

# 一种轻量级的多尺度特征人脸检测方法

朱 鹏<sup>1</sup>, 陈 虎<sup>1\*</sup>, 李 科<sup>2</sup>, 程宾洋<sup>3</sup>

(1. 四川大学 计算机学院, 四川 成都 610065;

2. 四川大学 视觉合成图形图像技术国防重点学科实验室, 四川 成都 610065;

3. 四川川大智胜软件股份有限公司, 四川 成都 610045)

**摘 要:**目前,各种 start-of-the-art 的人脸检测算法被提出,但是在不断提高精度的同时,却忽略了检测的实时性和可应用性。针对这一问题,提出了一种轻量级、实时性和 Single-Stage 的高精度人脸检测算法。通过将主干网络的不同尺度的特征输出到对应的检测模块中进行检测,实现不同大小人脸的检测,提高算法精度;主干网络运用深度可分离卷积将传统的 $3\times 3$ 卷积核分为一个深度卷积核和 $1\times 1$ 卷积核,减少计算量;检测模块中使用特征融合获得更多的上下文信息和更大的感受野,并包含目标分类和框体回归操作;采用全卷积神经网络,减少内存量,并使得网络可以输入不同尺寸的图像。实验结果表明,算法在 Fddb 数据集上有着 93.52% 的准确率,同时在不同尺度、姿态、装扮和光照等环境下具有较好的鲁棒性,并且能够达到实时检测。

**关键词:**人脸检测;轻量级;多尺度特征;特征融合;全卷积神经网络;实时检测

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2020)04-0001-07

doi:10.3969/j.issn.1673-629X.2020.04.001

## A Face Detection Method with Lightweight and Multi-scale Feature

ZHU Peng<sup>1</sup>, CHEN Hu<sup>1\*</sup>, LI Ke<sup>2</sup>, CHENG Bin-yang<sup>3</sup>

(1. School of Computer Science, Sichuan University, Chengdu 610065, China;

2. National Key Laboratory of Fundamental Science on Synthetic Vision, Sichuan University, Chengdu 610065, China;

3. Wisisoft Co. Ltd., Chengdu 610045, China)

**Abstract:** At present, various start-of-the-art face detection algorithms have been proposed, but while continuously improving the accuracy, the real-time and applicability of detection have been neglected. Aiming at this problem, we propose a lightweight, real-time and Single-Stage high-precision face detection algorithm. We input different scale features of the backbone network into different detection modules to realize the detection of different size faces and improve the detection accuracy of the algorithm. The backbone network uses the deep separable convolution to divide the traditional  $3\times 3$  convolution kernel into a deep convolution kernel and a  $1\times 1$  convolution kernel, reducing the amount of computation. The detection module uses feature fusion for more contextual information and greater receptive fields, including target classification and box regression operations. In addition, we use a full convolutional neural network to save memory and enable the network to input images of different sizes. The experiment shows that the proposed algorithm has 93.52% accuracy in Fddb dataset and better robustness in different scales, poses, dressing and illumination, which can achieve real-time detection.

**Key words:** face detection; lightweight; multi-scale feature; feature fusion; full convolutional neural network; real-time detection

## 1 概 述

早期人脸检测算法大多使用的是模板匹配技术。其中具有代表性的是 Rowley 等人提出的基于神经网络的正面人脸检测系统<sup>[1]</sup>,通过滑动窗口检测图像,并

确定每个窗口是否包含面部,用 $20\times 20$ 的人脸和非人脸图像训练了一个多层感知器模型,该系统在多个网络之间进行判断,从而提高单个网络的性能,主要解决近似正面的人脸检测问题。在后续 Rowley 对算法的

收稿日期:2019-05-15

修回日期:2019-09-17

网络出版时间:2019-12-05

基金项目:国家重点研发计划(NO.2016YFC0801100)

作者简介:朱 鹏(1994-),男,硕士研究生,研究方向为计算机视觉;陈 虎,博士,讲师,硕导,通讯作者,研究方向为模式识别、图像处理、人脸识别等。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20191205.1146.076.html>

更新中,加入了角度估计器<sup>[2]</sup>,可对待检测区域进行旋转从而再进行正面人脸检测。不过由于复杂的分类器设计和密集的滑动窗口采样需要大量的运算,使得检测速度差强人意。

