



# 基于关键特征增强机制的 3D 人脸识别

王 奇<sup>1\*</sup>, 钱伟中<sup>1,2</sup>, 雷 航<sup>1</sup>, 王旭鹏<sup>1</sup>

(1. 电子科技大学 信息与软件工程学院, 成都 610054; 2. 喀什地区电子信息产业技术研究院, 喀什 844000)

**摘要** 3D 人脸识别是计算机视觉领域的重要组成部分, Pointnet 依靠深度学习解决了点云的无序性, 实现了 3D 点云的全局特征提取, 但由于点云数据缺乏细节纹理, 仅靠全局特征很难实现复杂情况下的人脸识别。针对以上问题, 基于 Pointnet 提出了一种局部特征描述子, 用于描述点云局部空间的几何特征, 并引入关键特征增强机制, 通过特征概率分布增强人脸关键信息, 该机制能减少不必要特征对任务的干扰, 有效提升模型的准确率。在公共数据集 CASIA-3D、Lock3DFace、Bosphorus 上进行实验测试, 结果表明该方法能很好地应对表情变化、部分遮挡以及头部姿态的干扰, 在弱光环境下其准确率高于 RP-Net 1.1%, 并具有良好的实时性。

**关键词** 3D 人脸识别; 深度学习; 局部特征描述子; 特征增强; 点云数据

中图分类号 TP391 文献标志码 A DOI 10.12178/1001-0548.2023012

## 3D Face Recognition Based on Key Feature Enhancement Mechanism

WANG Qi<sup>1\*</sup>, QIAN Weizhong<sup>1,2</sup>, LEI Hang<sup>1</sup>, and WANG Xupeng<sup>1</sup>

(1. School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China;

2. Kash Institute of Electronics and Information Industry, Kash 844000, China)

**Abstract** 3D face recognition is an important part of the field of computer vision. Pointnet relies on deep learning to solve the disorder of point clouds and realize the global feature extraction. However, due to the lack of detailed texture of point clouds, it is difficult to realize face recognition in complex situations only by global features. In deal with the above problems, a local feature descriptor is proposed to describe the local spatial geometric features of the point clouds, and a key feature enhancement mechanism is introduced to enhance the key features of the face through the probability distribution, which can reduce the interference of unnecessary features and effectively improve the accuracy of the model. Experiments were carried out on public data sets CASIA-3D, Lock3DFace and Bosphorus. The results show that our method can deal well with the change of expression, partial occlusion and interference of head pose, especially in weak light conditions, compared with RP-Net, the accuracy is improved by 1.1 percent, and the method also has good real-time performance.

**Key words** 3D face recognition; deep learning; local feature descriptor; feature enhancement; point cloud

人脸识别作为计算机视觉领域的重要组成部分, 被广泛应用于日常生活中。但普通 RGB 图像的成像质量依赖于光照条件, 这使得在一些缺乏光照的场景很难通过普通相机获取有效信息, 如夜间驾驶、监视和跟踪等<sup>[1]</sup>。随着三维相机的发展, 如 Kinect, 获取高质量三维图像变得更加便捷, 基于红外的成像原理, 使其对光照变化保持稳定, 且每个像素点代表了传感器到目标的距离, 能有效地刻画目标的空间信息。通过简单的坐标转化, 能实现

三维图像到三维点云数据的转变。基于点云的人脸识别有以下两个优势: 一是数据本身对光照变化保持稳定, 使其能有效地应用于缺乏光照或光照变化大的场景; 二是能简单地根据距离信息实现背景分离, 减少背景对目标的干扰, 使任务能更好地聚焦于对象本身<sup>[1]</sup>。

文献 [2] 提出了 Pointnet, 该网络解决了点云的无序性以及深度学习中的应用, 有效地通过深度神经网络提取点云的全局特征并应用于点云的分

收稿日期: 2023-01-06; 修回日期: 2023-07-02

基金项目: 国家自然科学基金 (61802052); 喀什地区科技计划项目 (KS2023029)

作者简介: 王奇, 男, 博士生, 主要从事计算机视觉、三维人脸分析技术方面的研究。

\*通信作者 E-mail: wangqi1210@std.uestc.edu.cn

类以及分割任务。文献 [3] 提出 Pointnet++ 通过点云的空间位置信息对点云分组, 实现了点云特征的分层次提取。文献 [4] 通过融合局部和全局特征提出了 PPFNet, 该模型丰富了点云特征的表达方式。文献 [5] 通过 X-conv 算子解决了点云在卷积神经网络中的应用。文献 [6] 基于图卷积网络构建了一个收敛快、计算复杂度低的点云特征提取网络。近年来, 更多的网络模型被应用于三维点云, 如文献 [7] 为点云设计了自注意层, 并使用这些自注意层来构建语义场景分割和分类等任务的深度学习网络。

