文章编号:1007－6735(2025)01－0089－11 DOI: 10.13255/j.cnki.jusst.20231017003

改进多层级特征损失及全局注意力的三维人脸重建算法

何亚岚[1]， 魏国亮[2]， 武俊珂[2]
（1. 上海理工大学 理学院，上海 200093；2. 上海理工大学 管理学院，上海 200093）

摘要：针对人脸重建算法在细节重建能力、精度以及遮挡影响方面存在的不足，提出一种改进多层级特征损失及全局注意力的三维人脸重建算法。在输入层添加人脸关键点与遮罩的面部先验信息，引导模型关注人脸的重要区域；设计了全局感知金字塔注意力模块，增强模型对重要特征的关注程度，同时充分融合不同层级的特征信息；提出人脸掩膜一致性损失与结构一致性损失，并设计多层级特征损失对模型进行训练优化，提升算法对遮挡情况的重建稳健性，并使输入图像与重建结果在结构上更趋近于一致，丰富模型的特征表示。实验结果表明，重建出的人脸模型具有更多的细节特征，显著增强了遮挡情况下的面部细节重建效果，大幅提高了现有方法的重建精度与鲁棒性能。

关键词：三维人脸重建；深度学习；人脸建模；三维形变模型
中图分类号：TP 391 **文献标志码：**A

Three-dimensional facial reconstruction algorithm based on enhanced multi-level feature loss and global attention

HE Yalan[1], WEI Guoliang[2], WU Junke[2]
(1. *College of Science, University of Shanghai for Science and Technology, Shanghai 200093, China*; 2. *Business School, University of Shanghai for Science and Technology, Shanghai 200093, China*)

Abstract: Aiming at the shortcomings of facial reconstruction algorithms in terms of detail reconstruction capability, accuracy, and the impact of occlusions, a three-dimensional facial reconstruction algorithm was proposed, incorporating improved multi-level feature loss and global attention. Facial landmarks and facial mask priors were added at the input layer to guide the model to

收稿日期: 2023−10−17
基金项目: 国家自然科学基金资助项目 (62273239)
第一作者: 何亚岚(1999−)，男，硕士研究生．研究方向：三维重建．E-mail：heyalan87@163.com
通信作者: 魏国亮(1973−)，男，教授．研究方向：机器视觉、非线性系统优化、随机系统．E-mail：guoliang.wei@usst.edu.cn

引文格式: 何亚岚，魏国亮，武俊珂．改进多层级特征损失及全局注意力的三维人脸重建算法[J]．上海理工大学学报，2025，47(1)：89-99.
Citation: HE Yalan, WEI Guoliang, WU Junke. Three-dimensional facial reconstruction algorithm based on enhanced multi-level feature loss and global attention[J]. Journal of University of Shanghai for Science and Technology, 2025, 47(1): 89-99.

focus on the important facial regions. The global relation-aware pyramid attention module was designed to enhance the model's attention to important features and effectively integrate feature information from different levels. The face mask consistency loss and structural consistency loss were introduced, and the multi-level feature losses were designed to optimize model training, improve robustness to occlusions, make the input image and the reconstructed result approach each other in terms of structure, and enrich the feature representation of the model. Experimental results demonstrate that the reconstructed facial model exhibites more detailed features, significantly enhances facial detail reconstruction under occlusions, and greatly improves the reconstruction accuracy and robustness of existing methods.

Keywords: *three-dimensional face reconstruction*; *deep learning*; *face modeling*; *three-dimensional morphable mode*

三维人脸重建是计算机视觉领域的重要研究方向。相较于二维人脸，三维人脸能够提供更为真实和准确的人脸表情、姿态和动作信息，在人脸识别[1-2]、虚拟现实[3]、整形美容手术[4]等领域都具有广泛的应用前景。在三维人脸重建领域，基于单张人脸图像的方法不仅无需复杂的设备，而且操作简单，是三维人脸重建领域的研究热点。

单张图像的三维人脸重建问题，本质上是在特定的先验条件下，将二维人脸图像像素对应到三维坐标的问题。早期广泛采用基于形变模型的方法进行三维人脸重建。Blanz等[5]提出了三维人脸形变模型（three-dimensional morphable models, 3DMM），该模型运用主成分分析将三维人脸表征转化为线性模型，通过调节形状参数和纹理参数实现相应人脸的模型重建。由于采用线性基表示以及训练数据类型和数量的限制，3DMM对细节的重建效果并不理想。为了解决上述问题，Blanz等[6]利用光流算法将数据稠密对应，使得人脸顶点保持一一对应。此外，Booth等[7]提出的大规模人脸模型（large scale facial model, LSFM）采用固定拓扑结构作为模版对人脸数据进行非刚性配准，进一步提高了三维人脸模型的细节质量。随着深度学习的快速发展，许多学者将卷积神经网络应用到三维人脸重建领域。Tran等[8]对生成的大量合成数据进行训练，并使用非对称欧几里得损失监督估计系数与真实值之间的误差。Genova等[9]通过将3DMM与编码解码器结构相结合，并引入面部识别网络作为训练的损失函数从而保留个体身份信息。Guo等[10]在轻量级骨干网的基础上提出了一种元联合（meta-joint）优化策略，对3DMM参数进行动态回归，提高了重建速度和精度。

近年来，许多学者开始研究人脸的细节重建。2020年，Yang等[11]提出了一个大规模的三维人脸数据集FaceScape，并通过预测多张位移图进行人脸动态细节的生成。2021年，Feng等[12]提出详细的表情捕捉与动画方法（detailed expression capture and animation, DECA），引入了一致性损失来重建更细节的人脸信息。2021年，Gecer等[13]通过监督身份特征并训练生成对抗网络（generative adversarial network, GAN）重建面部纹理和形状。Wu等[14]利用3D界标和3DMM参数之间的关系设计了SynergyNet算法。2022年，Zielonka等[15]提出的MICA算法在模型中加入身份判别器，使重建后的三维人脸保持其原有的身份信息。2023年，申冲等[16]将人脸细节分为表情相关细节和表情无关细节，并根据两种细节的不同特性分别设计生成网络。

在实际应用场景中，由于人脸姿态变化大且易受遮挡影响，现有算法的鲁棒性相对较弱，难以适应多种场景。此外，现有算法的重建结果也存在着个性化不足的问题，无法准确保留人脸个体特征（如皱纹等），导致细节提取的能力有限。

