基于沙漏网络的人脸面部特征点检测

赵威驰[1]，赵其杰[1,2*]，江俊晔[1]，卢建霞[1]

[1]上海大学机电工程与自动化学院，上海 200444；

[2]上海市智能制造及机器人重点实验室，上海 200444

摘要 针对头部姿态变化较大、脸部遮挡等情况下，由面部特征类型多样和尺度不同造成的面部特征点检测准确度较低的问题，提出了一种面部分组特征线条化和点热图回归相结合的人脸特征点检测方法，并设计了两段式堆叠沙漏网络深度学习模型来实现图像特征分析与特征点定位。利用提出的方法开发了检测算法，并利用该领域几个典型的公共图像数据集，将所提方法与其他方法进行实验对比。结果表明，提出的方法可以适应姿态变化和脸部部分遮挡的应用，相比其他方法，具有检测误差较小、人脸面部特征点检测准确度较高的优势。

关键词 机器视觉；人脸特征点；堆叠沙漏网络；特征线条化；热图回归

中图分类号 TP301 **文献标识码** A **doi**：10.3788/AOS201939.1115003

New Method for Face Landmark Detection Based on Stacked-Hourglass Network

Zhao Weichi[1], Zhao Qijie[1,2*], Jiang Junye[1], Lu Jianxia[1]

[1] *School of Mechatronic Engineering and Automation, Shanghai University, Shanghai* 200444, *China*;

[2] *Shanghai Key Laboratory of Intelligent Manufacturing and Robotics, Shanghai* 200444, *China*

Abstract A method that combines facial dividing feature line and point heatmap regression is proposed to address the problem of low accuracy of face landmark detection caused by different facial feature types and scales in the cases of large posture changes and occlusion. A deep learning model based on two-stage stacked hourglass network is designed to realize feature analysis and landmark location. Based on the proposed method, the detection algorithm is developed, and the proposed method is compared with other methods by experiments based on several common image datasets. The experimental results show that the proposed method can adapt to the applications of large posture changes and face partial occlusion. Compared with other methods, the proposed method has less detection error and higher accuracy in face landmark detection.

Key words machine vision; face landmarks; stacked-hourglass network; feature line; heatmap regression

OCIS codes 150.1135; 100.4999; 100.4996; 200.4260

1 引 言

人脸面部特征检测是人脸识别、人脸认证等信息处理的关键。面部特征检测主要检测人脸的整体结构和几何特征，即人脸的关键特征点。人脸的面部特征具有不同的尺度和类型，如眼睛和鼻子的图像尺度和特征类型相差较大，如果直接从整体上求取面部不同特征的关键点，则在一定的姿势变化和遮挡干扰下无法确保特征点的精度。

面部特征点检测方法包括基于主动外观模型(AAM)和主动形状模型(ASM)的方法[1-2]、基于级联回归和深度网络的方法，以及近期提出的热图回归方法。ASM 和 AAM 利用形状变化模型与纹理变化模型检测特征点，容易受到形状特点变化的影响。在基于级联回归和深度网络的方法中：基于级联的姿态回归(CPR)[3]利用级联回归框架解决姿态估计问题，使用形状索引的关键点特征和级联回归器来预测形状残差；利用稀疏分布存储器模型(SDM)[4]提取尺度不变特征变换(SIFT)特征，并采用更简单的线性回归量。近年来，深度网络在级联回归框架下的面部特征点检测领域取得了较大的进步：Sun 等[5]提出级联深度卷积神经网络(DCNN)

收稿日期：2019-05-07；**修回日期**：2019-06-20；**录用日期**：2019-07-15

基金项目：上海汽车工业科技发展基金(1735)

* **E-mail**：zqj@shu.edu.cn

来逐步预测面部特征点；从粗到细的端到端递归卷积系统(MDM)[6]与 DCNN 类似，但其每个阶段都将前一阶段的隐藏层特征作为输入；Lü 等[7]将面部分成几个部分以减轻面部部分特征的变化，并分别回归不同部分的坐标，使坐标回归模型具有无需任何后处理即可推断特征点坐标的优点。这类方法虽然精度比 AAM 和 ASM 高，但无法适应较大的姿态变化，人脸特征点检测表现不如热图回归深度学习模型。热图回归模型分别为每个特征点生成概率热图，在人脸特征点检测中表现优异。Newell 等[8]使用热图回归模型，并设计出堆叠沙漏网络(Stacked-Hourglass)，从多尺度提取特征来估计人体姿态关键点。堆叠沙漏网络可以反复获取不同尺度下图像所包含的信息，比较适用于人脸特征点检测。Yang 等[9]使用标准化面部的监督变换和堆叠沙漏网络来获得预测热图，取得了较好的效果，证明了堆叠沙漏网络在面部特征检测上的优越性。Wu 等[10]使用人脸边界热图代替人脸特征点热图来表达人脸几何结构，证明了边界信息的重要性。但是，上述研究只是部分地解决了面部特征类型不同和面部特征尺度不同的问题，而且组合过多的堆叠沙漏网络会影响检测速度。

本文提出并设计了一种有效的人脸特征检测算法，解决了面部特征尺度不同和特征类型不同情况下，具有姿势变化、光照变化和遮挡干扰的人脸面部信息的提取问题，并提升了检测的精度。该研究主要利用面部分组特征，并将其全简化为线条来解决特征尺度不同和特征类型不同的问题。面部特征分开检测，特别适用于初步解决特征尺度不同的问题；特征线条化则能进一步解决特征类型不同的问题，将复杂多变的多维面部特征简化为简单同一的低维面部特征线条。用分组的面部特征线条组成整张全面部特征线条图，在此基础上进行点热图回归。这样可以避免大量干扰特征，相比直接在面部图像上进行检测，所提方法更加高效。也就是说，通过面部分组特征线条化和点热图回归相结合的方法(FDL-PHR)可求解面部特征点。

