文章编号:1674－2869(2019)05－0489－05

基于单一神经网络的实时人脸检测

熊寒颖,鲁统伟*,闵 峰,蒋冲宇
武汉工程大学计算机科学与工程学院,湖北 武汉 430205

摘 要:由于人脸尺度多样性使得人脸检测算法在CPU上运行速度受限,提出了一种新的基于单一神经网络的实时人脸检测算法。首先在网络初始卷积层和池化层中设置较大的卷积核尺寸和步长,缩小输入图像尺寸利于实时检测;然后网络将浅层特征图和深层特征图相融合,增强上下文联系和减少重复检测;最后在多个卷积层上预测人脸位置,利用预测框重叠策略,实现多尺度的人脸检测来提升图像中小尺寸人脸的检测精度。在人脸检测数据集基准和野外标注人脸数据集上测试实验结果表明,本文算法模型精度能够达到92.1%和95.4%。与此同时,本文算法在CPU上实现21帧/s的检测速度。

关键词:卷积神经网络;多尺度人脸检测;特征图融合;CPU
中图分类号:TP391.4 **文献标识码**:A **doi**:10.3969/j.issn.1674-2869.2019.05.015

Real-Time Face Detection Based on Single Neural Network

XIONG Hanying, LU Tongwei, MIN Feng, JIANG Chongyu*
School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430205, China

Abstract: To improve the limited speed of face detection algorithm on central processing unit (CPU) caused by the diversity of the facescales, we proposed a real-time face detection method based on a single neural network. Firstly, a large convolution kernel and step size were used in the initial convolution and pooling layers, which were able to reduce the size of input images. Then, the shallow and deep feature maps were merged to enhance the context-connection and reduce repeated boxes. Finally, we predicted the face location based on the output of different convolution layers. By using the strategy of overlapping prediction boxes, our method is able to improve the detection accuracy of the smaller size face of input images. Experimental results on face detection dataset and benchmark and annotated face dataset in the wild achieve accuracies of 92% and 95.4%, respectively. Above all, our face detection technique can achieve a high detection speed of 21 frames per second on CPU, which can satisfy real-time detection requirements.

Keywords: convolution neural network; multi-scaleface detection; feature map fusion; CPU

随着深度学习和计算机视觉技术的飞速发展,人脸检测技术被广泛应用于生活的各个角落。例如拍照美颜、安防监控、视频会议等,其中人脸检测技术是人脸识别[1]中最开始的一步。由于人脸尺度多样性,使得人脸检测模型在CPU上很难达到实时检测速度,所以如何让模型在不降低精度的同时保障运行速度,依旧是巨大的挑战。

在深度学习爆发之前,人脸检测主要使用浅层模型完成。人们利用算法把那些看上去抽象的信息变得易于处理,最后人工设计处理得到的半成品再交给模型去学习[2],这种方法严重影响了人脸检测算法的检测速度和精度。Viola等[3]使用Haar(Haar-based)的级联分类器来检测对象,使人脸检测算法有很大的改进。Viola等将Haar特征

收稿日期:2019-05-09
基金项目:武汉工程大学第十届研究生教育创新基金(CX2018193)
作者简介:熊寒颖,硕士研究生。E-mail:1069759052@qq.com
***通讯作者**:鲁统伟,博士,副教授。E-mail:lutongwei@wit.edu.cn
引文格式:熊寒颖,鲁统伟,闵峰,等.基于单一神经网络的实时人脸检测[J].武汉工程大学学报,2019,41(5):489-493.

