管先吉, 桑高丽, 张楠楠. 面向低质量三维人脸识别的高判别特征提取方法[J]. 智能计算机与应用,2025,15(7):149-154.
DOI:10.20169/j.issn.2095-2163.250722

面向低质量三维人脸识别的高判别特征提取方法

管先吉[1], 桑高丽[1,2], 张楠楠[1]
(1 浙江理工大学 计算机科学与技术学院(人工智能学院), 杭州 310018;
2 嘉兴学院 信息科学与工程学院, 浙江 嘉兴 314001)

摘 要: 针对因噪声干扰导致低质量三维人脸图像有效特征提取困难等问题,提出一种基于注意力机制的多尺度特征融合方法,旨在通过提取高判别性特征以提升低质量三维人脸的识别精度。为了增强网络对图像特征的提取能力,引入注意力多尺度特征融合模块。该模块在提取不同尺度特征的同时,利用注意力机制自动识别网络中的关键特征,从而在复杂的低质量数据中提取高判别性的信息。此外,通过使用不同大小卷积核获取不同感受野的特征,使网络捕捉的信息更丰富。与现有最优方法相比,在 Lock3DFace 的测试集上平均识别准确率提高了 0.97%;在 Extended-Multi-Dim 的测试集 A 和测试集 B 上平均识别准确率分别提高了 0.7%和 2.4%。实验结果表明,基于注意力机制的多尺度特征融合方法,在低质量三维人脸识别任务中取得显著性能提升。

关键词: 低质量; 三维人脸; 人脸识别; 多尺度特征融合; 注意力
中图分类号: TP319.4 **文献标志码**: A **文章编号**: 2095-2163(2025)07-0149-06

High-discriminative feature extraction method for low-quality 3D facial recognition

GUAN Xianji[1], SANG Gaoli[1,2], ZHANG Nannan[1]
(1 School of Computer Science and Technology(School of Artificial Intelligence), Zhejiang Sci-Tech University, Hangzhou 310018, China; 2 School of Information Science and Engineering, Jiaxing University, Jiaxing 314001, Zhejiang, China)

Abstract: A multi-scale feature fusion method based on attention mechanism is proposed to address the difficulty of effective feature extraction in low-quality 3D face images caused by noise interference. The aim is to improve the recognition accuracy of low-quality 3D faces by extracting highly discriminative features. In order to enhance the network's ability to extract image features, an attention multi-scale feature fusion module is introduced. This module utilizes attention mechanism to automatically identify key features in the network while extracting features of different scales, thereby extracting highly discriminative information from complex low-quality data. In addition, by using convolution kernels of different sizes to obtain features of different receptive fields, the network can capture richer information. Compared with the existing optimal methods, the average recognition accuracy on the Lock3DFace test set has been improved by 0.97%; The average recognition accuracy on test set A and test set B of Extended Multi Sim increased by 0.7% and 2.4%, respectively. The experimental results show that the multi-scale feature fusion method based on attention mechanism achieves significant performance improvement in low-quality 3D face recognition tasks.

Key words: low-quality; 3D face; face recognition; multi-scale feature fusion; attention

0 引 言

三维人脸数据包含人脸的深度信息,不受光照和表情变化的约束[1],同时具备较高的抗欺骗性,难以被照片攻击或面具欺骗[2]。因此,三维人脸识别受到越来越多的关注。近年来,深度学习技术的发展有效提高了三维人脸识别的精度,这些方法通常使用的是高精度扫描仪获取的高质量三维人脸数据,成本高、采集时间长,难以在现实世界中得到应用[3]。

基金项目: 浙江省教育厅科研项目(Y202249424)。
作者简介: 管先吉(1997—),男,硕士,主要研究方向:计算机视觉与模式识别。
通信作者: 桑高丽(1986—),女,博士,副教授,主要研究方向:模式识别,人工智能。Email:glsang@zjxu.edu.cn。
收稿日期: 2023-08-05