针对上述问题，本文提出改进多层级特征损失及全局注意力的三维人脸重建算法，能够从单张人脸图像中直接还原出带有细节的三维人脸模型。本文的主要贡献如下：在输入层添加包含人脸关键点与人脸遮罩的面部先验信息模块，引导模型关注人脸的重要区域；设计了全局感知金字塔注意力模块，增强模型对重要特征的关注程度和特征提取能力，并充分融合不同层级的特征信息，保留更多的原始细节信息；提出人脸掩膜一致性损失提升对常见遮挡情况的重建稳健性，提

出结构一致性损失使输入图像与重建结果在结构上更趋近于一致，并设计多层级特征损失对模型进行训练优化，丰富模型的特征表示。最后，本文采用 NoW基准对人脸重建模型进行评估，重建出的人脸模型具有更多的细节特征，显著增强了遮挡情况下的面部细节重建效果，大幅提高了现有方法的重建精度与鲁棒性能。

1　人脸模型相关理论

由于人脸的结构化信息呈现出高度复杂性，很难从单张二维图像还原出相应的三维模型，因此本研究采用了铰接模型和表情学习的人脸模型[17]（faces learned with an articulated model and expressions，FLAME）进行人脸重建。FLAME 模型作为一种通用的三维人脸模型，能够连接颈部、下巴和眼球等部位，从而实现对完整头部和颈部的重建，在重建过程中具有较高的准确性与表现力。

1.1　几何外观模型

FLAME 通过使用线性变换来描述与身份、姿势、表情相关的形状变化，并且包含 5023 个顶点与 4 个关节。FLAME 模型生成的顶点可表示为

$$\boldsymbol{M}(\beta,\theta,\psi)=W(T_{\mathrm{P}}(\beta,\theta,\psi),J(\beta),\theta,\delta) \tag{1}$$

式中：β为身份参数；θ为姿势参数；ψ为表情参数；T_{P}为零姿态下的平均模板T添加了形状、姿势、表情偏移的模型；J为关节；δ为混合权重。

模型通过混合蒙皮函数$W(T,J,\theta,\delta)$围绕关节J旋转T中的顶点，通过混合权重δ进行线性平滑。T_{P}表达式为

$$T_{\mathrm{P}}(\beta,\theta,\psi)=T+B_{\mathrm{S}}(\beta,\Gamma)+B_{\mathrm{P}}(\theta,\Lambda)+B_{\mathrm{E}}(\psi,\Omega) \tag{2}$$

式中：Γ、Λ、Ω分别为 3 个形变函数的基；B_{S}为形状形变函数；B_{P}为姿势形变函数，B_{E}为表情形变函数。B_{S}、B_{P}、B_{E}分别控制身份、姿势和表情相关的形状变化。

由于 FLAME 缺少外观模型，因此，将巴塞尔人脸模型[2]（Basel face model, BFM）的线性反照率子空间转换到 FLAME 的 UV 空间，并通过纹理参数α生成 UV 反照率图$A(\alpha)$使其与 FLAME 兼容，得到外观模型D_A。

1.2　相机模型

现有人脸数据集中的照片通常采用远距离拍摄，在进行三维网格到二维图像空间的投影时，常使用正交相机模型。人脸顶点投射到图像中的方式用公式表达为

$$v=s\boldsymbol{P}(\boldsymbol{M}_i)+t$$

式中：$\boldsymbol{M}_i$为$\boldsymbol{M}$中的一个顶点；$\boldsymbol{P}$为三维到二维的正交投影矩阵；s和t分别为尺度变换和平移变换，均归为相机参数c。

1.3　光照模型

为符合更加真实的光照情况，本文采用基于球面谐波的照明模型[18]。假设光源是遥远的，人脸表面为朗伯面，具有光照阴影的纹理贴图的计算方法如下：

$$\boldsymbol{F}(\alpha,l,\boldsymbol{N})_{i,j}=A(\alpha)_{i,j}\odot\sum_{k=1}^{9}l_kH_k(\boldsymbol{N}_{i,j}) \tag{3}$$

式中：$A(\alpha)$为反照率图；$\boldsymbol{N}$为表面法向量；$\boldsymbol{N}_{i,j}$为像素点 (i,j) 对应的表面法向量；$\boldsymbol{F}$ 为阴影纹理，以 UV 坐标表示；H_k为球谐函数的基；l为光照系数；$\odot$为哈达玛德积。

给定几何参数$\beta\theta$、ψ，纹理参数α，相机参数c，光照系数l，经过可微分渲染器E_{c}得到渲染后的二维图像I_{r}为

$$I_{\mathrm{r}}=\mathcal{R}(\boldsymbol{M},\boldsymbol{F}(\alpha,l,\boldsymbol{N}_{uv}),c) \tag{4}$$

式中：$\mathcal{R}$为渲染函数；$\boldsymbol{N}_{uv}$为 $\boldsymbol{N}$ 在 UV 空间上进行转换的结果。

将$\boldsymbol{M}$及其表面法线$\boldsymbol{N}$在 UV 空间上进行转换，然后通过解码器得到$\boldsymbol{M}'_{uv}$和$\boldsymbol{N}'_{uv}$，再进行细节渲染得到：

$$I_{\mathrm{r}}'=\mathcal{R}(\boldsymbol{M},\boldsymbol{F}(\alpha,l,\boldsymbol{N}'_{uv}),c) \tag{5}$$

2　改进多层级特征损失及全局注意力的三维人脸重建算法

本文构建了一个改进多层级特征损失及全局注意力的人脸图像三维重建网络。该网络以 ResNet50[19] 作为主干网络，增加人脸先验信息模块与全局关系感知金字塔注意力模块，并提出包含人脸掩膜一致性损失和结构一致性损失的多层级特征损失来训练优化网络。

2.1　先验信息模块

人脸图像经常包含环境背景，同时携带其他非人脸特征的冗余信息（如帽子、眼镜、头发等），这增加了精确提取人脸的难度。高质量的人脸提取对于人脸重建至关重要。为了进一步优化

模型对人脸的准确提取和重建能力，本研究利用人脸关键点和人脸遮罩作为先验信息模块（prior information module，PIM）以协助网络获取更准确的人脸特征信息。

使用人脸关键点和人脸遮罩作为先验信息能够引导模型关注人脸的重要区域，降低冗余信息对提取过程的干扰。人脸关键点提供了面部重要结构的位置信息，而人脸遮罩则确定了人脸区域和非人脸区域的边界。这种先验信息有助于网络更好地理解人脸的结构和特征，从而实现更精准、高效的人脸提取和重建。本文采用了 Bulat 等[20]提出的关键点检测算法以及 Nirkin 等[21]提出的人脸分割算法，用以提取关键点区域（包括面部轮廓、眼镜、鼻子等）的 68 个二维人脸关键点，并获得相应的人脸分割掩码（非人脸区域的权重设为0），最后将人脸图像与人脸先验信息送入网络进行训练。