与Adaboost[4]算法相结合实现人脸检测算法。Ahonen等[5]利用局部二值特征实现人脸检测算法。这些传统的人脸测算法在速度上具有一定优势,但是人脸图像易受光照不均、姿态多样性和遮挡等情况的影响,实际应用中检测精度不高。

随着深度学习的发展,人们提出了使用深层模型来实现人脸检测。深度学习舍弃了人工提取特征的步骤,让模型更好地根据数据的原始状态学习,因此更容易学到数据中有价值的信息[6]。Yang等[7]提出通道特征(aggregation channel feature,ACF)算法,将传统方法和神经网络相结合,利用多通道特征实现人脸检测。Chen等[8]提出Jiont Cascade算法,将人脸关键点检测与人脸检测相结合,提高人脸检测算法精度。Ghiasi等[9]提出高分辨率可变形部件模型(multires hierarchial deformable pot model,MultiresHPM),利用级联神经网络进行人脸检测和关键点定位,实现多角度和遮挡情境下的人脸检测。Zhan等[10]提出多任务级联神经网络(multi-task cascaded convolutional network,MTCNN),利用三层级联神经网络实现人脸检测和关键点对齐算法。Zhang等[11]提出Faceboxes算法,基于CPU的快速准确人脸检测。

现有的人脸检测网络可分为级联神经网络和单一神经网络两种。其中级联神经网络适合检测单个人脸图像,当图像中存在多个人脸时会增加检测时间,且训练方法复杂。单一神经网络的人脸检测算法,可快速检测出一张图像中多个人脸,且结构简单易于训练。为实现实时的人脸检测,本文选择单一神经网络。在网络的前2个卷积层中设置较大的步长,使输入图像尺寸快速减小;为防止图像中小尺寸的人脸信息丢失,将浅层特征信息和深层特征信息相融合,增加小尺寸人脸的信息并减少重叠检测框;由于图像中存在多尺度[12-13]的人脸,利用Inception[14]结构和重叠框预测策略,增加小尺寸人脸的检测概率;使用多级损失函数,分别预测人脸框和人脸类别。

1 基于单一神经网络的实时人脸检测

基于单一神经网络的实时人脸检测网络结构如图1所示。网络的输入为1 024*1 024像素大小的图像,当输入图像尺寸小于该尺寸时,用值为0的像素将图像自动填充成到1 024*1 024像素大小;Conv1、Pool1、Conv2和Pool2层采用较大的卷积核和步长,快速缩小输入图像尺寸,保证人脸检测的实时性;Conv3和Conv4层引入浅层特征信息,实现上下文特征融合,提高网络对细节信息的感知能力;Inception1、Inception2和Inception3层实现人脸的多尺度检测,利用多尺度卷积和预测框重叠策略,减少尺度变化对检测效果的影响;利用多任务损失函数,加快模型的收敛速度。

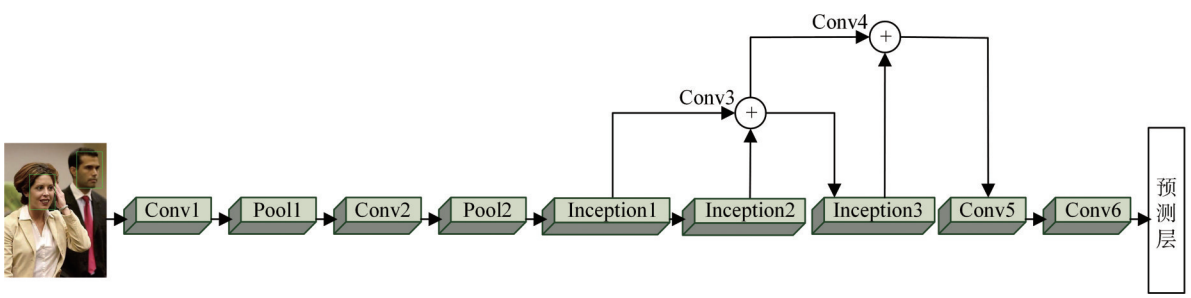

图1 人脸检测网络结构图

Fig. 1 Structure diagram of face detection network

1.1 快速下降卷积

当输入网络的检测特征图尺寸较大时,网络卷积的时间会增加,在CPU上的运行时间会加长。本文网络使用输入尺寸为1 024*1 024像素的彩色图,为快速缩小输入特征图尺寸,在Conv1、Pool1、Conv2、Pool2这4层设置较大的卷积核步长,分别为4、2、2和2,经过这4层卷积操作使输入空间快速缩小32倍。Conv1层和Conv2层卷积核的大小分别为7*7和5*5,特征图像填充为3。Pool1和Pool2层的卷积核大小都为3*3,无图像填充。为提高检测速度,在Conv1和Conv2层采用了C.ReLU[15]激活函数。C.ReLU激活函数应用在ReLU之前简单地连接否定输出,在保证输出维度不变的情况下减少卷积核数量,提高检测速度。C.ReLU结构如图2所示。