2001 年 Viola 和 Jones 设计了一种基于 AdaBoost 分类器的人脸检测算法<sup>[3]</sup>,使用简单的 Haar-like 特征和级联的 AdaBoost 分类器对滑动窗口中的候选框进行判定,同时为了提高效率,引用了积分图(integral image)机制。在深度学习出现之前,工程中使用的人脸检测大多都是基于该算法的延伸。但是由于 Haar-like 是相对简单的一种特征,所以其稳定性较低;同时 AdaBoost 弱分类器采用简单的决策树,对正面人脸的检测效果较好,但是对于各种人脸姿态、表情和遮挡等特殊复杂情况检测效果并不理想;并且在该算法的分类器设计中,对一个样本的评价不会基于该样本在之前阶段中的表现,所以分类器的鲁棒性差。基于该方法的后续研究中,使用了其他特征,比如局部的二进制模式(LBP)<sup>[4]</sup>和颜色特征<sup>[5]</sup>,使得检测效果得到了很大的改进,并且在此基础上用于人脸追踪<sup>[6]</sup>。

2008 年 Felzenszwalb 提出基于 DMP (deformable part model) 模型的人脸检测算法<sup>[7]</sup>,采用 FHOG (对 HOG 的改动)进行特征的提取,该算法在多姿态、多角度和遮挡等场景下的检测效果要优于 Viola 和 Jones 提出的算法效果,但是由于该模型的计算量过大,达不到实时性的要求,所以在实际工程中很少采用。

自 2012 年 AlexNet 在 ILSVRC (imagenet large scale visual recognition challenge) 的图像分类比赛中取得突破性成绩后,卷积神经网络很快运用到图像处理和目标检测中,并同样取得重大进展<sup>[8]</sup>。通过卷积神经网络能够让网络更好地获取到目标各个层级的特征,从而增强了网络对外界环境的鲁棒性。

Cascade CNN 是传统检测向深度网络过渡的代表<sup>[9]</sup>,由 Li Haoxiang 等人提出,和 Viola-Jones 的人脸检测器相似,采用级联的方式组织多个分类器,不同的是每一级的分类器由卷积神经网络组成,并且构建图像金字塔,通过滑动窗口将待检测区域送入三个网络中进行回归校正,通常第一个网络就可以剔除绝大多数负样本,同时每个网络间采用 NMS (non-maximum suppression) 算法去除高度重叠的区域。之后的 MTCNN 网络<sup>[10]</sup>也是借鉴了 Cascade CNN 模型的思想。随后的研究<sup>[11]</sup>中,使用 L2 损失与三元组损失相结合的总损失函数,减少漏检;在文献[12]中还使用了两层级联网络;在文献[13]中,级联网络之间加入了图像金字塔;这些改进都取得了较好的检测效果。

文中提出的模型是基于 SSH (single stage headless)<sup>[14]</sup>的人脸检测模型,SSH 最大的特点就是尺

度不相关性(scale-invariant),MTCNN 是对待检测图像的不同尺度分别进行预测,如此一来计算量会大大增加,而 SSH 只需要一个尺度的图像就可以预测,实现了 one stage 的人脸检测。其原理是利用 3 个检测模块(M1,M2,M3)分别对主干 VGG 网络的 3 个不同特征图进行预测,每个检测模块都包含检测和分类,通过对不同尺度的特征图进行预测,实现了多尺度的人脸检测,但是传统 VGG 网络的计算量相对较大,网络参数也相对较多。文中的主干网络参考了当前先进 Mobile-Net<sup>[15]</sup>结构的思想,设计出了一种轻量级网络结构,该网络主要由多个 3×3,1×1 卷积核和传统卷积核构成,将一个标准卷积分解为一个深度卷积和一个点卷积(1×1 卷积核),在保证卷积效果的同时大大缩减了计算复杂度,从而提高了检测效率。此外,模型的检测模块融合多个卷积特征,可以获得更多的上下文信息以及更大的感受野,从而提升检测性能。

## 2 设计方法

### 2.1 整体框架

该方法是在 caffe 深度学习框架下设计和训练的,该框架具有简洁快速的特点。网络模型由 1 个主干网络和 3 个检测模块构成。主干网络未使用标准卷积,而是将一个标准卷积分解为一个深度卷积和点卷积(1×1 卷积核),从而减少运算量和参数,提高运算效率,降低内存消耗;同时主干网络在每一层卷积后都会加入批量归一化(batch normal),加快模型收敛速度,避免网络的梯度消失或爆炸问题,并减少网络对初始参数的依赖;3 个检测模块(M1,M2,M3)分别检测主干网络中 3 个不同尺度的特征图,以检测不同大小的目标,其中检测模块包含目标分类和框体回归操作,由一个普通的 3×3 卷积,一个 context Module 和两个 1×1 卷积组成。context Module 为了获得更多的上下文信息和更大的感受野,使用了 5×5 和 7×7 的卷积核分别进行卷积操作,然后使用 concat 融合特征;另外网络去除了全连接层,使用 1×1 卷积核代替全连接层,使得网络的输入可以像 Faster RCNN<sup>[16]</sup>一样输入不同尺寸的图片,同时减少模型参数。总体来说,文中提出的模型考虑了效率和准确度,既能保证检测精度,又能达到人脸的实时检测,可应用于实际的工程中。