2.2 全局关系感知金字塔注意力模块

在实际场景中，由于人脸姿态变化较大，并且存在物品遮挡面部的问题，使得面部无法完整地进行重建。本文设计全局关系感知金字塔注意力模块（global relation-aware pyramid attention module，GRPAM），在 ResNet50 的 4 个 Block 之后添加基于空间和基于通道维度的 RGA 注意力机制[22]，对 Conv2、Conv3、Conv4、Conv5 层中提取最后一个残差 block 层的特征，采用上采样和相加的方法进行不同层级的特征融合，最后进行 3×3的卷积操作生成特征图，结构如图 1 所示。图中：ξ表示步长；{$C2$, $C3$, $C4$, $C5$}表示特征金字塔的 4 个层级；{$P2$, $P3$, $P4$, $P5$}表示全局关系感知金字塔注意力模块的 4 个层级特征；{RGA–S1, RGA–S2, RGA–S3, RGA–S4}表示基于空间维度的注意力机制；{RGA–C1, RGA–C2, RGA–C3, RGA–C4}表示基于通道维度的注意力机制。

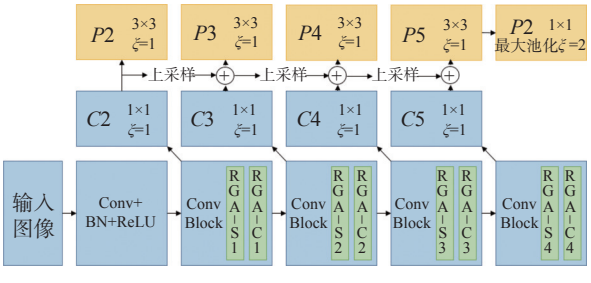


图 1 全局关系感知金字塔注意力模块

Fig.1 Global relation-aware pyramid attention module

注意力通常使用局部卷积进行学习，而局部卷积会忽略全局信息和隐藏信息的关系。对于每一个特征节点，结合该位置与所有其他特征位置之间的关系（即成对相关性），以及该特征本身的信息，紧凑地表示该节点的全局结构信息，再进行注意力卷积运算。

如图 2 所示，首先将特征本身与关系进行融合，对于每个特征节点$x_i(i=1,\cdots,N)$，全局结构信息为

$$y_i=[x_i,\boldsymbol{r}_i] \tag{6}$$

式中，$\boldsymbol{r}_i$为节点x_i与其他节点间的关系向量。

$$\boldsymbol{r}_i=[r_{i,1},\cdots,r_{i,N},r_{1,i},\cdots,r_{N,i}] \tag{7}$$

$a_i(i=1,\cdots,N)$表示节点x_i的权重，根据特征的重要性进行加权。

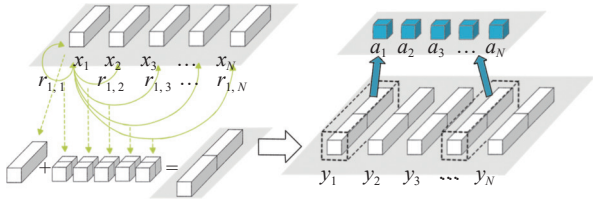


图 2 关系感知结构图

Fig.2 Structure diagram of relation-aware module

根据 Zhang 等[22]的工作，基于空间维度的 RGA-S 注意力机制的成对相关性如下：

$$r_{i,k}=f_{\mathrm{s}}(x_i,x_k)=\boldsymbol{\tau}_{\mathrm{s}}(x_i)^{\mathrm{T}}\boldsymbol{\varsigma}_{\mathrm{s}}(x_k) \tag{8}$$

式中，$\boldsymbol{\tau}_{\mathrm{s}}(x_i)$和$\boldsymbol{\varsigma}_{\mathrm{s}}(x_k)$分别为由1×1空间卷积批量归一化和 ReLU 激活实现的嵌入函数。

基于通道维度的 RGA-C 注意力机制的成对相关性如下：

$$r_{i,k}=f_{\mathrm{c}}(x_i,x_k)=\boldsymbol{\tau}_{\mathrm{c}}(x_i)^{\mathrm{T}}\boldsymbol{\varsigma}_{\mathrm{c}}(x_k) \tag{9}$$

与空间关系类似，$\boldsymbol{\tau}_{\mathrm{c}}(x_i)$和$\boldsymbol{\varsigma}_{\mathrm{c}}(x_k)$是由1×1空间卷积批量归一化和 ReLU 激活实现的两个嵌入函数。

为了使编码器能够获取更丰富的信息，融合多层特征信息并在不同的特征层进行输出，提取网络中 Conv2、Conv3、Conv4、Conv5 层最后一个残差 block 层的特征。利用特征金字塔[23]进行融合，从而增加模型对细节特征的恢复能力，进一步增强人脸细节重建效果。

具体而言，在自底向上结构中，每个阶段对应全局关系感知金字塔注意力模块的一个层级。从 Conv2、Conv3、Conv4、Conv5 层中提取最后一个残差 block 层的特征形成金字塔的 4 个层级{$C2$，$C3$，$C4$，$C5$}。如图 3 所示，为了实现不同

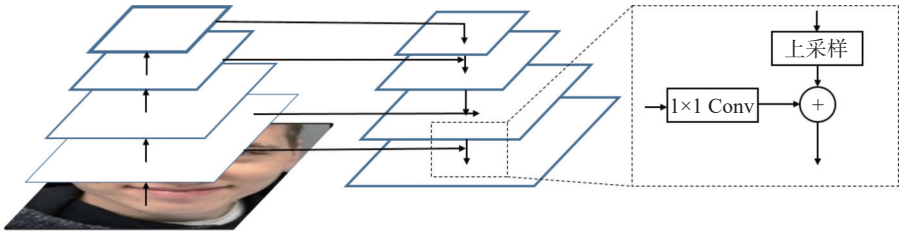


图 3　特征融合结构图

Fig.3　Structure diagram of feature fusion

层级特征的融合，采用上采样和相加的方法，上一层级的特征经过上采样后与当前层级分辨率一致的特征通过相加的方式进行融合。

为了保持所有层级特征的通道数一致，使用1×1的卷积核对特征进行卷积，得到全局关系感知金字塔注意力模块的 4 个层级特征{$P2$，$P3$，$P4$，$P5$}。最后，将每个层级特征进行一个 3×3 的卷积操作生成特征图。

