



文章编号:1001-9081(2018)07-2064-06

DOI:10.11772/j.issn.1001-9081.2017122928

# 基于尺度自适应与增量式学习的人脸对齐方法

陈平, 龚勋\*

(西南交通大学 信息科学与技术学院, 成都 611756)

(\*通信作者电子邮箱 xgong@swjtu.edu.cn)

**摘要:**针对传统基于回归的人脸对齐算法在人脸尺度归一化时会造成纹理的损失,以及为了提升算法模型的泛化能力必须扩充数据集重新训练而导致训练时间增加,甚至出现不收敛、不可计算等问题,提出一种基于尺度自适应与增量式学习(IL)的人脸对齐方法来提高定位精度。首先,建立初始人脸形状与标准人脸形状的映射关系;然后,通过映射关系实现纹理特征在原图上的提取和人脸尺度的归一化;最后,利用算法模型在新的数据集上进行增量式的学习,快速提高原模型的泛化能力。实验结果表明,与传统回归方法相比,所提方法有更高的对齐精度,特别是在AFW数据集(68个特征点)上提高了2~4个百分点;在10万级别的大数据集(5个特征点)上,所提方法的鲁棒性比基于深度学习的方法高1~2个百分点。同时,所提的增量式学习方法不仅适用于人脸对齐场景下的回归模型求解,还适用于其他应用场景下回归模型的求解。

**关键词:**人脸对齐; 增量式学习; 尺度自适应; 级联回归; 鲁棒性

**中图分类号:** TP391.413    **文献标志码:**A

## Face alignment method based on scale self-adaption and incremental learning

CHEN Ping, GONG Xun\*

(School of Information Science and Technology, Southwest Jiaotong University, Chengdu Sichuan 611756, China)

**Abstract:** To solve the problems of traditional regression-based methods such as the loss of texture caused by face scale normalization, costing more time for retraining expanded data set to improve the generalization ability of the original model, and even potential non-convergence and incomputability, a face alignment method based on scale self-adaption and Incremental Learning (IL) was proposed. Firstly, the mapping relationship between the initial face points and the standard face points was established. Secondly, the extraction of the texture features on the original image and the normalization of the face scales were achieved via the mapping relationship. At last, incremental learning was applied to the new data set by using the existing models, which improved the generalization ability of the original model quickly. The experimental results show that the proposed method performs higher alignment accuracy than traditional regression-based methods. On the AFW (Annotated Faces in the Wild) dataset (68 feature points), the accuracy is increased by 2 to 4 percentage points; and on a 100 000-level large dataset (5 feature points), the robustness is increased by 1 to 2 percentage points compared to the methods based on Deep Learning (DL). In addition, the proposed method is not only suitable for face alignment regression model, but also applicable for solving other regression models.

**Key words:** face alignment; Incremental Learning (IL); scale self-adaption; cascaded regression; robustness

## 0 引言

人脸对齐是在给定的人脸图像上自动精确定位人脸特征点,如瞳孔、嘴角、鼻尖等。人脸对齐的应用很广泛,比如人脸识别、表情识别以及人脸动画自动合成等。人脸对齐大致可以分成三个子问题:1)对人脸表观图像建模;2)对人脸形状建模;3)建立人脸表观图像与人脸形状的关联。基于模型的人脸对齐研究也围绕着这三个方面开展:基于统计学习的主动形状模型(Active Shape Model, ASM)<sup>[1]</sup>提出利用点分布模型表示人脸形状信息,并结合主成分分析(Principal Component Analysis, PCA)<sup>[2]</sup>进行降维,建立了可以通过加权表示人脸形状变分段化的形状模型。主动外观模型(Active Appearance Model, AAM)<sup>[3]</sup>方法考虑如何建模整张人脸的表

观信息,此方法是在ASM模型的基础上建立形状模型和纹理模型,并将两个模型进行有机的结合。局部约束模型(Constrained Local Model, CLM)<sup>[4]</sup>方法分别继承了ASM的高效性和AAM的准确性,使得人脸对齐效果得到了进一步提升。随着人脸对齐技术的发展,基于级联形状回归模型级联姿态回归(Cascaded Pose Regression, CPR)被Dollár等<sup>[5]</sup>在2010年首次提出,并用于人脸形状的预测。接着,具有鲁棒性的级联姿态回归(Robust Cascaded Pose Regression, RCPR)<sup>[6]</sup>在实现特征点定位的同时,还可以对特征点的可见性进行判断。

近些年来,监督下降方法(Supervised Descent Method, SDM)<sup>[7]</sup>从考虑如何使用监督梯度下降的方法来求解非线性最小二乘问题,推动了人脸对齐在现实工程中的应用。SDM认