随着基于点云的深度学习模型的发展, 更多的 3D 点云被应用于人脸分析任务, 文献 [8-9] 利用 Pointnet++ 提取 3D 人脸全局特征, 实现了高效的人脸姿态估计。文献 [10] 构建全新的孪生网络用于姿态干扰下的人脸验证。文献 [11-12] 采用二维特征互补三维点云信息用于人脸识别。文献 [13] 提出了一个系数成分分析网络用于三维人脸特征提取。文献 [14] 引入注意力机制结合上下文信息用于面部表情分析。大量基于点云的人脸分析工作取得了突破性进展, 但是, 由于点云数据自身缺乏细节纹理, 如何在面部表情和姿态干扰以及部分遮挡等复杂情况下有效地提取特征或强化关键信息依然是该领域研究的重点。

针对以上问题, 本文采用三维点云数据作为输入, 构建一个新的深度学习模型用于三维人脸识别。首先将点云数据进行分组, 利用一个局部特征描述子获取每组点云的空间几何信息, 再利用 Pointnet 提取局部特征, 然后通过最大池化层整合局部特征, 获取全局特征。

为增强关键特征的比重, 本文采用关键特征强化机制将获取的全局特征映射为特征空间中的概率分布并通过增强系数强化特征的辨识度, 从而减少面部表情以及头部姿态等对任务的干扰。由于点云数据本身对光照变化保持稳定, 本方法同样适用于缺乏光照或光照变化大的场景。

## 1 人脸特征提取

3D 点云的本质是一组三维空间中的点构成的几何模型, 表示为  $n \times (x_i, y_i, z_i)$  的矩阵, 其中  $n$  为点的数量,  $(x_i, y_i, z_i)$  为点在三维空间的位置坐标。但相同对象特征点的顺序不一定相同, 如图 1 所示, 等号左右的特征点构成相同空间几何形状但点的顺

序并不一致, 因此点云数据不能像规则的二维图像及三维体素一样具有有序的索引序列从而实现权重共享进行卷积操作。解决点云的无序性并进行有效的特征提取是人脸识别任务的关键因素。

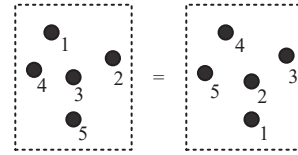


图1 点云无序性示意图

### 1.1 点云特征提取

为解决点云的无序性以及深度学习中的应用, 文献 [2] 根据对称函数的思想构建了基于点云的特征提取网络 Pointnet。

定义  $l$  维空间中的一组点  $\chi \in \mathbb{R}^l$ , 函数  $g: \mathbb{R}^k \rightarrow \mathbb{R}^N$  是不同特征空间之间的映射, 则存在以下定理。

假定  $f: \chi \rightarrow \mathbb{R}^M$ ,  $g: \mathbb{R}^k \rightarrow \mathbb{R}^N$  均为对称函数 (函数映射结果与自变量输入顺序无关), 则:

$$f(x_1, x_2, \dots, x_n) \approx \gamma \circ g(h(x_1), h(x_2), \dots, h(x_n)) \quad (1)$$

式中,  $h: \chi \rightarrow \mathbb{R}^k$ ,  $\gamma: \mathbb{R}^N \rightarrow \mathbb{R}^M$ 。

根据此定理, 对于无序点云  $\{p_1, p_2, p_3, \dots, p_N\}$ , 为实现  $f: p \rightarrow \mathbb{R}^M$  全局特征提取, Pointnet 采用多层感知机与最大池化层构建式 (1) 中右边部分, 其网络结构如图 2 所示。其中  $n$  为特征点的数量,  $l$  为点的特征维度, 多层感知机如式 (1) 中  $h$  函数将每个点通过三层卷积核为  $1 \times 1$  (由于点云的无序性, 卷积核无法对其周围特征点进行有序索引, 只能采用  $1 \times 1$  的卷积核对单个特征点进行逐点式的特征提取) 的卷积层进行特征提取, 输出通道数分别为 64, 64, 128 (本文根据文献 [8] 中的方法, 采用两层通道数为 64 的卷积层对特征点维度进行提升, 同时根据 FaceNet<sup>[15]</sup> 输出特征维度, 将点云映射到特征空间  $\mathbb{R}^{128}$  中)。最大池化层如式 (1) 中对称函数  $g$  取每个特征维度最大值 (与特征输入顺序无关) 获取点云全局特征。最后根据所需特征维度, 将全局特征通过多层感知机映射到特定的特征空间  $\mathbb{R}^M$  中。