相比于高精度数据,低质量三维人脸数据通过便携式三维传感器采集,具有使用方便,价格低廉等优势。但是低质量人脸图像具有较低的数据精度和大量的噪声,识别难度更高,图 1 对比了传统高质量和低质量三维人脸数据。在早期,低质量三维数据通常被用作附加信息,以增强二维人脸识别系统的性能[4]。目前,有些研究尝试只使用低质量三维数据而不依赖于二维信息[5],可以更加安全有效地进行人脸识别,有着更高的研究价值和广阔的应用前景,是该领域的研究热点。

图 1 低质量和高质量三维人脸数据对比图

Fig. 1 Comparison of high-quality and low-quality 3D facial data

低质量三维人脸的研究始于 2010 年,随着便携式三维采集设备 Kinect v1 和 RealSense 的出现,三维人脸数据的获取变得更方便,也更能满足实际应用需求。研究主要分为数据质量提升和有效特征提取两个方面。

关于数据质量提升,早期主要使用一些简单的预处理方法提高人脸数据质量。例如,对称填充、数据插值、平滑处理等[6]。随着大型低质量三维人脸数据集 Lock3DFace[7] 和 Extended-Multi-Dim[8] 的发布,越来越多的工作采用深度学习方法对其研究。在数据质量提升方面,Xiao 等[9]提出使用软阈值模块去噪方法,通过网络学习,自动设置阈值实现轻微的去噪效果。Lin 等[10]提出高质量人脸数据生成和高低质量人脸数据融合的识别方法,利用多质量融合网络 MQFNet,融合高低质量人脸数据提升识别性能。此外,桑高丽等[11]提出了一种联合软阈值去噪和视频数据融合的低质量三维人脸识别方法,有效融合多个视频帧提升人脸数据的质量。

在特征提取方面,早期大多采用传统提取方法,如局部二值模式(Local Binary Patterns,LBP)[12]、主成分分析(Principal Component Analysis,PCA)[13]、迭代最近点(Iterative Closest Point,ICP)[14]和方向梯度直方图(Histogram of Oriented Gradients,HOG)[15]等,这些方法在当时已有的数据集上获得了较好的效果,但使用的数据集所涉及的身份数量并不多,数据规模不够大,所涉及的变化也较少,因此并不具有足够的说服力。Extended-Multi-Dim 采用多种高质量三维人脸引导低质量三维人脸识别模型训练的方法,减轻了低质量三维人脸特征提取难度,但该网络需要同时使用低质量和高质量三维人脸,数据获取难度较大。龚勋等[16]针对低质量三维人脸难以提取有效特征的问题,提出了基于空间注意力机制的 Dropout(SAD)和类间正则化损失函数(IR Loss),有效提升了不完整三维人脸数据的识别精度,但 IR Loss 存在 2 个超参数,参数设置依赖经验,具有一定的局限性。

卷积神经网络(CNN)具有分层的体系结构,由多个卷积层堆叠而成,每层独立学习不同的信息。其中,较低层捕捉基本颜色和边缘等低级元素,较高层则编码抽象的语义信息。多尺度特征融合是指将来自不同尺度的特征信息进行整合,以提取更全面、更丰富的特征表示。早期的多尺度特征融合方法 FPN[17] 通过自顶向下的反馈路径和横向连接,将高层语义信息与低层细节信息进行融合。Liu 等[18]提出了 PANet,在 FPN 的基础上引入自底向上的上采样路径,以进一步增强不同尺度特征的融合效果。但上述方法只是在通道维度的简单拼接,没有关注带有身份信息的关键特征。

综上所述,本文提出了一种基于注意力机制的多尺度特征融合方法,在专注特征融合的同时关注身份信息,有效提取高判别性特征,提高低质量三维人脸识别率。

本文主要有以下贡献:

1)为了减轻噪声干扰,突出人脸特征,提出注意力多尺度特征融合模块,使用注意力机制对提取特征加权融合,增强与身份相关信息的特征表示,抑制噪声影响。

2)为了进一步提高特征判别能力,引入 Inception 模块,使用不同的卷积核分支获取多种感受野的特征输出,得到更加多样化的特征表示。

1 方 法

本文提出的低质量三维人脸识别模型的整体结构如图 2 所示。首先,将数据经过与 Led3D[19] 相同的预处理,具体包括线性插值、鼻尖点校正、离群点

剔除、面部投影、孔洞填充和法线图估计等处理，得到三维人脸的法线图作为网络的输入。主干网中插入软阈值去噪模块（Soft Thresholding Module，STM）[9]对数据噪声进行过滤，并在末端添加 Inception 模块[20]增强网络的特征提取能力，然后将这些特征输入到注意力多尺度特征融合模块，最后进行分类。其中，每个 Block 块和 Inception 块中都包含 BatchNorm 层和 ReLU 层。

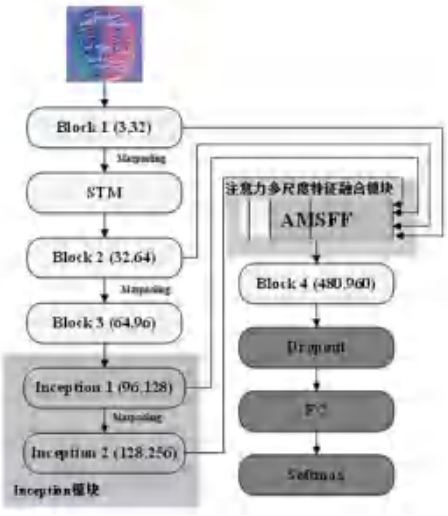

图 2　本文方法整体流程图

Fig. 2　Overall flowchart of the method

1.1　注意力多尺度特征融合模块

注意力机制允许模型对输入信息进行自适应的特征选择和加权，以便有效地关注与当前任务相关的部分。通过计算注意力权重，模型能够学习到输入中不同位置的重要性，从而实现有针对性地提取和利用特征。而多尺度特征融合能够从不同层次和尺度的特征表示中获取丰富的信息，帮助模型抵抗输入中的噪声、变形和尺度变化等干扰因素，使模型获得更鲁棒和稳定的表示。因此，通过联合使用这两种方法，模型可以更准确地关注感兴趣的区域和特征，提高语义建模能力，并增强模型的鲁棒性和泛化能力。

本文提出的注意力多尺度特征融合模块（Attention Multi-Scale Feature Fusion，AMSFF）的结构如图 3 所示。由图中可见，从骨干网中的 4 个卷积块（Block1、Block2、Inception1、Inception2）提取特征，作为注意力多尺度特征融合的输入，这些特征图分别对应不同感受野捕捉到的信息。先对不同层提取的特征图 $F \in R^{H\times W\times C}$ 分别添加注意力权重，强调不同位置像素的贡献。F 通过全局通道平均池化得到二维特征 $F_{\text{avg}} \in R^{H\times W}$，代表全局的特征信息。其中，$H$、$W$、$C$ 分别表示特征图 F 的高、宽和通道数。同时，为了得到 F 中最具有区分性的信息，使用全局通道最大池化得到二维特征 $F_{\text{avg}} \in R^{H\times W}$，代表 F 中最具有分辨能力的特征。将全局特征信息和最具辨别能力的特征信息进行融合，最终得到高宽不变，通道数为 1 的特征图 $F_{\text{fusion}} \in R^{H\times W}$：