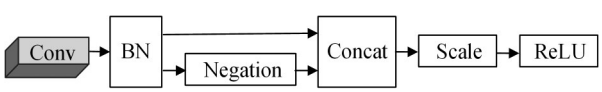

图2 C.ReLU结构图

Fig. 2 Structure diagram of C.ReLU

1.2　特征图融合

由于低层特征分辨率较高，包含更多位置和细节信息，但是其经过的卷积少，语义性低，噪声多；高层特征具有更强的语义信息，但是分辨率很低，对细节的感知能力较差[16]。因此将低层特征和高层特征融合增加不同层之间的联系，减少重复的人脸框，另一方面引入上下文信息可以提高小尺寸人脸的检测精度。本文提出的特征图融合模型如图3所示，图3(a)将Inception1和Inception2经过特定方式融合构成一个特征图Conv3；图3(b)将Conv3和Inception3经过特定方式融合构成一个特征图Conv4。Concat层可以将2个及以上的特征图按照通道数或特征维度进行拼接，以此融合输入层的特征信息；Conv1*1是卷积核大小为1*1的卷积层，可使特征图降维，减少网络计算量，加快检测速度。

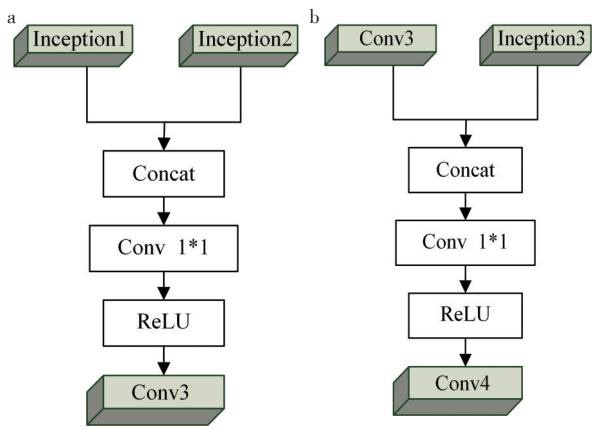

图3　特征图融合模块：(a)Inception1和Inception2，(b)Inception3和Conv3

Fig. 3　Modules of feature map fusion: (a)Inception 1 and 2, (b) Inception 3 and Conv 3

1.3　多尺度人脸检测

为了解决人脸的多尺度问题，采用多个卷积层预测人脸框位置。在网络中的Conv4、Conv5和Conv6层进行多尺度检测，利用不同大小的检测框和检测框密集策略来实现多尺度检测。Inception模块可用于检测多尺度的人脸，该模块由多个卷积核大小不同的卷积组成，针对网络宽度做多尺度设计，可以增加网络深度和宽度，减少网络参数。在图1人脸检测网络中使用了3个Inception结构，其结构如图4所示。

使用不同大小的检测框预测人脸位置，可以共享网络层参数，减少计算量提高人脸检测速度。利用人脸的形状特点将检测框设置成正方形，检测框内的任意输入都会影响输出结果。然而测试结果显示，中间位置的输入对输出结果的影响最大，整体呈现一种中心高斯分布形态。定义检测框密集公式，如公式(1)所示。A_{density}是预测框的密度，指的是检测框的长宽比；m是检测框密集次数，指检测框的重复次数；A_{scale}是预测框的尺寸，指的是对应检测框的像素大小；A_{stride}是预测框移动的位移量，指检测框移动的像素个数。

$$A_{\text{density}} = m \times A_{\text{scale}} / A_{\text{stride}} \tag{1}$$

为更好检测多尺度人脸，利用检测框重叠策略。将A_{density}的值设为4，这样不同尺度的人脸匹配到的检测框密度相同。当出现小尺度的预测框时，适当增加检测框密集次数m，使A_{density}的值等于4。多尺度人脸检测框的参数设置如表1所示。