2.3　粗略重建过程

本文算法首先在 FLAME 的模型空间中进行粗略重建：给定二维图像I作为输入，将图像编码为潜在代码，对其进行解码以合成渲染后的图像I_{r}。

为了最小化生成图像与输入图像之间的差异，丰富模型的特征表示，本文设计多层级的损失函数，通过最小化粗略重建损失函数L_{coarse}进行模型的训练，使模型能够更好地理解不同层次的特征，提高模型性能。粗略重建损失函数定义为

$$L_{\mathrm{coarse}} = L_{\mathrm{low}} + L_{\mathrm{mid}} + L_{\mathrm{deep}} + L_{\mathrm{reg}} \tag{10}$$

式中：L_{low}为浅层特征损失；L_{mid}为中层特征损失；L_{deep}为深层特征损失；L_{reg}为正则化损失。

2.3.1　浅层特征损失

为了使输入图像与渲染图像尽可能相近，设计浅层特征损失，包含像素损失与人脸掩膜一致性损失，即

$$L_{\mathrm{low}} = L_{\mathrm{pix}} + L_{\mathrm{mask}}$$

本文提出了人脸掩膜一致性损失L_{mask}（face mask consistency loss，FMCL），通过利用人脸掩膜信息，计算渲染图像I_{r}与输入图像I之间的损失，使模型仅关注面部区域的误差，减少背景和其他无关区域的影响，避免将人脸无关的遮挡信息带入损失计算中。同时提高人脸部位的重要性，对人脸的面部区域进行精细化约束，从而提高模型对人脸局部重建的准确性，以及对常见遮挡的稳健性，提高模型重建的准确性。人脸掩膜一致性损失定义为

$$L_{\mathrm{mask}} = \frac{\sum \rho_i \| I - I_{\mathrm{r}} \|_2}{\sum \rho_i} \tag{11}$$

式中，ρ_i为权重，如果像素在人脸内部则$\rho_i = 1$，在人脸外部则$\rho_i = 0$。

像素损失L_{pix}使用均方误差度量输入图像与渲染结果之间的像素差异：

$$L_{\mathrm{pix}} = \frac{1}{CHW \sum (I - I_{\mathrm{r}})^2} \tag{12}$$

式中，C、H和W分别为图像的通道数、高度和宽度。

2.3.2　中层特征损失

本文将关键点映射损失L_{kp}以及闭眼损失L_{eye}作为中层特征损失，$L_{\mathrm{mid}} = L_{\mathrm{kp}} + L_{\mathrm{eye}}$，引导网络模型参数的学习。

关键点映射损失能够衡量真实二维人脸关键点k_i与 FLAME 模型表面顶点$\boldsymbol{M}_i$上相应关键点之间的差异，由估计的相机模型投影到图像中。关键点映射损失定义为

$$L_{\mathrm{kp}} = \sum_{i=1}^{68} \| k_i - s\boldsymbol{P}(\boldsymbol{M}_i) - t \|_1 \tag{13}$$

闭眼损失计算上眼睑和下眼睑上的关键点k_i和k_h的相对偏移量，并测量投影到图像中的 FLAME 表面$\boldsymbol{M}_i$和$\boldsymbol{M}_h$上相应点偏移量的差异。闭眼损失定义为

$$L_{\mathrm{eye}} = \sum_{(i,h)\in G} \| k_i - k_h - s\boldsymbol{P}(\boldsymbol{M}_i - \boldsymbol{M}_h) \|_1 \tag{14}$$

式中，G为上下眼睑关键点的集合。

由于闭眼损失是平移不变的，因此与关键点映射损失相比更不容易受到投影后的三维人脸与图像未对准的影响。

2.3.3　深层特征损失

本文将新提出的结构一致性损失L_{SCL}、身份

损失L_{id}、细节一致性损失[11]L_{sc}作为深层特征损失，$L_{deep}=L_{SCL}+L_{id}+L_{sc}$，进一步约束网络模型的参数。

a. 结构一致性损失L_{SCL}。受结构相似性指数[24]的启发，本文提出的结构一致性损失（structural consistency loss，SCL），利用亮度、对比度和结构在输入图像与渲染图像的人脸部分计算损失，使输入的人脸图像与渲染结果更趋近于一致。同时，更加关注重建人脸部分的图像结构特征，优先保留与输入人脸图像相似的结构，使模型受到更强的结构保持约束，从而提高模型重建的准确性。结构一致性损失的定义为

$$L_{SCL}=1-\frac{2\mu(\chi_I)\mu(\chi_{I_r})+\eta_1}{\mu^2(\chi_I)+\mu^2(\chi_{I_r})+\eta_1}\cdot\frac{2\varsigma(\chi_I,\chi_{I_r})+\eta_2}{\varsigma^2(\chi_I)+\varsigma^2(\chi_{I_r})+\eta_2} \tag{15}$$

式中：χ为利用人脸遮罩信息选取对应图像的人脸部分像素；μ、ς^2、ς分别为相应图像的亮度、对比度和结构；η_1和η_2为稳定计算的常数。

b. 身份损失。使用训练好的人脸识别网络f[25]测量输出的渲染图像与输入图像的特征向量之间的余弦相似度，能够保留更多的身份信息。身份损失定义如下：

$$L_{id}=1-\frac{f(I)f(I_r)}{\|f(I)\|_2\cdot\|f(I_r)\|_2} \tag{16}$$

c. 细节一致性损失。根据 Feng 等[12]的工作，细节一致性损失函数能够确保同一身份的多张图像具有相同的身份参数，使渲染的图像更加真实。细节一致性损失定义如下：

$$L_{sc}=L_{coarse}(I_i,\mathcal{R}(\boldsymbol{M}(\beta_j,\theta_i,\psi_i),\boldsymbol{F}(\alpha_i,l_i,N_{uv,i}),c_i)) \tag{17}$$

式中，i、j为同一个人的两个输入图像的参数。

d. 正则化损失。为了防止三维人脸重建时出现形状的退化，分别对身份参数、表情参数和纹理参数进行正则化约束。具体表示为

$$L_{reg}=\omega_{id}\|\beta\|_2^2+\omega_{exp}\|\psi\|_2^2+\omega_{tex}\|\alpha\|_2^2 \tag{18}$$