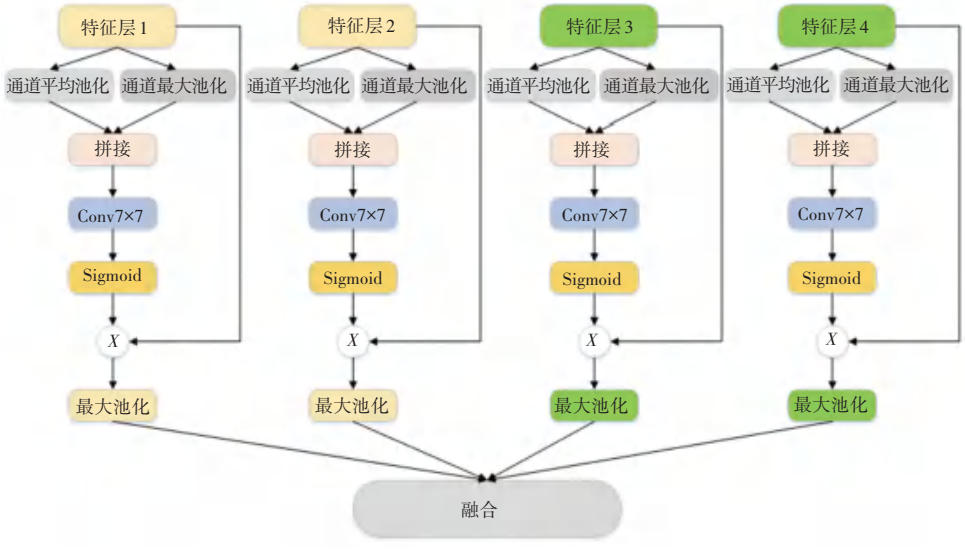

图 3　注意力多尺度特征融合模块结构图

Fig. 3　Attention multi-scale feature fusion module architecture diagram

$$F_{\text{fusion}} = Conv_{7\times 7}(F_{\text{avg}} \oplus F_{\text{max}}) \tag{1}$$

式中：$Conv_{7\times 7}()$ 表示使用 7 × 7 的卷积核计算，$\oplus$ 表示在通道维度拼接。为了达到给所有像素不同权重的目的，本文使用 *Sigmoid* 激活函数，对 F_{all} 按照

下式进行激活：

$$F_{Sigmoid} = \frac{1}{1 + e^{-F_{\text{fusion}}(x)}} \tag{2}$$

得到特征图 $F_{Sigmoid} \in R^{H\times W}$，作为特征图 F 的注意力权重图。对输入特征图 F 的每个通道按照下式计算，得到添加注意力之后的特征：

$$F_{\text{final}} = F \oplus F_{Sigmoid} \tag{3}$$

式中：⊗表示对 F 中每个通道都乘以 $F_{Sigmoid}$，最后将添加注意力后的特征图下采样到相同尺寸后拼接。其中，每个最大池化层所使用的卷积核大小、步长和填充分别为(33,16,16)、(17,8,8)、(9,4,4)、(3,2,1)。通过这种方式，有效地生成一个更具鉴别力的特征，来表示低质量三维人脸。

1.2 特征提取模块

感受野在神经网络中具有关键作用，对网络的感知范围和特征提取能力具有显著影响。通过调整感受野大小，能够实现层级特征提取、上下文信息整合、特征表达的层次性等功能。Inception 模块通过并行使用多个不同尺寸的卷积核和池化操作，从不同尺度下提取不同感受野的图像特征，并将其连接起来。这种并行操作允许网络同时捕捉不同尺度下的特征信息，提高了网络的感知能力和表达能力。此外，由于 Inception 模块中的操作是并行的，可以减少网络中的参数数量和计算复杂度，因此本文采用 Inception 模块作为网络的特征提取模块。

本文采用的 Inception 模块包含以下几个并行分支：1×1 卷积分支、3×3 卷积分支、5×5 卷积分支、最大池化分支。在每个 Inception 模块中，这些并行分支的输出特征图会沿着通道维度进行拼接，形成更多样化的特征表示。模块 Inception1 每个分支的输入和输出通道数分别为(32,32)、(48,64)、(8,16)、(16,16)；Inception2 的每个分支的输入和输出通道数分别为(64,64)、(96,128)、(16,32)、(32,32)。本文所提模型将 Inception 模块放在主干网的最后两层，是为了对输入数据进行非线性变换，使得模型能够更好地捕捉输入数据的非线性特征，为后续的 Inception 模块提取特征奠定基础。