2　基本原理

2.1　堆叠沙漏网络

堆叠沙漏网络可以反复获取不同尺度下图像所包含的信息，适用于检测人脸特征点。此模型基于残差模块。残差模块能够基于卷积运算提取高级特征，同时可以通过跳过路径保留原始信息。如图 1 所示，特征图经残差模块卷积输入(input)，之后进行池化下采样(max pooling)。下采样之前的特征图经过卷积得以保留，下采样之后的特征图则继续卷积下采样操作。当达到一定分辨率后，利用残差模块卷积和上采样对特征图进行处理，并将其与之前保留下来的具有相同分辨率特征的图像相融合(addition)，如此反复，一直到达到原始分辨率为止，形成输出(output)。

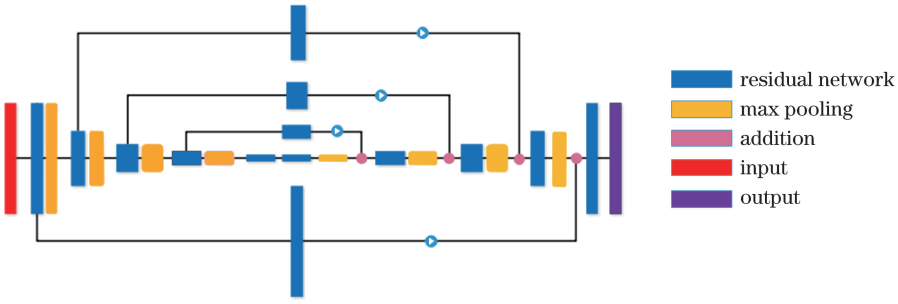

图 1　堆叠沙漏网络结构

Fig. 1　Structure of stacked-hourglass network

2.2　检测方法

检测过程框架如图 2 所示，主要包括 5 个阶段：输入原始图像，面部特征线条化，点热图回归，阈值化确定坐标和输出面部特征点。首先，将面部特征转化为特征线条。将面部特征线条分为脸部轮廓、鼻子、眉毛、眼睛和嘴唇等 5 个部位的特征线条图，其中，由于嘴唇轮廓十分接近，将其单独拆分成上下嘴唇外边、上嘴唇内边和下嘴唇内边。在得到 6 个面部特征线条图后，将其组合成整个面部的特征线条图，即全面部特征线条图。在全面部特征线条图中使用堆叠沙漏网络求取特征点热图。在点热图中进行阈值分割和处理后，就能得到特征点 j 的具体坐标(x_j^*, y_j^*)。

使用两段式堆叠沙漏网络提取脸部不同部位的特征线条，来解决遮挡和姿态变化引起的特征线条提取问题。脸部不同部位的特征线条可以互相参考

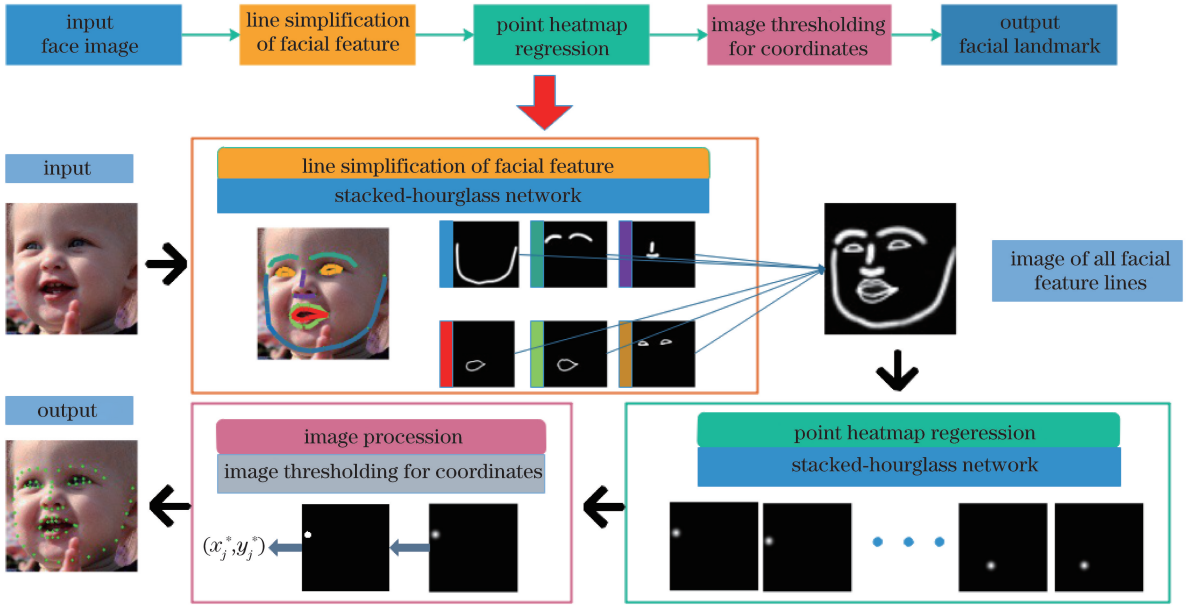

图 2　面部特征点检测方法的总体框架

Fig. 2　Overall framework of face landmark detection method

预测，即知道鼻子的特征线条，可以更好地预测嘴唇的特征线条，同时又可以预测眼睛的特征线条。线条组热图代表不同部位的特征线条，即包含了所有部位的相互关系，可以看作图模型。因此将前一段堆叠沙漏网络得到的热图作为下一段堆叠沙漏网络的输入，意味着下一段堆叠沙漏网络可以使用人脸不同部位关节点的相互关系，从而解决遮挡和姿态变化引起的特征线条提取问题，避免特征线条的缺失。