收稿日期:2017-12-14;修回日期:2018-02-10;录用日期:2018-02-28。    基金项目:国家自然科学基金资助项目(61772435)。

作者简介:陈平(1992—),男,四川宜宾人,硕士研究生,主要研究方向:人脸对齐、三维人脸重建; 龚勋(1980—),男,湖南湘西人,副教授,博士,CCF会员,主要研究方向:模式识别、三维人脸建模、人脸识别。



为相同特征点位置之间的尺度不变特征变换(Scale-Invariant Feature Transform, SIFT)<sup>[8]</sup>特征应该相似,从而利用牛顿法建立级联回归模型,在脸上取得较好的定位效果。接着,Ren等<sup>[9]</sup>用随机森林模型学习局部二值特征(Local Binary Feature, LBF);刘袁缘等<sup>[10]</sup>提出非约束人脸特征点精确定位;张海艳等<sup>[11]</sup>提出应用姿态估计的人脸特征点定位方法;贾项南等<sup>[12]</sup>提出改进的显式形状回归定位算法。除了基于回归的人脸对齐方法,基于深度学习的方法也大放光彩,如多任务级联卷积网络(Multi-Task Cascaded Convolutional Network, MTCNN)<sup>[13]</sup>,实现了高精度的人脸检测和特征点定位。

以上这些基于级联回归人脸对齐训练过程中都要对原始图片数据进行尺度变换的处理。一方面,图片的变换会导致纹理的损失和形状的微调,会影响到特征的准确提取;另一方面,坐标尺度选择不适合,会直接影响人脸对齐的精度。同时,当发现已有模型的泛化能力不够时,通常的解决办法是把新的数据集加入到原来的数据集中,重新训练模型。这样直接丢掉原有模型,不仅否定了自己以前的实验结果,而且重新训练带来的训练时长、模型不收敛等问题,甚至在面对庞大的数据集时,回归求解根本不可计算。针对这些问题,本文提出一种基于增量式学习(Incremental Learning, IL)和尺度自适应的人脸对齐方法:在增量式学习方面,通过充分利用原模型的预测结果与真实值带来的偏差,按照一定的学习率增加原模型的参数值,快速有效地提高原模型的泛化能力。在人脸尺度自适应方面,利用在人脸框内投影而来的初始人脸与标准人脸建立对应的映射关系,在迭代求解时,把任意尺度的人脸特征点映射归一到标准尺度,实现人脸尺度的自适应;在特征提取时,计算当前预测特征点在原始图片的位置,实现特征在原图上的提取。实验结果表明,本文所提方法有效地解决了已有模型泛化能力不够就要被丢弃的尴尬,并且可以大幅度提升人脸的对齐精度。

本文主要工作如下。

1) 利用不同人脸框的初始特征点与标准人脸特征点建立映射关系,将不同尺度的人脸统一到标准尺度,避免了图片归一化带来的纹理损失,同时分析不同尺度下对模型预测的影响,并给出了最佳的人脸尺度。

2) 提出一种增量式学习的回归训练方法。该方法在原有模型的基础上,在新的数据集上进行训练,避免了丢弃原模型和全部重新训练带来的模型训练时间长、不收敛甚至不可计算等问题,能够快速、有效地提高原模型的泛化能力。

## 1 人脸特征点与特征的关系建立

在人脸图像中,特征点是描述人脸形状的重要信息,特征描述子是根据特征点在图像位置提取而来。优秀的特征描述子具有人脸尺度、缩放、光照等不变性,因此不同人脸相同特征点位置的特征描述整体上应该相似。图1显示不同人脸在鼻尖处的特征直方图。

如图1所示,根据特征直方图可以看出不同人在相同位置提取出来的特征在整体上是相似的,因此推导出特征与特征点的关系函数:

$$h(\mathbf{x}) = \lim_{\Delta\mathbf{x} \rightarrow 0} h(\mathbf{x} + \Delta\mathbf{x}) \quad (1)$$

其中: $\mathbf{x}$ 为当前预测特征点(人脸形状), $h(\mathbf{x})$ 表示从 $\mathbf{x}$ 处提取特征的函数,当 $\Delta\mathbf{x}$ 趋近0的时候,相同位置的特征值相同,因此引入一个关系函数 $r(t)$ 消除特征与特征点的量纲: $\Delta\mathbf{x} =$

$h(\mathbf{x})(r(t))$ ,令 $\mathbf{R} = r(t)$ ,建立特征与特征点偏移量间的最小化目标函数:

$$\mathbf{R} = \arg \min_{\mathbf{R}} \|(\mathbf{x}_* - \mathbf{x}) - \mathbf{R}(h(\mathbf{x}))\|_2^2 \quad (2)$$

