

# 基于深度学习的多角度人脸检测方法研究

李欣, 张童, 厚佳琪, 张子昊

(中国人民公安大学 信息技术与网络安全学院, 北京 100038)

**摘要:** 基于多角度的人脸检测越来越受到关注, 特别是在公安领域侦破案件过程中, 通过捕捉人脸图像对犯罪嫌疑人进行检测识别被广泛应用。但是在实际图像采集过程中, 由于人脸姿势以及光照等环境因素的不确定性和多变性, 往往会导致人脸系统无法对该类人脸进行较为精确的定位。文中基于 DenseNet-201 对 YOLOV2 算法进行了改进, 提出了一种基于深度学习的多角度人脸检测方法。首先, 在 YOLOV2 算法的基础上, 使用 DenseNet-201 模型对人脸进行特征提取, 并结合带有锚点框的卷积层在主干网络提取到的人脸特征图上进行人脸定位; 然后, 通过在 DenseNet-201 模型中的过渡层中引入归一化层使模型收敛速度加快; 最后, 在 CelebA 和 FDDB 人脸数据集上对 YOLOV2 和改进的 YOLOV2 方法进行测试, 针对不同角度、不同光照、不同数据集对算法性能进行测试。实验结果表明, 改进后的 YOLOV2 算法对多角度人脸检测的准确性更高, 且具有更强的鲁棒性。

**关键词:** 多角度人脸检测; YOLOV2; DenseNet-201; 人脸特征提取; CelebA; FDDB

**中图分类号:** TP183

**文献标识码:** A

**文章编号:** 1673-629X(2020)09-0012-06

doi:10.3969/j.issn.1673-629X.2020.09.003

## Research on Multi-angle Face Detection Method Based on Deep Learning

LI Xin, ZHANG Tong, HOU Jia-qi, ZHANG Zi-hao

(School of Information Technology and Cyber Security, People's Public Security of China,  
Beijing 100038, China)

**Abstract:** Face detection based on multi-angle has attracted more and more attention, especially in the process of detecting cases in the field of public security, the detection and recognition of criminal suspects by catching face images has been widely used. However, in the actual image acquisition process, due to the uncertainty and variability of environmental factors such as face posture and lighting, the face system is often unable to locate this kind of face accurately. For this, the YOLOV2 algorithm is improved based on DenseNet-201, and a multi-angle face detection method based on depth learning is proposed. First of all, on the basis of YOLOV2 algorithm, DenseNet-201 model is used for the face feature extraction, and the convolution layer with the anchor frame is combined to carry on the face localization on the face feature map extracted by the backbone network. Then, by introducing the normalization layer into the transition layer of the DenseNet-201 model, the convergence speed of the model is accelerated. Finally, the YOLOV2 and the improved YOLOV2 are tested on CelebA and FDDB face datasets. The performance of the algorithm is tested for different angles, different lighting and different datasets. The experiment shows that the improved YOLOV2 algorithm is more accurate and robust to multi-angle face detection.

**Key words:** multi-angle face detection; YOLOV2; DenseNet-201; face feature extraction; CelebA; FDDB

## 0 引言

与其他生物特征相比,人脸特征信息具有非接触性、易采集性和易接受性等优点,所以人脸识别在模式识别领域一直备受关注<sup>[1]</sup>。近年来,通过捕捉人脸图像对犯罪嫌疑人进行检测识别,在公安领域侦破案件过程中广泛应用。人脸检测是人脸识别的基础,也是其中最重要的一个环节,常用于图像中人脸位置的定

位。但是在实际图像采集过程中,由于人脸姿势以及光照等环境因素的不确定性和多变性,往往会导致人脸系统无法对该类人脸进行较为精确的定位。因此,基于多角度的人脸检测越来越受到广大学者的关注。

传统的人脸检测方法是采用模板匹配技术,也就是将被检测图像与给定人脸模板图像各个位置进行对比,进而判断是否存在人脸,最终对人脸进行定位。后

收稿日期:2019-09-02

修回日期:2020-01-08

基金项目:国家重点研发计划(2018YFC0824300);公安部技术研究计划项目(2017JSYJB01)

作者简介:李欣(1977-),男,博士,中国计算机学会计算机安全专业委员会委员,研究方向为云计算、知识图谱;张童(1993-),男,硕士研究生,研究方向为模式识别、人工智能。

