基于人脸关键点与增量聚类的多姿态人脸识别

吴晓萍[1*]，管业鹏[1,2]

[1]上海大学通信与信息工程学院，上海 200444；

[2]上海大学新型显示技术及应用集成教育部重点实验室，上海 200072

摘要 人脸姿态变化复杂且对人脸识别性能影响明显，提出了一种融合 LCCDN（LSTM and CNN based Cascade Deep Network）与增量聚类的多姿态人脸识别方法。采用 LCCDN 模型定位人脸关键点，利用长短时记忆网络（LSTM）的记忆功能寻找人脸各关键点在空间上的全局上下文的依赖关系对人脸关键点初始化，并通过卷积神经网络模型，采用由粗到精的策略；定位人脸关键点；以人脸关键点作为人脸朝向描述子，同时为适应人脸姿态不断地动态更新，采用基于熵诱导度量机制的增量聚类方法，对头部姿态进行动态增量聚类，构建人脸姿态池。在此基础上，通过建立不同姿态的人脸识别分类模型实现多姿态人脸识别，在 CAS-PEAL-R1、CFP 和 Multi-PIE 三个数据集上的人脸识别准确率分别达到 96.75%，96.50%，97.82%。通过与同类人脸识别方法的客观定量对比，实验结果表明所提方法有效、可行。

关键词 图像处理；人脸识别；人脸关键点；增量聚类；多姿态

中图分类号 TP391.4 **文献标识码** A **doi**：10.3788/LOP56.141002

Multi-Pose Face Recognition Based on Facial Landmarks and Incremental Clustering

Wu Xiaoping[1*], Guan Yepeng[1,2]

[1] *School of Communication & Information Engineering, Shanghai University, Shanghai* 200444, *China*;

[2] *Key Laboratory of Advanced Display and System Applications, Ministry of Education, Shanghai University, Shanghai* 200072, *China*

Abstract Owing to complex changes in face pose and the obvious influence on face recognition performance, a new approach is proposed for multi-pose face recognition based on the fusion of the LSTM (long short term memory network) and convolutional neural network-based cascade deep network (LCCDN) and incremental clustering. First, a LCCDN is designed to locate facial landmarks, and the memory function of the LSTM in LCCDN is used to explore the spatial contextual information between facial landmarks; then, facial landmarks are initialized. A CNN network model is used to fine facial landmarks by employing a coarse-to-fine strategy. Next, we consider the facial landmarks as face orientation descriptors. Simultaneously, to adapt to the dynamic updating of the face pose, an entropy-induced metric-based incremental clustering method is used to construct a face-pose pool by dynamically clustering head poses. In this manner, multi-pose face recognition is realized by establishing various face classification models with different poses. The recognition accuracies using the CAS-PEAL-R1, CFP, and Multi-PIE datasets are 96.75%, 96.50%, and 97.82%, respectively. In addition, comparisons with existing multi-pose face recognition methods highlight the superior performance of the proposed method.

Key words image processing; face recognition; facial landmarks; incremental clustering; multi-pose

OCIS codes 100.4996; 150.1135

收稿日期：2018-12-29；**修回日期**：2019-01-16；**录用日期**：2019-02-17

基金项目：国家自然科学基金(11176016,60872117)

* **E-mail**：ypguan@shu.edu.cn

1 引　　言

近几十年来，人脸识别一直是计算机视觉中最受关注的研究主题之一。它在许多实际应用中具有巨大的潜力，如视频监控、社交网络、人脸解锁等。但是由于受到光照、姿态、遮挡和表情变化等因素的影响，人脸识别技术仍存在巨大的挑战，其中：丰富的姿态变化是影响人脸识别性能主要因素之一。

为解决人脸姿态变化对人脸识别的影响，基于多姿态的人脸识别方法相继被提出，主要分为两类。

第一类是将人脸标准化得到正脸图像。如 Ho 等[1]提出了利用马尔可夫随机场和信念传播算法来合成虚拟二维(2D)正脸图像。Ding 等[2]提出了一种稠密三维(3D)模型以生成正面姿势图像，但是这种方法将图像从 2D 转为 3D 时会引入噪声，因此难以建立准确的 3D 模型。在基于一些人脸标准化得到的正脸图像中，人脸关键点起到了关键性作用，如 Su 等[3]利用集成回归树(ERT)算法定位 68 个人脸关键点并对人脸进行对齐，利用仿射变换得到正脸图像。赵明华等[4]通过主动表现模型(AAM)进行人脸对齐，并基于线性回归方法寻求正、侧脸之间的关系进行姿态校正，用支持向量机(SVM)对校正后的人脸分类。但是这些方法均选用所有样本的平均值初始化所有人脸，导致初始定位不准确，且直接基于人脸关键点进行姿态校正会导致人脸变形。

另一类多姿态人脸识别方法是直接从人脸中提取具有姿态不变性的人脸特征。其中基于深度学习的方法是基于大量的人脸数据，使用复杂的网络结构和大量的网络参数，有效处理人脸因姿态变化引起的非线性变化，方法性能得到了显著的提高。如 Taigman 等[5]提出的 DeepFace 采用 3D 对齐，基于 9 层网络结构并用相似度来训练人脸分类器。龙鑫等[6]基于稠密连接卷积神经网络(CNN)，采用角度距离损失函数，使类内距离最小化，类间距离最大化。Parkhi 等[7]提出的 VGGFace 采用 37 层的 CNN 结构及 Softmax 在自建的人脸数据集 VggDataSet 上进行网络训练。这些方法在公开数据集上具有非常高的准确率，但是在姿态偏转角度较极端情况下(45°～90°)以及实际场景应用中，模型的识别性能明显下降。

