



# 基于级联网络的快速人脸检测算法

包晓安, 胡玲玲<sup>1</sup>, 张娜<sup>1</sup>, 吴彪<sup>2</sup>, 桂江生<sup>1</sup>

(1. 浙江理工大学信息学院, 杭州 310018; 2. 山口大学东亚研究科, 日本山口 753-8514)

**摘要:** 采用卷积神经网络可有效提高人脸检测算法的精度, 然而其模型参数过于复杂, 在一般设备上检测速度很慢。针对这个问题, 提出了一种三层网络级联的人脸检测算法, 利用级联方式实现网络小型化, 通过多任务方式提高人脸检测的精度。在网络的第一级采用金字塔结构网络, 结合 anchor 机制提取多尺度人脸建议框, 在此基础上结合卷积分解策略和网络加速的方法, 进一步增强网络特征提取的有效性并减少模型参数。实验结果表明: 在 FDDB 上该算法的检测精度和检测速度均优于 MTCNN; 在主频为 2.0 GHz 的八核设备上, 检测速度可以达到 80 fps。

**关键词:** 人脸检测; 金字塔网络; 网络加速; 小型化; 级联网络

**中图分类号:** TP181

**文献标志码:** A

**文章编号:** 1673-3851 (2019) 05-0347-07

## Fast face detection algorithm based on cascade network

BAO Xiaolan<sup>1</sup>, HU Lingling<sup>1</sup>, ZHANG Na<sup>1</sup>, WU Biao<sup>2</sup>, GUI Jiangshen<sup>1</sup>

(1. School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China; 2. Department of East Asian Studies, Yamaguchi University, Yamaguchi 753-8514, Japan)

**Abstract:** Convolutional Neural Network (CNN) can effectively improve the accuracy of face detection algorithm. However, the complex parameters in the model always cause the slow detection speed on common devices. To solve this problem, a face detection algorithm based on three-layer network cascaded was proposed in this paper. It can achieve miniaturization of the network by cascade system, and can improve the accuracy of face detection by using multitask mode. In the first stage of the network, the network with pyramid structure was applied, and multi-scale face prediction boxes were extracted by combining anchor mechanism. On this basis, convolutional decomposition strategy and network acceleration method were combined to further enhance effectiveness of network characteristic extraction and reduce model parameters. The results showed that, both detection precision and detection speed of the algorithm on FDDB were better than that of MTCNN. The speed could reach 80 fps on the eight-core device with the main frequency of 2.0 GHz.

**Key words:** face detection; pyramid network; network acceleration; miniaturization; cascade network

## 0 引言

人脸检测<sup>[1]</sup>是计算机视觉领域的热门研究方

向。随着生物特征技术的快速发展, 人脸检测技术已经被广泛应用于各个领域, 产生了一定的商业价值, 同时具有非常重要的学术价值。Viola 等<sup>[2]</sup>提

收稿日期: 2018-08-12      网络出版日期: 2018-12-03

基金项目: 国家自然科学基金项目(61502430, 61562015); 广西自然科学基金重点项目(2015GXNSFDA139038); 浙江理工大学 521 人才培养计划项目

作者简介: 包晓安(1973-), 男, 浙江东阳人, 教授, 硕士, 主要从事软件测试、智能信息处理方面的研究。

通信作者: 张娜, E-mail: zhangna@zstu.edu.cn

出的 haar 特征结合 AdaBoost 级联分类器的方法,可以快速检测正面人脸。此后,许多研究者致力于用更高级的特征来提高人脸检测精度,如局部二值模式(Local binary pattern, LBP)<sup>[3]</sup>、方向梯度直方图(Histogram of oriented gradient, HOG)<sup>[4]</sup>、尺度不变特征变换(Scale-invariant feature transform, SIFT)<sup>[5]</sup>等。有研究者采用分割策略将多个模型进行组合用于多角度人脸检测<sup>[6]</sup>,其中一种传统方法采用可变形组件模型(Deformable part model, DPM)<sup>[7]</sup>,即用面部各部分的集合来定义一张人脸,先分别检测五官位置,再判定是否为人脸。基于 DPM 的方法对于有遮挡的人脸检测具有较好的效果,但计算过于复杂,大多依赖手工设计的特征,缺乏代表性,且模型需要分步训练,即先提取特征,再统一利用特征和对应的标签训练出合适的分类器。