其中 $\mathbf{x}_*$ 代表人脸的手工标注点。式(2)通过不断最小化 $\Delta\mathbf{x}$ 使得预测特征点的坐标逼近手工标注点的坐标。由于这个方程式非二次的,不可能一步迭代完成。为了进行线性回归,令 $\Phi = h(\mathbf{x})$ ,建立线性回归方程:

$$\begin{cases} \Delta\mathbf{x}_k = \mathbf{x}_* - \mathbf{x}_k = \Phi_k \mathbf{R}_k \\ \mathbf{x}_{k+1} = \mathbf{x}_k + \Phi_k \mathbf{R}_k \end{cases} \quad (3)$$

其中: $k$ 代表第 $k$ 次迭代,根据式(3)可以利用特征 $\Phi$ 来进行特征点位置的预测。

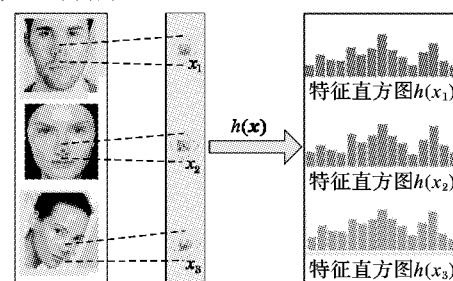


图1 不同人脸的相同位置有相似特征

Fig. 1 Similar feature with same position of different human face

但是进行算法训练时,为了统一人脸尺度对原始图片进行归一化处理,导致特征不能在原图上进行准确提取,因此,如何在统一人脸尺度的同时实现特征在原图上的提取是一个值得研究的问题。

## 2 尺度自适应与增量式学习对齐算法

为了在方法训练时,在保证人脸尺度统一的同时也能在原图上进行特征提取,本文提出一种任意尺度的人脸自动投影到标准尺度人脸的方法,简称人脸投影变换(Face Projection Transformation, FPT)。该方法巧妙地避开了图片预处理阶段,通过建立任意人脸与标准尺度人脸的特征点映射关系,实现人脸尺度的自适应和原图纹理的无损失。同时本文提出了一种基于增量式学习训练方法:在原有模型的基础上,在新的数据集上进行训练,这样避免了丢弃原模型和全部重新训练带来的模型训练时间长、不收敛甚至不可计算等问题,能够快速、有效地提高原模型的泛化能力。

### 2.1 FPT

FPT方法的核心是将任意尺度的人脸自动投影到标准尺度人脸。在实现人脸尺度自适应时,为了探究尺度对人脸对齐的影响,分别选择 $100 \times 100$ 、 $200 \times 200$ 、 $400 \times 400$ 、 $600 \times 600$ 这4种不同的尺度人脸。对于每种尺度,选取多张带有手工标注点的相同尺度正脸图片,对标注点集进行对齐,采用Procrustes Analysis<sup>[14]</sup>算法,得到该尺寸下的标准人脸,如图2所示。

选取一个标准人脸,对于任意尺度人脸的特征点坐标进行投影变换。假定人脸的一个特征点为 $\mathbf{p} = [u \ v \ 1]^T$ ,对应的标准人脸的特征点为 $\mathbf{p}_{std} = [u' \ v' \ 1]^T$ ,根据平面投影变换可以得到:

$$\mathbf{T}[u \ v \ 1]^T = [u' \ v' \ 1]^T \quad (4)$$

其中 $\mathbf{T}$ 为变换矩阵,当变换为局部投影变换时,令 $\mathbf{T} = \begin{bmatrix} a & -b & c \\ b & a & d \end{bmatrix}$ ,引入标准人脸 $\mathbf{x}_{std}$ 的坐标数据有:



$$\begin{bmatrix} u_1 & -v_1 & 1 & 0 \\ v_1 & u_1 & 0 & 1 \\ u_2 & -v_2 & 1 & 0 \\ v_2 & u_2 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ u_n & -v_n & 1 & 0 \\ v_n & u_n & 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} u_{std\_1} \\ v_{std\_1} \\ u_{std\_2} \\ v_{std\_2} \\ \vdots \\ u_{std\_n} \\ v_{std\_n} \end{bmatrix} \quad (5)$$