综上所述，目前多姿态人脸识别中存在关键点定位不准确，姿态校正导致人脸变形，姿态偏转角度较大情况下识别性能明显下降等问题。为了解决以上问题，本文提出一种融合 LCCDN (LSTM and CNN based Cascade Deep Network)与增量聚类的多姿态人脸识别方法。采用由粗到精的 LCCDN 模型，利用长短时记忆网络(LSTM)的记忆功能寻找人脸各关键点在空间上的全局上下文依赖关系并对人脸关键点进行初始化，通过 CNN 模型对局部关键点精定位，同时为每个样本定义了姿态权重用于解决数据集姿态不平衡问题；以人脸关键点作为人脸朝向描述子，同时为适应人脸姿态不断动态更新，采用基于熵诱导度量机制的增量聚类方法，对头部姿态进行动态增量聚类，构建人脸姿态池以避免传统姿态校正方法带来的人脸变形问题。在此基础上，基于 VGGFace 模型，通过建立不同姿态的人脸识别分类模型，实现多姿态人脸识别。

2 基于 LCCDN 模型定位人脸关键点

2.1 LCCDN 级联深度网络框架

由于人脸关键点 $\boldsymbol{x}=(a_1,b_1,\cdots,a_M,b_M,M$ 为人脸关键点数量)构成的形状向量能够有效反映人脸朝向和姿态，因此如何精准定位人脸关键点十分关键。提出一种由粗到精的 LCCDN 框架用于人脸关键点定位，分别以横向、纵向扫描图像结合使用 LSTM 寻找人脸各关键点在空间上的上下文依赖关系并对其粗定位，各局部区域通过 CNN 模型和独立的损失函数对局部人脸关键点进行精定位，如图 1 所示。

假定人脸图像在高维空间中分布于某个低维流形附近，LCCDN 分别在两阶段使用不同的网络结构进行非线性映射表达流形间的映射。在粗定位阶段，本文使用 LSTM 构建全局网络，并对人脸关键点进行定位。LSTM 通过记忆细胞单元存储时间序列的上下文依赖信息，而在图像中这种时间上的依赖性则被转换到了空间域。全局网络每一层隐藏层由 4 个 LSTM 组成，并分别以横向和纵向两种图像扫描方式来构建时间序列，以学习像素之间的上下文依赖信息。

将输入的图像表示为 m，其中，$m\in\mathbf{R}^{w\times h\times c}$，$w$ 表示图像的宽度，h 表示高度，c 表示通道数，以 $w_{\mathrm{p}}\times h_{\mathrm{p}}$ 将图像 m 不重叠地分割成图像块 $p_{i,j}$，然后分别从左到右、从右到左横向扫描各图像块得到 2 个序列。将 2 个序列分别输入 2 个 LSTM，LSTM 每个记忆细胞的当前状态根据当前输入序列、前一个记忆细胞输出和存储在该记忆细胞中的

