

基于知识蒸馏与三维先验引导的低光照人脸图像增强算法^①



程 亮, 张 玲

(武汉科技大学 计算机科学与技术学院, 武汉 430065)

通信作者: 张 玲, E-mail: zhling@wust.edu.cn

摘 要: 针对低光照人脸图像增强中真实配对数据不足和面部特征相关性较弱的问题, 提出了一种基于知识蒸馏与三维先验引导的半监督低光照图像增强算法. 首先, 构建了基于扩散模型的三维先验提取网络, 在三平面特征空间中扩散生成高质量的三维人脸先验, 为后续增强提供稳定的结构约束与细节参考. 随后, 设计了基于知识蒸馏的半监督图像增强网络, 联合使用配对与未配对数据, 实现低光照人脸增强. 其中, 提出的三维先验融合模块提升了模型对人脸空间结构的感知能力, 从而生成结构自然、细节逼真的增强结果. 同时, 无监督知识蒸馏策略结合网络内与网络间一致性, 有效提升模型的泛化性能. 实验结果表明, 该方法在合成和真实低光照人脸数据集上均取得了优异表现, 能够显著增强图像亮度与视觉真实感.

关键词: 人脸图像; 低光照增强; 知识蒸馏; 三维先验; 半监督学习

引用格式: 程亮, 张玲. 基于知识蒸馏与三维先验引导的低光照人脸图像增强算法. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/10146.html>

Low-light Face Image Enhancement Algorithm Based on Knowledge Distillation and 3D Prior Guidance

CHENG Liang, ZHANG Ling

(School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China)

Abstract: To address the scarcity of paired data and weak facial feature correlations in low-light face image enhancement, this study proposes a semi-supervised algorithm guided by knowledge distillation and 3D priors. First, a 3D prior extraction network based on a diffusion model is constructed to generate high-quality 3D facial priors in a tri-plane feature space, providing stable structural constraints and detailed guidance for subsequent enhancement. Then, a semi-supervised image enhancement network based on knowledge distillation is designed to jointly leverage paired and unpaired data for low-light face enhancement. The proposed 3D prior feature fusion module enhances the model's perception of facial spatial structures, thus generating enhanced results with more natural geometry and realistic details. Meanwhile, the unsupervised knowledge distillation strategy combines intra-network and inter-network consistency, effectively enhancing the model's generalization capability. Experimental results show that the proposed method achieves excellent performance on both synthetic and real low-light face datasets. It significantly improves image brightness and visual realism.

Key words: face image; low-light enhancement; knowledge distillation; 3D prior; semi-supervised learning

^① 基金项目: 湖北省自然科学基金 (2023AFB615)

收稿时间: 2025-09-29; 修改时间: 2025-10-28, 2025-12-02; 采用时间: 2025-12-05; csa 在线出版时间: 2026-03-09

在低光照条件下采集的人脸图像常伴随可见度不足、细节丢失和噪声干扰。这不仅降低了视觉质量,还会影响成像系统及高级视觉任务(如人脸识别^[1,2]、人脸检测^[3,4]等)的性能表现。因此,低光照人脸图像增强已成为计算机视觉领域一个重要且具有挑战性的任务。

现有基于深度学习的方法^[5-8]从数据中学习光照调整与细节恢复的映射。然而,它们大多依赖真实的成对数据训练。在实际中获取大规模的真实配对人脸数据极为困难,因此现有方法常使用合成数据训练,但容易造成过拟合并限制泛化能力。部分方法^[9-11]尝试摆脱对配对数据的依赖,但在真实场景中常出现过度或欠增强、噪声放大等问题。针对这些不足,本文结合配对与未配对数据进行学习,以弥补各自的局限性。

在低级视觉任务中,一些研究^[12,13]利用知识蒸馏在特征或预测层面建立一致性。一类方法将大型模型压缩为轻量化模型,但依赖严格配对的数据,在真实场景下泛化性有限。另一类方法蒸馏通用模型的知识以帮助专一模型适应多种退化,但主要关注特征幅值差异,可能导致曝光不当或颜色偏移。在高级视觉任务中,知识蒸馏常用于半监督学习^[14,15],通过有效利用标注与未标注数据提升泛化性能。受此启发,本文从知识蒸馏角度出发,联合使用配对与未配对数据,以克服单一监督模式的局限。此外,现有一致性正则化方法主要约束同一输入在不同扰动下的预测输出或概率分布,在特征层约束较弱,难以捕捉人脸图像中的结构与光照差异,容易导致颜色失真或细节丢失。

大多数低光照增强方法未能充分利用人脸先验知识,容易导致肤色不自然、五官细节恢复不佳。部分研究^[16]引入人脸解析图等二维先验辅助增强,但其仅反映面部的平面布局和局部特征,难以完整刻画面部空间几何关系。因此,在复杂光照条件或极端姿态下难以准确恢复面部细节。相比自然场景图像,人脸具有高度结构化的三维形态特征,这些三维结构不仅定义了面部的空间布局,还对光照分布具有显著影响。引入三维人脸先验可提供更丰富的空间信息,提升结构完整性与真实感。基于此,本文将三维人脸先验作为辅助增强条件,以改善模型在复杂光照下的表现。

本文提出了一种基于知识蒸馏与三维先验引导的半监督低光照图像增强方法,在恢复光照的同时提升结构保真度。首先,构建一个基于扩散模型的三维先验提取网络,通过条件三维去噪网络逐步去除噪声,以生

成高质量的三维先验。与传统二维扩散模型在像素空间建模噪声不同,本文在三平面特征空间中进行扩散建模,以显式建模并重建面部几何与光照分布。随后,我们设计了一个基于知识蒸馏的半监督图像增强网络,利用三维先验引导光照恢复与结构重建。为实现三维先验与二维特征的高效融合,提出了一个三维先验融合模块,通过自适应通道注意力与动态特征聚合缓解结构与纹理失真问题。此外,使用一种无监督蒸馏策略,联合使用配对与未配对数据,通过网络内与网络间的特征一致性约束提升模型在真实场景下的泛化能力。本文的主要贡献如下。

(1) 提出一种基于知识蒸馏与三维先验引导的半监督低光照图像增强方法,实现了光照自然、结构真实的增强效果。

(2) 构建一个基于扩散模型的三维先验提取网络,利用条件三维去噪网络生成符合真实分布的三维人脸先验,为图像增强提供几何约束。

(3) 设计一个基于知识蒸馏的半监督图像增强网络,包含三维先验融合模块和无监督蒸馏策略,可以保持人脸结构的一致性及提升模型的泛化性能。

1 相关工作

1.1 低光照自然图像增强

现有低光照增强方法主要分为两类:传统方法与深度学习方法。传统方法^[17,18]通常建立退化模型并估计相关参数,可在一定程度上改善亮度、对比度和噪声,但易受手工先验限制而产生光照不均或色偏。近年来,深度学习已成为低光照增强的主流方法,通常依赖于成对的低光和正常光照图像进行训练。Lore等人^[5]提出了第一个深度学习方法LLNet,采用叠加稀疏去噪自动编码器来增强低光照图像。Wang等人^[6]设计了一种通用的Transformer框架用于高分辨率图像增强。Liu等人^[7]将传统的去噪扩散过程解耦为残差扩散和噪声扩散。Yan等人^[8]提出了一个HVI颜色空间实现亮度与颜色的分离建模。Zhang等人^[12]提出了一种基于蒸馏的轻量级低光照图像增强网络以解决过曝和噪声残留问题。虽然这些方法可以获得良好的性能,但由于训练依赖严格配对数据,限制了它们在真实低光照环境下的泛化能力。为了解决这个问题,一些研究尝试使用无配对的数据进行训练。Jiang等人^[9]采用了带注意力的生成对抗网络学习非配对映射。Yang等人^[10]

以无监督的方式使用双循环协作对抗约束过程对增强模块进行训练. Shi 等人^[11]考虑了去噪与增强之间的耦合关系, 利用图像的固有信息增强低光照图像. 尽管这些方法在低光照自然图像增强中取得了显著进展, 但由于未考虑人脸的独特先验知识, 在人脸细节恢复和光照一致性方面仍表现有限.