3　深度网络模型

3.1　总体网络

如图 3 所示，所使用的深度学习模型可以分为 4 部分：图像初始化模块（图 3 中 1），面部分组特征线条化模块（图 3 中 2），点热图回归模块（图 3 中 3）和坐标点预测模块（图 3 中 4）。初始化指图像经过卷积、残差模块和池化下采样操作后，初步获得图像特征并将其作为下一部分的输入。面部分组特征线条化指使用两段堆叠沙漏网络多尺度地提取面部分组特征线条，得到分组的面部特征线条热图并将其组成人脸线条热图。在面部特征线条热图上，点热图模型同样使用两段堆叠沙漏网络，得到人脸特征点热图。在进行预测时，将所求得的人脸特征点热图进行阈值分割，就可以得到人脸特征点坐标。

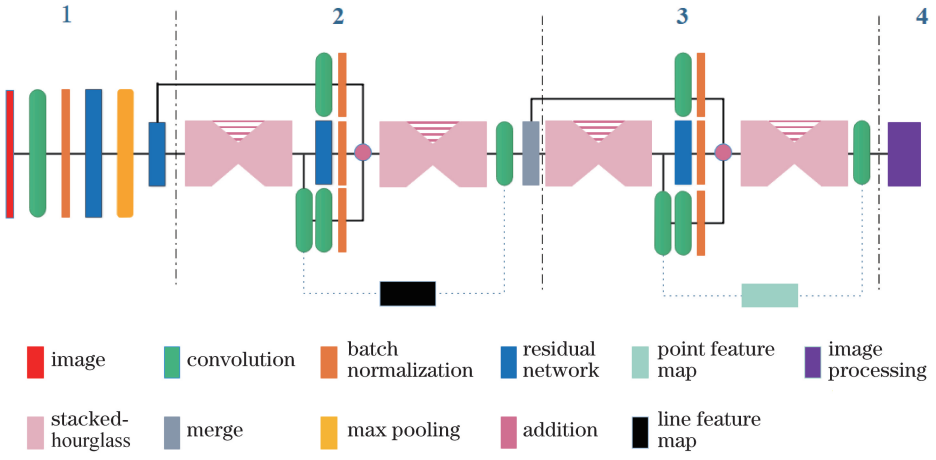

图 3　深度网络模型结构图

Fig. 3　Structural diagram of depth network model

3.2　面部分组特征线条化

面部分组特征线条化回归模型网络使用两段堆叠沙漏网络(图 3)。当组合两个堆叠沙漏网络时,堆叠沙漏网络输出分成两支:一支经过 1×1 卷积形成一个线条热图的集合,使用热图来代替原来的全连接层,热图表示的是各个特征在该像素出现的概率;另一支经过残差模块卷积提取特征。这两支的信息与堆叠沙漏网络的输入信息融合,形成下一个堆叠沙漏网络的输入。第二段堆叠沙漏网络的输出经过 1×1 卷积,形成一个线条热图的集合。将第二段堆叠沙漏网络所得的面部特征线条热图集合进行加权合并(即 merge 模块),构成全面部特征线条热图,将其作为特征点热图的输入。

$$\boldsymbol{f}_{\mathrm{ace}}=\sum_{i=1}^{n_{\mathrm{line}}}\boldsymbol{M}'_{\mathrm{line},i}, \tag{1}$$

式中:$\boldsymbol{f}_{\mathrm{ace}}$为全面部特征线条图,分为 6 个部位的特征线条组;$\boldsymbol{M}_{\mathrm{line},i}$表示一个部位的特征线条组,$i$ 表示特征线条组的索引。$\boldsymbol{M}_{\mathrm{line},i}$由所在面部部位的多个特征线条$\boldsymbol{H}_{\mathrm{line},k}$组成:

$$\boldsymbol{M}_{\mathrm{line},i}=\sum_{k=1}^{i}\boldsymbol{H}_{\mathrm{line},k}。 \tag{2}$$

以左眼为例,其特征线条组由上眼部轮廓特征线条和下眼部轮廓特征线条组成。设$\boldsymbol{M}'_{\mathrm{line},i}$是预测的特征线条组热图,$\boldsymbol{M}_{\mathrm{line},i}$是真实的特征线条组热图,$n_{\mathrm{line}}$是线条组数量(设为 6),下标 line 表示线条。定义线条组热图$\boldsymbol{M}'_{\mathrm{line},i}$时,需要定义单个线条热图,即

$$\boldsymbol{H}_{\mathrm{line},k}=\begin{bmatrix} f_{\mathrm{line},k}(0,0) & f_{\mathrm{line},k}(0,1) & \cdots & f_{\mathrm{line},k}(0,n)\\ f_{\mathrm{line},k}(1,0) & f_{\mathrm{line},k}(1,1) & \cdots & f_{\mathrm{line},k}(1,n)\\ \vdots & \vdots & & \vdots\\ f_{\mathrm{line},k}(m,0) & f_{\mathrm{line},k}(m,1) & \cdots & f_{\mathrm{line},k}(m,n) \end{bmatrix}, \tag{3}$$

式中:$\boldsymbol{H}_{\mathrm{line},k}$表示真实特征线条 k 的热图;$f_{\mathrm{line},k}(x,y)$表示 $\boldsymbol{H}_{\mathrm{line},k}$中坐标$(x,y)$的像素值;$m$ 表示图片$\boldsymbol{H}_{\mathrm{line},k}$的高;$n$ 表示图片$\boldsymbol{H}_{\mathrm{line},k}$的宽。线条热图是面部特征线条在图像上的数值化表现形式,表示面部线条特征的概率,即特征线条在图中出现的概率,使用像素值表示此概率值。线条热图中每个像素的响应由其到特征线条的距离决定,像素越接近线条出现位置,像素值越大,即图中像素的值越大,表示线条越有可能出现在该位置。