卷积神经网络(Convolutional neural network, CNN)在 ILSVRC2012 取得了巨大成功,被应用于各个研究领域。近年来,目标检测领域出现了很多基于 CNN 的方法,在挑战性极高的人脸数据集(Face detection data set and bench-mark, Fddb)上拥有较好表现力的检测方法均采用 CNN 实现,其中精度较高的方法均使用多任务学习的策略。除此之外,基于多任务级联卷积神经网络(Multi-task cascaded convolutional networks, MTCNN)<sup>[8]</sup>的人脸检测算法采用多层网络级联的方法,其网络包含人脸分类、边框回归和人脸特征点检测三项任务。Ranjan 等<sup>[9]</sup>进一步将人脸检测、特征点定位、姿势估计以及性别识别四种任务融合到一个网络中,在四项任务中都提高了精度。

然而,一般的基于 CNN 的方法算法复杂且计算量大,从而限制了它的应用环境,例如在手机这类的移动设备,就很难应用这类算法。为了解决该问题,研究人员提出了一系列的算法。Ren 等<sup>[10]</sup>提出了一种两级结构的目标检测框架,其中包含一个全卷积网络<sup>[11]</sup>用于提取建议框和一个深层网络用于调整建议框。为解决尺度不变性, R-CNN 摒弃了传统的图像金字塔结合滑动窗口的耗时方法,而是采用两个网络共享参数,再结合 anchor 模板,避免额外的参数量,提高了检测速度。因为这种方法需要一个深层网络来保证检测的精度,加大了计算量,不能达到实时效果。由此, Li 等<sup>[12]</sup>提出了级联 CNN 的人脸检测方法,该方法不同于 R-CNN 系列,使用浅层网络,从而减少了计算量。该算法在采用 Titan GPU 时速度达到了 100 fps。MTCNN 中

的多任务学习<sup>[13]</sup>方法再一次提升了精度,但速度仍没有提升。

在实际应用中,人脸检测算法不仅要满足光照变化、各种表情及遮挡有鲁棒性,同时还需要满足现实场景需求。之前的研究验证了基于级联方式的人脸检测算法在速度上的优势,但该算法在平衡精度和速度上仍然未达到较好的效果。级联中的第一级网络生成建议框仍然是整个网络最耗时部分。因此,本文提出了一种三层网络级联的人脸检测方法,通过第一层网络提取出大量的人脸建议框,排除大量的背景区域;第二、三层网络对上一层输出的建议框进行进一步分类,获得最终的建议框。该方法利用网络级联、多任务学习,在达到实时应用的情况下获得很好的精度,以期在移动端或无网络的状态下可以快速准确检测人脸。

## 1 人脸检测的级联网络

人脸检测的主要任务是确定图像中检测框的位置以及是否为人脸,而目前各种基于 CNN 的人脸检测算法中,级联 CNN 的人脸检测方案是速度最快的<sup>[14]</sup>。通常级联 CNN 的检测方法包含两个部分,第一部分称为预测阶段,第二部分称为回归阶段。在第一部分需要产生大量的物体建议框,将大部分的背景区域排除,这部分达到的效果越好、越快,一般整个网络在检测精度和检测速度上也是越好、越快。而第二部分则是对第一部分产生的建议框进行更优的回归提取,得到最终的目标检测框及其所属分类。

### 1.1 网络结构设计

#### 1.1.1 第一级网络的设计

针对多尺寸人脸检测问题,传统方法是采用图像金字塔结合滑动窗口,但是生成图像金字塔,即生成网络的多个输入,是一种很耗时的操作。MTCNN 是基于 Cascade CNN 改进,级联三个多任务网络,预处理时采用了图像金字塔的方法,精度明显提升但是速度并未提升。Faster R-CNN 使用了区域预测网络(Region proposal net, RPN)实现端到端的训练,采用 anchor 机制解决尺度不变的问题,并将 RPN 与 Fast R-CNN 网络进行特征共享,实现检测速度的提升。但是 RPN 和检测网络所共享的是深层特征。针对浅层网络,本文设计了一种基于 anchor 的粗略图像金字塔网络,包含一组 anchor 覆盖一个小范围的目标尺度,例如,目标大小从  $12 \times 12$  到  $96 \times 96$ ,第一级网络的结构如图 1 所示。