2 实验结果与分析

为了评估本文方法的有效性，进行了大量实验，并将其与最先进的方法进行比对。

2.1 数据集

Lock3dFace[7] 是一个综合的低质量 3D 人脸数据集，采集自 Kinect V2，其规模庞大且内容丰富。该数据集包含了 509 个不同个体的 5 671 个视频序列，每个序列又包含 59 帧图像。

该数据集的每个个体都包含中性 NU(Neutral)、表情 FE(Face Expression)、遮挡 OC(Occlusion)、姿态 PS(Pose)和时间 TM(Time)5 个子集。其中，中性指每个个体在没有任何表情和遮挡情况下的正面姿态图像；表情子集分为快乐、愤怒、悲伤、惊讶、恐惧和厌恶；遮挡子集中每个个体随机遮挡人脸的一部分；姿态子集要求每个个体在俯仰和偏航两个方向移动头部；时间子集要求 169 人在 7 个月后，使用与第一次相同的配置再次参加数据采集。因此，Lock3DFace 是目前最具挑战性的 3D 人脸识别数据集之一，数据集样例如图 4 所示。

图 4 Lock3DFace 数据集样例图

Fig. 4 Lock3DFace dataset sample images

Extended-Multi-Dim[8] 是目前规模最大的低质量三维人脸数据集之一。该数据集包含了由 RealSense 捕获的彩色图像、相应的低质量深度图像，以及高质量 3D 扫描仪扫描的高质量 3D 人脸形状。数据集包含了约 3 027 个视频，涵盖了 902 个个体的表情和姿态变化。其中，表情有 8 种变化，姿态包含俯仰和偏航两个方向，数据集样例如图 5 所示。

图 5 Extended-Multi-Dim 数据集样例图

Fig. 5 Extended-Multi-Dim dataset sample images

2.2 设置和协议

为了验证所提出模型的有效性，本文按照与文献[20]相同的协议，在真实的低质量数据集

Lock3DFace 上进行实验。从 509 个个体的中性表情视频中选择第一个视频，以等间隔选择 6 帧作为训练集，并对这些帧进行数据增强。剩余视频中的所有帧被用作测试集，并按照表情、遮挡、姿态和时间 4 个子集进行划分。在测试阶段，对每个视频的每一帧进行标签预测，选择在所有数据帧中出现次数最多的结果作为该视频的最终预测标签。

为了实验公平起见，本文采用与文献[9]相同的参数对 Extended-Multi-Dim 数据集的协议进行设置。训练数据由 128×128 像素处理后的深度图像组成，共涵盖了 430 个对象的 299 k 张图像。测试数据分为 A、B 两个子集。测试集 A 包含了 256 k 张图像，涵盖了 275 个对象，分为中性、表情、姿态 1 和姿态 2 共 4 个子集；测试集 B 包含了 60 k 张图像，涵盖了 228 个对象，每个对象的所有图像作为一个集合。

2.3　训练参数和实验环境

本次实验所有的输入数据图像大小都被调整到 128×128 像素，训练轮次为 100 轮，批大小为 128，初始学习率为 0.000 5，衰减因子为 0.000 05。模型先在高质量三维人脸数据集上进行预训练，然后在 Lock3DFace 或 Extended-Multi-Dim 数据集上进行微调，使用 Adam 优化器进行优化。

实验环境为 windows11、Pytorch1.10.1、CUDA11.3；显卡为 NVIDIA GeForce RTX 3090；处理器为 12th Gen Intel(R) Core(TM)i9-12900K 3.19 GHZ。