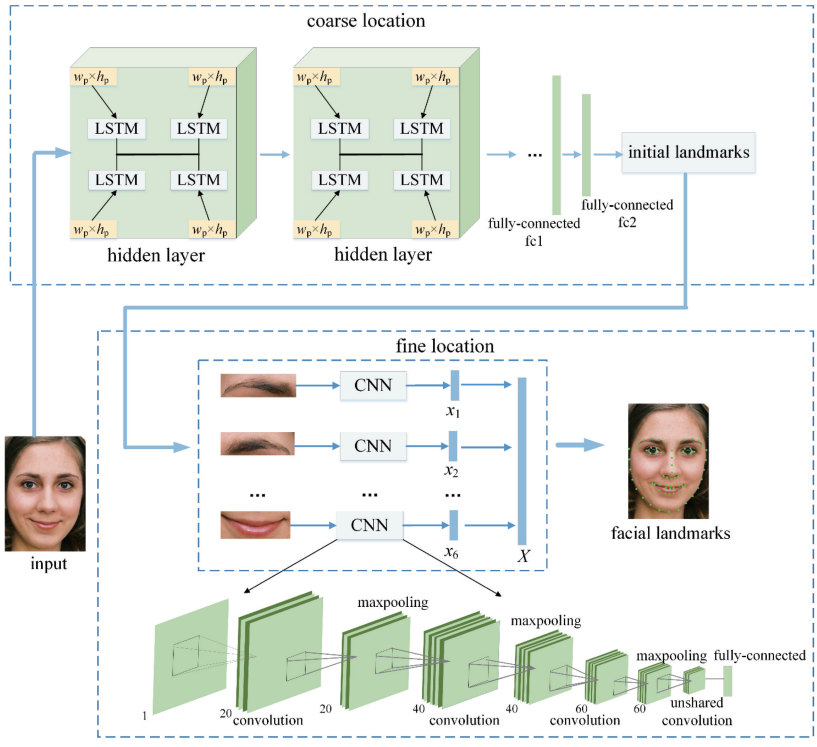

图 1 基于 LCCDN 模型的人脸关键点定位框架

Fig. 1 Framework of face landmark location based on LCCDN model

先前状态进行更新[8]，最后得到 LSTM 的输出为 h_{lf} 和 h_{lr}，然后将 h_{lf} 和 h_{lr} 串联得到 h_{l}。其中：$h_{\mathrm{l}} \in \mathbf{R}^{2e}$，$e$ 表示 LSTM 细胞单元的个数。同样地，再分别从上至下、从下至上纵向扫描图像并输入 LSTM 得到 h_{rf} 和 h_{rr}，串联得到 h_{r}，h_{r} 即为该隐藏层最后输出。多层堆叠的隐藏层后，通过 2 个全连接层将特征向量铺平并进行关键点预测，损失函数采用 L2 Loss。

全局网络的结构和参数如图 2 所示，图像输入大小为 112×112，将图像块 $p_{i,j}$ 的大小 $w_{\mathrm{p}} \times h_{\mathrm{p}}$ 设为 2×2，LSTM 长短时记忆网络细胞单元个数 e 设为 256，第 1 个全连接层的神经元个数为 1024，第 2 个全连接层作为预测层，神经元个数为 136，堆叠的隐藏层数量为 4。

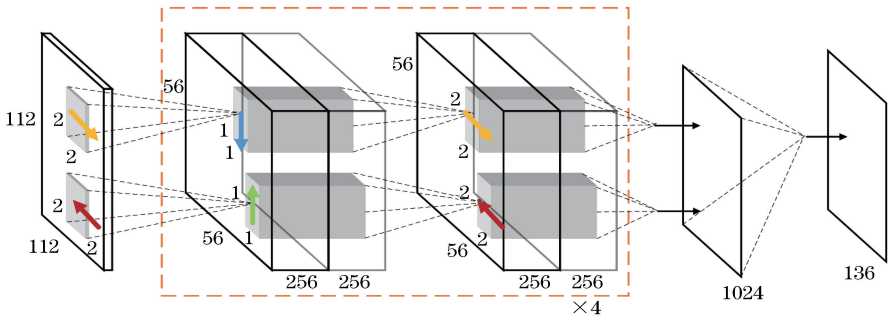

图 2 LSTM 全局网络结构图

Fig. 2 Structural diagram of LSTM global network

由于脸部不同区域关键点定位的难度存在差异，对于眉毛的定位误差比较大，而对眼睛的定位精度会比较高，因此眉毛的定位误差在损失函数中占有较大的比重[9]，在粗定位阶段使用整张人脸提取关键点无法有效避免训练不平衡的问题。因此 LCCDN 在精定位阶段使用浅层 CNN 分别对左眉毛、右眉毛、左眼、右眼、鼻子、嘴巴单独训练 6 个网络，每个区域有独立的损失函数。受文献[9]的启发，每个区域的浅层卷积神经网络结构如表 1 所示，损失函数采用 L2 Loss，其中最后一层全连接层的输出大小根据不同区域的关键点数设置。

表 1　浅层卷积神经网络结构

Table 1　Structure of shallow CNN

Network layer	Type	Filter	Output sizes	Others
Input	Input	—	40×40	—
Convolution 1	Convolution	5×5×20	36×36×20	—
Maxpooling 1	Max-pooling	2×2	18×18×20	—
Convolution 2	Convolution	3×3×40	16×16×40	—
Maxpooling 2	Max-pooling	2×2	8×8×40	—
Convolution 3	Convolution	3×3×60	6×6×60	—
Maxpooling 3	Max-pooling	2×2	3×3×60	—
Unshared Conv	Convolution	2×2×80	2×2×80	—
Fully-connected fc1	Fully-connected	—	120	—
Dropout1	Dropout	—	120	Keep_ratio is 0.6
Predicttion	Fully-connected	—	—	—

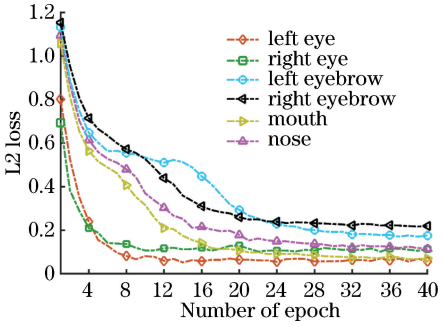

图 3　不同人脸区域的 CNN 损失函数下降曲线图

Fig. 3　Descent curves of CNN loss functions in different face regions

图 3 表示不同区域的 CNN 网络损失函数下降曲线图，数据集中全部样本训练完一次称为一个 epoch。由图 3 可知，左眼与右眼在迭代次数达到 8 个 epoch 时损失函数就开始收敛，而左眉毛和右眉毛在约 20 个 epoch 时，该区域的 CNN 网络损失函数才收敛。每个区域根据各自的损失函数下降来控制迭代次数，有效避免了网络过拟合，解决了粗定位阶段网络训练不平衡的问题。