来随着深度学习的出现,通过卷积网络来对图像进行特征提取便应用到了人脸检测领域。其中最具代表性的两种方法是:基于区域的人脸检测和基于滑动窗口的人脸检测。前者是通过搭建物体建议产生器(object proposal generators)或选择性搜索算法(selective search)来对人脸进行定位。后者是按照给定比例计算特征图上每个位置的人脸得分值,实现人脸边界框的定位和回归。Zhang 等<sup>[2]</sup>采用多任务的级联卷积神经网络将人脸检测和人脸对齐两个任务结合起来,提出了 MTCNN 模型。Joseph Redmon 等<sup>[3]</sup>在 YOLOV1 的基础上提出了 YOLOV2 检测方法,将原本主干网络换成 Darknet-19 网络<sup>[4]</sup>,同时在人脸框定位方面结合带有锚点框的卷积层大幅度地提高了定位精度。但是由于二者应用的特征提取的卷积网络层数不深,通过卷积网络提取到的特征鲁棒性和准确性并不高。

文中在 YOLOV2 算法的基础上提出了一种新的人脸检测算法。该算法使用 DenseNet 作为人脸特征提取器,使提取到的特征具有更高的鲁棒性和准确性。在人脸定位方面采用带有锚点框的卷积层对提取到的特征进行定位,同时通过引入归一化层,使模型的收敛速度加快。

## 1 基于 YOLOV2 的人脸检测

### 1.1 YOLOV1 算法

首先,将输入图像划分为  $n \times n$  的网格。如果某个物体中心位置坐标落在哪个网格内,就由哪个网格对该物体进行定位检测。每个单元格生成  $B$  个预测边界框以及对应的置信度,通过阈值限定以及非极大值抑制的方式对边界框进行筛选,得到最终的人脸框。在网络结构方面,YOLOV1<sup>[5]</sup>用  $1 \times 1$  卷积层加  $3 \times 3$  卷积层来代替 GoogLeNet 中的 inception module<sup>[6]</sup>并行模块,同时在网络模型的末端使用全连接层来输出类别。

### 1.2 YOLOV2 算法

YOLOV2 是 YOLOV1 的改进版本。对定位准确度和召回率进行了改进。YOLOV1 算法存在两个缺点:(1)对 bounding box<sup>[7]</sup>的定位不够准确,即对小目标检测效果不佳;(2)相对于 region proposal 方法,YOLOV1 的召回率较低。于是,YOLOV2 在 YOLOV1 的基础上进行了改进。具体改进方法如下:

(1)带有锚点框的卷积。

YOLOV2 中去掉全连接层,使用锚点框来预测边界框。具体方法如下:

YOLOV2 使用卷积层进行下采样,使得最终的特征图的尺寸为  $13 \times 13$ 。最终的特征图的宽和高都为奇数,使得特征图只有一个中心点。较大的物体通常占据图像的中心位置,通过这种操作使用一个中心点来

预测图像位置,而不是使用相邻的 4 个单元格进行预测。YOLOV1 使用每个单元格预测物体类别,使用边界框预测坐标值,而 YOLOV2 中使用锚点同时预测类别和边界框的坐标。使用锚点进行预测会使得准确率稍微下降,但是召回率有明显的提升。

(2)使用 K-means 聚类。

YOLOV2 使用 K-means<sup>[8]</sup>的方式对训练数据集中的边界框进行聚类分析,从而找到最优的锚点框。由于欧氏距离会因 bbox 的大小产生不同程度上的误差,而 IOU<sup>[9]</sup>与 bbox 尺寸无关,所以选择用 IOU 来计算距离,公式如下:

$$d(\text{box}, \text{centroid}) = 1 - \text{IOU}(\text{box}, \text{centroid}) \quad (1)$$

IOU 表示网络预测的边界框和图片标签中实际的边界框的重合率,计算公式如下:

$$\text{IOU}_{\text{pred}}^{\text{truth}} = \frac{\text{truth} \cap \text{pred}}{\text{truth} \cup \text{pred}} \quad (2)$$

分子表示预测边界框和实际边界框的交集部分,分母表示预测边界框和实际边界框的并集部分。

(3)直接预测边界框位置。

YOLOV2 中延续 YOLOV1 中直接预测的方法对边界框进行预测。网络在特征图中每个单元格中预测 5 个边界框,每个边界框预测 5 个值:  $t_x, t_y, t_w, t_h$  和  $t$ ,其中  $t_x$  和  $t_y$  表示相对于单个网络的边界框的中心坐标的偏移值,  $t_w$  和  $t_h$  表示相对于整幅图像的边界框的宽和高,  $t$  表示置信度,即表示预测的边界框和真实边界框的重合率。

$$\text{confidence} = \text{Pr}(\text{Object}) \times \text{IOU}_{\text{pred}}^{\text{truth}} \quad (3)$$