2.4　方法比较

2.4.1　Lock3DFace 协议实验

为验证本文方法的性能，本文与近年所提出的低质量三维人脸识别模型和一些经典分类网络进行对比，表 1 展示了上述方法在 Lock3DFace 数据集上的实验结果。

表 1　Lock3DFace 数据集的测试结果

Table 1　Test results of the Lock3DFace dataset　%

方法	表情	遮挡	姿态	时间	平均
ResNet34	62.83	20.32	22.56	20.70	32.23
InceptionV2	80.48	32.17	33.23	12.54	44.77
Led3D[20]	86.94	48.01	37.67	26.12	54.28
SAD[16]	87.02	**65.47**	45.17	23.52	54.83
STD[9]	89.88	45.32	47.04	38.76	59.03
MQFNet[10]	90.55	52.81	44.64	22.65	61.04
Sang[11]	**93.06**	46.31	48.12	39.95	59.43
本文	91.48	56.91	**48.64**	**39.99**	**62.01**

由表 1 中数据可见，本文所提出的模型达到了最高的平均识别准确率 62.01%，证明本文方法的有效性。其中，Sang 的方法融合多帧低质量三维人脸图像提高了数据质量，在表情子集获得最高的准确率，而本文方法使用单帧图像进行预测。SAD 方法重点关注遮挡问题，在遮挡子集表现出最高的准确率，在其他测试子集上效果不佳。值得注意的是，相比于平均识别准确率第二的网络 MQFNet，本文所提出的方法在表情、遮挡、姿态和时间方面都有较大提升。

2.4.2　Extended-Multi-Dim 协议实验

为了进一步验证本文提出方法的有效性，分别在 Extended-Multi-Dim 测试集 A 和测试集 B 与其他方法进行对比，实验结果见表 2。实验结果表明，本文提出方法在所有子集上均取得了最高的识别率。特别是在测试集 A 的中性和表情子集上，本文方法准确率达到 99%以上。这是因为添加的注意力多尺度特征融合模块和 Inception 模块，使网络能够更加关注上下文信息，理解低质量三维人脸数据，从而提高了特征判别性。

表 2　Extended-Multi-Dim 数据集的测试结果

Table 2　Test results of the Extended-Multi-Dim dataset　%

方法	中性	A-表情	A-姿态 1	A-姿态 2	A-平均	B-平均
ResNet34	94.6	90.8	67.2	49.6	80.4	70.3
InceptionV2	90.6	86.5	57.3	41.4	74.6	64.5
Hu[8]	94.2	90.7	60.8	42.3	80.0	66.4
Led3D	96.2	94.8	67.9	47.2	81.8	71.6
STD	98.8	97.2	83.9	65.0	90.5	79.1
本文	**99.5**	**99.0**	**86.5**	**65.7**	**91.2**	**81.5**

2.5　消融实验

为验证所提方法中每个模块的有效性，本文在 Lock3DFace 数据集上进行实验。按照提出模块的添加情况，预设以下 4 个网络：网络 A 仅包含 STM 和 5 个卷积层的基准网络，网络 B 具有 AMSFF，网络 C 具有 Inception 模块，网络 D 同时具有 AMSFF 和 Inception 模块。

通过表 3 可以看出，与基准网络相比，AMSFF 和 Inception 模块提高了识别性能。一方面，AMSFF 通过对不同层次的信息进行梳理，提取更具判别性的特征；另一方面，Inception 模块进一步增强网络对特征的提取能力，使模型更好地获取不同尺度下的图像信息，提高对图像细节和全局结构的感知能力。在最终结果中，结合 AMSFF 和 Inception 模块

的网络模型取得最佳性能,且结合之后提升效果最明显。

表 3 不同模块的识别准确率比较

Table 3 Comparison of recognition accuracy among different modules %

网络	模块	表情	遮挡	姿态	时间	平均
A	无	87.53	52.90	46.46	36.25	58.24
B	AMSFF	89.65	51.50	47.25	36.72	58.77
C	Inception	89.31	56.31	44.48	37.24	59.30
D	AMSFF+ Inception	**91.48**	**56.91**	**48.64**	**39.99**	**62.01**