1.2 低光照人脸图像增强

近年来, 针对低光照人脸图像增强任务, 已有多种方法被提出. 早期研究主要集中于针对特定且可建模的低光照退化进行增强. 例如, SeLENet^[19]使用球谐函数对人脸光照进行建模, 从而实现光照补偿与图像重建. Fan 等人^[16]提出了面向人脸识别的增强网络 Low-FaceNet, 通过联合优化低级和高级视觉任务, 实现了增强与识别的协同提升. 目前为应对低光照环境下人脸图像更加复杂的退化问题, 研究逐渐转向同时恢复光照与结构细节. IC-FSRDENet^[20]采用两阶段策略, 第1阶段进行光照调整与结构恢复, 第2阶段利用扩散模型进行细节增强. FRR-Net^[21]在建模过程中综合考虑多种退化因素, 以进一步提升增强质量. 然而, 这些方法普遍使用合成的低光照人脸数据进行训练, 且忽略三维几何信息的重要性, 从而在空间感知能力以及真实场景下的泛化性能方面存在一定局限.

2 基于扩散模型的低光照人脸三维先验建模

从低光照人脸图像中重建高质量的三维人脸先验任务面临诸多挑战. 一方面, 低光照图像中常伴随纹理退化、边缘模糊和噪声干扰, 增加了三维结构恢复的不确定性. 另一方面, 传统的卷积回归方法^[22]生成的结果往往缺乏高频纹理信息. 为此, 本文引入扩散模型作为三维先验建模的核心框架. 扩散模型通过学习从高斯噪声逐步还原真实样本的过程, 能够更稳定地建模低光照图像与真实三维结构之间的复杂映射关系.

本文使用去噪扩散概率模型 (DDPM)^[23]框架从低光照人脸图像中提取三维先验. 扩散模型包含一个前向扩散过程与一个反向去噪过程. 在推理时, 可以将低光照图像作为输入, 通过逐步去除噪声, 最终生成对应的三维先验.

- 前向扩散: 前向扩散过程以正常光照图像对应的三平面特征^[24]作为前向马尔可夫过程的起点, 这些特征由一个预训练的三维特征生成器^[25]获得. 考虑到渲染操作依赖相机参数, 本文将相机参数展平并广播,

与三平面特征拼接形成最终的扩散目标. 随后, 通过 T 步迭代逐步添加高斯噪声, 每一步定义如下:

$$x_t = \sqrt{1-\beta_t}x_{t-1} + \sqrt{\beta_t}\varepsilon \quad (1)$$

其中, $t = 1, 2, \dots, T$, $\beta_t \in (0, 1)$ 是预定义的噪声方差调控因子, $\varepsilon \sim \mathcal{N}(0, \mathbf{I})$ 为标准高斯噪声, \mathbf{I} 表示单位矩阵. 该过程可以进一步简化^[26]:

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\varepsilon \quad (2)$$

其中, $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$, $\alpha_i = 1 - \beta_i$.

- 反向去噪: 反向去噪过程的目标是从一个初始高斯噪声样本 $x_T \sim \mathcal{N}(0, \mathbf{I})$ 出发, 逐步去噪重建出高质量的三平面特征. 每一步可表示为:

$$x_{t-1} = \mu_t(x_t, x_0) + \sigma_t\varepsilon \quad (3)$$

其中, 均值 $\mu_t(x_t, x_0) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \varepsilon \right)$, 方差 $\sigma_t^2 = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t} \beta_t$. ε 表示 x_t 中的噪声成分, 是唯一的不确定变量, 由去噪网络 $\varepsilon_\theta(x_t, t)$ 进行估计. 在第 t 个时间步, 去噪网络可表示为 $\varepsilon_\theta(x_t, I, t)$, I 为低光照图像, 用作条件引导. 通过引入条件图像, 模型能够更准确地从噪声中恢复对应的人脸结构. 将式 (3) 中的 ε 用 $\varepsilon_\theta(x_t, t)$ 进行替换, 并将方差设置为 $1 - \alpha_t$, 可以得到最终的采样公式:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \varepsilon_\theta(x_t, t) \right) + \sqrt{1-\alpha_t}\varepsilon_t \quad (4)$$

经过 T 次反向迭代后, 获得最终的三平面特征. 之后, 利用预训练的三维渲染器生成三维渲染图和深度图. 具体地, 首先对三平面特征应用双线性插值, 以实现空间特征重采样并获得聚合特征. 随后, 通过轻量级多层感知机对聚合特征进行非线性映射, 将高维几何特征转换为颜色参数和密度参数. 最后, 使用体渲染沿着相机视线方向进行离散采样, 对每个采样点的颜色和密度进行数值积分, 生成对应视角下的三维渲染图和深度图. 之后, 对三维渲染图和深度图执行两次下采样操作, 获得 3 个尺度的三维图像特征 T_1 、 T_2 、 T_3 和深度特征 D_1 、 D_2 、 D_3 .

3 低光照人脸图像增强算法

本文提出了一种基于知识蒸馏与三维先验引导的半监督低光照人脸图像增强方法. 整体框架如图 1 所

示,主要包括两个部分:基于扩散模型的三维先验提取网络与基于知识蒸馏的半监督图像增强网络.首先,三维先验提取网络从低光照图像中估计对应的三维人脸先验,为后续增强过程提供几何与细节指导.随后,在三维先验的引导下,半监督图像增强网络对低光照图

像进行亮度提升与细节恢复.该网络采用知识蒸馏策略,教师网络与学生网络具有相同的结构.在训练阶段,合成与真实低光照图像共同输入两个网络,学生网络通过蒸馏学习教师网络的特征表达能力.在推理阶段,仅使用学生网络用于图像增强.