如上所述, 对于不同的人脸点云对象, 由于面部轮廓及头部大小不同, 其构成的点的数量也不相同, 为保证网络具有相同的特征输入维度, 本文根据文献 [3], 采用最远点采样法为每个对象采样固定数量的特征点, 即对于输入点集  $\{p_1, p_2, \dots, p_N\}$ , 随机选取  $p_i$  作为起始点, 遍历点集, 找到最远点

$p_{ij}$ 放入新的点集中,并在剩余点集中再次以 $p_i$ 为起始点,迭代以上过程,并获取固定数量的特征点 $\{p_{i1}, p_{i2}, \dots, p_{i4096}\}$ (本文每个对象固定采样 4 096 个特征点)。相较于随机采样法,最远点采样法能更均匀地对样本进行采样,即采样点能更好地覆盖整个点集。

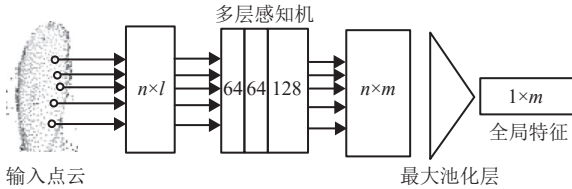


图 2 点云特征提取网络

## 1.2 局部特征描述子

相较于普通 RGB 图像,点云数据缺乏细节纹理,仅靠全局特征很难有效区分不同对象。文献 [3] 采用区域点簇的位置信息作为局部特征并采用分层采样的方法进行特征提取,该方法能有效应对点云物体的分类与分割任务,但是缺乏对局部特征的描述,因此该方法局部特征的表达非常孱弱,而人脸整体轮廓形状相似,点云的细粒表达(Fine-Grained Representation)是人脸识别任务的关键。相较于文献 [3] 中的方法,本文根据区域点对的相对位置与法线夹角构造一个四维描述子,该描述子对局部区域内点对间的关系进行手工标记(Hand-Crafted Feature),并组成局部特征描述子对局部区域的几何特征进行细粒表达且该描述子满足平移不变性的要求(采用相对位置与法线夹角)。 $(p_i, p_j)$ 为局部区域内的一对特征点,其空间几何关系为:

$$\psi_{ij} = (\|d\|, \angle(\mathbf{n}_i, \mathbf{d}), \angle(\mathbf{n}_j, \mathbf{d}), \angle(\mathbf{n}_i, \mathbf{n}_j)) \quad (2)$$

式中, $\mathbf{d}$ 用向量表示两点之间的差值; $\|\cdot\|$ 为欧氏距离; $\mathbf{n}_i$ 和 $\mathbf{n}_j$ 为该局部区域内点 $p_i$ 和 $p_j$ 的法向量; $\angle(\cdot, \cdot)$ 表示向量间的夹角:

$$\angle(\mathbf{n}_i, \mathbf{n}_j) = \text{atan2}(\|\mathbf{n}_i \times \mathbf{n}_j\|, \mathbf{n}_i \cdot \mathbf{n}_j) \quad (3)$$

该四维描述子描述了点对间的空间几何特征。在以 $p_i$ 为中心、 $k$ (也被称为局部特征描述子的感受野)为半径的球体内,包含的所有点 $\{p_1, p_2, p_3, \dots, p_j\}$ 与中心点 $p_i$ 构成 $j$ 个点对,该局部区域的编码方式为:

$$F_i = [p_1, \mathbf{n}_1, p_2, \mathbf{n}_2, \dots, p_j, \mathbf{n}_j, \psi_{i1}, \psi_{i2}, \dots, \psi_{ij}] \quad (4)$$

式中, $\mathbf{n}_j$ 为点 $p_j$ 在该局部区域的法向量, $\psi_{ij}$ 为 $p_j$ 与中心点 $p_i$ 间的四维特征描述子。如图 3 所示, $F_i$ 通