根据所有特征点的位置进行三次样条插值后,以插值生成的插值点和原来的特征点作为真实特征线条 k 的坐标,(x_k,y_k)表示线条 k 的坐标集合,$f_{\mathrm{line},k}(x,y)$的值由坐标(x,y)到线条 k 所有坐标(x_k,y_k)的距离求和决定,即

$$f_{\mathrm{line},k}(x,y)=\sum_{i_k}^{n_k}\exp\left(-\frac{|x-x_{i_k,k}|^2+|y-y_{i_k,k}|^2}{2\sigma^2}\right),\quad x_{i_k,k}\in x_k,y_{i_k,k}\in y_k, \tag{4}$$

式中:i_k 为线条 k 上坐标点的索引;n_k 为线条 k 上坐标点的数量;$x_{i_k,k}$为线条 k 上坐标点的横坐标;$y_{i_k,k}$为线条 k 上坐标点的纵坐标;σ 为标准差,一般设为 2。将各个面部部位的$\boldsymbol{H}_{\mathrm{line},k}$组成真实特征线条组热图$\boldsymbol{M}_{\mathrm{line},i}$。

3.3　点热图回归

全面部特征线条热图的点热图模型网络使用两段堆叠沙漏网络(图 3)。真实特征点 j 的热图可表示为

$$\boldsymbol{H}_{\mathrm{point},j}=\begin{bmatrix} f_{\mathrm{point},j}(0,0) & f_{\mathrm{point},j}(0,1) & \cdots & f_{\mathrm{point},j}(0,n)\\ f_{\mathrm{point},j}(1,0) & f_{\mathrm{point},j}(1,1) & \cdots & f_{\mathrm{point},j}(1,n)\\ \vdots & \vdots & & \vdots\\ f_{\mathrm{point},j}(m,0) & f_{\mathrm{point},j}(m,1) & \cdots & f_{\mathrm{point},j}(m,n) \end{bmatrix}, \tag{5}$$

式中:$f_{\mathrm{point},j}(x,y)$表示 $\boldsymbol{H}_{\mathrm{point},j}$中坐标$(x,y)$的像素值;$(x_j,y_j)$表示特征点 j 的坐标。$f_{\mathrm{point},j}(x,y)$由坐标$(x,y)$到特征点坐标$(x_j,y_j)$的距离决定,越靠近特征点,其值越大,其表达式为

$$f_{\mathrm{point},j}(x,y)=\exp\left(-\frac{|x-x_j|^2+|y-y_j|^2}{2\sigma^2}\right)。 \tag{6}$$

一般地,点热图模型需要使用很多段堆叠沙漏网络,但因为面部线条热图上只保留了面部特征线条,提取了关键特征,所以面部点热图模型只需使用较少的堆叠沙漏网络就可以进行点热图回归。

3.4　损失函数

图 3 中的深度学习网络采用中间监督的方式,即模型整体输出特征线条热图和特征点热图,并计算其与目标值的损失值。面部分组特征线条化图模型采用均方差计算线条热图集合与目标线条特征热图集合的距离,并将其作为线条热图的损失值。特征点热图模型也采用均方差计算点热图集合与目标点特征热图集合的距离,并将其作为特征点热图的损失值。将两个损失值加权相加,形成最终的总体损失函数,即

$$L_{\text{oss}}=\frac{1}{n_{\text{line}}}\sum_{i=1}^{n_{\text{line}}}\left\|\boldsymbol{M}'_{\text{line},i}-\boldsymbol{M}_{\text{line},i}\right\|^2+\frac{1}{n_{\text{point}}}\sum_{j=1}^{n_{\text{point}}}\left\|\boldsymbol{H}'_{\text{point},j}-\boldsymbol{H}_{\text{point},j}\right\|^2,\tag{7}$$

式中：$\boldsymbol{M}'_{\text{line},i}$和$\boldsymbol{M}_{\text{line},i}$分别表示预测线条组热图和真实线条组热图；$\boldsymbol{H}'_{\text{point},j}$和$\boldsymbol{H}_{\text{point},j}$分别表示预测点热图和真实点热图；$n_{\text{line}}$和$n_{\text{point}}$分别是线条组和点的数量，$n_{\text{line}}=6$，$n_{\text{point}}=68$。

3.5　坐标点预测

在所得点热图$\boldsymbol{H}'_{\text{point},j}$（图 2）上进行处理，$\boldsymbol{H}'_{\text{point},j}$表示特征点 j 的热图，对其进行阈值分割，得到二值化图像$G_{\text{point,j}}$，其中，$g_{\text{point},j}(x,y)$表示$G_{\text{point},j}$中坐标为(x,y)的像素值，其表达式为

$$g_{\text{point},j}(x,y)=\begin{cases}1, & f'_{\text{point},j}(x,y)\geqslant c\\ 0, & f'_{\text{point},j}(x,y)<c\end{cases},\tag{8}$$

式中：$f'_{\text{point},j}(x,y)$表示$\boldsymbol{H}'_{\text{point},j}$中坐标为$(x,y)$的像素值。

在二值化图像中遍历像素值为 1 的像素，W_j表示这些像素的坐标集合：

$$W_j=\{(x,y)\mid g_{\text{point},j}(x,y)=1\}。\tag{9}$$

对大量样本的测试情况进行统计分析，基于设定的不同阈值和分割确定的坐标情况确定图像阈值的选取范围。图像阈值 c 的范围为 190～210，取图像阈值 c 为 200，经测试该阈值满足分割需要。对坐标集合W_j求平均，得到特征点的坐标点(x_j^*,y_j^*)为

$$(x_j^*,y_j^*)=\left(\frac{1}{n_w}\sum_{i_w}^{n_w}x_{i_w,j},\frac{1}{n_w}\sum_{j}^{n_w}y_{i_w,j}\right),\ (x_{i_w,j},y_{i_w,j})\in W_j,\tag{10}$$