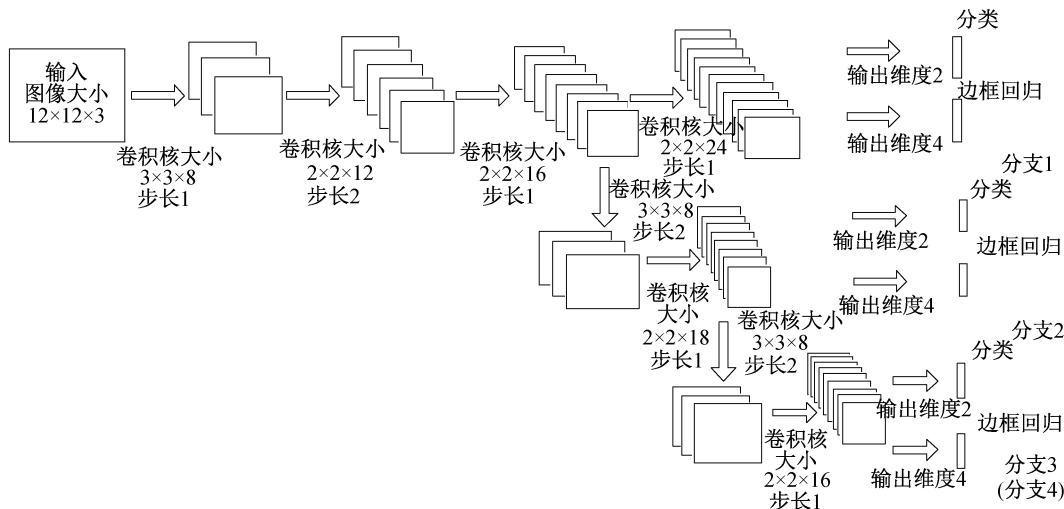


图 1 第一级网络的结构

该级网络采用了双任务，即人脸分类和边框回归，并包含 4 个分支的全卷积网络(FCN)。第一个分支包含 5 个卷积层，每层卷积后紧跟一个 ReLU 层，检测窗口大小为  $12 \times 12$ 。第二个分支检测窗口采用  $24 \times 24$  的大小，网络保留前三层的权重参数，并将第三层卷积的输出作为该分支的输入，因此相较于原图做网络的输入拥有更好的目标特征，且对于较小的分支网络易于训练，分类精度也更好。第三个分支检测窗口大小为  $48 \times 48$ ，如上一个分支一样，将第二个分支的第二层卷积的输出作为这一级分支的输入。第四个分支结构设计和第三分支相同，输入检测窗口大小为  $96 \times 96$ 。每层网络的第一个卷积步长均为 2，其余卷积层的步长设置为 1。

1.1.2 第二、三级网络的设计

卷积层是对局部区域的特征提取，对全局特征不敏感。全卷积网络多用于语义分割，在人脸检测中用于确定建议框的位置。而一般的 CNN 网络是带有全连接层的神经网络，全连接层有别于卷积层，

感应区域为全局的特征信息，因此对于需要实现分类的网络来说，全连接层是必不可少的。而本文的级联网络的第二、三级主要是对第一级网络产生的大量建议框进行回归分类，因此第二、三级网络采用全连接层。而有全连接层的网络要求网络的输入大小是一致的，因此第二、三级网络的输入均需要将上一级网络生成的建议框缩放至下一级网络的输入大小。