2.2　人脸多姿态偏差重加权

LCCDN 结构中，在分别训练 LSTM 全局网络和 CNN 时，当前的公开人脸关键点数据集（如 300-W[10]）中存在以下的问题：正脸姿态的样本所占比重较大，而大角度人脸姿态样本占比较少，这是由于样本的不平衡性可能导致网络对姿态偏转角度较大的样本欠拟合。为了解决这一问题，基于样本姿态偏差对样本进行加权，样本姿态偏差（P_{dev}）通过样本中已标注的形状向量 $\boldsymbol{x}=(a_1,b_1,\cdots,a_M,b_M)$ 与正脸之间的距离来量化，样本姿态偏差（P_{dev}）计算公式为

$$P_{\mathrm{dev}}(\boldsymbol{x})=\frac{1}{M}\frac{\sum_{j=1}^{M}\sqrt{(a_j-\bar{a}_j)^2+(b_j-\bar{b}_j)^2}}{d},\tag{1}$$

式中：$\bar{a}_j$、$\bar{b}_j$ 表示正脸的各关键点坐标值；d 表示左右眼瞳孔的距离，通过眼间距进行归一化以排除人脸实际大小和摄像头变焦等因素的干扰。

根据样本姿态偏差，每个样本的姿态权重可表示为

$$w_{\mathrm{s}}=\frac{\exp[\eta\cdot P_{\mathrm{dev}}(\boldsymbol{x})]}{Z},\tag{2}$$

式中：η 为常数。Z 为归一化因子，表达式为

$$Z=\sum_{i=1}^{N_{\mathrm{b}}}\exp[\eta\cdot P_{\mathrm{dev}}(\boldsymbol{x}_i)],\tag{3}$$

式中：N_{b} 表示网络训练一次的样本数（即 batch size）。

最终，使用样本权重修改 L2 Loss 损失函数 C_{L2}，最终的损失函数

$$C=w_{\mathrm{s}}\cdot C_{\mathrm{L2}}。\tag{4}$$

使用样本姿态权重 w_{s} 对损失函数进行加权，在训练过程中姿态较大样本根据（4）式可以动态提高损失函数中的权重，以解决样本姿态不平衡的问题。

3　人脸姿态动态自适应增量聚类

基于 LCCDN 能提取稳健的人脸关键点，为避免直接基于人脸关键点进行姿态校正导致的人脸变形问题，以人脸关键点作为特征向量，将不同姿态自动聚类成多个子集，构建人脸姿态池模型。同时为适应实际的身份识别系统中人脸姿态不断动态更新的情况，使用增量聚类算法构建人脸姿态池，利用前一次的结果来加速本次聚类过程。应用广泛的贝叶斯自适应共振理论（BART）[11]，因使用各类簇中所有数据计算最大样本容量，故易引入噪声误差。为了解决这一问题，本文提出一种基于相关熵诱导度量（CIM）的 CBART 增量聚类算法，使用 CIM 代替协方差矩阵来计算簇的最大样本容量。

在 CBART 算法中，类簇的形状由多元高斯分布的密度函数表示，每个类簇的高斯密度函数由均值向量、协方差矩阵和先验概率组成。根据贝叶斯概率公式，样本 $\boldsymbol{x}$ 在类簇 j 中的后验概率可表示为

$$P(\boldsymbol{y}_j \mid \boldsymbol{x})=\frac{P(\boldsymbol{x} \mid \boldsymbol{y}_j)P(\boldsymbol{y}_j)}{\sum_{k=1}^{K}P(\boldsymbol{x} \mid \boldsymbol{y}_k)P(\boldsymbol{y}_k)}, \tag{5}$$

式中：$\boldsymbol{x}$ 表示人脸关键点样本；$P(\boldsymbol{y}_j)$表示类簇 j 的估计先验概率；K 表示类簇的数量；$P(\boldsymbol{x}\mid\boldsymbol{y}_j)$则由类簇 j 的多元高斯函数估计得到，即

$$P(\boldsymbol{x} \mid \boldsymbol{y}_j)=\frac{1}{(2\pi)^{\frac{K}{2}}\mid\boldsymbol{M}_j\mid^{\frac{1}{2}}}\cdot \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{V}_j)^{\mathrm{T}}\boldsymbol{M}_j^{-1}(\boldsymbol{x}-\boldsymbol{V}_j)\right], \tag{6}$$

其中 $\boldsymbol{V}_j$ 和 $\boldsymbol{M}_j$ 分别表示类簇 j 的均值向量和协方差矩阵。

最大后验概率值的对应类簇则为获胜类簇 J，表达式为

$$J=\underset{k\in K}{\operatorname{argmax}}[P(\boldsymbol{y}_j \mid \boldsymbol{x})]。 \tag{7}$$

为了有效限制簇内人脸关键点样本数量的增长，设置警戒阈值 $S_{\max}$。使用 CIM 来代替协方差矩阵来计算簇的最大样本容量 S_j。协方差矩阵因使用所有数据计算 S_j 易引入噪声，CIM 是一种基于核函数的相似性度量方法，可直接量化新增样本和聚类簇的概率分布之间的相似性。因此，在高维和噪声样本的环境下，使用 CIM 对样本和聚类簇进行相似度匹配，拥有更好的稳健性。