1.4　损失函数

算法预测了人脸框的坐标和人脸的类别信息，所以采用多级损失函数。根据默认检测框和真实检测框位置做Jaccard相似度计算，把相似度大于0.5的默认框设置为正样本，其它为负样本。使用2级Softmax损失函数进行分类，用Smooth1Loss进行回归。损失函数公式如下。

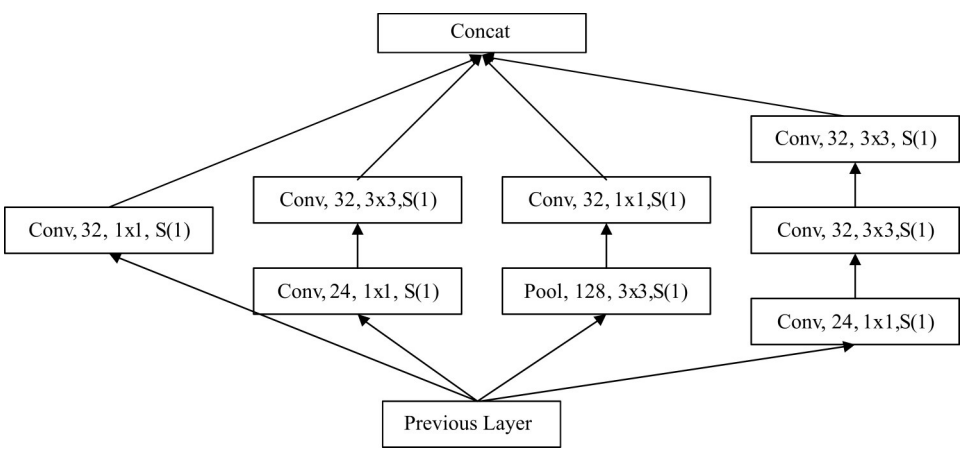

图4　Inception结构图

Fig. 4　Structure diagram of inception

表1 多尺度人脸检测框参数

Tab. 1 Parameters of face detection in multi-scale

特征层	检测框密度	密集数/次	检测框大小/像素	位移量/像素
Conv4	1、2、4	4、2、1	32、64、128	32
Conv5	4	1	256	64
Conv6	4	1	512	128

$$L\left(\{p_i\},\{u_i\}\right)=\frac{1}{N_{\text{cls}}}\sum_i L_{\text{cls}}\left(p_i,p_i^*\right)+\lambda\frac{1}{N_{\text{reg}}}\sum_i p_i^* L_{\text{reg}}(t_i,t_i^*) \tag{2}$$

公式(2)中，p_i是目标感受野的概率；p_i^*是标签，p_i^*为0时，表示为负样本，p_i^*为1时，表示为正样本；$\boldsymbol{t}_i=\{\boldsymbol{t}_x,\boldsymbol{t}_y,\boldsymbol{t}_w,\boldsymbol{t}_h\}$是一个向量，表示预测框的4个参数坐标；$\boldsymbol{t}_i^*$是正确的目标感受野的坐标向量；$L_{\text{cls}}(p_i,p_i^*)$是目标和非目标的对数损失，$L_{\text{cls}}\left(p_i,p_i^*\right)=-\log\left[p_i^* p_i+(1-p_i^*)(1-p_i)\right]$；$L_{\text{reg}}(\boldsymbol{t}_i,\boldsymbol{t}_i^*)$是回归损失，$L_{\text{reg}}\left(\boldsymbol{t}_i,\boldsymbol{t}_i^*\right)=R(\boldsymbol{t}_i-\boldsymbol{t}_i^*)$。

2 实验部分

2.1 实验数据和实验方法

人脸检测模型首先在广泛人脸数据集(wider face dataset，WIDERFACE)上训练，然后在人脸检测数据集基准(face detection dataset and benchmark，FDDB)和野外标注人脸数据集(annotated face in the wild，AFW)上验证。WIDERFACE数据集包含3万多个身份，其中人脸图像有40多万张，该数据库还标记了所有的人脸位置坐标。若人脸图片太小，在训练人脸检测模型时会降低模型收敛速度，所以先将尺寸小于20*20像素的人脸图像过滤掉再进行网络训练。