网络模型的总体结构如图 1 所示。

其中 M1、M2 和 M3 分别表示检测模块 1、检测模块 2 和检测模块 3,分别用于检测小目标、中目标和大目标。

### 2.2 主干网络设计

文中提出的主干网络参考了深度可分离卷积<sup>[17]</sup>的方法,该方法的核心是将一个传统卷积分离为深度

卷积 (depthwise convolution) 和一个  $1 \times 1$  的卷积 (pointwise convolution), 在尽量提取特征的同时, 减少网络参数和网络计算量。结合多尺度特征提取的思想和人脸检测的实际场景应用, 设计出了一种传统卷积

和深度可分离卷积相结合的, 轻量级的, 具备实时性的主干网络。现有一个输入为  $W \times W \times C \times 1$  的特征, 如图 2 所示将传统卷积分解成深度可分离卷积。

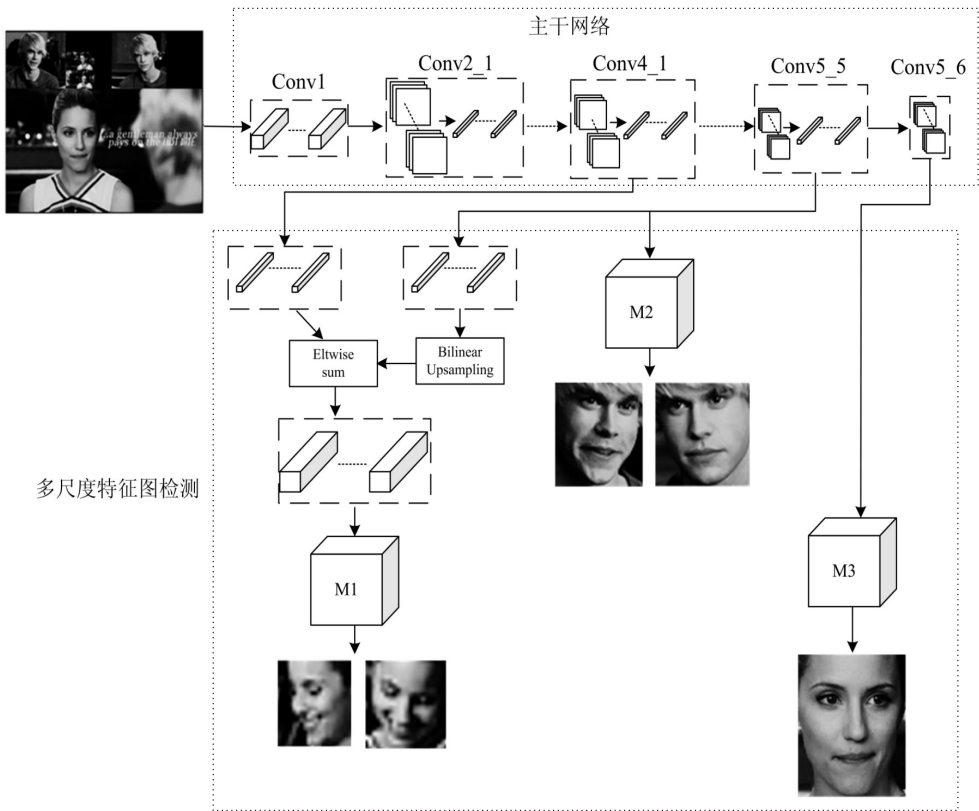


图 1 模型总体结构

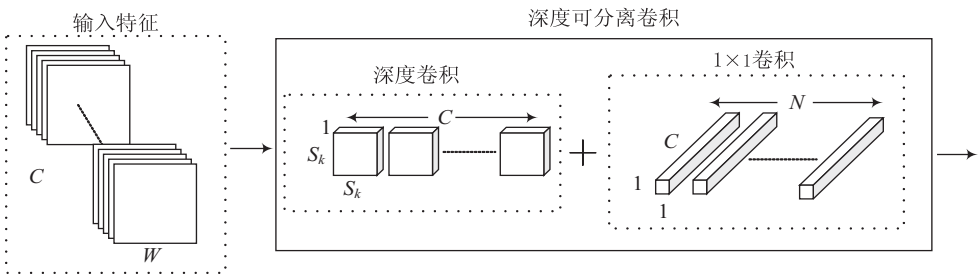


图 2 深度可分离卷积

其中  $k$ 、 $S$ 、 $C$  和  $N$  分别为卷积核的标识、卷积核的尺寸、输入特征图的深度和输出特征图的深度。深度卷积中每个卷积核的通道数为 1, 并且分别与输入特征图的单个相应通道做卷积运算, 即对每个输入特征的通道应用一个深度卷积核, 如第一个深度卷积核和输入特征的第一个通道做卷积; 现假设有一个  $D_{\text{input}} \times D_{\text{input}} \times C$  的输入特征图 ( $D_{\text{input}}$  为输入特征图的尺寸), 经过 stride 为 1, 卷积核为  $S_k \times S_k \times C \times N$  的卷积操作输出  $D_{\text{output}} \times D_{\text{output}} \times N$  的特征图 ( $D_{\text{output}}$  为输出特征图的尺寸), 令  $D_{\text{output}} = D_{\text{input}}$ , 则传统卷积操作的计算量  $\text{Amount}(g)$  为:

$$\text{Amount}(g) = S_k \times S_k \times C \times N \times D_{\text{input}} \times D_{\text{input}} \quad (1)$$

深度可分离卷积的计算量  $\text{Amount}(\text{ds})$  为:

$$\text{Amount}(\text{ds}) = S_k \times S_k \times D_{\text{input}} \times D_{\text{input}} + C \times N \times D_{\text{input}} \times D_{\text{input}} \quad (2)$$

那么深度可分离卷积和传统卷积的计算量之比为:

$$\frac{\text{Amount}(\text{ds})}{\text{Amount}(g)} = \frac{1}{N} + \frac{1}{S_k^2} \quad (3)$$

综上, 若输入  $224 \times 224 \times 64$  的特征图, 经过  $3 \times 3$  的卷积核输出  $224 \times 224 \times 128$  的特征图, 那么两者的计算量之比约为 0.118 9, 也就是说使用深度可分离卷积操作, 计算量可缩减为传统卷积的 1/9 左右, 运算速度大幅提升。

主干网络中的一组深度可分离卷积的结构如图 3 所示。

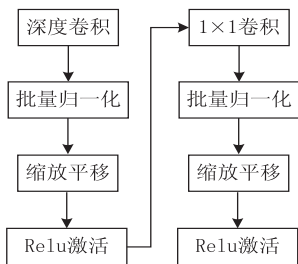


图 3 一组深度可分离卷积