对于有限长的人脸关键点样本 $\boldsymbol{X}=(x_1,x_2,\cdots,x_L)$和 $\boldsymbol{Y}=(y_1,y_2,\cdots,y_K)$，CIM 可表示为

$$M_{\mathrm{CI}}(\boldsymbol{X},\boldsymbol{Y})=\left\{\frac{1}{L}\sum_{l=1}^{L}[k_\sigma(0)-k_\sigma(\boldsymbol{x}_l-\boldsymbol{Y})]\right\}^{\frac{1}{2}}, \tag{8}$$

式中：k_σ 表示核函数，本文使用应用最广泛的高斯核函数；L 表示输入人脸关键点样本的数量。则基于 CIM 的最大样本容量 S_j 表达式为

$$S_j=[k_{\sigma_{\mathrm{cim}}}(0)-k_{\sigma_{\mathrm{cim}}}(\parallel\boldsymbol{x}-\boldsymbol{V}_j\parallel)]^{\frac{1}{2}}, \tag{9}$$

式中：σ_{cim}表示高斯核函数的带宽；$\boldsymbol{V}_j$ 表示类簇 j 的均值向量(聚类中心)。类簇样本容量 S_j 须满足

$$S_j\leqslant S_{\max}。 \tag{10}$$

获胜类簇 J 若满足(10)式，则根据新增的人脸关键点样本 x 更新均值向量 $\boldsymbol{V}_{J,\mathrm{new}}$和协方差矩阵 $\boldsymbol{M}_{J,\mathrm{new}}$，更新 (11)式和(12)式，否则寻找下一个类簇直至满足条件为止。如寻找失败，则根据输入的人脸关键点样本建立新的簇，新簇的均值向量 $\boldsymbol{V}_j=\boldsymbol{x}$，初始化协方差为一个极小值矩阵并满足(10)式。

$$\boldsymbol{V}_{J,\mathrm{new}}=\frac{N_J}{N_J+1}\boldsymbol{V}_{J,\mathrm{old}}+\frac{1}{N_J+1}\boldsymbol{x}, \tag{11}$$

$$\boldsymbol{M}_{J,\mathrm{new}}=\frac{N_J}{N_J+1}\boldsymbol{M}_{J,\mathrm{old}}+\frac{1}{N_J+1}(\boldsymbol{x}-\boldsymbol{V}_{J,\mathrm{new}})(\boldsymbol{x}-\boldsymbol{V}_{J,\mathrm{new}})^{\mathrm{T}}, \tag{12}$$

式中：N_j 表示类簇 J 的人脸关键点样本数量。

4 实验结果及讨论

4.1 CBART 增量聚类算法参数讨论

本文软件运行环境为 Linux、python、opencv 开源视觉库和 tensorflow 开源机器学习库，硬件配置为 NVIDIA GeForce GTX 1070 Ti GPU，16G RAM 内存。在人脸姿态池的基础上，使用预训练好的 VGGFace 模型进行微调训练，对不同姿态子集分别建立人脸识别模型。为验证本文方法的有效性，采用五折交叉验证分别在多姿态人脸识别数据集 CAS-PEAL-R1[12]、CFP[13]和 Multi-PIE[14]上训练。

CBART 增量聚类 $S_{\max}$的取值决定了姿态池的数量和大小。选取合适的 $S_{\max}$ 值使得召回率(TPR)较高，虚警率(FPR)较低，可以同时降低误检和漏检的数量，从而得到良好的识别效果。

$$R_{\mathrm{TP}}=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FN}}}, \tag{13}$$

$$R_{\mathrm{FP}}=\frac{N_{\mathrm{FP}}}{N_{\mathrm{FP}}+N_{\mathrm{TN}}}, \tag{14}$$

式中：R_{TP}表示召回率；R_{FP}表示虚警率；N_{TP}代表分类识别中正确识别为正样本的数量；N_{FN}表示被识别为负样本的正样本数量；N_{TN}表示为被正确识别为负样本的数量。

由图 4 可知，随着 $S_{\max}$值的增大，TPR 和 FPR 逐渐上升，当 $S_{\max}=0.30$ 和 0.35 时，TPR 基本达到最大值，接近 0.97，而 FPR 持续递增。选择 $S_{\max}=0.30$，以均衡 TPR 和 FPR 曲线使得 TPR 较高的同时 FPR 较低，并在后续的实验中保持不变。

4.2 人脸关键点定位实验结果

为验证本文所提 LCCDN 人脸关键点定位框架的效果，使用 300-W[10] 作为数据集对网络进行训练。该数据库中包含了 3837 张图像，每个人脸标定了 68 个关键点。采用五折交叉验证方法进行训练，采用平均误差和失败率评估实验结果。平均误差指预测的人脸关键点和数据集中标注好的关键点之间的平均距离，表达式为

$$E=\frac{1}{N}\sum_{n=1}^{N}\frac{\frac{1}{M}\sum_{m=1}^{M}\sqrt{(a_{n,m}^{\mathrm{p}}-a_{n,m}^{\mathrm{g}})^2+(b_{n,m}^{\mathrm{p}}-b_{n,m}^{\mathrm{g}})^2}}{d_n}, \tag{15}$$

