射去除的多级曲率引导网络”, IEEE Trans. 电路系统视频技术, 卷. 32, 没有. 10, 第 6515–6529 页, 2022 年。2[30] Y. Li, Q. Yan, K. Zhang 和 H. Xu, “通过上下文特征融合金字塔和任务驱动正则化去除图像反射”, IEEE Trans. 电路系统。视频技术, 卷. PP, 不. 99, pp. 1–1, 2021. 2[31] G. Fu, Q. Zhu, L. Zhu, P. Li, and C. Xiao, “一种用于联合镜面高光检测和去除的多任务网络”, 载于过程。IEEE 计算会议。维斯。模式识别。(CVPR), 2021 年, 第 7752–7761 页。2, 4, 9[32] G. Fu, Q. Zhang, L. Zhu, C. Xiao, and P. Li, “利用大规模合成数据实现高质量镜面高光去除”, in Proc. IEEE/CVF 国际。会议。计算。维斯。(ICCV), 2023, pp. 12 857–12 865. 2, 8, 9, 10[33] Z. 吴, J. 郭, C. 庄, J. 肖, D.-M. Yan 和 X. Zhang, “通过 unet-transformer 在单个图像中进行联合镜面高光检测和去除”, 计算视觉媒体, 卷. 9, 不. 1, 第 141–154 页, 2023 年。2, 8, 9, 10 [34] H. 钟, Y. 洪, S. 翁, J. 梁, 和 B. 石, “语言引导图像反射分离,” 在 Proc. IEEE 会议计算。维斯。模式识别。(CVPR), 2024. 2[35] N. Capece, F. Banterle, P. Cignoni, F. Ganovelli, R. Scopigno 和 U. Erra, “Deepflash: 将闪光灯自拍照变成工作室肖像”, 信号处理: 图像通信, 卷. 77, pp. 28–39, 2019. 2[36] R. Ju, D. Weng, and B. Liang, “学习真实高分辨率肖像图像的高光分离”, 2021 年第四届控制与计算机视觉国际会议, 2021, 第 18–23 页。2[37] X. Ren, J. Deng, Y. Cheng, J. Guo, C. Ma, Y. Yan, W. Zhu, and X. Yang, “单目身份条件面部反射重建”, in Proc. IEEE 会议计算。维斯。模式识别。(CVPR), 2024 年 6 月。2

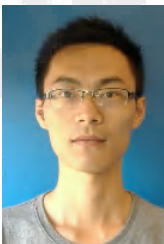


[38] Y. Han, J. Lyu 和 F. Xu, “在家中高质量面部几何和外观捕捉”, Proc. IEEE 会议计算. 维斯. 模式识别. (CVPR), 2024. 2[39] C. Li, K. Zhou 和 S. Lin, “通过基于物理的内在图像层操作来模拟化妆”, Proc. IEEE 会议计算. 模式识别. (CVPR), 2015 年, 第 4621–4629 页. 2[40] 苏涛, 周宇, 于宇, 杜思, “多视点人脸图像高光去除”, 《传感器》, 2017 年 22、没有. 17, p. 6656, 2022. 3[41] Z. Wang, M. Lu, F. Xu 和 X. Cao, “通过生成对抗网络进行野外面部高光去除”, CAAI 国际人工智能会议. 施普林格, 2021 年, 第 311–322 页. 3[42] 朱涛, 夏世, 下泽, 陆春, “人脸图像中的高光去除”, 中国模式识别与计算机视觉学术会议 (PRCV). 施普林格, 2020 年, 第 422–433 页. 3[43] W. Xu, C. Long 和 Y. Nie, “学习用于艺术风格迁移的动态风格内核”, 载于 IEEE/CVF 计算机视觉与模式识别会议论文集, 2023 年, 第 10 083–10 092 页. 3[44] R. Yi, P. Tan 和 S. Lin, “利用多视图图像集进行无监督内在图像分解和高光分离”, 载于 AAAI 人工智能会议论文集, 第 1 卷. 34、没有. 2020 年 07 月, 第 12 685–12 692 页. 3[45] J. Suo, D. An, X. Ji, H. Wang 和 Q. Dai, “从单个图像中快速且高质量地去除高光”, IEEE 图像处理交易, 卷. 25、没有. 11, 第 5441–5454 页, 2016 年. 3[46] S. A. Shafer, “使用颜色分离反射分量”, Color Research & Application, 卷. 10、不. 4, 第 210–218 页, 1985 年. 3[47] E. Reinhard, M. Adhikhmin, B. Gooch 和 P. Shirley, “图像之间的颜色传输”, IEEE 计算机图形与应用, 卷. 21、没有. 5, 第 34–41 页, 2001 年. 4[48] P. Kumar, S. Henikoff 和 P. C. Ng, “使用 sift 算法预测编码非同义突变对蛋白质功能的影响”, 《自然协议》, 卷. 4、没有. 7, 第 1073–1081 页, 2009 年. 4[49] A. Meka, M. Maximov, M. Zollhoefer, A. Chatterjee, H.-P. 赛德尔, C. Richardt 和 C. Theobalt, “Lime: 实时内在材料估计”, Proc. IEEE 会议计算. 维斯. 模式识别. (CVPR), 2018 年, 第 6315–6324 页. 4, 9[50] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane 和 M. Jager-sand, “U2-net: 使用嵌套 u 结构更深度地进行显著目标检测”, Pattern 承认, 卷. 106, p. 107404, 2020. 5[51] Z. Wu, C. Shen 和 A. Van Den Hengel, “更广泛或更深入: 重新审视用于视觉识别的 resnet 模型”, 模式识别, 卷. 90 页. 119–133, 2019. 5[52] Y. N. Dauphin, A. Fan, M. Auli 和 D. Grangier, “使用门控卷积网络进行语言建模”, 国际机器学习会议. PMLR, 2017 年, 第 933–941 页. 5[53] H. Nishimura 和 K. Sakai, “基于对比周围环境的边界所有权的确定”, 神经计算, 卷. 58, 第 843–848 页, 2004 年. 5[54] H. V. Nguyen 和 L. Bai, “用于面部验证的余弦相似度度量学习”, 亚洲计算机视觉会议. 施普林格, 2010 年, 第 17 页. 709–720. 6[55] J. Hu, L. Shen 和 G. Sun, “挤压和激励网络”, Proc. IEEE Conf. 计算. 维斯. 模式识别. (CVPR), 2018 年, 第 7132–7141 页. 6[56] Z. Yi, Q. Tang, S. Azizi, D. Jang 和 Z. Xu, “超高分辨率图像修复的上下文残差聚合”, Proc. IEEE 会议. 计算. 维斯. 模式识别. (CVPR), 2020 年, 第 7508–7517 页. 6[57] 体育. 丹尼尔森, “欧几里得距离映射”, 计算机图形学和图像处理, 卷. 14、没有. 3, 第 227–248 页, 1980 年. 6[58] Y. Wu 和 Y. Liu, “鲁棒截断铰链损失支持向量机”, 美国统计协会杂志, 卷. 102, 第 479 期, 第 974–983 页, 2007 年. 7[59] F. Schroff, D. Kalenichenko 和 J. Philbin, “Facenet: 用于人脸识别和聚类的统一嵌入”, Proc. 102, 第 479 期, 第 974–983 页, 2007 年. IEEE 会议计算. 模式识别. (CVPR), 2015 年, 第 815–823 页. 7[60] R. Zhang, P. Isola, A. A. Efros, E. Shechtman 和 O. Wang, “深度特征作为感知度量的不合理有效性”, in Proc. IEEE 会议计算. 维斯. 模式识别. (CVPR), 2018 年, 第 586–595 页. 8[61] S. Fu, N. Tamir, S. Sundaram, L. Chai, R. Zhu, T. Dekel 和 P. Isola, “Dreamsim: 使用合成数据学习人类视觉相似性的新维度”, arXiv 预印本 arXiv:2306.09344, 2023. 8[62] H. Liu, Z. Wan, W. Huang, Y. Song, X. Han, and J. Liao, “Pd-gan: 用于图像修复的概率多样化 gan”, Proc. IEEE 计算会议. 维斯. 模式识别. (CVPR), 2021 年, 第 9371–9381 页. 8