式中，ω_{id}、ω_{exp}、ω_{tex}分别为各项的权重。

2.4 精细重建过程

为了进一步提升模型的人脸细节重建能力，在粗略重建的基础上进行精细重建，增加细节的*UV*位移图。与粗略重建类似，训练编码器E_d，得到128维的细节参数δ，然后将其与粗略重建中的参数送入解码网络F_d中得到*UV*位移图。通过最小化精细重建损失函数L_{detail}进行模型的训练：

$$L_{detail}=L_{mrf}+L_{sym}+L_{dc}+L_{regD} \tag{19}$$

式中：L_{mrf}为感知损失；L_{regD}为细节正则化损失；L_{sym}为软对称损失；L_{dc}为细节一致性损失。

由 Wang 等[26]的工作，感知损失为

$$L_{mrf}=2L_M(\Theta_{4,2})+L_M(\Theta_{3,2})$$

式中：L_M为将 ID-MRF 损失应用于从$\Theta_{3,2}$和$\Theta_{4,2}$中提取的I_r'和I的特征块；$\Theta_{4,2}$为网络中第4个阶段的第2个卷积层；$\Theta_{3,2}$为网络中第3个阶段的第2个卷积层。

通过添加软对称损失修正不可见的面部区域：

$$L_{sym}=\|V_{uv}\odot(D-\mathrm{flip}(D))\|_{1,1} \tag{20}$$

式中：V_{uv}为*UV*空间中的面部遮罩；flip为水平翻转操作；D为位移图。

与粗略重建类似，根据 Feng 等[12]的工作，细节一致性损失能进一步增强细节提取能力，即

$$L_{dc}=L_{detail}(I_i,\mathcal{R}(\boldsymbol{M}(\beta_j,\theta_i,\psi_i),A(\alpha_i),\boldsymbol{F}_d(\delta_j,\psi_i,\theta_{jaw,i}),l_i,c_i)) \tag{21}$$

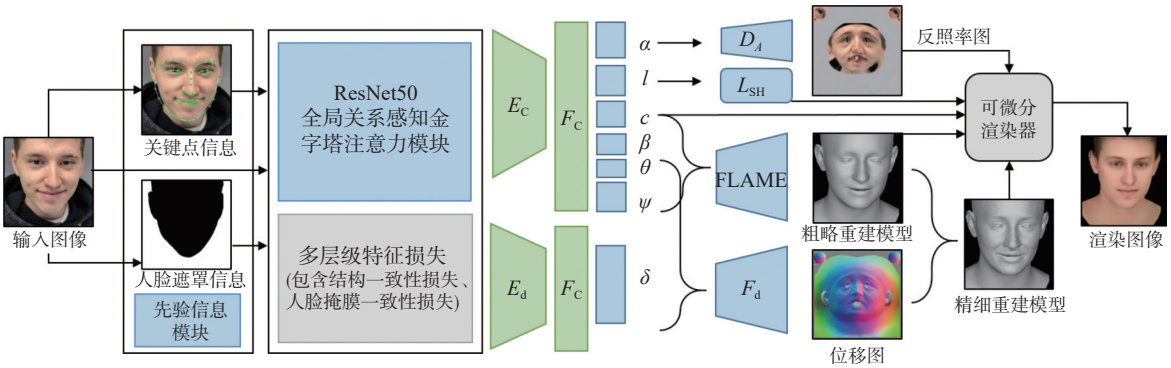
细节正则化损失$L_{regD}=\|D\|_{1,1}$，对细节位移进行正则化以减少噪声，θ_{jaw}为下巴姿势参数。

2.5 算法总结

本文算法以二维人脸图像作为输入，利用先验信息模块得到人脸特征点信息和面部遮罩信息，再将全局关系感知金字塔注意力模块融合至ResNet50主干网络。通过编码器得到潜在代码，再由解码器将潜在代码合成为渲染图像，并通过包含FMCL与SCL的多层级特征损失来优化网络。本文的重建过程分为粗略重建与精细重建。

a. 粗略重建：特征编码器E_C由融合全局关系感知金字塔注意力模块的ResNet50网络和全连接层F_C组成，通过多层级特征损失进行训练得到低维潜在代码。该潜在代码由FLAME的参数β、ψ、θ，反照率系数α，相机参数c和照明参数l组成。

b. 精细重建：通过*UV*位移图增强FLAME几何形状的细节表达。与粗略重建类似，精细重建编码器E_d将输入图像编码为128维潜在代码δ。然后潜在代码δ与FLAME的表情ψ和下巴姿势参数θ_{jaw}连接，并由F_d解码为D，最后得到三维人脸模型。算法具体流程如图4所示。



E_C 粗略重建编码器； E_d 精细重建编码器； F_C 全连接层； α 纹理参数； c 相机参数； D_A 外观模型；
L_{SH} 光照模型； δ 细节参数； F_d 细节解码器； l 光照参数； β 身份参数； θ 位姿参数； ψ 表情参数

图 4　算法流程图

Fig.4　Algorithm flowchart

3　实验结果及分析

为了更好地评估本文的重建结果，采用定性和定量两种方式验证所提算法的有效性。针对重建结果的定性评估，对 Aflw2000 数据集[27] 和 NoW 数据集[28] 中的部分结果进行可视化展示。为了进一步验证算法的性能，采用 NoW 基准进行度量与非度量的定量评估。

3.1　实验设置

3.1.1　数据集

本文旨在进行基于单张图像的人脸三维重建研究，为保证数据质量与多样性，需要采用每个身份具有多张图像的人脸数据集，因此，使用 VGGFace2 数据集[25] 及 BUPT-Balancedface 数据集[29] 作为训练集。

VGGFace2 数据集综合考虑了姿势、年龄、照明、职业等因素，选取了来自 9131 名受试者的约 331 万张图像。然而该数据集主要涵盖白种人群，为了平衡各种族人脸数据的分布，采用 BUPT-Balancedface 数据集。该数据集包含来自亚洲人、印度人、非洲人及白种人的约 130 万张图像。

