文章编号：1005-5630(2020)02-0045-05 **DOI**: 10.3969/j.issn.1005-5630.2020.02.008

基于注意力模型的人脸关键点检测算法

秦晓飞[1]，盛 凯[2]，朱 玥[1]，杨 勇[1]，赵 刚[3]，
贾 程[3]，李成名[3]，鲁小东[4]，周坚风[4]
（1. 上海理工大学 光电信息与计算机工程学院，上海 200093；
2. 上海理工大学 机械工程学院，上海 200093；
3. 杭州亿美实业有限公司，浙江 杭州 310000；
4. 杭州亿美光电科技有限公司，浙江 杭州 310000）

摘要：人脸关键点定位因受到表情、光照、姿态等的影响，常常会出现大的误差。为了准确地定位到人脸的关键点，提出了一种基于注意力模型的人脸关键点检测算法。先是利用可变型模型（DPM）算法检测出图片中的人脸区域，然后结合残差网络（ResNet）和收缩激励网络（SeNet）对该区域进行人脸关键点定位。实验结果表明，该算法在人脸数据集上获得了较高的准确率，证明了该算法的有效性。

关键词：人脸关键点检测；注意力模型；DPM 人脸检测
中图分类号：TP 391 **文献标志码**：A

Detection algorithm for key points on face based on attention model

QIN Xiaofei[1]，SHENG Kai[2]，ZHU Yue[1]，YANG Yong[1]，ZHAO Gang[3]，
JIA Cheng[3]，LI Chengming[3]，LU Xiaodong[4]，ZHOU Jianfeng[4]
（1. School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China;
2. School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China;
3. Hangzhou Yimei Industrial Co., Ltd., Hangzhou 310000, China;
4. Hangzhou Yimei Photoelectric Technology Co., Ltd., Hangzhou 310000, China）

Abstract: Due to the influence of expressions, illuminations, gestures, etc., large errors often occur when positioning key points of a face. In order to accurately locate the key points of the face, a detection algorithm for key points on a face based on attention mechanism is proposed. Firstly, the deformable part model（DPM）algorithm is used to detect the face region in the picture, and then the focal point of the face is located in the region using ResNet and SeNet. The experimental results show that the algorithm achieves good accuracy on the face dataset, and prove the effectiveness of the algorithm.

收稿日期：2019-05-08
基金项目：国家重点研究发展计划 (2016YFF0101400)
作者简介：秦晓飞（1982—），男，高级工程师，研究方向为人工智能算法。E-mail：xiaofei.qin@usst.edu.cn

Keywords: face key point detection；attention model；DPM face detection

引　言

人脸关键点检测是计算机视觉领域的一个重要而且具有挑战性的任务，它涉及到检测人脸的中心、眼角、鼻尖等，其关键是预测出给定人脸在图像像素空间中的坐标。人脸关键点检测对视频中的人脸设别、医学诊断中的畸形面部体征检测等起着重要的作用。因此，如何快速、准确地检测人脸关键点，受到人们广泛的关注。

人脸关键点检测方法大致分为三种：基于 ASM（active shape model）[1] 和 AAM（active appearnce model）[2-3] 的传统方法、基于级联形状回归的方法[4] 和基于深度学习的方法[5-10]。目前，应用最广、精度最高的是基于深度学习的方法。2013 年，Sun 等[5] 首次将 CNN（convolutional network）应用到人脸关键点检测，提出一种级联的 CNN（拥有三个层级）即DCNN（deep convolutional network），此种方法属于级联回归方法。通过精心设计拥有三个层级的级联卷积神经网络，本文不仅改善了初始不当导致陷入局部最优的问题，而且借助于 CNN 强大的特征提取能力，获得了更为精准的关键点检测。

2017 年，Kowalski 等work[10] 提出一种新的级联深度神经网络 DAN（deep alignment network），以往级联神经网络输入的是图像的某一部分，而 DAN 各阶段的输入均为整张图片。当网络均采用整张图片作为输入时，DAN 可以有效地克服头部姿态以及初始化带来的问题，从而得到更好的检测效果。DAN 之所以能将整张图片作为输入，是因为其加入了关键点热图。