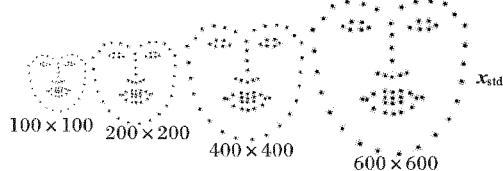


图2 不同标准人脸的尺度

Fig. 2 Different reference face scale

根据式(5)得到任意尺度人脸与标准尺度人脸间的变换矩阵  $T$ , 建立任意尺度人脸在标准尺度人脸下的表示, 如式(6):

$$x' = FPT(x) = Tx \quad (6)$$

其特征点在标准尺度人脸下的形象表示如图3所示。

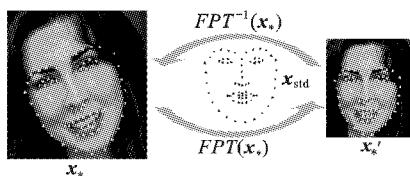


图3 FPT坐标变换

Fig. 3 Coordinate transformation of FPT

从图3可以看出,原始人脸的特征点在经过FPT后,不仅在坐标尺度和标准人脸  $x_{std}$  的坐标尺度一致,而且在形状上逼近了标准人脸  $x_{std}$ 。

## 2.2 基于FPT的人脸自适应对齐

本文将FPT应用到人脸对齐,简称人脸投影变换对齐(Face Projection Transformation Alignment, FPTA)。在训练阶

段,选择标准尺度下的人脸  $x_{std}$ ,引入FPT方法。令  $x_*' = FPT(x_*)$ ,  $x_k' = FPT(x_k)$ ,根据式(3)得到FPTA的线性回归方程:

$$\begin{cases} \Delta x_k' = x_*' - x_k' = \Phi_k R_k \\ x_{k+1}' = x_k' + \Phi_k R_k \end{cases} \quad (7)$$

其中:  $x_{k+1}' = FPT^{-1}(x_{k+1}')$ ;  $\Phi_k$  为当前预测的特征点在原图提取出来的特征;  $R$  为特征与偏移量间的关系矩阵,该关系矩阵的最优化求解函数如下:

$$R_k = \arg \min_{R_k} \sum_{i=1}^N \| (x_k^{i'} - x_{k-1}^{i'}) - R_k(h(x_{k-1}^i)) \|_2^2 \quad (8)$$

其中:  $N$  代表训练图片的数量;  $h(x_{k-1}^i)$  表示在第  $i$  张原图上提取对应的特征,该函数可以利用最小二乘直接求解。初始的  $x_0$  是平均人脸  $\bar{x}$  在对应人脸框下的表,  $\bar{x}$  是所有人脸对应的标注点与人脸框归一化后平均而来。FPTA 算法的具体描述见算法 1。

### 算法 1 FPTA 算法。

**输入:** 原始人脸图片集和其对应的人脸标注特征点和平均人脸  $\bar{x}$ 。

**输出:** 级联回归的模型  $\{R_k\}_1^K$ 。

- 1) 对于每张原始人脸图片,把平均人脸  $\bar{x}$  投放到人脸检测框中得到的初始预测点  $x_0$ ,  $k = 0$ ;
- 2) 用式(5)建立人脸特征点  $x_*$  在标准人脸  $x_{std}$  下的映射关系及其表现形式  $x_*'$ ;
- 3) 利用  $x_k$  在原图上进行特征提取;
- 4) 利用式(8)求得对应的  $R_k$  和  $x_{k+1}'$ ;
- 5) 利用  $x_{k+1}' = FPT^{-1}(x_{k+1}')$  求出特征点在原图的位置;
- 6) 重复步骤 3) ~ 5) 直到达到指定次数  $K$ 。

通过实验仿真发现,当  $K \geq 4$  时算法收敛。

从图4看出,FPTA 算法在计算  $\Delta x$  时利用 FPT 方法实现人脸尺度的自适应,在迭代更新的时候利用 FPT 逆运算得到预测特征点在原图的位置,从而实现特征在原图上的提取。通过实验仿真发现,当迭代次数达到 4 时算法收敛。



图4 人脸对齐实例

Fig. 4 Examples of face alignment

## 2.3 基于增量式学习的FPTA人脸对齐

通过式(8)可以看出,当模型泛化能力不够时,只能把新的数据集加入到原有数据集中,重新训练,以得到满足要求的

新模型;但是数据量的增大会直接导致模型训练时间增加而不收敛,甚至当数据集过大的时候,求解时,庞大的矩阵将会出现不可计算等问题,因此在 FPTA 算法的模型基础上,进行



增量式学习(IL)训练:

$$\begin{cases} \Delta \mathbf{x}_k' = \mathbf{x}_* - \mathbf{x}_k' = \Phi_k(\mathbf{R}_k + \Delta \mathbf{R}_k) \\ \mathbf{R}_k = \mathbf{R}_k + lr \cdot \Delta \mathbf{R}_k \\ \mathbf{x}_{k+1}' = \mathbf{x}_k' + \Phi_k \mathbf{R}_k \end{cases} \quad (9)$$