3.1.2　实验细节

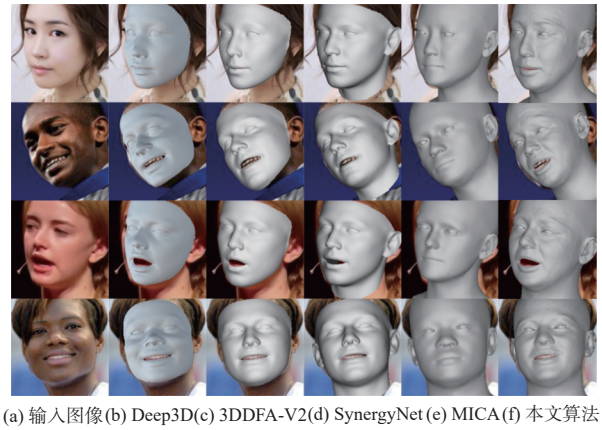
实验的硬件环境为 Intel®Core i9-10920X CPU，NVIDIA RTX 3090 GPU，软件环境为 Ubuntu20.04，算法实现编程语言为 Python 3.7，基于 PyTorch 架构进行开发。为了优化算法的训练过程，选用 Adam 作为优化器，并将初始学习率设置为 0.0001，batch 大小设置为 18，使用 PyTorch3D 作为可微分渲染器。为了进一步提升训练速度，使用 CUDA 和 cuDNN 对 GPU 模型的学习进行加速。

3.2　实验结果及分析

3.2.1　定性分析

为验证本文算法的细节重建能力和鲁棒性，本研究选择了多个先进的人脸重建算法进行对比。由图 5 可知，相较于 MICA[15]、SynergyNet[14]、3DDFA-V2[10] 等方法，本文算法能够更完整地重建头部结构，包括人脸形状和面部表情。同时，该算法对重要部位（眼部、鼻部、唇部等）重建出了更多的细节特征，且重建结果更完整地覆盖了输入图像的面部区域。由此可知，全局关系感知金字塔注意力模块和人脸先验信息模块是有效的。此外，由实验结果可以看出，模型对眼周的皱纹、法令纹等细节特征具有更高的表现力，说明了多层级特征损失能够丰富模型的特征表示。图 6 展示了本文算法在细节重建方面的效果。

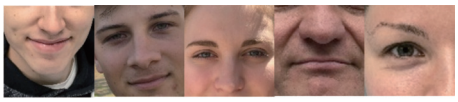
由图 7 可知，本文模型在眼镜、头发以及其他物品遮挡情况下的重建结果表现得更加真实且自然。此外，如图 8 所示，针对同一身份在不同



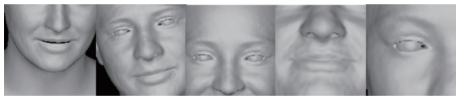
(a) 输入图像 (b) Deep3D (c) 3DDFA-V2 (d) SynergyNet (e) MICA (f) 本文算法

图 5　不同算法的重建效果对比

Fig.5　Comparison of reconstruction results of different algorithms



(a) 细节部分特写

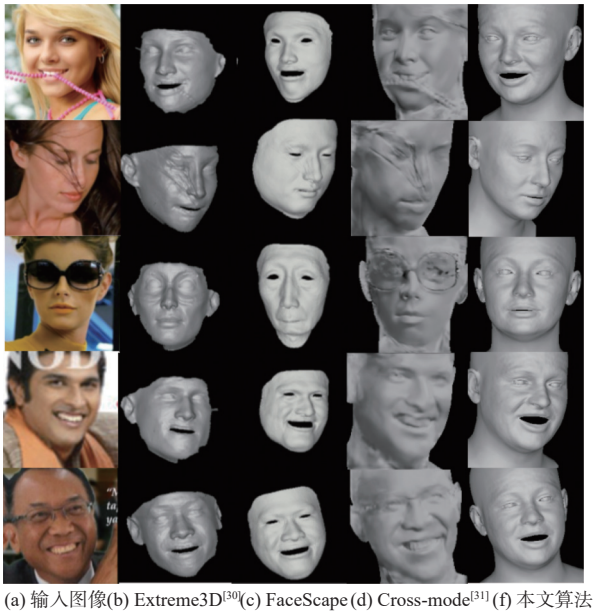


(b) 细节重建部分特写

图 6 细节重建效果

Fig.6 Detail reconstruction results

姿势、光照、表情和遮挡下的情况，重建结果与输入图像的结构更加相似。综上，本文算法不仅具备对人脸细节进行重建的能力，而且增强了对于遮挡情况以及不同光照等变化的鲁棒性。



(a) 输入图像(b) Extreme3D[30](c) FaceScape (d) Cross-mode[31] (f) 本文算法

图 7 不同算法在遮挡情况下的重建效果对比

Fig.7 Comparison of reconstruction performance of different algorithms under occlusion

图 8 同一身份在不同姿势、光照、表情、遮挡下重建效果

Fig.8 Reconstruction results of the same identity under different poses, illumination, expressions and occlusion

3.2.2 定量分析

NoW 基准[28] 引入了一个标准评估指标，用于衡量人脸图像重建算法在视角、照明和常见光照遮蔽下的准确性与鲁棒性。该基准数据集由 100 名受试者的 2054 张人脸图像组成，分为验证集(20 名受试者)和测试集(80 名受试者)。每个受试者都进行了 3D 面部扫描，图像包括室内和室外图像、中性表情和富有表现力的面部图像、部分遮挡(眼镜、头发、帽子等)的面部以及从正面视图到侧面视图的不同视角，以及自拍图像。NoW 基准使用真实扫描与界标对准，将初始化重建网格之间的扫描网格距离(即每个扫描顶点和网格表面中最近点之间的绝对距离)作为误差进行计算。其中：度量评估是指使用 3D 扫描仪对人脸进行扫描，然后与重建的人脸网格进行比较，计算重建误差；非度量评估是指使用刚性对齐的方法，将重建的人脸网格与参考扫描进行比较，再计算误差。

表 1 与表 2 分别展示了不同算法在 NoW 基准上的度量重建误差和非度量重建误差，从中可看出，对比最新的 FOCUS[32]、SynergyNet[14]、Wood 等[33]、DECA 方法[12] 的度量误差，本文算法的重建精度分别提升了 12.05%，45.6%，8.8%，8.14%。对于非度量误差，相比 SynergyNet[14]、Dib 等[34]、申冲等[16]，本文算法的重建精度分别提升了 11.8%，11.1%，6.25%。

表 1 NoW 基准下不同方法的度量误差

Tab.1 Metrical reconstruction error of different methods on NoW benchmark mm

方法	中位数	平均值	标准差
3DMM-CNN[8](Tran 等, 2017)	3.91	4.84	4.02
SynergyNet[14](Wu 等, 2021)	2.28	2.86	2.39
Dib 等[34](2021)	1.59	2.12	1.93
3DDFA-V2[10](Guo 等, 2020)	1.53	2.06	1.95
RingNet[28](Sanyal 等, 2019)	1.50	1.98	1.77
FOCUS[32](Li 等, 2022)	1.41	1.85	1.70
DECA[12](Feng 等, 2021)	1.35	1.80	1.64
Wood 等[33](2022)	1.36	1.73	1.47
本文算法	**1.24**	**1.59**	**1.39**