通过实验发现，人脸检测时对关键点检测很重要。为了避免不同人脸检测算法带来的影响，不管从哪里来的图片，都经过同一个人脸检测算法后再输入到后面的关键点检测中，这样效果就会变好。本文先是利用可变型模型（deformable part model, DPM）算法检测出图片中的人脸区域，即先利用 Sobel 梯度算子计算出图片的梯度方向直方图，并用支持向量机（support vector machine, SVM）算法对梯度进行分类，检测出人脸区域，然后采用残差网络（residual network, resNet）和收缩激励网络（squeeze-and-excitation networks, SENet）相结合的注意力机制深度神经网络对该区域进行人脸关键点定位。该算法在人脸数据集上获得了较好的准确率，证明了算法的有效性。

1　注意力机制算法

1.1　DPM 人脸检测算法

DPM 是一种基于组件的检测算法，先对初始图像计算梯度方向直方图，然后用 SVM（surpport vector machine）训练得到物体的梯度模型，最后利用该模型来做人脸与非人脸的分类。其检测效果如图 1 所示。

图 1　DPM 人脸检测效果图

Fig. 1　DPM face detection effect map

1.2　梯度方向直方图

梯度在图像中对应的就是其一阶导数。模拟图像 $f(x,y)$ 中任一像素点 (x,y) 的梯度是一个矢量，可表示为

$$\nabla f(x,y)=\left[G_x G_y\right]^{\mathrm{T}}=\left[\frac{\partial f}{\partial x}\frac{\partial f}{\partial y}\right]^{\mathrm{T}} \tag{1}$$

式中：G_x 为沿 x 方向上的梯度；G_y 为沿 y 方向

上的梯度。梯度的幅值及方向角可表示如下：

$$\begin{cases}|\nabla f(x,y)|=\mathrm{mag}(\nabla f(x,y))=\left(G_x{}^2+G_y{}^2\right)^{1/2}\\ \Phi(x,y)=\arctan(G_y/G_x)\end{cases} \tag{2}$$

数字图像中像素点的梯度是用差分来计算的，即

$$\nabla f(x,y)=\left\{\left[f(x,y)-f(x+1,y)\right]^2+\left[f(x,y)-f(x,y+1)\right]^2\right\}^{1/2} \tag{3}$$

一维离散微分模板可将图像的梯度信息简单、快速且有效地计算出来，即

$$\begin{cases}G_x=H(x+1,y)-H(x-1,y)\\ G_y=H(x,y+1)-H(x,y-1)\end{cases} \tag{4}$$

式中：G_x、G_y 分别为像素点(x, y)在水平方向及垂直方向上的梯度；$H(x,y)$为像素点的灰度值。像素点梯度的幅值及方向计算式如下：

$$G(x,y)=\sqrt{G_x(x,y)^2+G_y(x,y)^2} \tag{5}$$

$$\alpha(x,y)=\arctan\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \tag{6}$$

不同的梯度运算模板在其检测效果上也不一样，本文采用 Sobel 梯度算子，其形式如图 2 所示。

−1		1
−2		2
−1		1

(a) 垂直算子

1	2	1
−1	−2	−1

(b) 水平算子

图 2　Sobel 梯度算子

Fig. 2　Sobel gradient operator

从梯度计算式中可以看出，梯度幅值绝对值的大小容易受到前景与背景对比度及局部光照的影响，要减少这种影响得到较准确的检测结果就必须对局部细胞单元进行归一化处理。

最后计算目标窗口的梯度直方图。对于整个目标窗口，我们需要将其分成互不重叠、大小相同的细胞单元，然后分别计算出每个梯度的梯度信息，包括梯度大小和梯度方向。

1.3　对梯度特征进行 SVM 分类

图 3 为 SVM 分类梯度向量图，对于滑动窗口提取的 2 个窗口，分别计算出归一化的梯度特征，然后应用 SVM 实现是人还是背景的分类判定。SVM全称为支持向量机，是一种二分类的模型。其主要思想是找到空间中的一个能够将所有数据样本划开的超平面，并且使得本集中所有数据到这个超平面的距离最短。