在第一级网络输出的大量建议框中，仍然存在很多非人脸的建议框，因此往往会再训练一个或多个更深层的网络来进一步提炼建议框，从而降低第一级网络输出建议框中的困难样本数。本文的回归阶段采用类似 MTCNN 中方法，训练两个浅层网络，网络的输入大小分别为 24 和 48。第二、三级网络的设计如图 2 所示。不同于 MTCNN，在第二级网络训练包含两个任务，相比 MTCNN 中的三个任务，可以将模型的参数进一步缩减，第三级网络采用三个任务训练。后两级网络中均不使用池化层，对应将所有的卷积核步长设置为 2。

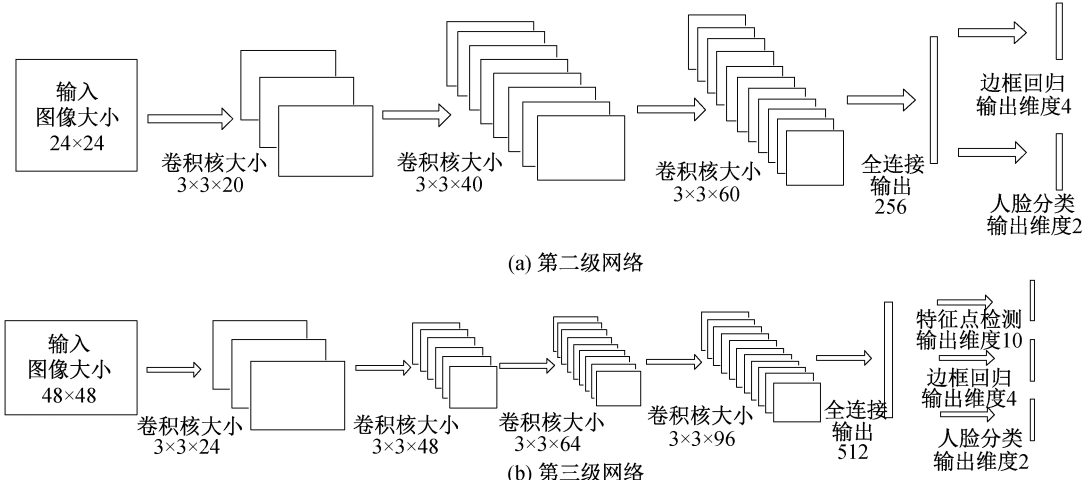


图 2 第二、三级网络的结构

1.2 网络优化

1.2.1 网络速度优化

目前基于 CNN 的深度学习方法在实现速度上主要存在三个问题:需要大量的数据集,且训练花费时间太长;前向传播<sup>[15]</sup>时,卷积层的计算花费时间太长;全连接层的参数太多。为此,本文网络采用网络加速方法,主要从三个方面提升网络速度,分别为用同等的参数量加速网络模型、分解卷积算子减少参数数量以及训练一个小模型且精度高。

在网络中引入 BN(Batch normalize)层<sup>[16]</sup>,可以加快网络训练速度,同时降低权重初始化的成本。训练时,固定 BN 层的权重参数训练整个网络,训练结束后,将每个 BN 层权重和它相邻的前一层卷积权重进行合并,实际检测时的网络不包含 BN 层,从而减少网络运行时间。卷积层和 BN 层的表达式为:

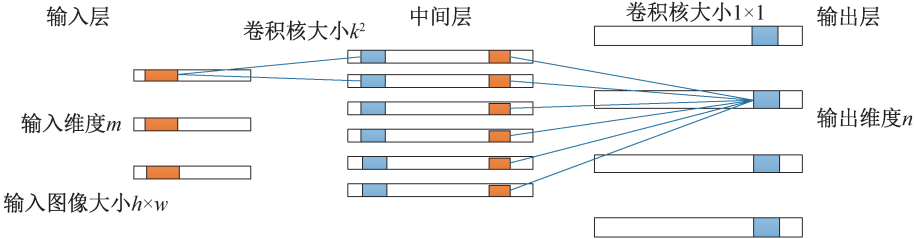


图3 加入基层结构后的卷积层结构

原始卷积核为  $k \times k$  的原始卷积层的计算量为  $k^2 \times m \times n \times h \times w$ ,参数量为  $k^2 \times m \times n$ 。加入基层结构后的卷积层的计算量为  $k^2 \times h \times w \times m + h \times w \times m \times n$ ,参数量为  $k^2 \times m + m \times n$ 。对比可以看出,加入基层结构的卷积层计算次数和参数量均小于传统卷积层,也表示这种卷积构成的网络的运行速度和模型大小也均小于传统卷积构成的网络。