主干网络输入为  $224 \times 224 \times 3 \times 1$  的图像数据,在网络中的参数变化如表 1 所示;主干网络总共 26 个卷积层,没有池化层。网络最终输出特征图的尺寸为  $7 \times 7 \times 512 \times 1$ 。

表 1 主干网络参数

卷积/步长/填充	卷积核	输入尺寸
Conv1/s2/p1	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3 \times 1$
Conv2_1(dw)/s1/p1	$3 \times 3 \times 1 \times 32$ (dw)	$112 \times 112 \times 32 \times 1$
Conv2_1(sep)/s1/p0	$1 \times 1 \times 32 \times 64$ (sep)	$112 \times 112 \times 32 \times 1$
Conv2_2(dw)/s1/p1	$3 \times 3 \times 1 \times 64$ (dw)	$112 \times 112 \times 64 \times 1$
Conv2_2(sep)/s1/p0	$1 \times 1 \times 32 \times 64$ (sep)	$112 \times 112 \times 64 \times 1$
Conv2_3(dw)/s2/p1	$3 \times 3 \times 1 \times 64$ (dw)	$112 \times 112 \times 64 \times 1$
Conv2_3(sep)/s1/p0	$1 \times 1 \times 64 \times 128$ (sep)	$56 \times 56 \times 64 \times 1$
Conv3_1(dw)/s1/p1	$3 \times 3 \times 1 \times 128$ (dw)	$56 \times 56 \times 128 \times 1$
Conv3_1(sep)/s1/p0	$1 \times 1 \times 128 \times 128$ (sep)	$56 \times 56 \times 128 \times 1$
Conv3_2(dw)/s2/p1	$3 \times 3 \times 1 \times 128$ (dw)	$56 \times 56 \times 128 \times 1$
Conv3_2(sep)/s1/p0	$1 \times 1 \times 128 \times 256$ (sep)	$28 \times 28 \times 128 \times 1$
Conv4_1(dw)/s1/p1	$3 \times 3 \times 1 \times 256$ (dw)	$28 \times 28 \times 256 \times 1$
Conv4_1(sep)/s1/p0	$1 \times 1 \times 256 \times 256$ (sep)	$28 \times 28 \times 256 \times 1$
Conv4_2(dw)/s2/p1	$3 \times 3 \times 1 \times 256$ (dw)	$28 \times 28 \times 256 \times 1$
Conv4_2(sep)/s1/p0	$1 \times 1 \times 256 \times 512$ (sep)	$14 \times 14 \times 256 \times 1$
Conv5(dw)/s1/p1	$3 \times 3 \times 1 \times 512$ (dw)	$14 \times 14 \times 512 \times 1$
Conv5(sep)/s1/p0	$1 \times 1 \times 512 \times 512$ (sep)	$14 \times 14 \times 512 \times 1$
Conv5_6(dw)/s2/p1	$3 \times 3 \times 1 \times 512$	$14 \times 14 \times 512 \times 1$