过以 $p_i$ 为球心、 $k$ 为半径的局部区域的所有点与中点 $p_i$ 的空间几何关系描述了该局部区域的空间几何特征。

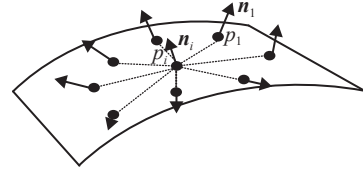


图 3 局部空间几何特征示意图

## 1.3 局部到全局特征的提取

如 1.1 节所述,每组点云对象通过最远点采样法,均匀地采样 $N_1$ 个特征点 $\{p_1, p_2, \dots, p_{N_1}\}$ ,将每个特征点作为中点,以 $k$ 为半径得到 $N_1$ 个局部区域,对每个局部区域根据式(4)的局部特征描述子进行编码,得到一组局部空间几何特征 $\{F_1, F_2, \dots, F_{N_1}\}$ ,再对每一个 $F_{N_i}$ 采用 1.1 节中的点云特征提取网络进行局部特征提取,并获取一组高维特征空间中的特征 $\{f_1, f_2, \dots, f_{N_1}\}$ 。如图 4 所示,将该组特征送入最大池化层获取该点云对象从局部到全局的特征 $f_{LtoG}$ 。由于 $f_{LtoG}$ 是通过局部特征整合提取,因此包含了丰富的局部几何信息。

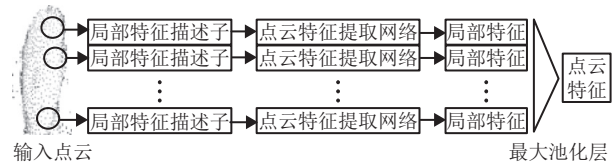


图 4 局部特征到全局特征提取结构示意图

## 2 关键特征增强机制

理论上,对点云局部区域进行编码并通过最大池化层可获取高维特征 $f_{LtoG}$ 。但实际中,并不是每个维度的特征都对识别任务起相同的作用。如某一维度的特征值较大,则该维度上的特征差异越大,辨识度越高,则称该特征为关键特征;反之,某一维度的特征值越小,则该维度上的特征差异越小,辨识度越低,则称该维度特征为非关键特征<sup>[14]</sup>。本文提出了一种关键特征增强机制,用于增强关键特征的比重,提高模型识别准确率。该机制的结构如图 5 所示。首先根据式(5),采用双曲正切函数将特征 $f_{LtoG}$ 归一化到 $[-1, 1]$ ,并得到归一化后的特征 $f_{norm}$ 。再根据式(6)和式(7),采用 softmax 函数将 $f_{norm}$ 映射为 $0 \sim 1$ 的概率分布。



$$f_{\text{norm}} = \tanh f_{LtoG} = \frac{e^{f_{LtoG}} - e^{-f_{LtoG}}}{e^{f_{LtoG}} + e^{-f_{LtoG}}} \quad (5)$$

$$f_{\text{en}} = \text{softmax}(f_{\text{norm}}) \quad (6)$$

$$\text{softmax}(f_i) = \frac{e^{f_i}}{\sum_{k=1}^n e^{f_k}} \quad (7)$$

式中,  $f_{\text{en}}$  表征了  $f_{\text{norm}}$  每个维度的特征值在特征空间中的概率分布。如式 (7) 所示, 在特征概率分布  $f_{\text{en}}$  中, 较大的特征值其概率值呈指数增长, 而较小的特征值其概率值呈指数减小。该方法能有效增强关键特征在概率分布中的辨识度。

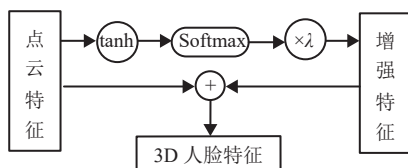


图5 关键特征增强机制结构示意图

如上所述,  $f_{\text{norm}}$  为通过深度学习网络提取并归一化后的 3D 点云特征,  $f_{\text{en}}$  为关键特征通过概率分布增强后的特征。本文通过增强参数  $\lambda$  将  $f_{\text{en}}$  与  $f_{\text{norm}}$  线性相加, 获取增强后的三维人脸特征  $f_{\text{total}}$ :

$$f_{\text{total}} = f_{\text{norm}} + \lambda f_{\text{en}} \quad (8)$$

如式 (8) 所示, 增强参数  $\lambda$  决定了  $f_{\text{en}}$  对点云特征  $f_{\text{norm}}$  的增强程度。