式中：n_w为聚集的特征点数目；i_{w}表示聚集点的索引；$x_{i_w,j}$表示聚集点的横坐标；$y_{i_w,j}$表示聚集点的纵坐标。

4　实　　验

4.1　实验内容

为了评估所提出的方法，运用 300VW 数据集[11]、300W 数据集[12]、300 竞赛测试集[12]、Menpo 挑战数据集进行实验。300VW 数据集包含 114 段视频，标注 68 个点，将其转成约 210000 张图像作为预训练数据。300W 数据集是来自 5 个数据集（LFPW[13]，HELEN[14]，AFW[15]，IBUG[12] 和 300W 比赛测试集）的图像汇编，数据集中的每个图像都使用 68 个标记进行注释，并伴随由人脸检测器生成的重叠框。将 300W 数据集划分为训练和测试部分：训练部分包括 AFW 数据集，以及 LFPW 和 HELEN 的训练子集，总共产生 3148 个图像；测试数据包括剩余的数据集，如 IBUG、300W 比赛测试集、LFPW 测试集、HELEN。为了便于与以前的方法进行比较，将此测试数据拆分为三个子集：由 LFPW 和 HELEN 的测试子集组成（554 张图像）common 集；由 IBUG 数据集组成 challenge 集（135 幅图像）；由 common 集和 challenge 集组成 full 集（689 张图像）。300W 竞赛测试集包含室内和室外共 600 张图像，其中包括各种光照、遮挡干扰和较大姿态变换的图像。Menpo 挑战数据集由来自 FDDB[16] 和 AFLW[17] 数据集的图像。该图像使用与 300W 竞争数据相同的 68 个标记集注释，但没有面部检测器边界框。

在测试集上使用累积误差分布曲线函数的面积除以对应误差阈值的值（N_α）和失败率评估所提方法。人脸特征点的统计平均误差，是指规范化后的预测坐标与真实坐标之间的平均距离，其表达式为

$$E=\sum_{s}^{S}e_s=\frac{\frac{1}{n_{\text{point}}}\sum_{j_s=1}^{n_{\text{point}}}\left\|\boldsymbol{X}'_{s,j_s}-\boldsymbol{X}_{s,j_s}\right\|^2}{d_s},\tag{11}$$

式中：e_s是第 s 个测试样本的误差；n_{point}是特征点的数量；$\boldsymbol{X}'_{s,j_s}$是第 s 个测试样本点 j_s 的预测坐标矩阵；$\boldsymbol{X}_{s,j_s}$是第 s 个测试样本点 j_s 的真实坐标矩阵；d_s是归一化因子，目的是使性能度量与实际面部大小无关，如 d_s 可为外眼角距离、瞳孔间距离或人脸检测框宽度；S 是测试样本数量。

N_α 为在误差阈值 α 下累积误差分布曲线函数 $f(e)$的面积除以误差阈值 α 的值，其表达式为

$$N_\alpha=\frac{1}{\alpha}\int_0^\alpha f(e)\mathrm{d}e,\tag{12}$$

式中：e 为误差。根据不同测试样本，α 取值一般为 0.1，0.08，0.05，结合三个测试数据集的特点将 α 取值为 0.08 和 0.05，并与相关方法进行对比分析。N_α 取值范围为 0～1，1 表示最好的成绩。如果测试样本误差大于 α，则认为失败。

本文实验使用 python 编程语言在 tensorflow 1.9.0环境下构建深度学习模型并进行训练，训练完成后，此模型可在 NVIDIA GTX1060 的 GPU 显卡加速下以 36 frame/s 运行。

4.2　实验比较分析

表 1 表示不同方法在 300W common 集、challenge 集和 full 集的平均误差。测试所有（68 个）点坐标，将所提方法（FDL-PHR）与 MDM[6]

等其他方法进行比较，分别选择两眼中心距离和两外眼角间距离作为归一化因子，可以看出，所提方法产生的平均误差最小。

表2显示了不同方法在300W full集上的N_α值和失败率。选择两外眼角距离作为归一化因子，所提方法的N_α为0.6893，失败率为2.35%，所提方法与以前方法相比结果更加精准。

表1　面部特征点检测方法在300W测试集上的误差

Table 1　Error of face landmark detection methods on the 300W test set　%

Condition	Method	Common subset	Challenging subset	Full set
Inter-pupil normalization	Method in Ref. [18]	6.65	19.79	9.22
	Method in Ref. [19]	5.50	16.78	7.69
	Method in Ref. [20]	5.28	17.00	7.58
	Method in Ref. [4]	5.60	15.40	7.52
	Method in Ref. [21]	5.25	13.62	6.40
	Method in Ref. [22]	4.95	11.98	6.32
	Method in Ref. [23]	4.51	13.80	6.31
	Method in Ref. [24]	4.73	9.98	5.76
	Method in Ref. [25]	4.80	8.60	5.54
	Method in Ref. [26]	4.12	8.35	4.94
	Method in Ref. [27]	3.67	7.62	4.44
	FDL-PHR	3.22	7.92	4.14
Inter-ocular normalization	Method in Ref. [27]	3.67	7.62	4.44
	Method in Ref. [6]	3.33	6.99	4.05
	Method in Ref. [28]	3.34	6.60	3.98
	Method in Ref. [29]	3.34	6.56	3.97
	FDL-PHR	3.11	5.71	3.62

表2　选择两外眼角距离作为归一化因子，面部特征点检测方法在300W full集上的$N_{0.08}$和错误率

Table 2　$N_{0.08}$ and failure rate of face landmark detection methods on the 300W full test set by inter-ocular normalization

Condition	Method	$N_{0.08}$	Failure /%
Inter-ocular normalization	Method in Ref. [4]	0.4294	10.89
	Method in Ref. [20]	0.4312	10.45
	Method in Ref. [24]	0.4987	5.08
	Method in Ref. [6]	0.5212	4.21
	FDL-PHR	0.6893	2.35