其中:lr 是微调系数,称为学习率,初始由人工设定。式(9)可以通过最小二乘法解决  $\Delta \mathbf{R}_k$ ,进行模型的微调:

$$\begin{cases} \Delta \mathbf{R}_k = \arg \min_{\Delta \mathbf{R}_k} \sum_{i=1}^N \| \mathbf{x}_k^{i'} - \mathbf{x}_{k-1}^{i'} - (\mathbf{R}_k + \Delta \mathbf{R}_k) h(\mathbf{x}_{k-1}^i) \|_2^2 \\ \mathbf{R}_k' = \mathbf{R}_k + lr \cdot \Delta \mathbf{R}_k \end{cases} \quad (10)$$

通过式(10)的求解可以很快让原模型在新的数据集上进行多次微调,从而到达对新姿态也有较好泛化能力的模型。本文将此方法简称为增量式学习的 FPTA 算法,具体描述如算法 2。

### 算法 2 基于增量式学习的 FPTA 算法。

输入:原始人脸图片集和其对应的人脸标注特征点和原始模型  $\mathbf{R}_k$  以及平均人脸  $\bar{\mathbf{x}}$ 。

输出:级联回归的模型  $\{\mathbf{R}_k'\}_1^K$ 。

- 1) for  $k = 1$  to  $K$  do
- 2) 对于每张图片利用  $\mathbf{R}_k$  和式(7) 进行特征点  $\mathbf{x}_k'$  的估计;
- 3) 利用式(10) 计算  $\Delta \mathbf{R}_k$ ,并求出  $\mathbf{R}_k'$ ;
- 4) 利用式(9) 和  $\mathbf{R}_k'$  评估出  $\mathbf{x}_{k+1}'$ ;
- 5) 利用  $\mathbf{x}_{k+1} = FPT^{-1}(\mathbf{x}_{k+1}')$  反解算出特征点在原图的位置。

通过实验发现,有时在新数据集上进行一次模型的微调并不能达到模型对该数据集的泛化要求时,可以设置较小的学习率对模型进行多次微调。

## 3 实验结果与分析

本章通过 5 个实验来说明本文的主要工作与贡献。所有实验都是 Intel i5-7500 CPU@3.40 GHz 的 Windows 10 平台下进行的,实验选取的数据集如表 1。

表 1 实验用人脸数据集

Tab. 1 Face datasets for experiment

数据集	人脸图片数	点数	大姿态( $>60^\circ$ )占比/%
LFPW	1035	68	3.0
Helen	2330	68	2.2
AFW	337	68	12.1
IBUG	135	68	15.2
CelebA	202 599	5	2.1

其中 LFPW(Labeled Face Parts in the Wild)包含 811 张训练图片和 224 张测试图片;Helen 有 2 000 张训练图片和 330 张测试图片;AFW 和 IBUG 具有大姿态的人脸;CelebA(CelebFaces Attributes)包含 20 多万名人的大规模人脸数据集,其图像涵盖大姿态、复杂的背景,非常适合用于算法鲁棒性的检测。

同时,实验特征提取可以使用 SIFT、方向梯度直方图(Histogram of Oriented Gradient, HOG)<sup>[15]</sup>等特征,本文从速度考虑,实验选用 HOG 特征。在求解最小二乘优化时可以使用卷积神经网络(Convolutional Neural Network, CNN)进行求解。实验发现当迭代次数  $K=4$  时,算法收敛效果较好。

### 3.1 人脸归一化尺度选择

为了快速研究不同归一化尺度对人脸对齐算法影响,本节分别选择尺度为  $100 \times 100$ 、 $200 \times 200$ 、 $400 \times 400$  和  $600 \times 600$  的标准人脸,用 FPTA 算法在 LFPW 训练集上进行训练,单次训练迭代 4 次用时 1.5 h,并在 Helen、AFW、LFPW 上进行测试分析。测试的分析结果如表 2 所示。

表 2 不同标准人脸下的 FPTA 结果

Tab. 2 FPTA results for different reference faces

尺度	平均差			标准差		
	Helen	AFW	LFPW	Helen	AFW	LFPW
$100 \times 100$	6.08	6.49	5.07	2.73	2.54	2.47
$200 \times 200$	5.46	5.55	4.98	2.85	2.62	2.51
$400 \times 400$	<b>5.04</b>	<b>5.27</b>	<b>4.95</b>	<b>2.34</b>	<b>2.43</b>	<b>2.21</b>
$600 \times 600$	5.70	5.64	6.32	2.66	2.63	2.55

根据表 2 的结果可以分析出,坐标尺度在  $400 \times 400$  时对齐精度最好。通过实验数据分析发现,当选取的尺度过小时,人脸框与标准人脸产生的初始人脸特征点集存在更多的重合,会导致模型泛化能力减弱;当尺度过大的时候,初始人脸特征点集尺度离散较大,导致求解最小二乘时,模型的参数数值差过大,从而影响到模型的鲁棒性,所以在后面的实验中都选择  $400 \times 400$  的标准人脸。