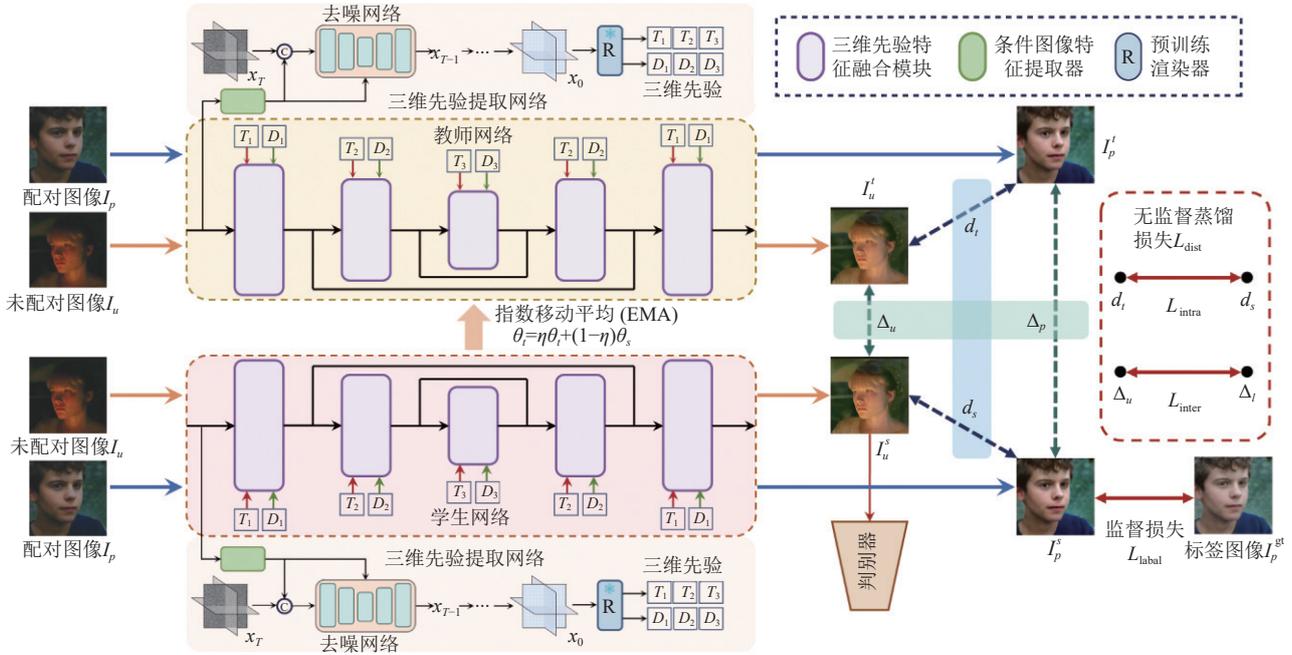


图1 整体框架图

3.1 基于扩散模型的三维先验提取网络

与自然场景图像相比,人脸具有高度结构化的三维形态.而二维先验(如人脸解析图、肤色图)难以提供完整的空间结构信息,限制了模型在复杂光照下的表现.为增强模型的空间感知能力,本文引入三维人脸先验,包括三维渲染图和深度图.渲染图模拟标准光照下的面部外观,为光照校正和纹理补全提供细节参考;深度图刻画面部几何和空间分布,有助于提升轮廓与立体感的建模能力.通过融合这两种三维先验,模型能够在保持结构一致性的同时恢复更多细节.

为了获取稳定且高保真的三维先验,本文基于扩散模型构建了一个三维先验提取网络.在反向去噪过程中,使用一个条件三维去噪网络估计噪声,逐步恢复高质量的三平面特征.与在像素空间中建模的二维扩散方法不同,我们在三平面特征空间中进行扩散建模.三平面^[23]是一种高效紧凑的三维表示形式,通过3个正交平面有效编码人脸的三维几何结构,在保留强几何表达能力的同时具有较低的计算开销.本文采用的

三平面特征的维度为 $3 \times 32 \times 256 \times 256$.

3.1.1 条件三维去噪网络

为从噪声三平面特征中逐步恢复高质量的三维人脸表示,本文设计了一个条件三维去噪网络,如图2所示.该网络由主干去噪网络和条件图像特征提取器组成,旨在利用低光照图像中的结构与纹理信息引导扩散过程中的噪声估计,从而提升三平面特征的重建质量与结构稳定性.不同于现有二维扩散模型仅简单拼接条件图像与噪声图像的方式,本文的去噪网络引入多级条件引导机制,在不同层级中动态注入条件特征,对扩散重建过程进行逐步补偿与校正,实现几何一致性与细节恢复的协同优化.

条件图像特征提取器由3个串联的残差块构成,用于从低光照图像中提取与三维结构相关的深层特征,其输出与三平面特征在空间维度上保持一致,以增强结构感知能力.主干去噪网络采用3级U-Net架构,每一级均包含一个残差块,用于提取和恢复多尺度特征.该网络以当前扩散步的噪声特征和条件特征为输入,

预测相应的噪声分量. 在 U-Net 的第 2、3 级中, 通过条件特征注入模块引入不同尺度的引导信息. 为避免简单拼接造成特征失衡或噪声传播. 本文设计了一个通道注意力引导的自适应融合块. 具体地, 首先将主干特征与条件特征拼接后进行全局平均池化 (global

average pooling, GAP), 再经过两个卷积层, 并通过 Softmax 得到通道注意力分布, 对拼接特征进行加权增强. 最后将加权特征与主干特征逐元素相加. 该机制在保持主干网络去噪能力的同时, 有效引入条件图像的结构与纹理信息, 提升了噪声估计的准确性和鲁棒性.