从累计误差图 9 和图 10 可以看出，相较于其他算法，本文算法给出了最优的效果，具有最低的均值误差、中值误差和标准差误差。这是由于本文模型充分考虑了图像的多个特征层次，同时有效约束了网络的训练过程，使模型在不同的特征层次上进行有效学习，并能够在更高层次和更

表 2　NoW 基准下不同方法的非度量误差

Tab.2　Non-metrical reconstruction error of different methods on NoW benchmark　　mm

方法	中位数	平均值	标准差
3DMM-CNN[8](Tran 等, 2017)	1.84	2.33	2.05
PRNet[35] (Feng 等, 2018)	1.50	1.98	1.88
3DDFA-V2[10](Guo 等, 2020)	1.23	1.57	1.39
RingNet[28](Sanyal 等, 2019)	1.21	1.53	1.31
Deep3D[36](Deng 等, 2019)	1.23	1.54	1.29
SynergyNet[14](Wu 等, 2021)	1.27	1.59	1.31
Dib 等[34](2021)	1.26	1.57	1.31
申冲 等[16](2023)	1.19	1.46	1.22
本文算法	**1.12**	**1.39**	**1.18**

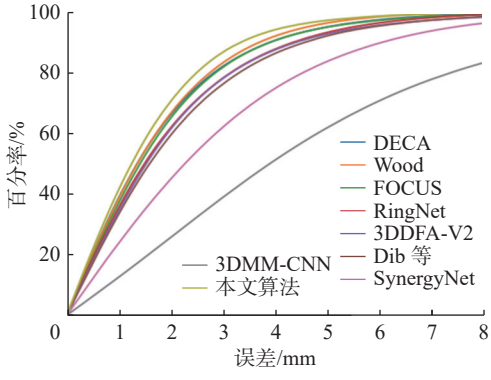


图 9　度量累计误差图

Fig.9　Metrical cumulative error chart

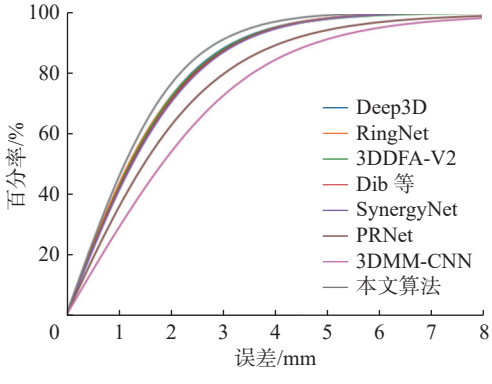


图 10　非度量累计误差图

Fig.10　Non-metrical cumulative error chart

抽象的特征上捕捉人脸图像的细节信息，同时保持对低层次细节的敏感性。

3.2.3　消融实验

为了验证本文新提出的模型，以及人脸掩膜一致性损失、结构一致性损失的有效性，在不同情况下对模型进行训练，使用 NoW 基准的非度量误差得出消融实验结果，如表 3 所示。表中“—”表示不使用这个模块。实验中把 ResNet50 网络与去除人脸掩膜一致性损失与结构一致性损失的多层级特征损失作为基础模型（表 3 实验 a），分别在基础模型上引入面部先验信息模块（PIM）、全局关系感知金字塔注意力模块（GEPAM）、包含人脸掩膜一致性损失与结构一致性损失的多层级特征损失（LOSS）进行实验。

表 3　消融实验结果

Tab.3　Ablation experiment results　　mm

实验	PIM	GEPAM	LOSS	中位数	平均值	标准差
a	—	—	—	1.28	1.64	1.25
b	—	√	—	1.18	1.46	1.25
c	√	—	—	1.25	1.60	1.40
d	—	—	√	1.16	1.46	1.24
e	√	√	—	1.17	1.45	1.24
f	√	—	√	1.15	1.42	1.20
g	—	√	√	1.14	1.44	1.23
h	√	√	√	**1.12**	**1.39**	**1.18**

由表 3 的实验 b 可见，与实验 a 相比，采用 GEPAM 模块能够使中值误差减小 0.10 mm，可见该模块能够提升模型的精度。由实验 c 可见，PIM 模块也能在一定程度上提升模型精度，但相对于 GEPAM 模块提升较小，原因是先验信息模块无法提取出图像更多的特征信息，进一步提升模型精度仍需在网络中提取更多的特征信息。由实验 d 可见，LOSS模块能够降低 0.12 mm 的误差，对模型精度提升贡献最大，原因是包含人脸掩膜一致性损失与结构一致性损失的多层级损失能够有效地对模型训练进行约束。需要注意的是，人脸掩膜一致性损失需要使用人脸掩膜信息，虽然人脸掩膜信息并不十分精准，但由于人脸掩膜能够让模型更加关注人脸部分，这与不使用人脸掩膜信息相比，模型的精度提高是显著的。由实验 e、g、f 可以看出，任意两个模块的结合都有助于提升模型精度。由实验 h 可见，本文算法整合了面部先验信息模块、全局关系感知金字塔注意力模块、包含人脸掩膜一致性损失与结构一致性损失的多层级特征损失，完整模型能使误差降低到最低值 1.12 mm，相较于基础模型误差降低了 0.16 mm，证实了本文所提算法的优越性。

4　结　论

为了提高三维人脸重建算法的准确性、加强

细节重建能力并减小遮挡的影响，本文提出了改进多层级特征损失及全局注意力的三维人脸重建算法。该模型将人脸关键点和遮罩的面部先验信息嵌入 ResNet50 网络模型中，引导模型关注人脸的重要区域。同时，设计了全局关系感知金字塔注意力模块，以增强模型对重要特征的关注程度和特征提取能力，并融合不同层级的特征信息，保留更多的原始细节信息。本文还提出了人脸掩膜一致性损失函数以提升算法对遮挡情况的重建稳健性，提出结构一致性损失函数使输入图像与重建结果在结构上更趋近于一致。此外，设计了多层级特征损失函数，以丰富模型的特征表示。最终通过网络预测的人脸参数生成三维人脸模型，并采用 NoW 基准对本文算法进行定性与定量评估。实验证明，本文的重建人脸模型具有更丰富的细节特征，显著增强了在遮挡情况下的面部细节重建效果，同时极大提高了现有方法的重建准确度与鲁棒性能。未来将进一步深入研究人脸表情重建算法，使模型更加真实自然，并持续提升人脸重建的准确度。

参考文献：

[1] BLANZ V, VETTER T. Face recognition based on fitting a 3D morphable model[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(9): 1063−1074.