文献 [15] 中, 谷歌公司采用卷积神经网络提取特征并引入三元组损失构建了著名的 FaceNet 网络, 用于二维 RGB 人脸识别。本文采用三维特征  $f_{\text{total}}$  替换 FaceNet 中卷积神经网络提取的二维特征, 并将其作为该网络的特征嵌入用于三维人脸识别任务。

### 3 实验与分析

本文在 3 个数据集上进行实验分析, 通过消融实验验证特征增强机制的有效性并测试了  $\lambda$  的最优值, 最后选取最优值与近几年的先进方法进行实验分析对比。

#### 3.1 数据集

1) CASIA-3D: 该数据集由中国科学院自动化研究所模式识别与安全技术研究中心采集创建<sup>[16]</sup>。CASIA-3D 基于 Minolta vivid910 3D 数字扫描仪采集了 123 名不同受试者。每名受试者包含 37 或 38 张脸部图像, 每张 RGB 图像对应一张 3D 人脸,

其中包含不同的面部表情、头部姿态以及光照强度对样本的影响。该数据集共计 4 626 个人脸模型。

2) Lock3DFace: 文献 [17] 通过 Kinect V2 采集了 509 位受试者的 5 671 组样本 (RGB-D 图像)。该数据集主要根据表情变化 (FE)、普通面部 (NU)、部分遮挡 (OC)、头部姿态变化 (PS) 划分了 4 个子集, 能应用于不同场景下的人脸分析任务。

3) Bosphorus: 文献 [18] 为研究多种 2D 和 3D 人脸处理任务而创建该数据集, 包括表情识别、面部动作检测、面部动作强度估计、不利条件下的人脸识别、可变形人脸建模和 3D 人脸重建。该数据集基于结构光的 3D 系统采集了 105 个受试者共计 4 666 帧面部数据, 其中三分之一受试者为职业演员, 每位受试者为数据集提供了 35 种类型的表情。如图 6 所示, 3 个子图分别为该数据集上面部表情、头部姿态和部分遮挡下的干扰示意图。

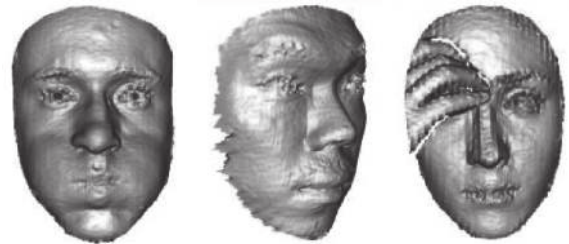


图6 Bosphorus 数据集干扰示意图

#### 3.2 描述子感受野

如 1.2 节所述, 局部特征描述子通过点对间的四维描述符对以  $p_i$  球心,  $k$  为半径区域内的点集进行几何特征描述, 其中半径  $k$  的大小决定了局部区域的范围。类似于卷积神经网络中的卷积核, 本文将  $k$  定义为局部特征描述子的感受野。而感受野的大小决定了网络对局部特征的感知能力。如表 1 所示, 本节列出了不同感受野在 CASIA-3D 数据集上的实验结果 (为体现网络自身性能, 此时将增强参数  $\lambda$  置 0)。

表1 感受野  $\lambda$  与模型准确率

感受野 $k$	准确率/%
0.2	86.6
<b>0.4</b>	<b>89.2</b>
0.6	82.7
0.8	77.9
1	73.6

由表 1 可知, 当感受野过小时, 网络获取的局部信息过多且细粒程度偏高, 导致部分全局特征的丢失; 而当感受野偏大时, 忽略部分局部特征, 同样会影响网络的感知能力。  $k = 0.4$  时, 网络准确率最高。因此, 本文将 0.4 设为感受野  $k$  的固定值, 并进行后续实验。

### 3.3 增强参数 $\lambda$

根据式 (8) 所示, 参数  $\lambda$  决定了特征增强机制对特征的增强程度,  $f_{en}$  为特征空间中每个维度的概率分布, 其值域为  $(0, 1)$ 。  $f_{norm}$  为双曲正切函数归一化后的 3D 点云特征, 其值域为  $[-1, 1]$ 。  $f_{en}$  与  $f_{norm}$  在相同的特征空间, 但取值范围不同, 而增强参数  $\lambda$  的取值决定了两者的耦合程度, 并对模型的性能起着至关重要的作用。