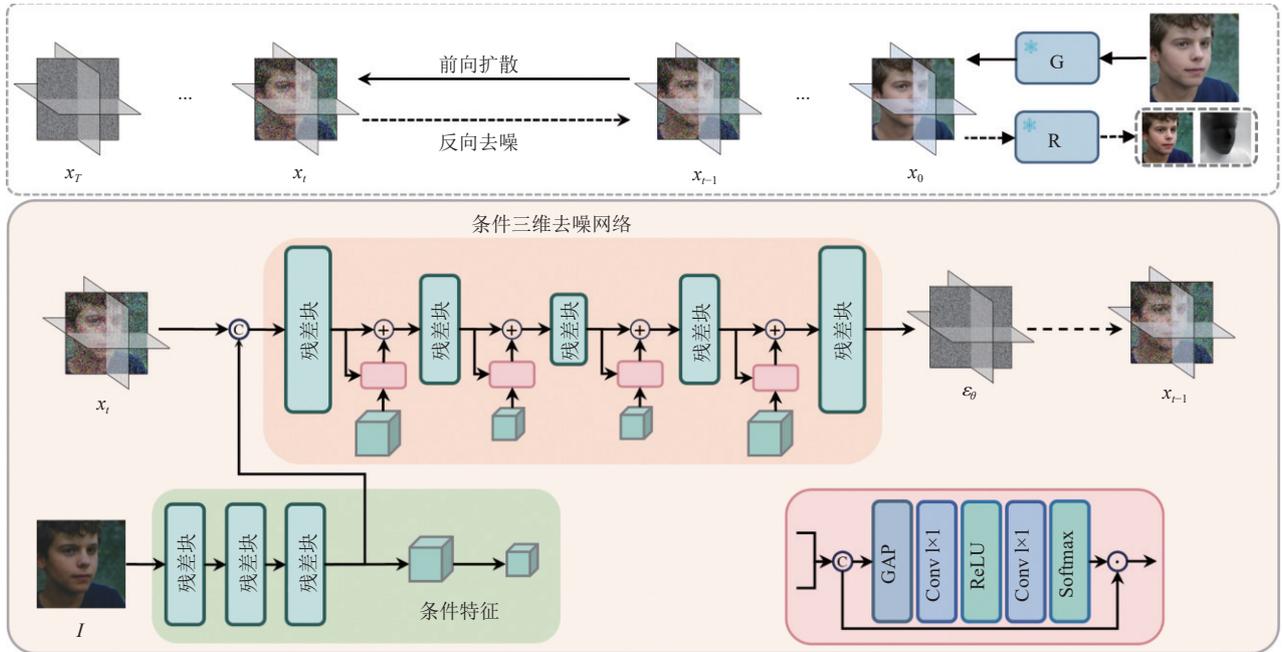


图 2 基于扩散模型的三维先验提取网络

3.2 基于知识蒸馏的半监督图像增强网络

在低光照人脸图像增强任务中, 现有方法常受到配对数据不足和无监督学习不稳定的限制. 因此, 本文提出了一种基于知识蒸馏的半监督图像增强网络, 联合使用配对数据和未配对数据, 在监督与无监督学习之间取得平衡. 该网络由结构相同的教师网络和学生网络组成, 均采用 3 级编码器-解码器结构. 每一级均包含一个三维先验融合模块, 利用三维图像特征与深度特征增强几何一致性. 编码器部分通过逐层下采样在空间分辨率减半的同时将通道数扩展 2 倍. 解码器则采用对称结构, 通过逐层上采样恢复分辨率, 并与对应编码器特征进行跳跃连接, 以提升细节重建能力. 为避免小批量训练引起的不稳定性并确保推理阶段的一致性, 该网络未使用批归一化层. 此外, 本文使用无监督知识蒸馏策略, 通过一致性约束挖掘配对与未配对数据的互补信息, 从而提升泛化性.

3.2.1 三维先验融合模块

如果直接将三维先验与图像特征进行拼接或简单

相加融合, 往往会造成特征分布不均衡, 导致三维先验信息被弱化, 进而降低其对光照恢复与细节重建的引导作用. 为此, 本文设计了一个三维先验融合模块, 如图 3 所示. 该模块旨在充分挖掘三维先验与二维特征的互补关系, 通过自适应通道注意力引导与动态特征聚合协同优化结构一致性与细节重建. 与传统基于拼接或简单注意力的融合方式不同, 该模块不仅区分三维图像特征与深度特征的语义层次, 通过通道重加权使网络聚焦于与光照和面部几何相关的关键通道, 还将三维先验用作动态调制参数, 用于校正结构失真和增强细节恢复.

首先, 对输入特征 F 进行预处理. 具体包括层归一化操作以确保数据分布稳定, 接着通过 Swish 激活函数引入非线性增强表达能力, 随后使用 3×3 卷积提取局部空间特征. 然后, 预处理后的特征 F 利用深度特征 D 对其进行调整, 以增强空间感知能力. 接着, 三维图像特征 T 输入通道注意力单元以获得通道注意力权重. 该单元由全局平均池化、 1×1 卷积、ReLU 激活函

数、另一个 1×1 卷积和 Sigmoid 激活函数组成,使模型可以关注人脸图像中的关键结构通道.之后,利用注意力权重,分别对输入特征和三维图像特征进行逐元素乘法,得到融合特征 F_1 和 F_2 . 随后将 F_1 和 F_2 拼接,

通过一个 1×1 卷积、GELU 激活函数和另一个 1×1 卷积(以下将其简化描述为 MLP)进行特征融合,以提高特征的紧凑性和有效性.最后,将融合结果与初始结果 F 进行逐元素加法,获得中间输出 F_m .

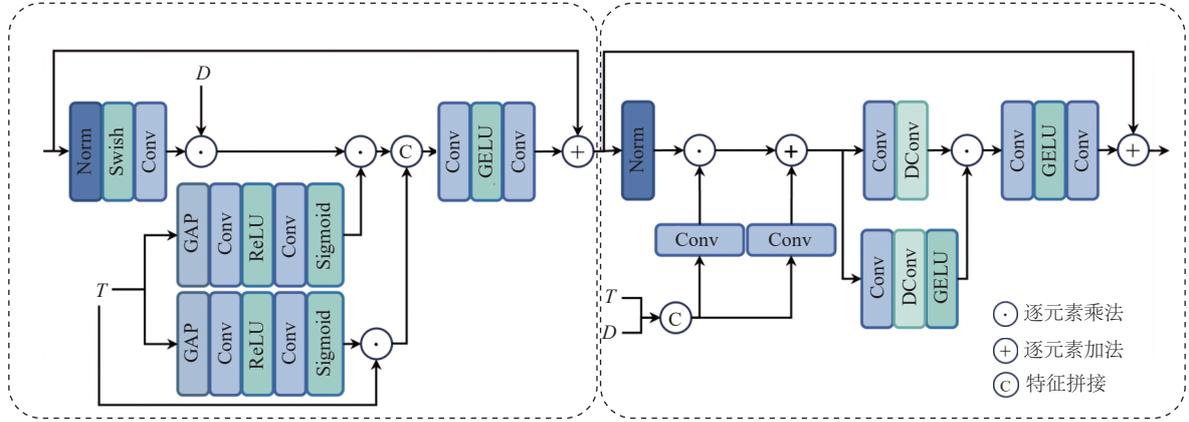


图3 三维先验融合模块

中间输出 F_m 被输入到下一个单元,再次进行层归一化操作,以保持特征分布的一致性.随后,将三维图像特征 T 和深度特征 D 拼接,通过卷积获得两个仿射参数 γ 和 β ,并对 F_m 执行逐像素仿射变换,得到重构特征.之后,该重构特征一路经过 1×1 卷积和 3×3 深度可分离卷积提取空间细节,获得特征 F_3 ;另一路经过 1×1 卷积、 3×3 深度可分离卷积和 GELU 激活函数,得到带非线性的细节特征 F_4 .将 F_3 与 F_4 进行逐元素乘法,实现特征的互补增强.接着,融合后的特征经过 MLP 进一步优化.最后,优化后的特征和 F_m 进行逐元素加法,获得最终输出特征.

3.2.2 无监督蒸馏策略

为充分利用有限配对数据之外的未配对数据,本文使用一种无监督蒸馏策略,通过网络内和网络间一致性蒸馏,在配对与未配对数据间建立特征级约束.与常见的一致性正则化方法不同,该策略不仅关注教师与学生网络在相同输入下的输出一致性,还进一步从不同输入与不同网络两个层面约束特征分布的一致性,使学生网络能够获得更稳定且细粒度的监督信号.

如图1所示,本文将配对数据和非配对数据统一纳入教师-学生框架.配对的低光照图像 I_p 与未配对的低光照图像 I_u 同时输入教师网络与学生网络,得到预测结果 I_p^t 、 I_p^s 、 I_u^t 和 I_u^s .网络内一致性蒸馏使用余弦距离约束教师网络与学生网络在配对与未配对输入下的

特征方向一致性,避免幅值敏感问题,增强特征表征的稳定性.网络间一致性蒸馏利用 Gram 矩阵计算通道相关性,保持特征结构的一致性.