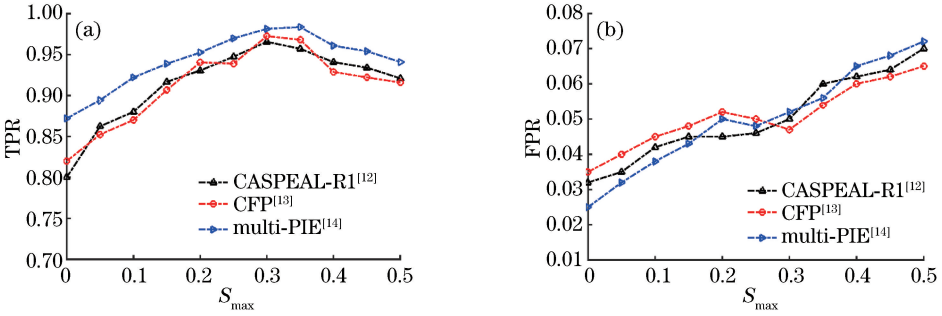

图 4 多姿态人脸识别在不同 S_{max} 下的分类识别性能。(a) TPR;(b) FPR

Fig. 4 Classification performance of multi-pose face recognition under different S_{max}. (a) TPR; (b) FPR

式中:M 为人脸关键点数量;N 为样本数量;a^p 和 b^p 表示预测出的关键点的坐标;a^g 和 b^g 表示人工标注的关键点的坐标;d_n 表示人工标注的左右眼的距离。若单张人脸误差大于 0.1,则认为该张人脸关键点定位失败。若输入 N 张图像,定位失败的图像数量为 n_0,则失败率为

$$R_F = \frac{n_0}{N} \times 100\%。 \tag{16}$$

为了评价本文所提 LCCDN 人脸关键点定位方法的有效性,在 300-W 数据集上与其他人脸关键点定位方法进行对比,实验结果如表 2 所示。

表 2 不同人脸关键点定位方法实验对比

Table 2 Experimental comparison of different facial landmark location methods

Method	Mean error /10^{-2}	Failure rate /%
ERT[3]	7.96	13.06
AAM[4]	7.58	12.56
CFCNN[9]	6.31	10.20
TCDCN[15]	4.60	6.59
LCCDN	4.06	5.26

由表 2 的对比实验结果可知,在 300-W 数据集上,本文方法的平均误差为 0.0406,失败率为 5.26%,均低于其他人脸关键点定位方法,这主要是因为 LCCDN 能够充分利用人脸各关键点在空间上的上下文依赖关系初始化关键点,而 ERT[3] 和 AAM[4] 中均用所有样本的平均值初始化所有人脸,导致人脸关键点初始定位不准确。LCCDN 在精定位阶段使用不同损失函数独立训练 6 个局部区域网络,避免不同区域的人脸训练不平衡,而 TCDCN (Tasks-Constrained Deep Convolutional Network)[15] 对所有关键点使用同一个损失函数,易导致网络出现过拟合。

本文在自己采集的监控场景身份视频库中进一步验证 LCCDN 的人脸关键点定位效果,该视频库采用海康威视 1080P 高清网络相机进行采集,视频库中共包含 99 个行人,每个行人在场景中的运动包含正对相机及左右偏转 30°、60°、90°的直线行走及一段任意角度随意行走,部分采集片段如图 5 所示。

基于 LCCDN 的人脸关键点定位部分定性结果如图 6 所示。其中:第 1 行是在人脸关键点数据集 300-W 上的部分检测结果;第 2 行是在多姿态人脸识别数据集 CFP 上的部分检测结果;第 3 行是在监控视频场景下部分检测结果。

图 5 监控场景身份视频库中部分采集片段示例

Fig. 5 Examples of some clips in surveillance video dataset

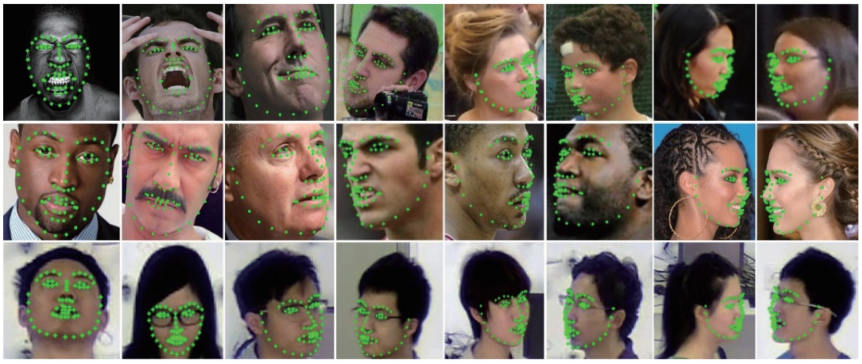

图 6 基于 LCCDN 提取的人脸关键点定位在不同姿态下的部分定性结果

Fig. 6 Qualitative results of facial landmark location based on LCCDN with various poses

4.3 客观定量评价与对比

为进一步验证本文采用 LCCDN 提取的人脸关键点作为人脸朝向描述子构建姿态池时对人脸识别的有效性，采用其他人脸朝向描述子基于 CBART 增量聚类构建姿态池作对照组。以人脸识别准确率 R_{acc} 作为评价指标，所得实验结果如表 3 所示。

$$R_{acc}=\frac{N_p}{N_p+N_w}\times 100\%, \tag{17}$$

式中：N_p 表示被识别正确的人脸数量；N_w 表示被识别错误的人脸数量。

表 3 基于不同人脸朝向描述子的多姿态人脸识别实验对比

Table 3 Experimental comparison of multi-pose face recognition based on different face orientation descriptors