网络训练和测试都是基于Caffe深度学习框架。使用学习率衰减策略，前8万次网络迭代使用的学习率为0.001，然后每训练2万次迭代学习率会降低0.1倍，一共训练12万次。为避免网络陷入局部最小，网络的动量设置为0.9。为避免过拟合，使用l_2正则化，权重衰减为0.000 5。在CPU检测图像可以达到21帧/s的速度，在GPU上测试可达到125帧/s的速度。

2.2 实验结果分析

2.2.1 模型合理性验证 为了验证本文网络模型的合理性，做了2个对比试验。实验一：没有加入特征融合模块，直接在Inception3、Conv5和Conv6层进行多尺度检测，人脸检测框参数与本文算法一致。实验二：在多个特征图上进行多尺度检测，Conv4设置检测框大小为32*32像素，检测框密集次数为4；Conv5层设置检测框大小为64*64像素，检测框密集次数为2；Inception3设置检测框大小为128*128像素，检测框密集次数为0；Conv5和Conv6层的检测框参数与本文算法一致。在AFW数据集上验证模型合理性，AFW数据集包含205张人脸图像。实验结果表明本文模型优于对比模型，如表2所示。

表2 不同融合方式的精确度比较

Tab. 2 Accuracy comparison of different fusion methods

模型名称	平均精确度
本文融合模型	0.954
对比模型1	0.931
对比模型2	0.948

2.2.2 模型检测速度对比 为验证本文方法的实时性，对比了ACF、Jiont Cascade、MultiresHPM、MTCNN和Faceboxes 5种人脸检测方法。从网络输入图像尺寸、网络检测人脸尺寸和检测速度做了对比。人脸检测算法在CPU环境下检测速度对比，如表3所示。由表3可知，在CPU下检测大小为640*480像素的图像，本文算法对比其他5种算法，网络检测到的人脸尺寸最小且速度最快。

表3 算法的检测速度和检测尺寸比较

Tab. 3 Comparison of detection speed and time with different algorithms

算法	人脸检测尺寸/像素	检测速度/(帧/s)
ACF	—	20
JointCascade	80*80	35
MultiresHPM	—	10
MTCNN	24*24	16
Faceboxes	20*20	20
本文模型	20*20	21

2.2.3 模型检测精度对比 为验证本文模型检测精度，在FDDB数据集与5种人脸检测算法进行对比。FDDB数据集包括2 845张图像，一共标注了5 171张人脸。FDDB是具有标准评估过程的数据集，使用椭圆框标注人脸位置。本文算法用矩形框标注，所以先把椭圆标注转化为矩形标注。我们遵循FDDB数据集的评估流程，用官方提供的工具箱测试本文人脸检测算法的精度。FDDB数据集的测试标准可分为离散评分和连续评分2种情况。离散评分以检测到的人脸框和真实人脸框的重合面积为评判标准，当2个框的重合面积大于0.5时认为检测到了人脸。连续评分是以检测框和重叠框的重合面积的比率为评判标准，重叠比

越大识别率越高。FDDB数据集检测评分如图5所示，图5(a)为离散评分，图5(b)为连续评分。

由图5可知，检测算法在FDDB数据集上的离散和连续情况下正确率分别为92.1%和71.1%，都达到了较为领先的检测效果。虽然算法检测精度低于MTCNN，但是算法在速度上是快于MTCNN，所以本文算法可实现实时高效的人脸检测。