其中,  $\text{Pr}(\text{Object})$  表示边界框是否包含物体。如果边界框包含物体,即人脸,那么  $\text{Pr}(\text{Object}) = 1$ ,如果边界框中不包含物体,则  $\text{Pr}(\text{Object}) = 0$ 。

假设此单元格距离图像左上角顶点的偏移量为  $c_x, c_y$ ,先验框的宽和高为  $p_w, p_h$ ,那么网络的预测结果将如式(4)~式(8)所示,预测的边界框参数如图 1 所示。

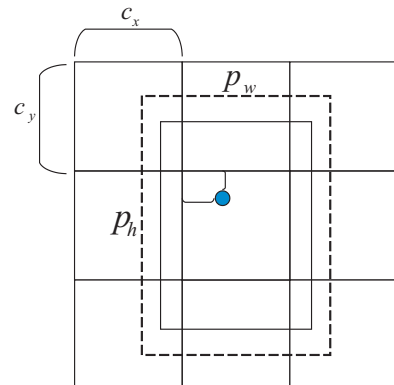


图 1 边界框预测参数图

$$b_x = \sigma(t_x) + c_x \quad (4)$$

$$b_y = \sigma(t_y) + c_y \quad (5)$$

$$b_w = p_w e^{t_w} \quad (6)$$

$$b_h = p_h e^{t_h} \quad (7)$$

$$\Pr(\text{Object}) \times \text{IOU}(b, \text{object}) = \sigma(t_o) \quad (8)$$

(4) 网络架构。

在 YOLOV2 中,使用 448×448 分辨率的图片代替原先的 224×224 作为预训练图像。采用 Darknet -

19<sup>[10]</sup> 模型来提取图像中的特征信息。Darknet-19 采用全局平均池化层做预测,并在 3×3 卷积之间使用 1×1 卷积来实现特征图通道数量的压缩,进而减少模型参数和计算量。最后,Darknet-19 在每个卷积层后面使用了 BN 层,以加快模型收敛速度,降低模型过拟合。

YOLOV2 人脸检测算法流程如图 2 所示。

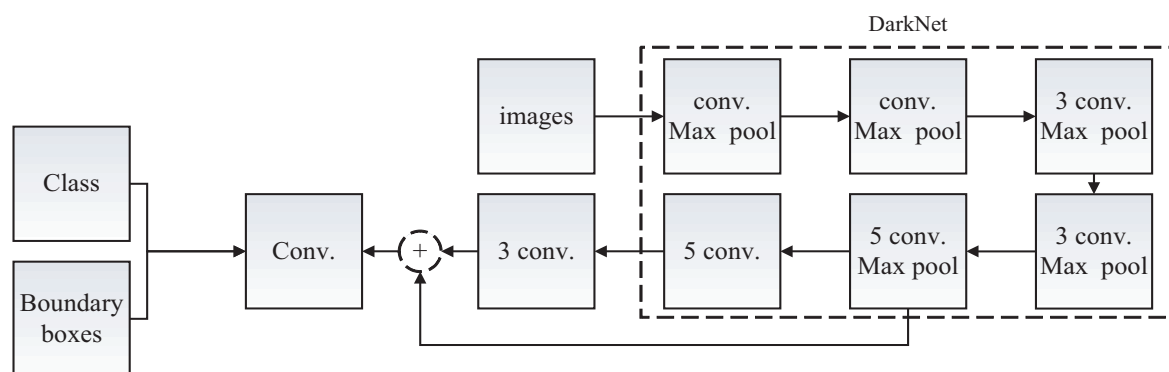


图 2 YOLOV2 人脸检测算法流程

## 2 基于 DenseNet-201 改进的 YOLOV2 人脸检测算法

YOLOV2 中使用的主干网络是 DarkNet-19,但是由于网络层数较浅,无法提取到更加细粒和有效的人脸特征信息。而 DenseNet<sup>[11]</sup> 网络在图像分类任务中取得了较高的准确率,所以文中采用 DenseNet-201 网络结构代替 DarkNet-19 模型用做人脸特征提取模块。

### 2.1 DenseNet 网络结构

在传统的卷积神经网络中,网络的连接方式是直接相连的,即:网络中存在  $L$  个层,那么网络就有  $L$  个连接。但是在 DenseNet 中,网络的连接方式是密集连接,使得网络中的特征图的个数显著增加。网络中的每一个层与其他层都直接相连。如果网络中存在  $L$  个层,那么网络的连接就有  $\frac{L(L+1)}{2}$  个。