Facial orientation descriptor	Accuracy /%		
	CASPEAL-R1[12]	CFP[13]	Multi-PIE[14]
D-LGBPH[16]	91.75	91.25	92.63
ASIFT[17]	91.53	90.77	91.03
CFCNN[9]	93.54	92.28	94.16
TCDCN[15]	93.89	94.24	94.82
LCCDN	96.75	96.50	97.82

由表 3 可知，使用人脸关键点作为人脸朝向描述子[9,15]进行多姿态人脸识别的准确率均高于其他特征描述子[16-17]，其中 LCCDN 提取的人脸关键点作为人脸朝向描述子的准确率最高。D-LGBPH[16]采用直方图表示特征能力，因局部二值模式种类较多导致直方图过于稀疏。ASIFT[17]对不同姿态的人脸具有仿射不变性，但姿态差异大于 60°时其仿射不变性无法保持。

为进一步验证本文采用的 CBART 增量聚类算法对多姿态人脸识别的有效性，以 LCCDN 提取的人脸关键点作为人脸朝向描述子，采用不同的聚类算法构建姿态池作对照组，所得实验结果如表 4 所示。

表 4 基于不同聚类算法的多姿态人脸识别实验对比

Table 4 Experimental comparison of multi-pose face recognition based on different clustering methods

Clustering method	Accuracy /%		
	CASPEAL-R1[12]	CFP[13]	Multi-PIE[14]
BART[11]	93.85	94.13	95.28
FCM[18]	90.47	91.28.	90.86
K-means[19]	91.39	92.05	93.90
CBART	96.75	96.50	97.82

由表 4 可知，基于 CBART 增量聚类算法构建人脸姿态池的人脸识别方法相对于 BART[11]算法识别效果提升，这主要是因为 BART 增量聚类中引入了样本的噪声误差。而模糊 C-均值(FCM)聚类算法[18]和 K 均值(K-means)[19]聚类算法需要事先确定聚类的个数，而不同的样本库由于姿态的差异，聚类的个数不能事先确定，且 FCM 聚类的中心是随机选取，其算法性能严重依赖聚类中心集的初始化，从而影响了人脸姿态池的性能。

为了验证本文方法的有效性，采用其他多姿态人脸识别方法作对比，所得实验结果如表 5 和图 7 所示。

表 5 不同人脸识别方法实验对比

Table 5 Experimental comparison of different face recognition methods

Method	Accuracy /%		
	CASPEAL-R1[12]	CFP[13]	Multi-PIE[14]
HPN[2]	90.15	89.17	89.39
VGGFace[7]	93.20	92.89	92.78
TPCNN[20]	90.89	90.53	91.39
DFLP[21]	92.56	91.25	92.16
Proposed	96.75	96.50	97.82

由表 5 可知，采用不同方法对人脸进行识别，本文方法效果最佳，在 CAS-PEAL-R1、CFP 和 Multi-PIE 三个数据集上的人脸识别准确率分别达到

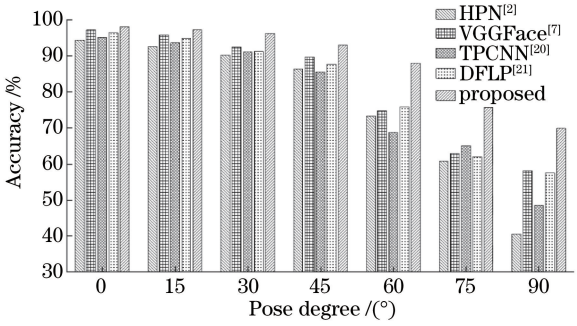

图 7 不同人脸识别方法在各姿态偏转角度下的准确率

Fig. 7 Accuracies of different face recognition methods with various poses

96.75%,96.50%,97.82%。由图 7 可知,在大姿态下(45°到 90°)本文方法的平均识别准确率明显高于其他方法。主要原因在于:HPN[2]方法使用 3D 模型生成正面姿势图像,易导致人脸变形。而本文方法直接对头部姿态进行自动增量聚类,有效地避免了直接生成正面虚拟图像导致的人脸变形现象。VGGFace[7]和 DFLP[21]方法使用 CNN 对人脸数据集进行训练,并分别通过优化损失函数提高准确率,但并未对人脸姿态信息进行约束,因此在大姿态下识别性能不佳。而本文方法基于人脸关键点将姿态自动聚类成不同的子集,对各子集分别建立人脸识别分类模型以适应各种不同的姿态变化。

5 结 论

提出一种基于融合 LCCDN 与增量聚类的多姿态人脸识别方法。采用由粗到精的 LCCDN 结构用于人脸关键点定位,以人脸关键点作为朝向描述子并基于熵诱导度量机制对头部姿态进行增量聚类,将人脸姿态会聚成不同的子集,在此基础上建立不同姿态子集的人脸识别分类模型。实验结果表明,本文方法能进一步提高多姿态人脸识别率。

参 考 文 献

[1] Ho H T, Chellappa R. Pose-invariant face recognition using Markov random fields[J]. IEEE Transactions on Image Processing, 2013, 22(4): 1573-1584.