### 3.2 基于增量式学习的 FPTA 与同类算法比较

本节共两个实验,第 1 个实验是把 LFPW、Helen、IBUG 结合为一个数据集(共 2946 张)进行单独的 FPTA 训练,得到了 modelA,第 2 个实验是先在 LFPW、Helen 训练集上进行 FPTA 算法的训练集,得到初始 modelB,考虑到 modelB 对大姿态人脸没有较好的泛化能力,所以利用增量式学习在 IBUG(135 张)上进行模型的微调,得到对大姿态有较好泛化能力的 modelC,并对 modelA、modelB、modelC 进行了实验分析。三个模型的对齐误差分析结果如表 3。

表 3 增量式学习的对齐平均误差分析

Tab. 3 Alignment average error analysis of incremental learning

模型	Helen	AFW	LFPW
modelA	3.55	3.68	3.49
modelB	4.10	4.81	3.94
modelC	<b>3.40</b>	<b>3.61</b>	<b>3.24</b>

通过表 3 分析可以看出,通过增量式学习得到的 modelC 比 modelB 具有更好的泛化能力,而且其泛化能力比直接利用全部数据集训练得到 modelA 的要更强一些;并且 modelC 的训练时间比 modelA 的更短。同时在保证实验环境相同的前提下,用开源的 LBF 算法和集成回归树(Ensemble of Regression Trees, ERT)<sup>[16]</sup>算法在 Helen、AFW、LFPW 测试集上进行 68 个特征点的对齐误差测试分析,分析结果如表 4 所示。

表 4 68 个特征点的平均误差比较

Tab. 4 Average error comparison of 68 feature points

方法	Helen	AFW	LFPW
LBF	5.20	7.82	5.71
ERT	4.12	5.82	3.82
SDM	5.79	—	3.40
本文方法	<b>3.40</b>	<b>3.61</b>	<b>3.24</b>

根据表 4 分析得出,本文方法在 3 个数据集上表现最好,



对齐误差都在4%以下。特别是在姿态变化较大的AFW数据集上,本文方法的对齐精度比ERT高2个百分点,比LBF高4个百分点,这充分说明了增量式学习的优势。图4为本文方法的一些定位实例,从图中看出,本文方法对姿态、遮挡有很好的鲁棒性。

### 3.3 鲁棒性测试

算法的鲁棒性是评价算法好坏的标准,鲁棒性分析是在对齐误差的基础上进行进一步的分析,因此,本节在3.2节的实验基础上选取了68个特征点中的15个关键点(每个眼睛边缘6个点、鼻尖1个、2个嘴角)和基于深度学习的MTCNN、由粗到精的自编码网络(Coarse-to-Fine Auto-encoder Networks, CFAN)<sup>[17]</sup>和深度神经网络(Deep Convolutional Neural Network, DCNN)<sup>[18]</sup>进行5个关键点(两个瞳孔、两个嘴角、一个鼻尖)的对齐误差分析。其中每个瞳孔的坐标由每个眼睛边缘的6个点平均得来。MTCNN和DCNN是目前基于深度学习的最主流的人脸对齐方法,CFAN是人脸识别框架Seetaface中的人脸对齐方法。对齐误差如图5所示。

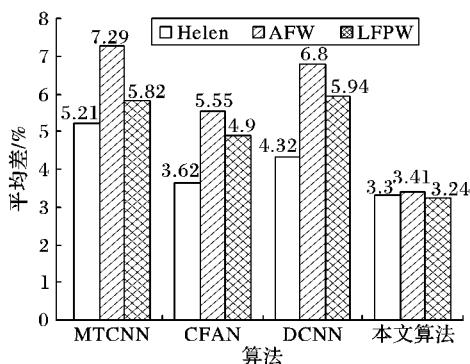


Fig. 5 Alignment error analysis

通过图5可以知道,在本文算法的对齐误差都稳定在0.03左右,MTCNN、CFAN和DCNN波动相对较大。汇总各数据集的测试结果后,FPTA的平均对齐精度比MTCNN高了2.79个百分点,比CFAN高了1.37个百分点,比DCNN高了2.37个百分点。在对齐误差的基础上,本文采用标准差对算法进行鲁棒性分析,图6为本文算法与基于深度学习的算法的分析结果。