3 结束语

三维人脸噪声对识别效果产生重要影响,导致模型难以提取高判别性人脸特征。本文提出一种基于注意力机制的多尺度特征融合方法,提高了人脸特征提取能力,有效减弱噪声影响。此外,引入的 Inception 模块,提高网络的表征能力和分类性能。消融实验验证了各模块的有效性。结果显示本文的方法优于其他低质量三维人脸识别方法,在公开的低质量三维人脸数据集 Lock3DFace 和 Extended-Multi-Dim 上取得了最高识别率。在未来的工作中,预将所提方法迁移到其他领域,探索其在不同任务中的适用性效果。

参考文献

[1] SOLTANPOUR S, WU Q M J. Weighted extreme sparse classifier and local derivative pattern for 3D face recognition[J]. IEEE Transactions on Image Processing, 2019, 28(6): 3020-3033.

[2] 罗常伟,於俊,于灵云,等. 三维人脸识别研究进展综述[J]. 清华大学学报(自然科学版),2021,61(1):77-88.

[3] CAI Y, LEI Y, YANG M, et al. A fast and robust 3D face recognition approach based on deeply learned face representation [J]. Neurocomputing, 2019, 363: 375-397.

[4] UPPAL H, SEPAS-MOGHADDAM A, GREENSPAN M, et al. Depth as attention for face representation learning[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 2461-2476.

[5] JING Y, LU X, GAO S. 3D face recognition: A survey[J]. arXiv preprint arXiv,2108.11082, 2021.

[6] LI B Y L, MIAN A S, LIU W, et al. Using kinect for face recognition under varying poses, expressions, illumination and disguise [C]//Proceedings of 2013 IEEE Workshop on Applications of Computer Vision (WACV). Piscataway, NJ: IEEE,2013: 186-192.

[7] ZHANG J, HUANG D, WANG Y, et al. Lock3dface: A large-scale database of low-cost kinect 3d faces[C]//Proceedings of 2016 International Conference on Biometrics (ICB). Piscataway, NJ: IEEE, 2016: 1-8.

[8] HU Z, GUI P, FENG Z, et al. Boosting depth-based face recognition from a quality perspective[J]. Sensors, 2019, 19(19): 4124.

[9] XIAO S, LI S, ZHAO Q. Low-quality 3D face recognition with soft thresholding [C]// Proceedings of the 15th Chinese Conference on Biometric Recognition. CCBR, 2021: 419-427.

[10] LIN S, JIANG C, LIU F, et al. High quality facial data synthesis and fusion for 3D low-quality face recognition[C]// Proceedings of 2021 IEEE International Joint Conference on Biometrics (IJCB). Piscataway, NJ: IEEE, 2021: 1-8.

[11] 桑高丽, 肖述笛, 赵启军. 联合软阈值去噪和视频数据融合的低质量 3 维人脸识别[J]. 中国图象图形学报, 2023, 28(5): 1434-1444.

[12] MIN R, KOSE N, DUGELAYJ L. Kinectfacedb: A kinect database for face recognition[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2014, 44(11): 1534-1548.

[13] GOSWAMI G, BHARADWAJ S, VATSA M, et al. On RGB-D face recognition using Kinect[C]// Proceedings of 2013 IEEE 6th International Conference on Biometrics: Theory, Applications and Systems (BTAS). Piscataway, NJ: IEEE, 2013: 1-6.

[14] BERRETTI S, DEL BIMBO A, PALA P. Superfaces: A super-resolution model for 3D faces[C]// Proceedings of Computer Vision - ECCV 2012 Workshops and Demonstrations: Florence. Cham:Springer, 2012: 73-82.

[15] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]// Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). Piscataway, NJ: IEEE, 2005: 886-893.

[16] 龚勋, 周炀. 面向低质量数据的 3D 人脸识别[J]. 电子科技大学学报, 2021, 50(1): 43-51.

[17] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ:IEEE, 2017: 2117-2125.

[18] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 8759-8768.

[19] MU G, HUANG D, HU G, et al. LED 3D: A lightweight and efficient deep approach to recognizing low-quality 3D faces[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ:IEEE, 2019: 5773-5782.

[20] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ:IEEE, 2015: 1-9.