通常情况下, 可以通过深度学习网络拟合最优参数  $\lambda$ , 但实际中, 网络学习获取的  $\lambda$  会使  $\lambda f_{en}$  更拟合于  $f_{norm}$  的中位数, 而  $f_{norm}$  每个维度的取值在值域空间中差异较大, 此时  $\lambda f_{en}$  并不能起到最优增强作用。因此, 本文将  $\lambda$  设为固定值, 通过消融实验在 CASIA-3D 数据集上探究  $\lambda$  的最优值, 其结果如表 2 所示。

表 2 增强参数与模型准确率

增强参数 $\lambda$	准确率/%	增强参数 $\lambda$	准确率/%
0	89.2	55	98.9
5	89.2	60	98.8
10	89.1	65	98.7
15	89.5	70	98.3
20	91.7	75	98.1
25	92.9	80	98.1
30	94.8	85	97.7
35	96.3	90	97.1
40	97.6	95	96.5
45	98.0	100	95.6

根据表 2, 当  $\lambda$  为 0 时, 根据式 (8), 特征增强机制没有起到作用, 仅依靠归一化后的  $f_{norm}$  作为 FaceNet 的嵌入特征, 此时模型的识别准确率为 89.2%, 并以此为准线进行实验对比。随着  $\lambda$  的增大特征增强机制开始作用于  $f_{total}$ , 模型准确率开始提升, 当  $\lambda$  取 55 时模型准确率达到最大值 98.9%, 此时  $\lambda f_{en}$  在特征空间中与  $f_{norm}$  具有最佳耦合度, 为  $f_{total}$  提供最佳的特征区分度。随着  $\lambda$  的增大,  $\lambda f_{en}$  与  $f_{norm}$  的耦合程度降低, 模型性能开始下降。当  $\lambda = 100$  时,  $\lambda f_{en}$  与  $f_{norm}$  出现数量级差异, 此时模型主要依靠特征增强机制的概率分布  $\lambda f_{en}$  作为 FaceNet

的嵌入特征, 但依然保持了较高的准确率, 证明特征增强机制的  $f_{en}$  能很好地通过概率分布表征点云在高维空间中的差异。

为进一步探究增强参数  $\lambda$  在不同数据集上的性能, 本文根据表 2 的参数设置, 在 Lock3DFace 和 Bosphorus 数据上进行实验, 其准确率变化曲线如图 7 所示。在 Bosphorus 和 Lock3DFace 数据集上, 增强参数  $\lambda$  分别等于 50 和 45 时, 准确率最高。在不同数据集上, 由于样本分布不同 (面部表情、姿态干扰和部分遮挡), 不同的  $\lambda$  会对准确率造成不同的影响。然而, 根据表 2 所示数据以及图 7 变化曲线分析, 当  $\lambda \in [45, 60]$  时, 网络准确率达到较为稳定的峰值。这是因为人脸特征整体轮廓相似, 不同于 2D 图像分辨率的差异, 点云数据均为空间中的一组点集, 其结构简单且数学表达统一。同时, 如 1.2 节所述, 本文所采用的局部特征描述子满足平移不变性的要求, 点云人脸特征具有一定的相似性, 因此, 在不同数据集上,  $\lambda$  在较固定的取值范围内可使网络准确率达到最高。

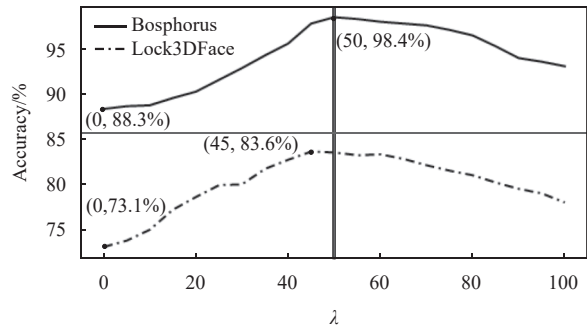


图 7 增强参数与准确率变化示意图

图 7 列出了  $\lambda = 0$  时, 网络在数据集 Bosphorus 和 Lock3DFace 上的准确率 (此时特征增强机制未起作用), 随着  $\lambda$  增加, 网络准确率明显提升, 准确率变化曲线证明特征增强机制能有效提升网络对点云特征的鉴别能力。