1.2.2 网络精度优化

通常模型的运行速度和表现力是对立的关系,即模型运行速度越快,模型的表现力将越差。针对该问题,在网络加速之后,本文所设计的网络在训练时将采用以下几种方法来保持模型的表现力甚至提高模型的表现力。

每级网络的多任务损失设置类似 MTCNN。人脸分类的损失函数采用传统的交叉熵<sup>[18]</sup>函数,边框回归和特征点检测的损失函数根据各个位置坐标采用交叉熵进行计算。而在模型训练过程中,经常会存在样本被判错的情况,例如,负样本和标记人脸有超过阈值的交叉区,此时,负样本会被误判为正样本;或者包含较大背景区域的正样本也会有很大可能被误判为负样本。然而,一般的 softmax 分类层

$$y_1 = wx + b \tag{1}$$

$$y_1 = \alpha \frac{y_1 - \mu}{\sigma} + \beta = \frac{\alpha w}{\sigma} + \frac{\alpha b - \alpha \mu + \alpha \beta}{\sigma} = w'x + b' \tag{2}$$

其中: $w$  和  $b$  是卷积层的权重和偏置值, $\mu$ 、 $\sigma$ 、 $\alpha$  和  $\beta$  分别代表 BN 层的均值、方差、刻度以及滑动系数。BN 层实际上是对卷积层的结果进行一次再处理,两层合并后,模型中的权重和偏置值替换为  $w'$  和  $b'$ 。

级联方法中的网络通常层数较少,训练困难,模型容易过拟合,收敛不稳定。为了解决这些问题,本文引入基层结构<sup>[17]</sup>,替换传统的卷积算子。加入基层结构后的卷积层结构如图 3 所示,初始卷积核的大小为  $k, h \times w$  为输入目标的大小, $m$  和  $n$  为输入、输出通道数,传统的卷积核大小为  $k \times k$ ,基层结构是引入中间结构,如此可以加深网络的深度。

会忽略掉样本之间的差距从而产生困难样本,这也是小型网络训练困难的一个重要原因。为了解决该问题,训练小型网络时应用知识提炼(Knowledge distillation)<sup>[19]</sup>。首先用 softmax 层训练一个紧密的网络,然后将所有训练样本的阈值重新计算,计算公式为:

$$p = \frac{e^{y'_0}}{e^{\frac{y'_0}{T}} + e^{\frac{y'_1}{T}}} \tag{3}$$

其中: $y'_0$  和  $y'_1$  分别代表紧密网络的输出, $T$  代表控制 soft 阈值分布的固定参数。本文实验中将其设置为 10,最终本文网络的 loss 函数的表达式为:

$$L = \alpha L^{\text{soft}} + (1 - \alpha) L^{\text{hard}} \tag{4}$$

$$L^{\text{soft}} = - \left[ p \log \frac{e^{\frac{y_1}{T}}}{e^{\frac{y_0}{T}} + e^{\frac{y_1}{T}}} + (1 - p) \log \frac{e^{\frac{y_0}{T}}}{e^{\frac{y_0}{T}} + e^{\frac{y_1}{T}}} \right] \tag{5}$$

其中: $L^{\text{hard}}$  代表 softmax 函数是对困难样本的交叉熵, $L^{\text{soft}}$  代表 soft 目标的交叉熵值, $y_0$  和  $y_1$  是当前轻量级网络的输出, $\alpha$  是权重系数。

一般目标检测算法中,网络输出的建议框之后需要应用非极大值抑制(Non-maximum

suppression, NMS), 即将网络输出的建议框按得分进行排序, 合并重叠面积大于阈值的建议框。但是该方法存在阈值难以确定的问题, 过低会增加漏检率, 过高又会增大误检率。NMS 表达式如下:

$$S_i = \begin{cases} S_i, iou(M, b_i) < N_t \\ 0, iou(M, b_i) \geq N_t \end{cases} \quad (6)$$