从300W数据集中选取头部姿态变化较大和脸部遮挡的图像进行对比实验，结果如图4和表3所示，其中，文献[21]中方法(ERT)是基于级联回归的方法，文献[6]中方法(MDM)是基于深度网络的方法。图4是从选取的图像中挑出的干扰特点较为明显的6张图像，从左到右分析可知：文献[21]中方法在实验比较中所检测的特征点偏离较为严重，例如第2张和第5张；文献[6]中方法在实验中从整体上看与所提方法差别不大，但从细节上来看，例如第4张图像，其所检测的特征点在下巴处没有紧贴脸部轮廓，在下嘴唇处也发生些许偏移；而所提方法充分发挥了面部特征线条化的优势，即让特征点仅与面部特征线条有关，效果优于文献[6]中方法。表3体现的是全部选取图像的检测结果平均误差，可以看出，所提方法(FDL-PHR)检测的特征点更为精准，比其他两种方法表现更好，由此表明所提方法更适于处理具有较大姿态和遮挡干扰的人脸图像。

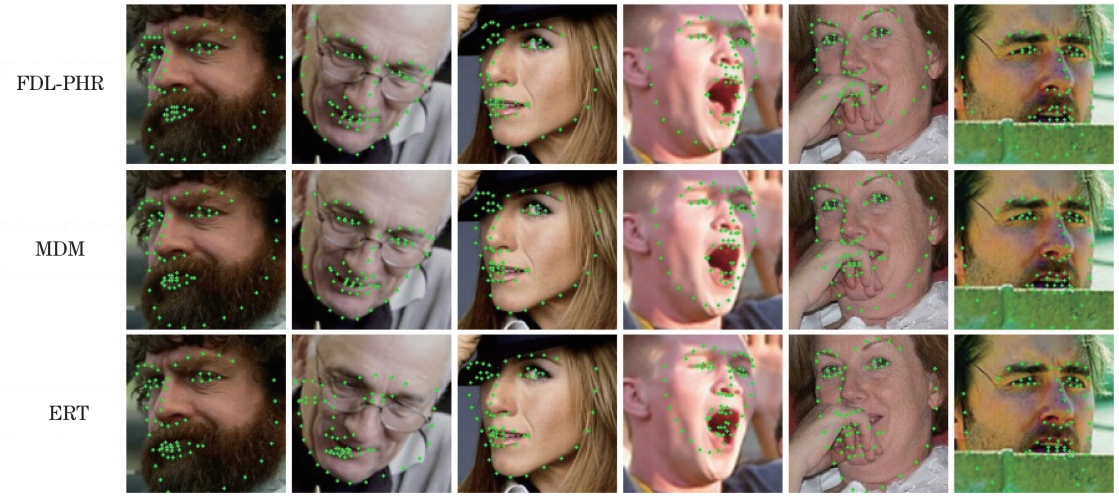

图4　在姿态变化较大和脸部遮挡的人脸图像上的特征点检测对比实验

Fig. 4　Comparative experiment of face landmark detection on images with large posture changes and face partial occlusion

表 3　选择两外眼角距离作为归一化因子，面部特征点检测方法在姿态变化较大和脸部遮挡的人脸图像上的误差

Table 3　Error of face landmark detection methods on face images with large posture changes and face partial occlusion by inter-ocular normalization

Condition	Method	Error /%
Inter-ocular normalization	Method in Ref. [21]	13.89
	Method in Ref. [6]	10.02
	FDL-PHR	8.32

表 4 显示了所提方法和 ESR、CFSS、MDM 等方法在 300W 比赛数据集上的评估指标。其中，所提方法的平均损失、$N_{0.08}$ 和失败率分别为 4.35%，0.5805和 2.86%。

表 4　选择两外眼角距离作为归一化因子，面部特征点检测方法在 300W 比赛数据集上的 $N_{0.08}$ 和错误率

Table 4　$N_{0.08}$ and failure rate of facial landmark detection methods on the 300W competition dataset by inter-ocular normalization

Condition	Method	$N_{0.08}$	Failure /%
Inter-ocular normalization	Method in Ref. [30]	0.1955	38.83
	Method in Ref. [20]	0.3235	17.00
	Method in Ref. [31]	0.3281	13.00
	Method in Ref. [32]	0.3497	12.67
	Method in Ref. [24]	0.3981	12.30
	Method in Ref. [6]	0.4532	6.80
	FDL-PHR	0.5805	2.86

图 5 显示了 300W 比赛数据集上各方法的累积误差分布(CED)曲线，其中，NRMSE 表示标准均方根误差，Image proportion 表示图片比例。平均损失选择两外眼角距离作为归一化因子。可以看出，所提方法在 300W 比赛数据集上比其他方法在各方面都表现得更好。

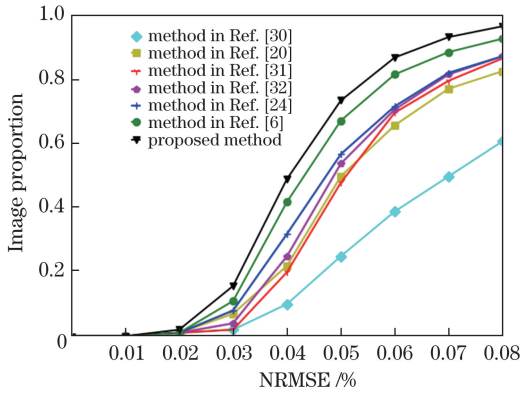

图 5　不同方法在 300W 竞赛数据集上的 CED 曲线图（以两外眼角点矩为归一化因子）

Fig. 5　CED of different methods for 300W competition dataset with inter-ocular normalization

图 6 为所提方法在 300W 竞赛数据集上求得的全面部特征线条热图，可以发现，面部分组特征线条化有助于简单和准确地提取面部特征，为点热图回归排除大量的干扰，如特征类型和尺度不同、光照引起的像素不均匀、遮挡引起的特征不全，以及较大姿态引起的特征仿射变化等问题。图 7 为模型所得的点坐标结果图，进一步说明了所提方法可以优异地处理此类问题。