### 3.4 实验结果分析

相比于 RGB 图像, 点云数据缺乏细节纹理, 为公平对比, 本文仅与采用三维数据的方法进行实验对比。

将 CASIA-3D 数据集按照相同的方式划分, 并在正常光照和弱光场景下与近几年的方法进行实验对比, 其结果如表 3 所示。

根据表 3 所示, 在正常光照和缺乏光照的场景下, 本文的准确率均高于其他方法, 且在弱光条件下, 本文的准确率与正常光照下的准确率非常接

近, 证明本文采用的模型与数据在光照变化条件下具有良好的鲁棒性。

表 3 在 CASIA-3D 数据集上的准确率 %

方法	正常光照	弱光环境
文献[19]	98.4	96.04
文献[13]	97.9	95.4
文献[11]	97.9	97.3
<b>本文方法</b>	<b>98.9</b>	<b>98.4</b>

为验证本文方法在更多场景下的性能, 本文在 Lock3DFace 数据集上, 根据其对表情变化 (FE)、普通面部 (NU)、部分遮挡 (OC) 和头部姿态变化 (PS) 划分的测试序列进行测试, 其结果如表 4 所示。

表 4 在 Lock3DFace 数据集上的准确率 %

方法	干扰序列下的准确率				总计
	FE	NU	OC	PS	
文献[20]	96.1	99.3	54.9	61.4	76.6
文献[21]	93.6	99.0	57.0	54.1	74.4
文献[22]	95.7	98.9	61.4	69.9	79.5
文献[23]中方法	98.1	99.6	78.1	70.4	84.2
<b>本文方法</b>	<b>98.5</b>	<b>99.2</b>	<b>78.7</b>	<b>70.0</b>	<b>83.6</b>

如表 4 所示, 本文的方法与文献 [23] 的方法在面部表情、普通面部、部分遮挡和姿态变化测试序列上, 识别准确率均高于其他方法, 实验结果证明本文方法能较好地应对表情变化、部分遮挡以及头部姿态变化所带来的影响。与文献 [23] 的方法相比, 本文方法虽然整体准确率低于该方法, 但在面部表情变化以及部分遮挡情况, 本文方法准确率分别高了 0.4% 和 0.6%, 且在其他测试序列下, 准确率非常接近。

表 5 列出了在 Bosphorus 数据集上的实验对比结果。根据表 5 所示, 在正常光照条件下 (数据集中高质量样本), 虽然本文方法的准确率略低于文献 [13] 中的方法, 但在弱光条件下 (数据集中低质量样本), 本文方法的准确率均高于其他方法。由于 Bosphorus 数据集丰富的样本数据, 98.3% 的准确率证明了本文方法能较好地应对面部表情、头部姿态和部分遮挡对任务的干扰, 且对光照变化保持稳定。

为验证本文方法的时效性, 与近几年其他三维头部特征提取网络进行时效对比, 其结果如表 6 所示。由于特征增强机制采用概率分布对点云特征进行增强, 并未额外增加网络参数, 因此在提升网络

鉴别力的同时依然保持了良好的实时性。

表 5 在 Bosphorus 数据集上的准确率 %

方法	正常光照	弱光环境
文献[24]	97.3	95.2
文献[25]	95.2	95.1
文献[13]	98.5	96.5
文献[11]中方法	98.0	97.2
<b>本文方法</b>	<b>98.4</b>	<b>98.3</b>

表 6 时效性对比

方法	数据值/fps
文献[13]	7
文献[22]	6
文献[9]	117
文献[8]	125
文献[23]	136
文献[10]	148
<b>本文方法</b>	<b>218</b>

与文献 [23] 相比, 在各种场景下, 本文准确率与该方法非常接近, 但如表 6 所示, 本文方法耗时更短, 实时性更好。

综上所述, 本文采用三维点云作为输入数据, 通过浅层网络提取点云人脸特征, 且提出的特征增强机制结构简单、计算复杂度低。简洁的网络结构在提高准确率的同时也保证了良好的实时性。

## 4 结束语

为了解决在光照变化、面部表情、头部姿态和部分遮挡下的人脸识别问题, 本文直接采用 3D 点云作为输入, 并引入局部特征描述子和关键特征增强机制提升了特征的辨识度, 有效地提高了模型的识别准确率, 并保持了良好的实时性。该方法对光照变化保持稳定, 可以应用于缺乏光照或光照变化较大的场景, 如夜间驾驶认证、夜间检测等。

后期将进一步研究新的算法, 提升三维人脸的特征辨识度, 并应用于其他复杂情况下的人脸分析任务, 如头部检测、姿态估计和人脸验证等。

## 参考文献

- [1] BORGHI G, FABBRI M, VEZZANI R, et al. Face-from-depth for head pose estimation on depth images[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2018, 42(3): 596-609.
- [2] QI C R, SU H, MO K, et al. Pointnet: Deep learning on point sets for 3D classification and segmentation[C]// Proceedings of the IEEE Conference on Computer Vision