其中:  $iou$  表示重合区域,  $b_i$  表示第  $i$  个建议框,  $M$  表示得分最高的建议框,  $N_t$  表示 NMS 的阈值,  $S_i$  表示第  $i$  个建议框的得分。

而高斯加权的非极大值抑制方法<sup>[20]</sup> (Soft-NMS) 不同于传统的 NMS 是采用的单一阈值, 该方法对与得分最高的建议框  $M$  重叠部分超过阈值  $N_t$  的建议框进行抑制, 表达式为:

$$S_i = S_i e^{-\frac{iou(M, b_i)^2}{\sigma}}, \forall b_i \notin D \quad (7)$$

其中:  $\sigma$  为高斯函数的方差,  $D$  表示最终保留的建议框集合。

### 1.3 算法步骤

本文设计的第一级网络为全卷积网络, 因此在实际检测时对输入图像的大小没有限制, 但考虑图像中目标的尺度不变性, 网络的输入不是单一的原图像。基于级联网络的快速人脸检测的流程如图 4 所示。

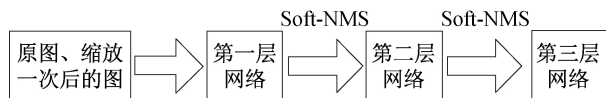


图 4 基于级联网络的快速人脸检测的流程

从图 4 中可以发现, 网络的输入是原图和经过一次缩放后的图, 将第一级网络输出的建议框经过 Soft-NMS 处理, 再进入第二级网络, 输出的建议框再经过 Soft-NMS 的处理, 然后进入最后一层网络, 得出最终的建议框位置及其对应的分类得分。

预先设定好的参数:  $\theta_1, \theta_2, \theta_3$  代表每一级网络的建议框类别得分阈值。  $B$  是  $b_i$  的集合, 代表输出的人脸建议框。

算法实现的具体流程步骤如下:

- 图像预处理;
- 图像输入第一级网络, 并保留输出人脸建议框中得分高于阈值  $\theta_1$  的  $b_i$ , 构成  $B$ ;
- 将  $B$  中每个建议框根据对应的回归值, 重新定位并对  $B$  做高斯 NMS 处理, 更新  $B$  中的建议框;
- 更新后的建议框缩放为  $24 \times 24$  并送入第二级网络;
- 保留第二级网络的输出人脸建议框中得分高于阈值  $\theta_2$  的  $b_i$ , 再次更新  $B$ ;

f) 重复步骤 c);

g) 更新后的建议框缩放设为  $48 \times 48$ , 送入第三级网络;

h) 保留第三级网络的输出人脸建议框中得分高于阈值  $\theta_3$  和  $b_i$ , 将  $B$  中每个建议框根据对应的回归值, 重新定位, 输出最终人脸检测框  $B$ 。

## 2 实验及结果分析

### 2.1 实验环境

本文中实验用到的硬件设备: 主机为华为荣耀 X2, CPU 为海思 Kirin 930, 主频为 2.0 GHz, RAM 为 3 G。软件条件: 安卓系统版本为 android 7.2, 采用 Android Studio 编写。Java development kit 版本为 8.0.144。

实验分别在 Wilder Face、CelebA 以及实验室自建的人脸数据集上运行测试。Wilder Face 数据集包含 32203 张图像, 标注了 393703 张人脸图像, 涵盖了各种尺度大小, 各种姿势以及遮挡, 是复杂度最高的人脸数据集。CelebA 包含 202599 张图像, 每张图像包含只包含一张人脸, 且标注了 5 个特征点的信息。而自建数据集则是一个视频数据集, 包含了 22 个人的样本, 每人 3 段 90 s 的视频, 视频所采用的编码为每秒 25 帧的 H.265 格式, 颜色空间为 YCrCb, 按 LFW 要求为每人每段视频的关键帧<sup>[21]</sup>标注人脸。同时为训练所得的模型编写接口用于安卓程序使用, 实验将从移动设备直接读取数据进行实时人脸检测。