2.3 检测模块设计

检测模块也是一个全卷积网络,分为三个检测模型 M1、M2 和 M3 分别抽取主干网络中 Conv4\_1、Conv5\_5 和 Conv5\_6 层的 3 个不同尺度特征图,对应的 stride 为 8、16 和 32,再分别对这三个尺度的特征图进行预测,包括人脸分类和人脸框体回归。

2.3.1 多尺度特征图

主干网络中 Conv4\_1、Conv5\_5 和 Conv5\_6 层生成的特征图尺寸分别为  $28 \times 28 \times 256$ 、 $14 \times 14 \times 512$  和  $7 \times 7 \times 512$ ,其中低层的感受野较小用于检测小目标,高层的感受野较大用于检测大目标,不同尺寸的特征图可

以获取多尺度的特征映射。

M2 和 M3 检测模块都是直接从主干网络中获取特征图,而 M1 检测模块融合了 Conv4\_1 和 Conv5\_5 层的特征图,首先对 Conv4\_1 和 Conv5\_5 层的输出采用  $1 \times 1$  卷积核,降维得到  $28 \times 28 \times 128$  和  $14 \times 14 \times 128$  尺寸的特征图,并对 Conv5\_5 采取双线性上采样输出尺寸为  $28 \times 28 \times 128$  的特征图,再将两个特征图采用对应元素相加的方式进行特征融合,融合后得到  $28 \times 28 \times 128$  的特征图,经过一个  $3 \times 3 \times 128 \times 128$  的卷积再送入 M1 检测模块。这样做的主要目的是为了增加 M1 检测模块的语义信息。

综上,通过检测不同尺度的特征,可以检测到不同大小的目标。如图 4 所示,浅层的 M1 检测模块可以更好地获取到小目标人脸特征,而更深层的 M2 检测模块则无法获取到小目标人脸特征。

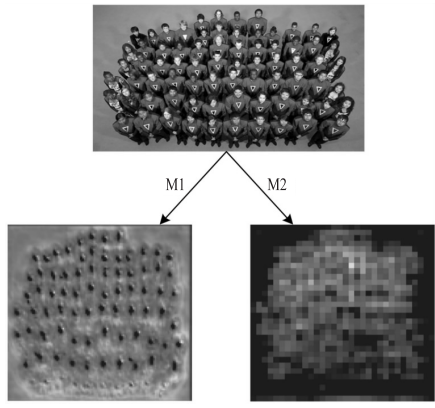


图 4 多尺度特征检测热力图

2.3.2 检测模块

M1、M2 和 M3 检测模块的结构相似,区别在于输入的特征图大小不同,对应的 stride 分别为 8、16 和 32,同时 M1 的输入特征通道数为 128,而 M2 和 M3 都是 512。检测模块结构如图 5 所示,每个检测模块都包含人脸分类和框体回归,进行框体回归的时候,定义一系列边界框(anchor),通过回归的方式逐渐使其靠近真实的人脸框。类似 Faster-RCNN 中的 RPN 思想使用密集的滑动窗去定义一系列的 anchor,在每个检测模块特征图的任一被滑动窗口覆盖的区域,定义具有和当前滑动窗口相同中心位置但尺度不同的  $K$  个 anchor,每个 anchor 对应一个尺度,原始的 RPN 中  $K$  个 anchor 具有不同的长宽比,文中所有的 anchor 长宽比都设为 1,因为人脸的长和宽差距不大;假设一个特征图的大小为  $W \times H$ ,那么最终就会得到  $W \times H \times K$  个长宽比为 1,尺度不同的 anchor。

如图 5 所示,首先检测模块对输入特征分别进行两种卷积,一种是普通的  $3 \times 3$  卷积,另一种是 Context Model;接着将普通的  $3 \times 3$  卷积和 Context Model 进行 channel 维度上的拼接,最后两个  $1 \times 1$  的卷积层用于框