本文中的教师网络和学生网络使用相同的数据集训练,但在优化方式上存在差异.学生网络参数 θ_s 通过反向传播直接优化更新,而教师网络参数 θ_t 则使用学生参数的指数移动平均 (EMA) 进行更新:

$$\theta_t = \eta \theta_t + (1 - \eta) \theta_s \quad (5)$$

其中, $\eta \in (0, 1)$ 为动量系数.这种方式能够稳定教师网络的输出,为学生网络提供平滑的一致性蒸馏信号.

3.3 损失函数

3.3.1 三维先验扩散模型的损失

三维先验扩散模型通过最小化网络预测的噪声与真实噪声之间的均方误差进行优化.

$$L_{\text{noise}} = \|\varepsilon_\theta - \varepsilon\|_2^2 \quad (6)$$

其中, ε_θ 为网络预测的噪声, ε 为真实噪声.

3.3.2 半监督增强框架的损失

优化半监督增强框架的损失函数由监督损失、无监督蒸馏损失和对抗损失组成,即:

$$L_{\text{enhance}} = L_{\text{label}} + \lambda_1 L_{\text{distill}} + \lambda_2 L_{\text{adv}} \quad (7)$$

其中, λ_1 和 λ_2 是权重参数.

监督损失用于衡量学生网络输出与配对真实标签图像的差异,以减少对蒸馏的过拟合并保证基本重建

质量. 它由重建损失和感知损失组成:

$$\begin{aligned} L_{\text{label}} &= L_{\text{rec}} + \lambda_3 L_{\text{per}} \\ &= \|I_p^s - I_p^{\text{gt}}\|_1 + \lambda_3 \|\Phi(I_p^s) - \Phi(I_p^{\text{gt}})\|_2^2 \end{aligned} \quad (8)$$

其中, I_p^s 为学生网络的增强结果, I_p^{gt} 为对应正常光照图像, $\Phi(\cdot)$ 表示预训练 VGG19 模型的特征提取器.

无监督蒸馏损失包含网络内一致性蒸馏损失 L_{intra} 和网络间一致性蒸馏损失 L_{inter} , 权重分别为 λ_4 和 λ_5 :

$$L_{\text{distill}} = \lambda_4 L_{\text{intra}} + \lambda_5 L_{\text{inter}} \quad (9)$$

网络内一致性蒸馏损失旨在保持同一网络在不同输入下预测的低层特征差异的一致性. 设 Φ_{ij} 表示 VGG19 的第 i 个残差块中的第 j 层特征, 则教师网络与学生网络的特征距离分别定义为:

$$\begin{cases} d_t^{ij} = c(\Phi_{ij}(I_p^t), \Phi_{ij}(I_u^t)) \\ d_s^{ij} = c(\Phi_{ij}(I_p^s), \Phi_{ij}(I_u^s)) \end{cases} \quad (10)$$

其中, $c(\cdot, \cdot)$ 表示余弦距离. 相比于常用的 L1 距离, 余弦距离更注重特征方向的一致性, 从而能够在忽略幅值变化的情况下, 更准确地保持特征间的结构化关系. 为了进一步对齐教师和学生网络的特征分布, 对二者的特征距离进行 Softmax 归一化后, 计算交叉熵损失:

$$L_{\text{intra}} = -\frac{1}{hw} \sum_{i,j,m,n} S(d_t^{ij}[m,n]) \log S(d_s^{ij}[m,n]) \quad (11)$$

其中, $[m, n]$ 为空间索引, $S(\cdot)$ 表示 Softmax 函数. 交叉熵损失能够提升预测结果的局部判别性与细节锐度.

网络间一致性蒸馏损失约束不同网络在相同输入下预测的低层特征变化保持一致. 对于同一输入的两组预测 $\{I_p^t, I_p^s\}$ 和 $\{I_u^t, I_u^s\}$, 首先计算它们的特征差异:

$$\begin{cases} \Delta_p^{ij} = \text{Gram}(\Phi_{ij}(I_p^t) - \Phi_{ij}(I_p^s)) \\ \Delta_u^{ij} = \text{Gram}(\Phi_{ij}(I_u^t) - \Phi_{ij}(I_u^s)) \end{cases} \quad (12)$$

其中, $\text{Gram}(\cdot)$ 表示 Gram 矩阵运算, 通过计算通道维度上的特征相关性, 有效消除了空间布局差异对对齐过程的影响. 最终的跨网络一致性蒸馏损失定义为:

$$L_{\text{inter}} = \sum_{i,j} \|\Delta_u^{ij} - \Delta_p^{ij}\|_{\mathcal{F}} \quad (13)$$

其中, $\|\cdot\|_{\mathcal{F}}$ 为 Frobenius 范数, 用于度量 Gram 矩阵的分布差异.

对抗损失用于最小化增强图像与正常光照图像在分布上的差距, 使输出图像更具自然感.

$$L_{\text{adv}} = \mathbb{E}[(D(I_u^s, I_p^{\text{gt}}) - 1)^2] \quad (14)$$

其中, I_u^s 为学生网络中未配对图像的增强结果, I_p^{gt} 为正常光照图像, $D(\cdot)$ 是判别器. 本文使用双通道颜色边缘判别器^[10]来区分图像.

4 实验结果与分析

4.1 实验准备

- 数据集: 本文在合成数据集 FFHQ^[27]和真实世界数据集 CelebV-Low 上进行训练和测试. 具体而言, 从 FFHQ 数据集中选取 5 000 张图像作为训练集, 另外选取 1 000 张图像作为测试集, 对应的低光照图像通过与 FRR-Net^[21]相同的退化策略合成. CelebV-Low 来源于大规模视频人脸数据集 CelebV-HQ^[28], 该数据集包含多样化的真实拍摄场景和光照变化. 为获得真实低光照样本, 本文根据亮度分布筛选 CelebV-HQ 中的帧, 选取平均亮度低于 0.2 的帧作为低光照样本. 为进一步保证数据质量, 人工剔除了部分存在严重模糊、遮挡或异常曝光的人脸图像. 最终, 随机选取 5 000 张图像作为训练集, 892 张图像作为测试集. 该数据集未包含配对的正常光照样本, 因此用于本文的半监督学习与泛化性能评估.

- 实验设置: 本文方法基于 PyTorch 实现, 并在 NVIDIA GeForce RTX 3090 上训练. 首先, 在 FFHQ 数据集上训练三维先验提取网络, 迭代次数为 1×10^6 . 优化器为 Adam, 动量参数为 $\beta_1 = 0.9$, $\beta_2 = 0.99$. 学习率使用余弦退火从 4×10^{-5} 逐渐衰减至 2×10^{-6} . 随后, 在 FFHQ 与 CelebV-Low 数据集上训练半监督图像增强网络 200 个 epoch. 优化器为 Adam, 动量参数为 (0.5, 0.999), 学习率使用余弦退火从 1×10^{-4} 衰减至 1×10^{-6} . 实验中, 预训练的三维渲染器的渲染分辨率设为 128×128 , 权重参数 λ_1 、 λ_2 、 λ_3 、 λ_4 、 λ_5 分别设置为 0.5、0.01、0.1、0.5 和 0.5. 所有输入图像均调整为 256×256 分辨率, 并通过随机水平翻转进行数据增强.