### 2.2 模型训练

本文人脸检测网络的任务主要有三个, 分别是人脸分类、边框回归以及人脸特征点检测。人脸分类和边框回归的目的是从图像中确定人脸的位置信息并根据数据集中的标签进行矫正, 训练数据主要来自 Wilder Face 数据集和自建的人脸数据集。人脸特征点检测目的是确定人脸五官的位置, 监督提高人脸分类的精度, 训练数据来自 CelebA 数据集。所有训练集中负样本与正样本的比率固定为 3:1。

第一级网络包含四个分支, 分别对数据集中图像随机截取大小为  $12 \times 12$ 、 $24 \times 24$ 、 $48 \times 48$  以及  $96 \times 96$  的区域, 根据与标准人脸的重叠区域分为正负样本, 本文实验中学习率设为 0.1。首先训练第一个分支, 训练好的模型去掉最后两层的参数, 其余保留做为第二级网络前半部分的初始权重。构建第二个分支结构并完成训练, 训练大小为  $24 \times 24$  的样本集, 完成第二个分支的训练。同理, 对第三和第四个

分支依次训练。尽管第一级网络包含四层,但是训练不会耗时很久。高层分支的输入是优化后的特征,因而网络的收敛速度会更快。

### 2.3 结果对比

MTCNN 是目前效果最好的人脸检测模型,因此,本文将训练好的网络模型与它在各级网络中的模型大小、网络运行速度及在验证集中准确率三个方面进行对比。参数量代表生成的模型的大小,1000 个建议框的提取时间,即生成固定数量的建议框所需的运行时间,代表网络运行的速度。

第一级网络各分支与 PNet 参数量、速度和精度见表 1。从表 1 中可以很明显地发现,本文所提出的模型的表现能力比 PNet 要好,多层分支结构的实现是并行运算,所以并不会因为分支数目的增多而影响速度;从而也可以看出该模型的速度更快,并且验证精度随着层数级别的增加而增高,这也表明了更大的窗口可以获得更高的分类精度,而传统的方法对小窗口和大窗口只能保持相同的分类精度。此外,实验中发现,随着分支数进一步增加,分类精度将渐渐不再提高,因为高层次的分支结构缺乏较大的人脸做正样本。因此本文只采用了四个分支的金字塔网络结构。

表 1 第一级网络各分支与 PNet 参数量、计算时间和精度

模型	参数量 /kb	10 <sup>3</sup> 个建议框 计算时间/s	精度
PNet	27.5	0.044	0.950
分支 1	18.1	0.020	0.953
分支 2	9.4	0.018	0.957
分支 3	7.8	0.015	0.961
分支 4	7.2	0.014	0.963

级联网络后两级和 MTCNN 的 RNet 以及 ONet 的参数量、速度和精度见表 2。从表 2 中对比可以明显发现在建议框数相同的情况下,本文方法的运行时间相较 MTCNN 有较大的提升。采用优化加速方法后的层网络在速度和精度上明显优于 MTCNN,模型也有缩小,这表明使用网络加速的方法可以实现网络的性能提升。

表 2 级联网络后两级和 MTCNN 的 RNet 以及 ONet 的参数量、计算时间和精度

模型	参数量 /kb	10 <sup>3</sup> 个建议框计 算时间/s	精度
RNet	399	0.598	0.972
第二级	322	0.128	0.976
ONet	1522	2.625	0.984
第三级	1172	0.615	0.989

在采用和不采用知识提炼的情况下,在 FDDB 上分别对本文网络不同层上的人脸建议框数进行了比较,结果见表 3。由表 3 可见,在每一级网络保持相同召回率<sup>[22]</sup>的条件下,采用知识提炼的效率比不采用的时候更高。因为知识提炼的方法可以提升网络分类的性能,所以人脸的建议框数在同样的召回率下会减少,也意味着检测速度会大大提升。

表 3 在 FDDB 上每一级采用知识提炼和不采用情况下人脸建议框数目

知识提炼	第一级/10 <sup>3</sup> 个	第二级/10 <sup>3</sup> 个	第三级/10 <sup>3</sup> 个
有	252	26	5.7
无	302	31	6.4