[63] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte 和 L. Van Gool, “重新绘制: 使用去噪扩散概率模型进行修复”, Proc. IEEE 会议计算. 维斯. 模式识别. (CVPR), 2022, 第 11 461–11 471 页. 8[64] Z. Guo, H. Zheng, Y. Jiang, Z. Gu 和 B. Zheng, “内在图像协调”, Proc. IEEE 会议计算. 维斯. 模式识别. (CVPR), 2021 年 6 月, 第 16 367–16 376 页. 10[65] W. Cong, J. Zhang, L. Niu, L. Liu, Z. Ling, W. Li 和 L. Zhang, “Dovenet: 通过域验证实现深度图像协调”, Proc. IEEE Conf. 计算. 维斯. 模式识别. (CVPR), 2020 年 6 月. 10[66] R. A. Bradley 和 M. E. Terry, “不完整模块设计的排名分析: I. 配对比较方法”, Biometrika, 卷. 39, 第 3/4 期, 第 324–345 页, 1952 年. 11[67] M. Kim, A. K. Jain 和 X. Liu, “Adaface: 人脸识别的质量自适应边缘”, Proc. 39, 第 3/4 期, 第 324–345 页, 1952 年. IEEE 会议计算. 维斯. 模式识别. (CVPR), 2022 年 6 月, 第 18 750–18 759 页. 13



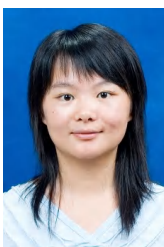
郑洪生获得学士学位2019年毕业于华侨大学工商管理学院, 获硕士学位. 2023年毕业于云南大学信息科学与工程学院, 获博士学位. 目前正在攻读博士学位. 武汉大学计算机学院博士, 研究方向为计算机视觉、面部高光去除、AIGC.



徐文举获得中国西安交通大学硕士学位. 他获得了博士学位. 美国劳伦斯塔萨斯大学博士学位. 他的主要研究兴趣是深度学习、计算机视觉和智能系统领域, 重点是图像合成、图像理解和语义分割. 徐先生还是20+国际顶级期刊和会议的审稿人.



王振宇, 毕业于中国武汉大学人工智能学院, 获得硕士学位. 他的研究兴趣包括深度学习、计算机视觉和面部高光去除.



肖鲁获得博士学位2015年毕业于湖南大学模式识别与智能系统专业, 获工学博士学位. 现为湖南师范大学工程设计学院副教授. 主要研究方向为机器学习、计算机视觉、图像等. 处理、机器人感知和决策.

肖春霞获得学士学位和硕士。分别于1999年和2002年获得湖南师范大学数学学士学位。他获得了博士学位。2006年获浙江大学CAD&CG国家重点实验室应用数学博士学位，2006年任武汉大学助理教授，2011年晋升教授，2006年10月至2007年4月在武汉大学从事博士后研究。香港科技大学计算机科学与工程系，香港。2012年2月至2013年2月，访问加州大学戴维斯分校。现任武汉大学计算机学院教授。他的主要兴趣包括计算机视觉、计算机图形学、虚拟现实和增强现实。他在期刊和会议上发表了 150 多篇论文。他是 IEEE 的高级会员。

ReadPaper.com