图 6　300W 竞赛数据集的全面部特征线条热图

Fig. 6　Face feature line heatmaps of 300W competition test set

在 Menpo 数据集上对不同方法进行比较(图 8，表 5)。Menpo 数据集包含侧脸图，即人物只露出半边脸，在这种情况下，眼间距离会变得非常小，对结果的影响非常大，不能作为合格的归一化因子。为此，在 Menpo 数据集上使用面部对角线距离作为归一化因子，使其对面部姿势的变化更稳健。

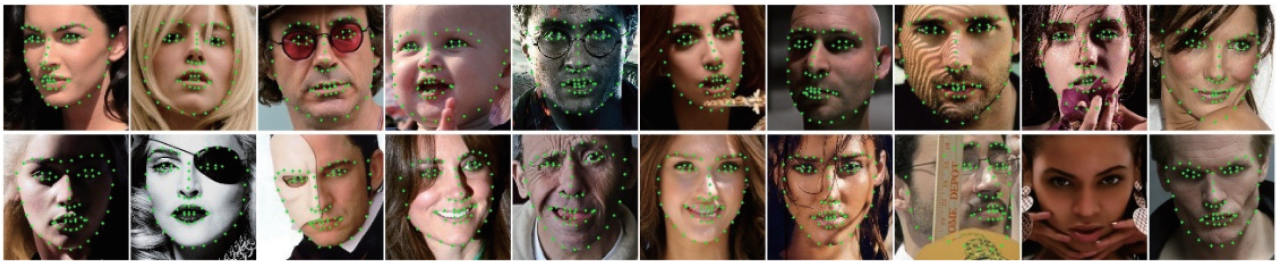

图 7　在 300W 比赛数据集上的检测结果

Fig. 7　Detection results on 300W competition dataset

图 8 是所得的 CED 曲线，所提方法的 CED 曲线明显好于其他方法，其 $N_{0.05}$ 为 0.8679。表 5 表示所有方法的误差平均值、标准差和最大误差，可以看出，所提方法的各项误差指标均较小，表明该模型可以更加准确地检测人脸坐标点。

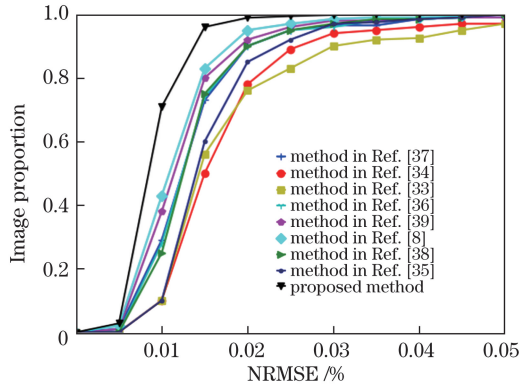

图 8　在 Menpo 比赛数据集上的 CED 曲线图
（以面部图片的对角线距离为归一化因子）

Fig. 8　CED for the Menpo competition dataset with face diagonal normalization

表 5　选择面部对角线距离作为归一化因子，面部特征点检测方法在 Menpo 比赛数据集上的误差分析

Table 5　Error analysis of face landmark detection methods on the Menpo competition dataset by face diagonal normalization

Condition	Method	Mean error	Standard deviation	Max error
Face diagonal normalization	Method in Ref. [33]	0.0205	0.0340	0.9467
	Method in Ref. [34]	0.0182	0.0179	0.4661
	Method in Ref. [35]	0.0165	0.0235	0.9612
	Method in Ref. [36]	0.0159	0.0201	0.6717
	Method in Ref. [37]	0.0200	0.0756	0.7290
	Method in Ref. [38]	0.0135	0.0095	0.5098
	Method in Ref. [29]	0.0138	0.0157	0.6312
	Method in Ref. [39]	0.0139	0.0260	0.9624
	Method in Ref. [9]	0.0120	0.0060	0.1453
	FDL-PHR	0.0199	0.0071	0.07184

综合以上对比分析情况，图 4 和图 6 的数据集中包括了脸部不同姿态、脸部遮挡等情况，从图 4 和图 7 的特征点检测定位情况来看，所提方法都取得了较好的结果，可以适应不同的图像。利用平均误差、N_a 和失败率等评估指标，在 300W 测试集、300W 竞赛数据集，以及 Menpo 比赛数据集上将所提方法与其他方法进行比较，从表 1～5、图 5 和图 8 的结果分析可以看出，利用面部分组特征线条化与点热图相结合的方法进行检测的误差比较小，在检测的精确性方面有较好的效果。

5　结　　论

将面部分组特征、特征线条化和点热图回归组合在一起来检测人脸特征点。面部分组特征可以解决面部特征类型不同的问题，特征线条化可以解决面部特征尺度不同的问题，利用点热图可以进一步地提高检测精度。将所提方法与领域内的其他方法进行实验对比，结果表明，所提方法在检测精确度方面有较为明显的提升，并且可以适应一定的面部姿态变化和遮挡干扰。研究结果为人脸面部特征点检测及应用提供了一种新的方法。

参　考　文　献

[1] Cootes T F, Edwards G J, Taylor C J. Active appearance models[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2001, 23(6): 681-685.

[2] Cootes T F, Taylor C J, Cooper D H, *et al*. Active shape models-their training and application[J]. Computer Vision and Image Understanding, 1995, 61(1): 38-59.

[3] Dollár P, Welinder P, Perona P. Cascaded pose regression[C]//2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 13-18, 2010, San Francisco, CA, USA. New York: IEEE, 2010: 1078-1085.

[4] Xiong X H, de la Torre F. Supervised descent

method and its applications to face alignment[C]// 2013 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2013, Portland, OR, USA. New York: IEEE, 2013: 532-539.