### 3 结 语

本文提出了一种基于轻量级联网络的人脸检测方法,第一阶段提出的人脸的预提取框是基于 anchor 的金字塔网络结构,无需构造图像金字塔,给网络多个输入,提高了检测速度。对比级联网络 MTCNN 的 PNet,建议框的准确度也有提升。第二阶段的两层网络采用知识提炼、BN 层结构、改进后的卷积层以及高斯 NMS 优化方法,对比 MTCNN 的 RNet 和 ONet 进一步提高了网络的精度和检测速度。尽管本文的网络对人脸的提取速度相较其他基于 CNN 的方法有明显提升,但是,第一阶段生成人脸建议框的过程仍然占用了大部分的检测时间。下一步将继续致力于进一步提高初始人脸建议框提取速度的研究,从而更进一步提高人脸检测速度。

### 参考文献:

[1] 梁路宏,海舟. 脸检测研究综述[J]. 计算机学报, 2002, 25(5):449-458.

[2] Viola P, Jones M J. Robust real-time face detection[J]. International Journal of Computer Vision, 2004,57(2): 137-154.

[3] Ojala T, Pietikäinen M, Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002,24(7): 971-987.

[4] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. IEEE, 2005, 1: 886-893.

[5] Ng P C, Henikoff S. SIFT: Predicting amino acid changes that affect protein function[J]. Nucleic acids research, 2003, 31(13): 3812-3814.

- [6] Wu B, Haizhou A I, Huang C, et al. Fast rotation invariant multi-view face detection based on real adaboost [C] //IEEE International Conference on Automatic Face and Gesture Recognition, IEEE, 2004: 79-84.
- [7] Benini L, Bogliolo A, De Micheli G. A survey of design techniques for system-level dynamic power management [J]. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2000, 8(3): 299-316.
- [8] 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.
- [9] Ranjan R, Patel V M, Chellappa R. Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition [J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017. (2017-12-08) [ 2018-08-12 ]. <https://doi.org/10.1109/TPAMI.2017.2781233>.
- [10] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks [J]. International Conference on Neural Information Processing Systems, 2015, 39(6): 91-99.
- [11] 郑云飞, 张雄伟, 曹铁勇, 等. 基于全卷积网络的语义显著性区域检测方法研究[J]. 电子学报, 2017, 45(11):2593-2601.
- [12] Li H, Lin Z, Shen X, et al. A convolutional neural network cascade for face detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2015: 5325-5334.
- [13] Xue Y, Liao X, Carin L, et al. Multi-task learning for classification with dirichlet process priors[J]. Journal of Machine Learning Research, 2007, 8(1): 35-63.
- [14] 伍凯, 朱恒亮, 郝阳阳, 等. 级联回归的多姿态人脸配准[J]. 中国图象图形学报, 2017, 22(2):257-264.
- [15] 李鸣, 张鸿. 基于卷积神经网络迭代优化的图像分类算法[J]. 计算机工程与设计, 2017, 38(1):198-202.
- [16] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [C]//Proceedings of the 32nd International Conference on International Conference on Machine Learning. JMLR, 2015: 448-456.
- [17] Chollet F. Xception: Deep learning with depthwise separable convolutions[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2017:1800-1807.
- [18] Shore J, Johnson R. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy [J]. IEEE Transactions on Information Theory, 1980, 26(1): 26-37.
- [19] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network[J]. Computer Science, 2015, 14(7):38-39.
- [20] Bodla N, Singh B, Chellappa R, et al. Soft-nms: Improving object detection with one line of code[C]// Computer Vision (ICCV), 2017 IEEE International Conference on. IEEE, 2017: 5562-5570.
- [21] 张娜, 周冬, 张俊为, 等. 基于 H.264 编码帧内预测模式的视频监控算法研究[J]. 浙江理工大学学报, 2017, 37(3):432-437.
- [22] Wessely S, Unwin C, Hotopf M, et al. Stability of recall of military hazards over time[J]. British Journal of Psychiatry, 2003, 183(4):314-322.

(责任编辑:康 锋)