体回归和人脸分类,分类器决定当前卷积核覆盖的中心位置的各个尺度的图像是否为人脸。最终输出  $W/S \times$

$H/S \times 2K$  个分类得分和  $W/S \times H/S \times 4K$  个回归坐标。

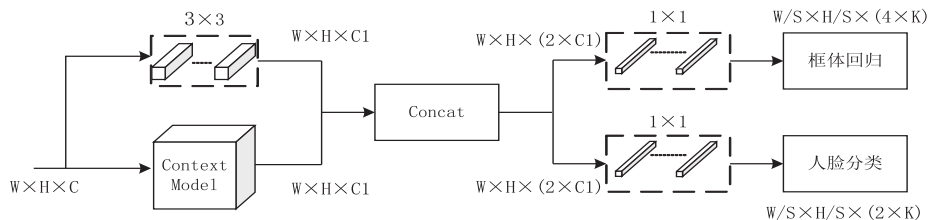


图5 检测模块结构

图5中 Context Model 的具体结构如图6所示。在 two-stage 检测中,通常会通过扩大候选区域的窗口来合并上下文信息,文中的 Context Model 通过简单的卷积层来模仿这种策略来合并上下文信息,由于 anchor 是以卷积方式进行分类和回归的,所以通过使用更大的卷积核就可以实现类似 two-stage 检测中的扩大候选框的方式。Context Model 采用  $5 \times 5$  和  $7 \times 7$

的卷积核,这样可以增加对应层的感受野,同时也增加了每个检测模块中的目标尺度;为了减少模型的参数量,并提高网络的非线性表达能力,采用两个  $3 \times 3$  的滤波器和三个  $3 \times 3$  的滤波器替代  $5 \times 5$  和  $7 \times 7$  的滤波器,该方法和文献[18]中提到的类似。Context Model 输出通道的个数为  $C1$ ,  $M1$ 、 $M2$  和  $M3$  的  $C1$  分别为 128、256 和 256。

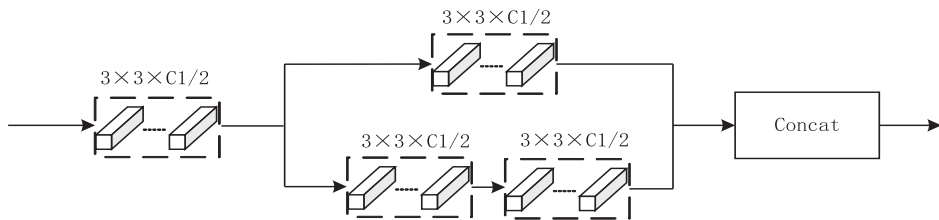


图6 Context Model

### 3 训练

使用动量梯度下降和权值衰减的方式训练网络。因为该网络通过3个检测模块来实现多尺度的人脸检测,所以训练过程中需要使用3个相互独立的 loss。为了使得3个检测模块针对相应的尺度进行检测,在训练的时候为每个模块分配特定尺度范围的 anchor,将最小尺度的 anchor 分配给  $M1$  模块,中等尺度的分配给  $M2$  模块,最大尺度的分配给  $M3$  模块,同时只将属于该检测模块尺寸的人脸的分类和回归损失值进行反向传播。另外当且仅当一个 anchor 与真实人脸之间的 IOU (预测框和真实框的重叠指标) 大于 0.5 时,才将该 anchor 与真实人脸区域进行对比,所以反向传播不会迭代与真实人脸区域不匹配的 anchor。

#### 3.1 损失函数

损失函数  $J$  如式(4)所示,其中  $l_c$  为人脸分类 loss,使用标准的多项 log 的 loss。 $k$  为检测模块的索引  $M_k$ ,并且  $A_k$  表示定义在  $M_k$  中的 anchor 集合。在  $M_k$  中第  $i$  个 anchor 和对应真实人脸区域的标签分别表示为  $p_i$  和  $g_i$ 。anchor 中的正样本与真实人脸区域的 IOU 超过 0.5,负样本与真实人脸区域的 IOU 小于 0.5。 $N_k^r$  为模块  $M_k$  中的 anchor 个数,会参与到分类 loss 的计算。

$$J = \sum_k \frac{1}{N_k^r} \sum_{i \in A_k} l_c(p_i, g_i) + \lambda \sum_k \frac{1}{N_k^r} \sum_{i \in A_k} I(g_i = 1) l_r(b_i, t_i) \quad (4)$$

$l_r$  表示边界框回归中的 loss,回归空间是基于框体维度的 log 空间迁移和尺度不变性的变换,使用  $l_r$  平滑 loss。在参数化空间中,  $b_i$  表示预测框的值,  $t_i$  表示真实人脸区域,  $I(\dots)$  表示指示函数,用于限制回归 loss 只对标记类别为 true 的 anchor 进行计算,同时  $N_k^r = \sum_{i \in A_k} I(g_i = 1)$ 。