[2] PAYSAN P, KNOTHE R, AMBERG B, et al. A 3D face model for pose and illumination invariant face recognition[C]//Proceedings of 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, Genova: IEEE, 2009: 296–301.

[3] 高翔, 黄法秀, 刘春平, 等. 3DMM 与 GAN 结合的实时人脸表情迁移方法 [J]. 计算机应用与软件, 2020, 37(4): 119−126.

[4] 毛爱华, 司徒亨哥. 图像驱动的三维人脸自动生成与编辑算法 [J]. 计算机辅助设计与图形学学报, 2019, 31(1): 17−25.

[5] BLANZ V, ROMDHANI S, VETTER T. Face identification across different poses and illuminations with a 3D morphable model[C]//Proceedings of the Fifth IEEE International Conference on Automatic Face Gesture Recognition. Washington: IEEE, 2002: 202–207.

[6] BLANZ V, SCHERBAUM K, SEIDEL H P. Fitting a morphable model to 3D scans of faces[C]//Proceedings of 2007 IEEE 11th International Conference on Computer Vision. Rio de Janeiro: IEEE, 2007: 1–8.

[7] BOOTH J, ROUSSOS A, PONNIAH A, et al. Large scale 3D morphable models[J]. International Journal of Computer Vision, 2018, 126(2): 233−254.

[8] TRAN A T, HASSNER T, MASI I, et al. Regressing robust and discriminative 3D morphable models with a very deep neural network[C]//Proceedings of 2007 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Honolulu: IEEE, 2017: 1493–1502.

[9] GENOVA K, COLE F, MASCHINOT A, et al. Unsupervised training for 3D morphable model regression[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 8377–8386.

[10] GUO J Z, ZHU X Y, YANG Y, et al. Towards fast, accurate and stable 3D dense face alignment[C]//Proceedings of the 16th European Conference on Computer Vision -ECCV 2020. Glasgow: Springer, 2020: 152–168.

[11] YANG H T ZHU H, WANG Y R, et al. FaceScape: a large-scale high quality 3D face dataset and detailed riggable 3D face prediction[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 598–607.

[12] FENG Y, FENG H W, BLACK M J, et al. Learning an animatable detailed 3D face model from in-the-wild images[J]. ACM Transactions on Graphics (TOG), 2021, 40(4): 88.

[13] GECER B, PLOUMPIS S, KOTSIA I, et al. Fast-GANFIT: generative adversarial network for high fidelity 3D face reconstruction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(9): 4879−4893.

[14] WU C Y, XU Q G, NEUMANN U. Synergy between 3DMM and 3D landmarks for accurate 3D facial geometry[C]//Proceedings of 2021 International Conference on 3D Vision. London: IEEE, 2021: 453–463.

[15] ZIELONKA W, BOLKART T, THIES J. Towards metrical reconstruction of human faces[C]//Proceedings of the 17th European Conference on Computer Vision-ECCV 2022. Tel Aviv: Springer, 2022: 250–269.

[16] 申冲, 刘川, 张满囤, 等. 基于弱监督学习的细节三维人脸重建 [J]. 燕山大学学报, 2023, 47(2): 144−151,163.

[17] LI T Y, BOLKART T, BLACK M J, et al. Learning a model of facial shape and expression from 4D scans[J]. ACM Transactions on Graphics (TOG), 2017, 36(6): 194.

[18] RAMAMOORTHI R, HANRAHAN P. An efficient representation for irradiance environment maps[C]// Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press, 2001: 497–500.

[19] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas IEEE, 2016: 770–778.

[20] BULAT A, TZIMIROPOULOS G. How far are we from solving the 2D & 3D face alignment problem? (and a dataset of 230, 000 3D facial landmarks)[C]//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017: 1021–1030.

[21] NIRKIN Y, MASI I, TRAN A T, et al. On face segmentation, face swapping, and face perception [C]//Proceedings of 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition. Xi' an: IEEE, 2018: 98–105.

[22] ZHANG Z Z, LAN C L, ZENG W J, et al. Relation-aware global attention for person re-identification[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 3183–3192.

[23] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017: 936–944.

[24] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600−612.

[25] CAO Q, SHEN L, XIE W D, et al. VGGFace2: a dataset for recognising faces across pose and age[C]//Proceedings of 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition. Xi'an: IEEE, 2018: 67–74.

[26] WANG Y, TAO X, QI X J, et al. Image inpainting via generative multi-column convolutional neural networks[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montreal: Curran Associates Inc. , 2018: 329–338.

[27] ZHU X Y, LEI Z, YAN J J, et al. High-fidelity pose and expression normalization for face recognition in the wild[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) . Boston: IEEE, 2015: 787–796.

[28] SANYAL S, BOLKART T, FENG H W, et al. Learning to regress 3D face shape and expression from an image without 3D supervision[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 7755–7764.

[29] WANG M, DENG W H, HU J N, et al. Racial faces in the wild: Reducing racial bias by information maximization adaptation network[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019: 692–702.

[30] TRAN A T, HASSNER T, MASI I, et al. Extreme 3D face reconstruction: seeing through occlusions[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 3935–3944.

[31] ABREVAYA V F, BOUKHAYMA A, TORR P H S, et al. Cross-modal deep face normals with deactivable skip connections[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 4978–4988.

[32] LI C L, MOREL-FORSTER A, VETTER T, et al. Robust model-based face reconstruction through weakly-Supervised outlier segmentation[C]//Proceedings of 2023 IEEE/CVF Conference on Computer Vision (CVPR). Vancouver: IEEE, 2023: 372–381.

[33] WOOD E, BALTRUŠAITIS T, HEWITT C, et al. 3D face reconstruction with dense landmarks[C]//Proceedings of the European Conference on Computer Vision. Tel Aviv: Springer, 2022: 160–177.

[34] DIB A, THEBAULT C, AHN J, et al. Towards high fidelity monocular face reconstruction with rich reflectance using self-supervised learning and ray tracing[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE, 2021: 12799–12809.

[35] FENG Y, WU F, SHAO X H, et al. Joint 3D face reconstruction and dense alignment with position map regression network[C]//Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018: 557–574.

[36] DENG Y, YANG J L, XU S C, et al. Accurate 3D face reconstruction with weakly-supervised learning: from single image to image set[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Long Beach: IEEE, 2019: 285–295.

（编辑：丁红艺）