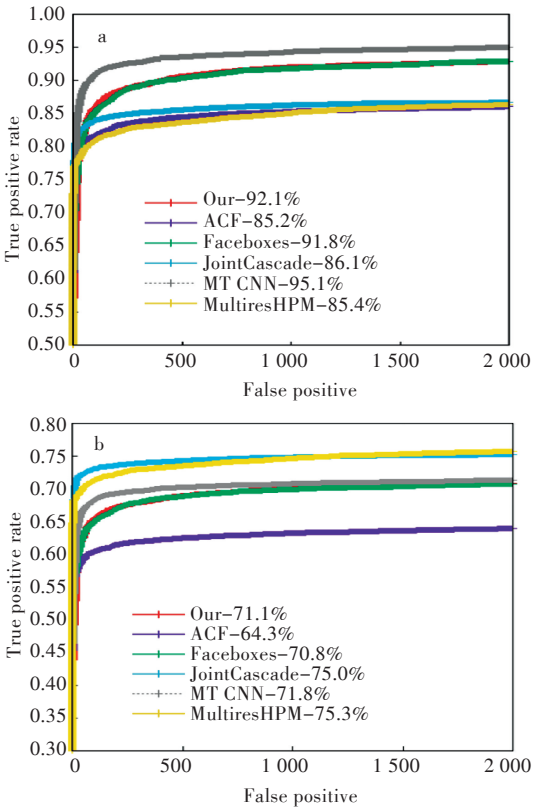

图5　FDDB数据集结果：(a)离散评分，(b)连续评分

Fig. 5　Evaluation on FDDB dataset: (a) discontinuous score, (b)continuous score

3　结　语

为提高人脸检测速度，在初始卷积层中使用较大的卷积核和移动步长，快速缩小输入图像尺寸；加入特征图融合模块，增强不同层之间的联系，减少人脸重复框；使用多尺度卷积层，检测不同尺度的人脸；利用多级损失函数，使模型训练更快收敛。该算法在FDDB和AFW数据集上达到了良好的检测效果，在CPU上图像的检测速度为21帧/s。但是该算法限制了最小检测的人脸尺寸，当检测的人脸图像尺寸小于20*20的像素时检测效果不佳。因此对小尺寸人脸图像的检测需要进一步提高其检测效果。

参考文献

[1]　夏平平，吕太之. 动态人脸识别系统的设计与实现[J]. 武汉工程大学学报，2011，33(10)：107-110.

[2]　冯超. 深度学习轻松学[M]. 北京：电子工业出版社，2018：4.

[3]　VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features [J]. CVPR (1), 2001, 1(3): 511-518.

[4]　阮锦新，尹俊勋. 基于人脸特征和AdaBoost算法的多姿态人脸检测[J]. 计算机应用，2010，30(4)：967-970.

[5]　AHONEN T, HADID A, PIETIKAINEN M. Face description with local binary patterns: application to face recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28 (12): 2037-2041.

[6]　薛超，于宏志，王景彬. 基于卷积神经网络的级联人脸检测[J]. 中国安防，2017(11)：93-96.

[7]　YANG B, YAN J J, LEI Z, et al. Aggregate channel features for multi-view face detection [C]//IEEE International Joint Conference on Biometrics. Florida: IEEE, 2014: 1-8.

[8]　CHEN D, REN S Q, WEI Y C, et al. Joint cascade face detection and alignment [C]//European Conference on Computer Vision. Zurich: ECCV, 2014: 109-122.

[9]　GHIASI G, FOWLKES C C. Occlusion coherence: detecting and localizing occluded faces [J]. Computer Science, 2015: 1-9.

[10]　ZHAN 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

[11]　ZHANG S F, ZHU X Y, LEI Z, et al. FaceBoxes: a cpu real-time face detector with high accuracy [C]// 2017 IEEE International Joint Conference on Biometrics (IJCB). Colorado: IEEE, 2017: 1-9.

[12]　卢涛，章瑾，陈白帆，等. 多尺度自适应配准的视频超分辨率算法[J]. 武汉工程大学学报，2016，38(2)：178-184.

[13]　汪家明，卢涛. 基于多尺度残差深度神经网络的卫星图像超分辨率算法[J]. 武汉工程大学学报，2018，40(4)：440-445.

[14]　SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. LasVegas: IEEE, 2016: 2818-2826.

[15]　SHANG W, SOHN K, ALMEIDA D, et al. Understanding and improving convolutional neural networks via concatenated rectified linear units [C]// International Conference on Machine Learning. New York: ICML, 2016: 2217-2225.

[16]　王成济，罗志明，钟准，等. 一种多层特征融合的人脸检测方法[J]. 智能系统学报，2018，13(1)：138-146.

本文编辑：陈小平