密集连接的操作需要特征图的大小统一,这就需

要引入池化层来进行限定,同时为了引入池化层,需要添加过渡层来将 Dense Block 连接起来,这样就把网络划分成了若干个不同的 Dense Bloc。其中每一个 Dense Bloc 为瓶颈结构,包含 3×3 和 1×1 两种类型的卷积。可以通过设置参数  $k$  来限定每一个 Dense Bloc 输出的特征图的数量,卷积表示为  $3 \times 3 \times k$ 。同时由于过渡层包括 BN 层,1×1 的卷积层以及 2×2 的平均池化,过渡层中的 BN 层会进一步对特征进行归一化,从而加快模型的收敛速度。

### 2.2 基于 DenseNet-201 改进的 YOLOV2 人脸检测

相比于 DarkNet-19, DenseNet-201 模型网络的层数明显得到增加。DenseNet-201 中的密集连接的方式不仅使得网络每一层提取到的人脸特征信息尽可能得到利用,并且在训练过程中可以学习到新的人脸特征信息。因此,文中使用 DenseNet-201 作为人脸检测器的特征提取部分,其他部分与 YOLOV2 算法相同。改进的 YOLOV2 人脸检测算法流程如图 3 所示。

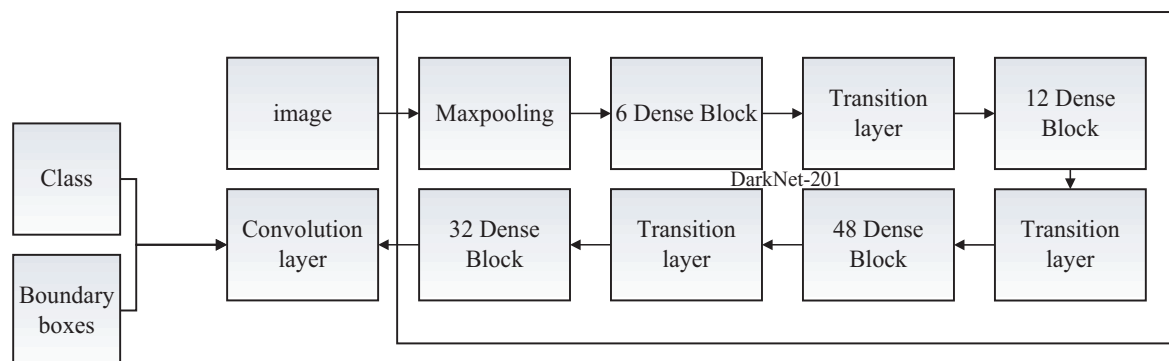


图 3 基于 DenseNet-201 的 YOLOV2 人脸检测算法流程

由于 DenseNet-201 采用的是密集连接的方式,会

使特征图的数量大大增加,同时参数的数量也会大大

增加。于是文中在 DenseNet-201<sup>[12]</sup>的部分提出了三点来控制网络参数和特征图的数量。具体参数设置过程如下:

第一,引入超参数  $k$  作为网络的增长率。网络的增长率表示每一个 Dense Block 输出的特征图的个数为  $k$  个,为了防止网络变得太宽,并且提升网络的计算效率,将  $k$  值限制成一个较小的整数。在 DenseNet 中,将每一个 Dense Block 输出的特征图的个数设置为 32,即  $k = 32$ 。实验证明,当  $k$  值是一个较小的正整数时,网络的表现性能也更好。

第二,引入瓶颈层。DenseNet 采取密集连接的方式,虽然每一层产生  $k$  个特征图,但是整个网络产生的特征图的数量是巨大的。为了控制 DenseNet 网络中特征图的数量,在网络的构建中引入了瓶颈层。

第三,对网络进行压缩。通过压缩过渡层中的特征图,增强模型的紧凑性。在 DenseNet 中,当 Dense Block 包含  $m$  个特征图时,网络中随后的过渡层产生的特征图的个数为  $\theta_m$ ,其中  $0 < \theta < 1$ 。于是,将  $\theta$  的值设置为 0.5,即过渡层将前一个 Dense Block 产生的特征图的数量减半。DenseNet-201 网络模型的具体设置细节如表 1 所示。

表 1 DenseNet-201 网络模型

Layers	DenseNet-201	Output size
Convolution	$7 \times 7$ , 64, stride 2	$208 \times 208$
Max Pool	$2 \times 2$ , stride 2	$104 \times 104$
Dense Block (1)	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 32 \end{bmatrix} \times 6$	$104 \times 104$
Transition Layer (1)	$1 \times 1, 128$ $2 \times 2$ Max Pool, stride 2	$104 \times 104$ $52 \times 52$
Dense Block (2)	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 32 \end{bmatrix} \times 12$	$52 \times 52$
Transition Layer (2)	$1 \times 1, 128$ $2 \times 2$ Max Pool, stride 2	$52 \times 52$ $26 \times 26$
Dense Block (3)	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 32 \end{bmatrix} \times 48$	$26 \times 26$
Transition Layer (3)	$1 \times 1, 128$ $2 \times 2$ Max Pool, stride 2	$26 \times 26$ $13 \times 13$
Dense Block (4)	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 32 \end{bmatrix} \times 32$	$13 \times 13$
Classification Layer	$13 \times 13$ global average pool	