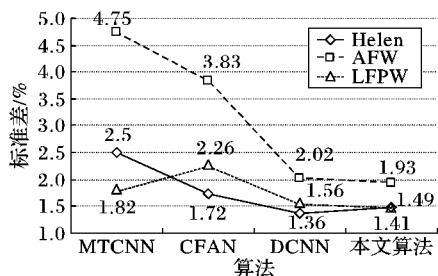


Fig. 6 Robustness analysis of different algorithms

从图6可以看出,基于深度学习的算法的标准差在AFW上都最大,且大于2%,但是本文算法的标准差还是在2%以下。结合其他两个数据集的结果,本文算法的鲁棒性比MTCNN高了1.39个百分点,比CFAN高了0.97个百分点,与DCNN近似,因此本文算法在68个点的训练模型中选取出来的5个特征点在小数据集上无论在对齐精度上还是算法的鲁

棒性上都要优于目前基于深度学习的人脸对齐算法。

为了进一步测试算法鲁棒性,本文在大数据集上作人脸的5个关键点测试。为了缩短大数据集的测试时间,本实验利用按照3.2节的第2个实验选取7个点(两个嘴角、四个眼角、一个鼻尖)进行增量式学习模型训练。在测试的时候把两个眼角的平均坐标作为该眼的瞳孔坐标。同时,在相同实验环境下,对开源的MTCNN和CFAN也作了相应的测试,与本文方法的对比分析如表5所示。

表5 大数据集上5个特征点的比较

Tab. 5 Comparison of five feature points on CelebA %

方法	平均差	标准差
MTCNN	4.87	3.28
CFAN	4.89	2.26
本文方法	5.15	1.47

本实验中,CelebA是标准化后的数据集,有202 599张,去掉人脸框检测错误带来定位误差比较大的人脸后,MTCNN、CFAN和本算法测试了相同的人脸一共181 509张。从表5结果可以看出,本文算法对齐精度略低于基于深度学习的对齐算法;但是从标准差来看,本文算法具有更强的鲁棒性。图7显示的是对齐误差较大的样例,实验发现,当样本姿态较大时,对齐误差相对较大。这是因为训练样本的大姿态占比较少,导致算法模型对大姿态人脸没有较好的泛化能力。



图7 CelebA中错误大的样本

Fig. 7 Fittings with the highest error from CelebA

## 4 结语

本文从不同人脸相同特征点位置的特征描述子大致相似出发,建立特征与特征点偏移量的关系方程,并提出一种基于尺度自适应与增量式学习的人脸对齐方法:通过利用人脸框生成初始预测点与标准的人脸点建立映射关系,实现纹理特征在原图上的提取和人脸尺度的归一化;接着建立增量式学习方程,利用已有的模型在新的数据集上进行增量式的学习,快速有效地增强了原模型的泛化能力。实验证明:本文的人脸尺度的选择与增量式学习的方式,在人脸对齐上取得了很好的效果;同时该方法不仅适用于人脸对齐,还适用于其他回归模型的求解。

### 参考文献 (References)

- [1] COOTES T F, TAYLOR C J, LANITIS A. Active shape models: evaluation of a multi-resolution method for improving image search [C]// Proceedings of the 1994 British Machine Vision Conference. [S. l.]: BMVC Press, 1994: 327–338.
- [2] HOTELLING H. Analysis of a complex of statistical variables into principal components [J]. British Journal of Educational Psychology, 1933, 24(6): 417–520.