- 评估指标: 为了全面评估所提方法的性能, 针对不同类型的数据集采用相应的评价指标. 对于配对数据集 FFHQ, 使用 PSNR^[29]、SSIM^[30]和 LPIPS^[31]衡量增强图像与真实图像之间的差异. 对于非配对数据集 CelebV-Low, 使用无参考图像质量指标 NIQE^[32]和 BRISQUE^[33]对结果的感知质量进行评估.

4.2 对比实验

为验证所提方法的有效性, 本文将其与一些主流的低光照增强方法进行比较. 为确保公平比较, 所有监

督方法在 FFHQ 数据集上重新训练, 无监督方法在 FFHQ 和 CelebV-Low 数据集上重新训练. 实验结果如表 1 所示, 可以看出, 本文方法在数据集 FFHQ 上取得了最好的 PSNR、SSIM 和 LPIPS 分数, 并在数据集 CelebV-Low 上取得最佳 NIQE 和 BRISQUE 表现.

为了进一步验证方法的效果, 本文在合成与真实数据集上进行了可视化对比, 结果如图 4 与图 5 所示. 当低光照图像出现严重退化时, EnlightenGAN^[9]容易产生面部畸变. ZERO-IG^[11]和 Low-FaceNet^[16]虽然推理速度较快, 但由于光照估计不准确, 常出现明显的光照伪影或不均衡增强. UHDformer^[6]借助长程依赖建模提升了整体亮度, 但容易导致细节模糊. DLIENet^[12]也采用知识蒸馏策略增强低光照图像, 但在未知场景中泛化能力有限, 容易出现欠增强. IC-FSRDENet^[20]利用扩散

模型对图像进行细化, 但增强结果仍存在光照不均的现象. 相比之下, 本文方法更好地保持了人脸结构完整性, 有效减轻光照伪影, 增强结果更加自然.

表 1 评估得分对比

方法	FFHQ数据集			CelebV-Low数据集	
	PSNR (dB)	SSIM	LPIPS	NIQE	BRISQUE
EnlightenGAN ^[9]	17.19	0.57	0.35	3.85	35.91
ZERO-IG ^[11]	18.28	0.58	0.42	3.94	40.85
UHDformer ^[6]	23.85	0.74	0.32	4.92	37.97
DLIENet ^[12]	24.03	0.75	0.29	4.51	34.57
Low-FaceNet ^[16]	17.33	0.54	0.43	3.77	39.42
IC-FSRDENet ^[20]	23.92	0.73	0.31	5.11	32.68
RDDM ^[7]	22.25	0.72	0.28	4.88	36.13
HVI-CIDNet ^[8]	24.01	0.75	0.24	4.19	35.84
本文方法	24.09	0.75	0.21	3.36	31.79

注: 加粗数据表示最佳结果

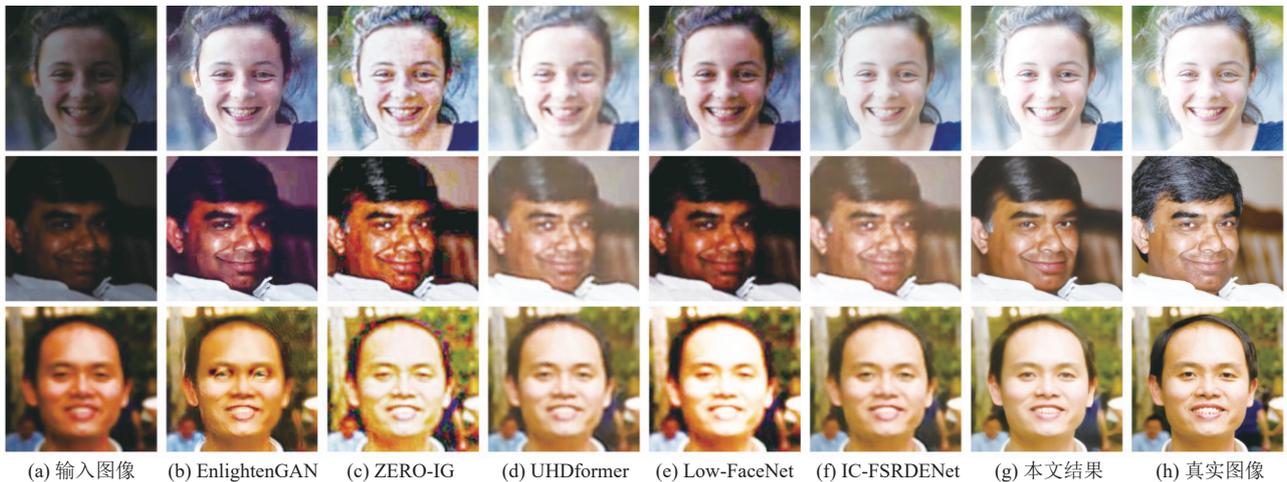


图 4 在合成数据集 FFHQ 上的视觉比较图



图 5 在真实世界数据集 CelebV-Low 上的视觉比较图

4.3 消融实验

为验证所提出方法各个组成部分的有效性, 本文构建了一系列变体, 各变体设置如下.

变体 1: 不使用知识蒸馏, 仅在 FFHQ 数据集上进行监督训练.

变体 2: 不使用无监督蒸馏策略, 仅将教师网络输出作为伪标签对学生网络进行监督.

变体 3: 移除网络内一致性蒸馏损失.

变体 4: 移除网络间一致性蒸馏损失.

变体 5: 不使用三维人脸先验引导.

变体 6: 移除基于扩散模型的三维先验提取网络, 使用 RDDM^[9]中的普通 U-Net 回归三平面特征.

变体 7: 将三维先验融合模块替换为普通残差块, 直接拼接融合三维先验.

本文在 FFHQ 与 CelebV-Low 数据集上训练并评估了这 7 个模型变体, 结果如表 2 所示. 图 6 展示了变体 1-4 在真实低光照图像上的可视化结果. 可以观察到, 变体 1 在完全去除蒸馏机制后, 增强结果易出现光

照失真或颜色偏差, 说明仅依赖监督训练难以应对真实场景中复杂多变的光照条件. 变体 2 使用伪标签虽然较变体 1 有所改善, 但仍明显不如完整方法, 反映出单一伪监督在泛化能力上的不足. 变体 3 (缺少网络内一致性) 和变体 4 (缺少网络间一致性) 则出现细节模糊或局部光照伪影, 说明无监督蒸馏策略在提升泛化性和图像自然度方面起到了关键作用.