#### 3.2 困难样例挖掘 (OHEM)

由于实际情况中正负样本不均衡,因此对每个检测模块进行 OHEM<sup>[19]</sup>,即使用负样本中与真实人脸区域的 IOU 最大的样本与正样本中与真实人脸区域的 IOU 最小(大于 0.5)的样本组成一个小批量进行训练,不过由于负样本通常会很多,所以这里选取 25% 的小批量作为正样本。

### 4 结果分析

文中提出的算法在 FDDB (face data set and benchmark) 数据集上进行评测,FDDB 数据集是评估人脸检测算法性能最常用的大型数据集之一,这个数据集拥有 2 845 张有人脸的互联网新闻图像,总共标

记了 5 171 张人脸。数据集在人脸姿态、表情、光照、清晰度、分辨率、遮挡程度等方面具有很大的多样性, 贴近真实应用场景, 因此成为最受欢迎、最权威的数据集之一。FDDDB 提出两种类型的图像检测评分, 分别是离散分数和连续分数。在第一种评分下, 关注的是检测框和标注框之间的 IOU 是否超过 0.5, 而第二种评分下关注的则是检测框和标注框之间的 IOU 越大

越好。文中提出的算法采用离散分数的方法进行评测。图 7 为此算法与当下表现较好的 DP2MFD<sup>[20]</sup>、FaceNess<sup>[21]</sup>、CCF、CascadeCNN<sup>[9]</sup>、Acf<sup>[22]</sup>、VJ 算法的对比结果, 其中 CCF 为级联分类器, 使用的是 Haar-like 矩形特征。其中横坐标表示假正类的个数, 纵坐标表示真正类率。可以看出, 文中提出的算法优于其他算法, 最终的真正类率可达到 93.52%。

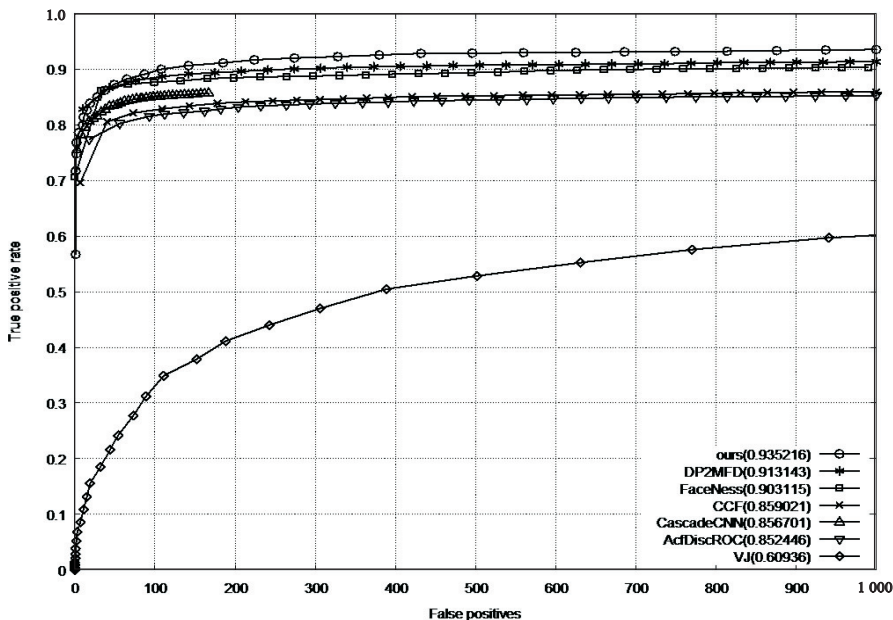


图 7 FDDDB 数据集 ROC 曲线

另外运用该算法分别对 AFW (annotated faces in the wild)、FDDDB 和 WiderFace 数据集的部分数据做检测, 检测效果如图 8 所示。可以看到, 测试数据中的人脸包括不同尺度, 不同姿态, 不同装扮和不同光照的情况, 但是算法鲁棒性强, 检测的准确度很高。



图 8 算法检测效果

不同尺度人脸的 640×480 的图像, 在 Intel (R) Core (TM) i5-7500 CPU @ 3.40 GHz×2 的 CPU 处理上的检测速度达到 2 FPS, 虽然在 CPU 上的表现较为平庸, 但是在普通的 NVIDIA GeForce GTX 1050Ti 4g 显卡上面的检测速度达到 40 FPS。这得益于深度可分离卷积的思想, 通过将普通卷积分解为一个深度卷积和 1×1 卷积, 减少运算量, 并且使用不同尺度的特征图, 没有使用图像金字塔。实验表明该算法在普通 GPU 可以达到实时的人脸检测, 可应用于实际场景。