- and Pattern Recognition. Honolulu: IEEE, 2017: 652-660.
- [3] QI C R, YI L, SU H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: [s.n.], 2017: 5105-5114.
- [4] DENG H, BIRDAL T, ILIC S. PPFNet: Global context aware local features for robust 3D point matching[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 195-205.
- [5] LI Y, BU R, SUN M, et al. PointCNN: Convolution on x-transformed[EB/OL]. [2023-01-04]. <https://proceedings.neurips.cc/paper/2018/file/f5f8590cd58a54e94377e6ae2ded4d9-Paper.pdf>.
- [6] XIE Z, CHEN J, PENG B. Point clouds learning with attention-based graph convolution networks[J]. *Neurocomputing*, 2020, 402: 245-255.
- [7] ZHAO H, JIANG L, JIA J, et al. Point transformer[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Beijing: IEEE, 2021: 16259-16268.
- [8] 肖仕华, 桑楠, 王旭鹏. 基于深度学习的三维点云头部姿态估计[J]. *计算机应用*, 2020, 40(4): 996-1001.  
XIAO S H, SANG N, WANG X P. 3D point cloud head pose estimation based on deep learning[J]. *Journal of Computer Applications*, 2020, 40(4): 996-1001.
- [9] XIAO S H, SANG N, WANG X P, et al. Leveraging ordinal regression with soft labels for 3D head pose estimation from point sets[C]//ICASSP 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona: IEEE, 2020: 1883-1887.
- [10] WANG Q, LEI H, WANG X. Deep face verification under posture interference[J]. *Journal of Computer Applications*, 2022, 43: 595-600.
- [11] CAO Y, LIU S, ZHAO P, et al. RP-Net: PointNet++ 3D face recognition algorithm integrating RoPS local descriptor[J]. *IEEE Access*, 2022, 10: 91245-91252.
- [12] XU K, WANG X, HU Z, et al. 3D face recognition based on twin neural network combining deepmap and texture[C]//IEEE 19th International Conference on Communication Technology (ICCT). Xi'an: IEEE, 2019: 1665-1668.
- [13] DUTTA K, BHATTACHARJEE D, NASIPURI M. SPPCANet: A simple deep learning-based feature extraction approach for 3D face recognition[J]. *Multimedia Tools and Applications*, 2020, 79(41): 31329-31352.
- [14] GERA D, BALASUBRAMANIAN S. Landmark guidance independent spatio-channel attention and complementary context information based facial expression recognition[J]. *Pattern Recognition Letters*, 2021, 145: 58-66.
- [15] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: A unified embedding for face recognition and clustering[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 815-823.
- [16] BIOMETRICS C. Note on CASIA-3D FaceV1[EB/OL]. [2023-01-04]. <http://biometrics.idealtest.org>.
- [17] ZHANG J, HUANG D, WANG Y, et al. Lock3DFace: A large-scale database of low-cost kinect 3D faces[C]//2016 International Conference on Biometrics (ICB). Halmstad: IEEE, 2016: 1-8.
- [18] SAVRAN A, ALYÜZ N, DIBEKLIOĞLU H, et al. Bosphorus database for 3D face analysis[C]//Biometrics and Identity Management: First European Workshop, BIOID 2008. Berlin: Springer, 2008: 47-56.
- [19] CHANDRAKALA M, RAVI S. Effective 3D face recognition technique based on Gabor and LTP features [J]. *International Journal of Engineering and Advanced Technology (IJEAT)*, 2018, 8(2): 284-290.
- [20] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 770-778.
- [21] SZEGEDY C, VANHOUCKE V, IOFFE S J, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 2818-2826.
- [22] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 4510-4520.
- [23] MU G, HUANG D, HU G, et al. Led3D: A lightweight and efficient deep approach to recognizing low-quality 3D faces[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Los Angeles: IEEE, 2019: 5766-5775.
- [24] SOLTANPOUR S, WU Q J. High-order local normal derivative pattern (LNDP) for 3D face recognition [C]//IEEE International Conference on Image Processing (ICIP). Beijing: IEEE, 2017: 2811-2815.
- [25] DUTTA K, BHATTACHARJEE D, NASIPURI M, et al. 3D face recognition based on volumetric representation of range image[J]. *Advanced Computing and Systems for Security*, 2019, 8: 175-189.