[2] Ding C X, Tao D C. Pose-invariant face recognition with homography-based normalization[J]. Pattern Recognition, 2017, 66: 144-152.

[3] Su Y, Gao X B, Yin X C. Fast alignment for sparse representation based face recognition[J]. Pattern Recognition, 2017, 68: 211-221.

[4] Zhao M H, Mo R Y, Shi Z H, *et al*. A novel method for recognition of pose invariant face with single image[J]. Journal of Xi′an University of Technology, 2017, 33(1): 18-23.
赵明华, 莫瑞阳, 石争浩, 等. 一种新的基于单视图的多姿态人脸识别方法[J]. 西安理工大学学报, 2017, 33(1): 18-23.

[5] Taigman Y, Yang M, Ranzato M, *et al*. DeepFace: closing the gap to human-level performance in face verification[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE, 2014: 1701-1708.

[6] Long X, Su H S, Liu G H, *et al*. A face recognition algorithm based on angular distance loss function and convolutional neural network[J]. Laser & Optoelectronics Progress, 2018, 55(12): 121505.
龙鑫, 苏寒松, 刘高华, 等. 一种基于角度距离损失函数和卷积神经网络的人脸识别算法[J]. 激光与光电子学进展, 2018, 55(12): 121505.

[7] Parkhi O M, Vedaldi A, Zisserman A. Deep face recognition[C]//Proceedings of the British Machine Vision Conference 2015, September 7-10, 2015, Swansea, UK. Durham, England, UK: BMVA Press, 2015: 41.

[8] Greff K, Srivastava R K, Koutník J, *et al*. LSTM: a search space odyssey[J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 28(10): 2222-2232.

[9] Zhou E J, Fan H Q, Cao Z M, *et al*. Extensive facial landmark localization with coarse-to-fine convolutional network cascade[C]//2013 IEEE International Conference on Computer Vision Workshops, December 2-8, 2013, Sydney, NSW, Australia. New York: IEEE, 2013: 386-391.

[10] Sagonas C, Tzimiropoulos G, Zafeiriou S, *et al*. 300 faces in-the-wild challenge: the first facial landmark localization challenge[C]//2013 IEEE International Conference on Computer Vision Workshops, December 2-8, 2013, Sydney, NSW, Australia. New York: IEEE, 2013: 397-403.

[11] Chin W H, Loo C K, Seera M, *et al*. Multi-channel Bayesian adaptive resonance associate memory for on-line topological map building[J]. Applied Soft Computing, 2016, 38: 269-280.

[12] Gao W, Cao B, Shan S G, *et al*. The CAS-PEAL large-scale Chinese face database and baseline

evaluations [J]. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2008, 38(1): 149-161.

[13] Sengupta S, Chen J C, Castillo C, *et al*. Frontal to profile face verification in the wild [C] // 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), March 7-10, 2016, Lake Placid, NY, USA. New York: IEEE, 2016: 7477558.

[14] Gross R, Matthews I, Cohn J, *et al*. Multi-PIE [J]. Image and Vision Computing, 2010, 28(5): 807-813.

[15] Zhang Z P, Luo P, Loy C C, *et al*. Learning deep representation for face alignment with auxiliary attributes [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(5): 918-930.

[16] Chu W B, Guan Y P. Identity verification based on facial pose pool and bag of words model [J]. Journal of Advanced Computational Intelligence and Intelligent Informatics, 2017, 21(3): 448-455.

[17] Oji R. An automatic algorithm for object recognition and detection based on ASIFT keypoints [J]. Signal & Image Processing, 2012, 3(5): 29-39.

[18] Zhu Z L, Wang J F. Image segmentation based on adaptive fuzzy C-means and post processing correction [J]. Laser & Optoelectronics Progress, 2018, 55(1): 011004.
朱占龙, 王军芬. 基于自适应模糊 C 均值与后处理的图像分割算法 [J]. 激光与光电子学进展, 2018, 55(1): 011004.

[19] Zhan Y J, Dai T D, Huang J J, *et al*. Synthetic aperture radar image filtering based on clustering three-dimensional block-matching [J]. Laser & Optoelectronics Progress, 2018, 55(4): 041004.
詹云军, 代腾达, 黄解军, 等. 基于聚类三维块匹配的合成孔径雷达影像滤波算法 [J]. 激光与光电子学进展, 2018, 55(4): 041004.

[20] Zhao L T, Lin J J. Pose-invariant face recognition with two-pathway convolutional neural network [J/OL]. Journal of East China University of Science and Technology (Natural Science Edition). http://kns.cnki.net/kcms/detail/31.1691.TQ.20180611.1032.011.html.
赵澜涛, 林家骏. 基于双路 CNN 的多姿态人脸识别方法 [J/OL]. 华东理工大学学报. http://kns.cnki.net/kcms/detail/31.1691.TQ.20180611.1032.011.html.

[21] Wen Y D, Zhang K P, Li Z F, *et al*. A discriminative feature learning approach for deep face recognition [M] // Leibe B, Matas J, Sebe N, *et al*. Lecture notes in computer science. Cham: Springer, 2016, 9911: 499-515.