[5] Sun Y, Wang X G, Tang X O. Deep convolutional network cascade for facial point detection[C]// 2013 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2013, Portland, OR, USA. New York: IEEE, 2013: 3476-3483.

[6] Trigeorgis G, Snape P, Nicolaou M A, *et al*. Mnemonic descent method: a recurrent process applied for end-to-end face alignment[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 4177-4187.

[7] Lü J J, Shao X H, Xing J L, *et al*. A deep regression architecture with two-stage re-initialization for high performance facial landmark detection[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 3691-3700.

[8] Newell A, Yang K Y, Deng J. Stacked hourglass networks for human pose estimation[M]// Leibe B, Matas J, Sebe N, *et al*. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9912: 483-499.

[9] Yang J, Liu Q S, Zhang K H. Stacked hourglass network for robust facial landmark localisation[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2025-2033.

[10] Wu W Y, Qian C, Yang S, *et al*. Look at boundary: a boundary-aware face alignment algorithm[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 2129-2138.

[11] Shen J, Zafeiriou S, Chrysos G G, *et al*. The first facial landmark tracking in-the-wild challenge: benchmark and results[C]// Proceedings of the IEEE International Conference on Computer Vision Workshops, December 11-18, 2015, Santiago, Chile. New York: IEEE, 2015: 50-58.

[12] 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.

[13] Belhumeur P N, Jacobs D W, Kriegman D J, *et al*. Localizing parts of faces using a consensus of exemplars[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(12): 2930-2940.

[14] Le V, Brandt J, Lin Z, *et al*. Interactive facial feature localization[M]// Fitzgibbon A, Lazebnik S, Perona P, *et al*. Computer vision-ECCV 2012. Lecture notes in computer science. Berlin, Heidelberg: Springer, 2012, 7574: 679-692.

[15] Zhu X X, Ramanan D. Face detection, pose estimation, and landmark localization in the wild[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI, USA. New York: IEEE, 2012: 2879-2886.

[16] Jain V, Learned-Miller E G. Fddb: a benchmark for face detection in unconstrained settings[R]. Amherst: UMass Amherst Technical Report, 2010.

[17] Köstinger M, Wohlhart P, Roth P M, *et al*. Annotated Facial Landmarks in the Wild: a large-scale, real-world database for facial landmark localization[C]// 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), November 6-13, 2011, Barcelona, Spain. New York: IEEE, 2011: 2144-2151.

[18] Asthana A, Zafeiriou S, Cheng S Y, *et al*. Robust discriminative response map fitting with constrained local models[C]// 2013 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2013, Portland, OR, USA. New York: IEEE, 2013: 3444-3451.

[19] Zhang J, Shan S G, Kan M N, *et al*. Coarse-to-fine auto-encoder networks (CFAN) for real-time face alignment[M]// Fleet D, Pajdla T, Schiele B, *et al*. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8690: 1-16.

[20] Cao X D, Wei Y C, Wen F, *et al*. Face alignment by explicit shape regression[J]. International Journal of Computer Vision, 2014, 107(2): 177-190.

[21] Kazemi V, Sullivan J. One millisecond face alignment with an ensemble of regression trees[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE, 2014: 1867-1874.

[22] Ren S Q, Cao X D, Wei Y C, *et al*. Face alignment at 3000 fps via regressing local binary features[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE, 2014: 1685-1692.

[23] Shi B G, Bai X, Liu W Y, *et al*. Deep regression for face alignment[J/OL]. (2014-09-18)[2019-05-06]. https://arxiv.org/abs/1409.5230.

[24] Zhu S Z, Li C, Loy C C, *et al*. Face alignment by coarse-to-fine shape searching[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 4998-5006.

[25] 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.

[26] Xiao S T, Feng J S, Xing J L, *et al*. Robust facial landmark detection via recurrent attentive-refinement networks[M]//Leibe B, Matas J, Sebe N, *et al*. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 57-72.

[27] Kumar A, Chellappa R. Disentangling 3D pose in a dendritic CNN for unconstrained 2D face alignment [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 430-439.

[28] Dong X Y, Yan Y, Ouyang W L, *et al*. Style aggregated network for facial landmark detection[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 379-388.

[29] Kowalski M, Naruniec J. Face alignment using K-cluster regression forests with weighted splitting[J]. IEEE Signal Processing Letters, 2016, 23(11): 1567-1571.

[30] Baltrusaitis T, Robinson P, Morency L P. Constrained local neural fields for robust facial landmark detection in the wild[C]//2013 IEEE International Conference on Computer Vision Workshops, December 2-8, 2013, Sydney, NSW, Australia. New York: IEEE, 2013: 354-361.

[31] 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.

[32] Yan J J, Lei Z, Yi D, *et al*. Learn to combine multiple hypotheses for accurate face alignment[C]// 2013 IEEE International Conference on Computer Vision Workshops, December 2-8, 2013, Sydney, NSW, Australia. New York: IEEE, 2013: 392-396.

[33] Zadeh A, Baltrusaitis T, Morency L P. Convolutional experts constrained local model for facial landmark detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2051-2059.

[34] Feng Z H, Kittler J, Awais M, *et al*. Face detection, bounding box aggregation and pose estimation for robust facial landmark localisation in the wild[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2106-2111.

[35] Shao X H, Xing J L, Lü J, *et al*. Unconstrained face alignment without face detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2069-2077.

[36] Xiao S T, Li J S, Chen Y P, *et al*. 3D-assisted coarse-to-fine extreme-pose facial landmark detection [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2060-2068.

[37] Chen X, Zhou E J, Mo Y C, *et al*. Delving deep into coarse-to-fine framework for facial landmark localization[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2088-2095.

[38] Wu W Y, Yang S. Leveraging intra and inter-dataset variations for robust face alignment[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2096-2105.

[39] He Z L, Zhang J, Kan M N, *et al*. Robust FEC-CNN: a high accuracy facial landmark detection system[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2044-2050.