表 2 消融实验结果

方法	FFHQ数据集		CelebV-Low数据集	
	PSNR (dB)	SSIM	NIQE	BRISQUE
变体1	22.58	0.69	4.23	40.07
变体2	22.99	0.72	3.92	39.08
变体3	23.73	0.74	3.44	32.59
变体4	23.51	0.73	3.67	36.13
变体5	21.97	0.69	4.07	38.97
变体6	22.53	0.68	3.71	36.29
变体7	23.08	0.73	3.47	32.86
本文方法	24.09	0.75	3.36	31.79

注: 加粗数据表示最佳结果

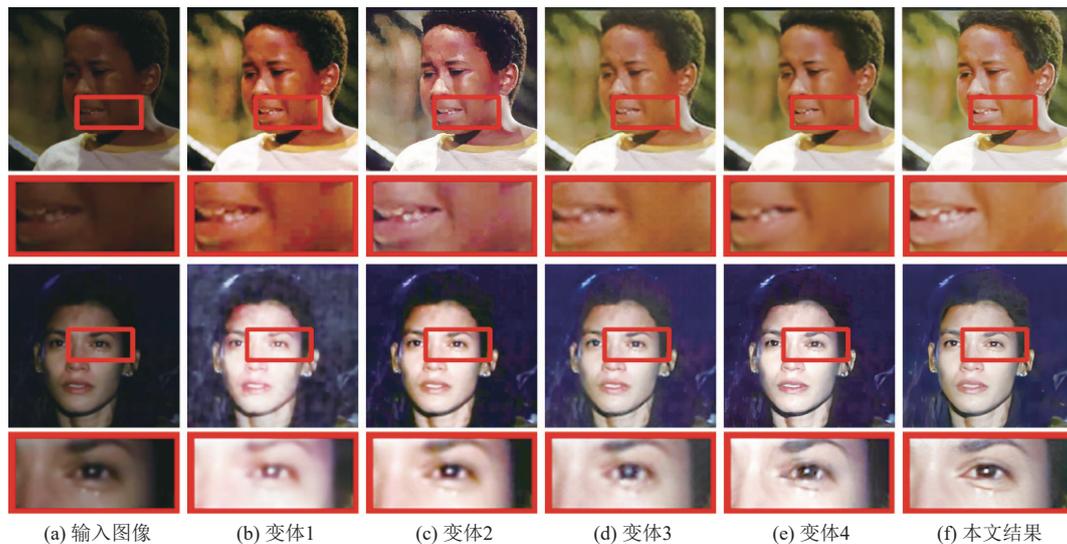


图 6 消融实验的可视化结果 (变体 1-4)

图 7 展示了变体 5-7 在合成低光照图像上的可视化结果. 变体 5 在缺少三维先验引导时, 会出现明显的噪声放大和结构失真, 表明三维几何信息在噪声去除与结构保持中不可或缺. 在变体 6 中, 由于普通 U-Net 难以准确预测三平面特征, 其生成的三维先验质量明显下降, 导致增强结果存在噪声残留. 图 8 中也可看到变体 6 所生成的三维渲染图与深度图存在较强噪声与细节缺失. 变体 7 采用直接拼接融合, 未能有效利用三

维先验的结构信息, 出现局部细节缺失. 综上所述, 无论从定量指标还是视觉结果来看, 完整方法均取得最佳表现. 知识蒸馏显著提升了模型的泛化能力与光照一致性, 而高质量三维人脸先验及其有效融合机制对结构恢复与细节增强发挥了重要作用.

4.4 更多讨论

为更全面评估本文方法的计算复杂度与资源消耗, 本文对比了多类代表性的低光照图像增强方法在参数

量、FLOPs 和显存占用方面的结果,如表 3 所示.对比方法包括基于 CNN 的方法 (EnlightenGAN^[9])、基于 Transformer 的方法 (UHDformer^[6])、零样本方法 (ZERO-

IG^[11]和 Low-FaceNet^[16]) 以及基于扩散模型的方法 (RDDM^[7]和 IC-FSRDENet^[20]).所有实验均在 NVIDIA RTX 3090 GPU 上进行,输入分辨率为 256×256.

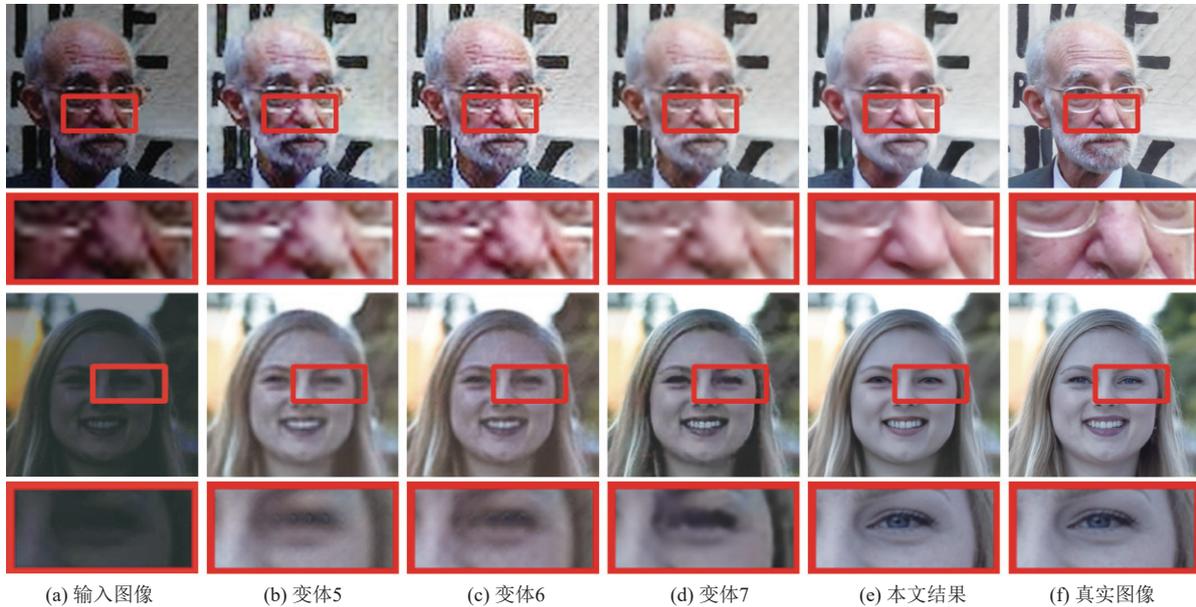


图 7 消融实验的可视化结果 (变体 5-7)

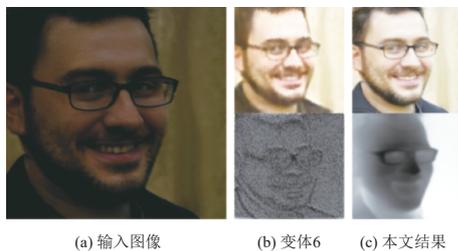


图 8 三维先验的可视化结果

表 3 计算复杂度结果

方法	参数量 (M)	FLOPs (G)	显存占用 (MB)
EnlightenGAN ^[9]	114.35	61.01	230.71
UHDformer ^[6]	13.35	164.69	932.33
ZERO-IG ^[11]	0.12	32.42	166.73
Low-FaceNet ^[16]	0.46	30.76	182.82
RDDM ^[7]	15.47	658.74	480.11
IC-FSRDENet ^[20]	19.67	9332.43	494.38
本文方法	9.29	411.86	225.36

尽管本方法在三平面特征空间中进行扩散建模,整体计算资源需求相对较高,但其复杂度仍低于现有基于扩散模型的方法.从结果可以看出,轻量级零样本方法具有更低的参数量与 FLOPs,但是此类方法在光照一致性和人脸细节恢复方面存在明显不足.与基于

CNN 和 Transformer 的方法相比,我们的方法由于扩散模型逐步去噪的推理机制, FLOPs 高于单次前向网络.然而在模型参数量与显存占用方面,本文方法更具优势.与基于扩散模型的方法相比,本方法在计算效率与显存占用上均表现更优.此外,我们的方法在增强质量方面表现优异,尤其在结构保真度和光照一致性方面显著优于现有方法.这些结果表明,本文方法在保持高质量增强效果的同时,兼具良好的计算可控性.