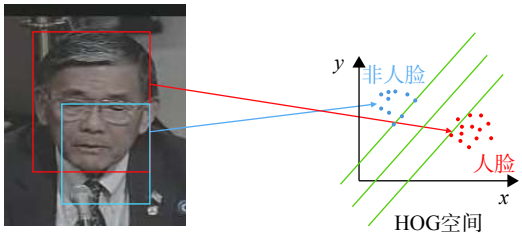

图 3　SVM 分类梯度向量

Fig. 3　SVM classification gradient vector

1.4　注意力模型

注意力模型来源于人脑对事物的观察，由于人脑在观察事物时，人眼睛聚焦的位置只是很小的一块，这时人脑会聚焦在这一小块图案上，此时，人脑对图的设别并不是均衡的，是有权重区别的。

我们在 ResNet 中嵌入 SE 模块，SeNet 会通过学习的方式来自动获取每个特征通道的重要程度，然后依照这个重要程度去提升有用的特征并抑制对人脸关键点检测任务用处不大的特征。我们使用全局平均池化作为 Squeeze 操作，紧接着两个全连接层组成一个 Bottleneck 结构去建模通道间的相关性，并输出和输入特征同样数目的权重。首先，将特征维度降低到输入的 1/16，并经过 ReLu 激活后再通过一个全连接层升回到原来的维度，这样做具有更多的非线性，可以更好地拟合通道间复杂的相关性，极大地减少了参数量和计算量；然后，通过一个 Sigmoid 的门获得 0～1 之间归一化的权重；最后，通过一个 Scale 的操作来将归一化后的权重加权到每个通道的特征上。具体框架如图 4 所示。

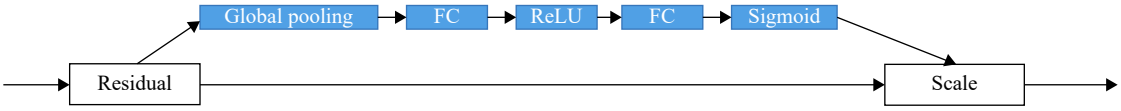

图 4　具有注意力机制的残差神经网络框架

Fig. 4　A framework of residual neural network with attention mechanism

2　实验测试

2.1　数据集来源

实验采用的操作系统是 Windows10 64 位，GTX1080Ti 显卡 GPU，32 GB 内存台式工作服务器，运行环境为 Pytorch 平台。

数据集的图像数据是从 YouTube 人脸数据集中提取的，其中包含 YouTube 视频中的人物视频。这些视频通过一些处理步骤进行输入，并转换为包含一个人脸和相关关键点的图像帧集。该人脸关键点数据集由 5 770 张彩色图像组成。所有这些图像都被分成训练数据集与测试数据集。这些图像中有 3 462 张是训练图像，用于预测关键点的模型，另外 2 308 张是测试图像，用于测试该模型的准确性。

2.2　评价指标

由于实验任务是一个回归问题，因此选用均方根损失函数（mean squared error, MSE）计算算法的误差。具体公式如下：

$$\mathrm{MSE}=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \tag{7}$$

式中：$\hat{y}$ 是数据集中人工标定的关键点坐标；y 是本文算法定位的关键点坐标。通过计算，本文算法的均方根误差为 1.229。图 5 为采用本文算法得到的人脸关键点检测的部分效果图。