- [3] COOTES T F, EDWARDS G J, TAYLOR C J. Active appearance models [ C]// Proceedings of the 1998 European Conference on Computer Vision. Berlin: Springer, 1998: 484 – 498.
- [4] CRISTINACCE D, COOTES T F. Feature detection and tracking with constrained local models [ C]// Proceedings of the 2006 British Machine Vision Conference. [ S. l. ]: BMVC Press, 2006: 929 – 938.
- [5] DOLLÁR P, WELINDER P, PERONA P. Cascaded pose regression [ C]// CVPR 2010: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2010: 1078 – 1085.
- [6] BURGOS-ARTIZU X P, PERONA P, DOLLAR P. Robust face landmark estimation under occlusion [ C]// Proceedings of the 2013 IEEE International Conference on Computer Vision. Washington, DC: IEEE Computer Society, 2013: 1513 – 1520.
- [7] XIONG X, TORRE F D L. Supervised descent method and its applications to face alignment [ C]// Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2013: 532 – 539.
- [8] LOWE D G. Distinctive image features from scale-invariant keypoints [ J]. International Journal of Computer Vision, 2004, 60(2): 91 – 110.
- [9] REN S, CAO X, WEI Y, et al. Face alignment at 3000 FPS via regressing local binary features [ C]// CVPR 2014: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2014: 1685 – 1692.
- [10] 刘袁缘, 谢忠, 周顺平, 等. 基于条件迭代更新随机森林的非约束人脸特征点精确定位 [ J]. 计算机辅助设计与图形学学报, 2017, 29(10): 1881 – 1890. ( LIU Y Y, XIE Z, ZHOU S P, et al. Conditional iteration updated random forests for unconstrained facial feature location [ J]. Journal of Computer-Aided Design and Computer Graphics, 2017, 29(10): 1881 – 1890.)
- [11] 张海艳, 高尚兵, 姜明新. 应用姿态估计的人脸特征点定位算法 [ J]. 计算机应用, 2017, 37(11): 3256 – 3260. ( ZHANG H Y, GAO S B, JIANG M X. Facial feature localization algorithm using pose estimation [ J]. Journal of Computer Applications, 2017, 37(11): 3256 – 3260.)
- [12] 贾项南, 于凤芹, 陈莹. 改进的显式形状回归人脸特征点定位算法 [ J]. 计算机应用, 2018, 38(5): 1289 – 1293. ( JIA X N, YU F Q, CHEN Y. Improved explicit shape regression for face alignment [ J]. Journal of Computer Applications, 2018, 38(5): 1289 – 1293.)
- [13] ZHANG K, ZHANG Z, LI Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks [ J]. IEEE Signal Processing Letters, 2016, 23(10): 1499 – 1503.
- [14] SLICE D E. Landmark coordinates aligned by Procrustes analysis do not lie in Kendall's shape space [ J]. Systematic Biology, 2001, 50(1): 141 – 149.
- [15] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [ C]// Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision & Pattern Recognition. Washington, DC: IEEE Computer Society, 2005: 886 – 893.
- [16] KAZEMI V, SULLIVAN J. One millisecond face alignment with an ensemble of regression trees [ C]// CVPR 2014: Proceedings of the 2014 Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2014: 1867 – 1874.
- [17] ZHANG J, SHAN S, KAN M, et al. Coarse-to-Fine Auto-Encoder Networks ( CFAN) for real-time face alignment [ C]// Proceedings of the 2014 European Conference on Computer Vision. Berlin: Springer, 2014: 1 – 16.
- [18] SUN Y, WANG X, TANG X. Deep convolutional network cascade for facial point detection [ C]// CVPR 2013: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2013: 3476 – 3483.

This work is partially supported by National Natural Science Foundation of China (61772435).

**CHEN Ping**, born in 1992. M. S. candidate. His research interests include face alignment, 3D face reconstruction.

**GONG Xun**, born in 1980. Ph. D., associate professor. His research interests include pattern recognition, 3D face modeling, face recognition.

(上接第 2043 页)

- [22] 袁悦. 基于复制粘贴操作的克隆代码一致性维护需求预测方法 [ D]. 哈尔滨: 哈尔滨工业大学, 2016: 30 – 32. ( YUAN Y. Prediction method of consistency maintaining clone code based on copy and paste operation [ D]. Harbin: Harbin Institute of Technology, 2016: 30 – 32.)
- [23] WANG W, GODFREY M W. Recommending clones for refactoring using design, context, and history [ C]// Proceedings of the 2014 IEEE International Conference on Software Maintenance and Evolution. Piscataway, NJ: IEEE, 2014: 331 – 340.
- [24] 赵玉武, 翟晔, 刘东升. 基于 bugs 的克隆代码稳定性评估方法 [ J]. 计算机应用研究, 2017, 34(4): 1109 – 1113. ( ZHAO Y W, ZHAI Y, LIU D S. Studies on the stability of clone code based on bugs [ J]. Application Research of Computers, 2017, 34(4): 1109 – 1113.)
- [25] 刘冬瑞, 刘东升, 张丽萍, 等. 基于贝叶斯网络预测克隆代码质量 [ J]. 计算机科学, 2017, 44(4): 165 – 168. ( LIU D R, LIU D S, ZHANG L P, et al. Prediction of clone code quality based on Bayesian

network [ J]. Computer Science, 2017, 44(4): 165 – 168.)

- [26] KOSKER Y, TURHAN B, BENER A. An expert system for determining candidate software classes for refactoring [ J]. Expert Systems with Applications, 2009, 36(6): 10000 – 10003.

This work is partially supported by the National Natural Science Foundation of China (61462071), the Natural Science Foundation of Inner Mongolia Autonomous Region (2016MS0612).

**SHE Rongrong**, born in 1991, M. S. candidate. Her research interests include software engineering, software analysis.

**ZHANG Liping**, born in 1974, M. S., professor. Her research interests include software engineering, software analysis.

**HOU Min**, born in 1972, M. S., lecturer. Her research interests include software analysis, computer aided instruction.

**YAN Sheng**, born in 1984, M. S., lecturer. His research interests include software analysis, parallel computing.