5 总结

本文提出了一种基于知识蒸馏与三维先验引导的半监督低光照人脸图像增强方法.该方法首先构建了基于扩散模型的三维先验提取网络以生成高质量的三维人脸先验.随后,提出了基于知识蒸馏的半监督图像增强网络,提升了增强结果的质量与稳定性.所提出的三维先验融合模块有效增强了人脸空间结构的感知能力,而无监督知识蒸馏策略则提高了在真实场景下的泛化性能.实验结果表明所提方法在低光照人脸图像增强方面的可行性和有效性.在未来工作中将继续优化该方法将其应用扩展到动态视频领域,并探索其在更广泛视觉任务中的应用潜力.

参考文献

- 1 吕军, 苒婉婷, 陈付龙, 等. 优化双线性 ResNet34 的人脸表情识别. 计算机系统应用, 2024, 33(11): 27–37. [doi: 10.15888/j.cnki.csa.009682]
- 2 陈公冠, 张帆, 王桦, 等. 区域增强型注意力网络下的人脸表情识别. 计算机辅助设计与图形学学报, 2024, 36(1): 152–160.
- 3 赵娅, 郜明超, 姚文达, 等. 基于深度学习的伪造人脸检测技术综述. 计算机系统应用, 2025, 34(4): 1–17. [doi: 10.15888/j.cnki.csa.009839]
- 4 刘本航, 王京京, 母亚双. 基于改进 YOLOv10 的多人脸检测算法. 计算机应用, 2025, 45(S1): 78–83.
- 5 Lore KG, Akintayo A, Sarkar S. LLNet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognition, 2017, 61: 650–662. [doi: 10.1016/j.patcog.2016.06.008]
- 6 Wang C, Pan JS, Wang W, *et al.* Correlation matching transformation Transformers for UHD image restoration. Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver: AAAI Press, 2024. 5336–5344.
- 7 Liu JW, Wang Q, Fan HJ, *et al.* Residual denoising diffusion models. Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024. 2773–2783.
- 8 Yan QS, Feng YX, Zhang C, *et al.* HVI: A new color space for low-light image enhancement. Proceedings of the 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2025. 5678–5687.
- 9 Jiang YF, Gong XY, Liu D, *et al.* EnlightenGAN: Deep light enhancement without paired supervision. IEEE Transactions on Image Processing, 2021, 30: 2340–2349. [doi: 10.1109/TIP.2021.3051462]
- 10 Yang SZ, Ding MX, Wu YM, *et al.* Implicit neural representation for cooperative low-light image enhancement. Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023. 12872–12881.
- 11 Shi YQ, Liu D, Zhang LG, *et al.* ZERO-IG: Zero-shot illumination-guided joint denoising and adaptive enhancement for low-light images. Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024. 3015–3024.
- 12 Zhang L, Li ZY, Cheng L, *et al.* DLIENet: A lightweight low-light image enhancement network via knowledge distillation. Pattern Recognition, 2026, 169: 111777. [doi: 10.1016/j.patcog.2025.111777]
- 13 Zhang YH, Lee S, Yao A. Pairwise distance distillation for unsupervised real-world image super-resolution. Proceedings of the 18th European Conference on Computer Vision. Milan: Springer, 2025. 429–446.
- 14 Barraco M, Stefanini M, Cornia M, *et al.* CaMEL: Mean teacher learning for image captioning. Proceedings of the 26th International Conference on Pattern Recognition (ICPR). Montreal: IEEE, 2022. 4087–4094.
- 15 Chen ZH, Zhu L, Wan L, *et al.* A multi-task mean teacher for semi-supervised shadow detection. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 5610–5619.
- 16 Fan YH, Wang YZ, Liang D, *et al.* Low-FaceNet: Face recognition-driven low-light image enhancement. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 5019413.
- 17 Guo XJ, Li Y, Ling HB. LIME: Low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing, 2017, 26(2): 982–993. [doi: 10.1109/TIP.2016.2639450]
- 18 Singh H, Kumar A, Balyan LK, *et al.* A novel optimally gamma corrected intensity span maximization approach for dark image enhancement. Proceedings of the 22nd International Conference on Digital Signal Processing (DSP). London: IEEE, 2017. 1–5.
- 19 Le HA, Kakadiaris IA. SeLENet: A semi-supervised low light face enhancement method for mobile face unlock. Proceedings of the 2019 International Conference on Biometrics (ICB). Crete: IEEE, 2019. 1–8.
- 20 Wang CY, Jiang JJ, Jiang K, *et al.* Low-light face super-resolution via illumination, structure, and texture associated representation. Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver: AAAI Press, 2024. 5318–5326.
- 21 Pouyanfar S, Sengupta S, Mohammadi M, *et al.* FRR-Net: A real-time blind face restoration and relighting network. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Vancouver: IEEE, 2023. 1240–1250.
- 22 Lan YS, Meng XY, Yang S, *et al.* Self-supervised geometry-aware encoder for style-based 3D GAN inversion. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 20940–20949.
- 23 Dhariwal P, Nichol A. Diffusion models beat GANs on image synthesis. Proceedings of the 35th International Conference on Neural Information Processing Systems.

- Curran Associates Inc., 2021. 672.
- 24 Chan ER, Lin CZ, Chan MA, *et al.* Efficient geometry-aware 3D generative adversarial networks. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 16102–16112.
- 25 Trevithick A, Chan M, Stengel M, *et al.* Real-time radiance fields for single-image portrait view synthesis. ACM Transactions on Graphics (TOG), 2023, 42(4): 135.
- 26 Nichol AQ, Dhariwal P. Improved denoising diffusion probabilistic models. Proceedings of the 38th International Conference on Machine Learning. PMLR, 2021. 8162–8171.
- 27 Karras T, Laine S, Aila T. A style-based generator architecture for generative adversarial networks. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 4396–4405.
- 28 Zhu H, Wu W, Zhu WT, *et al.* CelebV-HQ: A large-scale video facial attributes dataset. Proceedings of the 17th European Conference on Computer Vision. Tel Aviv: Springer, 2022. 650–667.
- 29 Johnson J, Alahi A, Li FF. Perceptual losses for real-time style transfer and super-resolution. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 694–711.
- 30 Dziembowski A, Nowak W, Stankowski J. IV-SSIM—The structural similarity metric for immersive video. Applied Sciences, 2024, 14(16): 7090. [doi: [10.3390/app14167090](https://doi.org/10.3390/app14167090)]
- 31 Mentzer F, Toderici G, Tschannen M, *et al.* High-fidelity generative image compression. Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020. 999.
- 32 Wu LY, Zhang XG, Chen H, *et al.* VP-NIQE: An opinion-unaware visual perception natural image quality evaluator. Neurocomputing, 2021, 463: 17–28. [doi: [10.1016/j.neucom.2021.08.048](https://doi.org/10.1016/j.neucom.2021.08.048)]
- 33 Yang SD, Wu TH, Shi SW, *et al.* MANIQA: Multi-dimension attention network for no-reference image quality assessment. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans: IEEE, 2022. 1190–1199.

(校对责编: 张重毅)