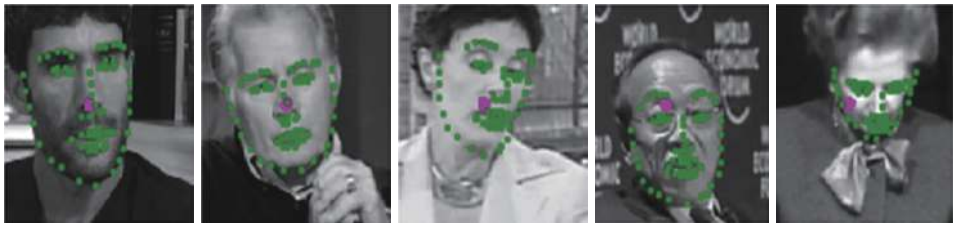

图 5　部分实验效果图

Fig. 5　Part of experimental results

2.3　与其他算法的对比

为了和其他算法作对比，本文算法和其他算法全都放在 YouTube 人脸数据集中的测试集下做实验。图 6 为本文算法和杨海燕等[12]的算法的对比，实验数据如表 1 所示。由表 1 可以看出，本文所提出的算法在同一测试集中表现出更好的性能，误差更低。

图 6　本文算法和杨海燕算法的对比

Fig. 6　Comparisons between the algorithm in this paper and Yang Haiyan's

表 1　不同算法误差对比
Tab. 1　Error comparison of different algorithms

方法	均方根误差
本文算法	**1.229**
Sun	2.221
ASEF[11]	2.812
杨海燕[12]	1.453

3　结束语

本文针对深度学习在人脸关键点检测上的应用进行了研究，并成功将注意力机制应用到该领域上。基于注意力的模型可以提取到普通深度网络不易学习到的人脸信息，可以提高人脸关键点检测的准确性。通过实验发现，本文算法在人脸数据集上获得了 98% 的准确率，同时均方根误差只有 1.229，证明了算法的有效性。下一步将针对算法的实时性对其做速度上的改进。

参考文献：

[1] 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.

[2] EDWARDS G J, COOTES T F, TAYLOR C J. Face recognition using active appearance models[C]// Proceedings of the 5th European conference on computer vision. Freiburg, Germany: Springer, 1998: 581 − 595.

[3] COOTES T F, EDWARDS G J, TAYLOR C J. Active appearance models[C]//Proceedings of the 5th European conference on computer vision. Freiburg, Germany: Springer, 1998: 484 − 498.

[4] DOLLÁR P, WELINDER P, PERONA P. Cascaded pose regression[C]//Proceedings of 2010 IEEE computer society conference on computer vision and pattern recognition. San Francisco, CA, USA: IEEE, 2010: 1078 − 1085.

[5] SUN Y, WANG X G, TANG X O. Deep convolutional network cascade for facial point detection[C]// Proceedings of 2013 IEEE conference on computer vision and pattern recognition. Portland, OR, USA: IEEE, 2013: 3476 − 3483.

[6] ZHOU E J, FAN H Q, CAO Z M, et al. Extensive facial landmark localization with coarse-to-fine convolutional network cascade[C]//Proceedings of 2013 IEEE international conference on computer vision workshops. Sydney, NSW, Australia: IEEE, 2013: 386 − 391.

[7] ZHANG Z P, LUO P, LOY C C, et al. Facial landmark detection by deep multi-task learning[C]//Proceedings of the 13th European conference on computer vision. Zurich, Switzerland: Springer, 2014: 94 − 108.

[8] WU Y, HASSNER T, KIM K, et al. Facial landmark detection with tweaked convolutional neural networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(12): 3067 – 3074.

[9] ZHANG K P, ZHANG Z P, LI Z F, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE Signal Processing Letters, 2016, 23(10): 1499 – 1503.

[10] KOWALSKI M, NARUNIEC J, TRZCINSKI T. Deep alignment network: a convolutional neural network for robust face alignment[C]//Proceedings of 2017 IEEE conference on computer vision and pattern recognition workshops. Honolulu, HI, USA: IEEE, 2017: 2034 − 2043.

[11] BOLME D S, DRAPER B A, BEVERIDGE J R. Average of synthetic exact filters[C]//Proceeding of 2009 IEEE conference on computer vision and pattern recognition. Miami, FL, USA: IEEE, 2009: 2105 − 2112.

[12] 杨海燕, 蒋新华, 聂作先. 基于并行卷积神经网络的人脸关键点定位方法研究 [J]. 计算机应用研究, 2015, 32(8): 2517 – 2519.

（编辑：刘铁英）