## 5 结束语

为了提高检测速度, 采用深度可分离卷积代替普通卷积, 减少计算量, 同时减少模型参数和内存, 最终模型参数内存为 25.4 M; 同时将主干网络中的不同尺度特征图输出到不同的检测模块, 实现不同尺度人脸的检测; 此外, 网络去除全连接层, 使用 1×1 卷积代替, 从而进一步减少模型参数, 并且网络可以输入不同大小的图像。另外该人脸检测算法实现 Single-Stage 检测, 便于训练和测试的部署; 模型基于 caffe 深度学习框架, 具体良好的可移植性。最后, 该算法在公开人脸数据库上的检测表现优异, 且在 GPU 上的检测速度满足实时性要求。不足之处在于该算法在 CPU 上的

最后, 对算法的运算速度进行实验, 选用包含 8 张

运算速度并不理想,后续会考虑融入更加轻量型算法的思想,提升 CPU 检测速度,从而运用到配置更低的设备中。

#### 参考文献:

- [1] ROWLEY H A, BALUJA S, KANADE T. Neural network-based face detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1998, 20(1): 23-38.
- [2] ROWLEY H A, BALUJA S, KANADE T. Rotation invariant neural network-based face detection[C]//Proceedings of 1998 IEEE computer society conference on computer vision and pattern recognition (Cat. No. 98CB36231). Santa Barbara, CA; IEEE, 1998: 963.
- [3] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. Kauai, HI, USA; IEEE, 2001: 1-511-1-518.
- [4] 陶西浩, 万定生, 杨志勇. 一种基于改进 LBP 特征的人脸检测方法[J]. 信息技术, 2018(2): 1-4.
- [5] 崔 鹏, 燕天天. 融合 YCbCr 肤色模型与改进的 Adaboost 算法的人脸检测[J]. 哈尔滨理工大学学报, 2018, 23(2): 91-96.
- [6] 史双飞, 张 震. 基于决策的人脸检测与自动跟踪方法[J]. 计算机技术与发展, 2019, 29(9): 35-39.
- [7] FELZENSZWALB P, MCALLESTER D, RAMANAN D. Adiscriminatively trained, multiscale, deformable part model[C]//2008 IEEE conference on computer vision and pattern recognition. Anchorage, AK; IEEE, 2008: 1-8.
- [8] 孙志远, 鲁成祥, 史忠植, 等. 深度学习研究与进展[J]. 计算机科学, 2016, 43(2): 1-8.
- [9] LI H, LIN Z, SHEN X, et al. A convolutional neural network cascade for face detection[C]//2015 IEEE conference on computer vision and pattern recognition (CVPR). Boston, MA; IEEE, 2015: 5325-5334.
- [10] 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.
- [11] 王灵珍, 赖惠成, 王 睿. 基于多任务级联 CNN 与三元组损失的人脸识别[J]. 激光杂志, 2019, 40(5): 17-22.
- [12] 张海涛, 李美霖, 董帅含. 两层级联卷积神经网络的人脸检测[J]. 中国图象图形学报, 2019, 24(2): 203-214.
- [13] 郑成浩, 刘 兵, 周 勇. 尺度无关的级联卷积神经网络人脸检测算法[J]. 计算机应用研究, 2019, 36(2): 593-597.
- [14] NAJIBI M, SAMANGOU EI P, CHELLAPPA R, et al. SSH: single stage headless face detector[C]//International conference on computer vision. Venice; IEEE, 2017: 4885-4894.
- [15] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[J]. The Computing Research Repository, 2017(4): 1-9.
- [16] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [17] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]//Computer vision & pattern recognition. Hawaii; IEEE, 2017: 1800-1807.
- [18] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Computer vision & pattern recognition. Boston; IEEE, 2015: 1-9.
- [19] SHRIVASTAVA A, GUPTA A, GIRSHICK R. Training region-based object detectors with online hard example mining[C]//Computer vision & pattern recognition. Las Vegas; IEEE, 2016: 761-769.
- [20] CHEN J, RANJAN R, KUMAR A, et al. An end-to-end system for unconstrained face verification with deep convolutional neural networks[C]//2015 IEEE international conference on computer vision workshop (ICCVW). Santiago; IEEE, 2015: 360-368.
- [21] YANG S, LUO P, LOY C C, et al. From facial parts responses to face detection: a deep learning approach[C]//2015 IEEE international conference on computer vision (ICCV). Santiago; IEEE, 2015: 3676-3684.
- [22] YANG B, YAN J, LEI Z, et al. Aggregate channel features for multi-view face detection[C]//IEEE international joint conference on biometrics. Clearwater, FL; IEEE, 2014: 1-8.