从以上表格可以看出,将每一个 Dense Block 的输出特征图的数量以及大小设置为  $3 \times 3 \times 32$ ,同时每经过一个过渡层输出的特征图尺寸以及数量均减小一半,

大大压缩了特征网络,使网络的运行速度加快。

### 3 仿真实验与结果分析

#### 3.1 数据预处理

文中使用 CelebA 和 FDDB 人脸数据集作为人脸检测的训练数据集和测试数据集。CelebA 人脸数据集包括 202 599 张人脸图像<sup>[13]</sup>。每一张图像的标签包括人脸边界框,5 个人脸关键点以及 40 个人脸属性。FDDB 人脸数据集包括 2 845 张人脸图像和 5 171 个人脸区域,数据集包括不同人脸姿态、不同分辨率以及旋转和遮挡的人脸图像。

文中选取 CelebA 人脸数据集中的 190 000 张图像作为训练集。将 12 599 张 CelebA<sup>[14]</sup>数据中的人脸图像和 FDDB<sup>[15]</sup>人脸数据集作为测试集,测试算法性能。

#### 3.2 仿真实验

实验一:在不同角度人脸图像上进行算法性能测试。

实验一中选择三种不同角度的人脸图像作为测试图像,分别为正面人脸图像,60 度人脸偏转图像和 90 度人脸偏转图像,衡量改进前和改进后不同人脸检测模型的性能,如图 4 所示。

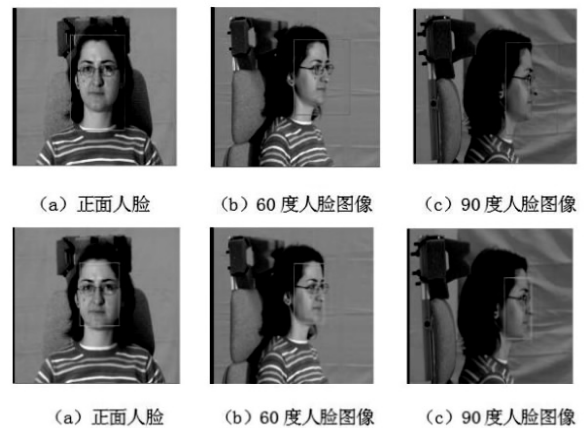


图 4 基于 YOLOV2 人脸检测算法和基于 DenseNet-201 的 YOLOV2 人脸检测算法效果示例(1)

由图 4 可以看出,两种人脸检测模型都可以检测出三种不同角度图像中的人脸。在人脸角度一致的前提下,基于 DenseNet-201 的 YOLOV2 算法检测效果最佳,其次是 YOLOV2 算法;在人脸角度不同的前提下,基于 DenseNet-201 的 YOLOV2 算法鲁棒性最强,其次是 YOLOV2 算法。实验结果表明,改进的 YOLOV2 人脸检测算法相较于原始算法,检测性能有明显的提升。

实验二:在不同光照人脸图像上进行算法性能测试。

实验二中选择三种不同光照环境下的人脸图像作为测试图像,衡量改进前和改进后人脸检测模型的性能,如图 5 所示。



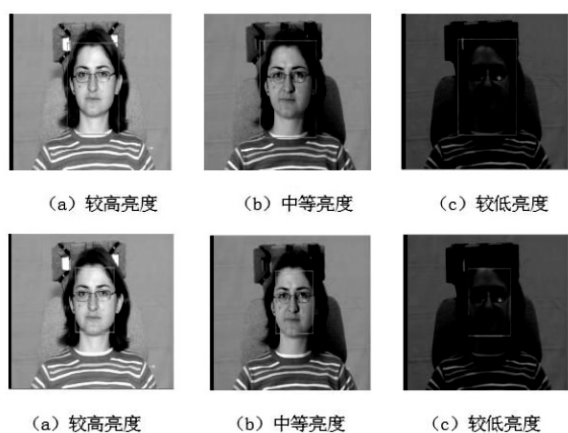


图 5 基于 YOLOV2 人脸检测算法和基于 DenseNet-201 的 YOLOV2 人脸检测算法效果示例(2)

表 2 不同人脸检测模型在 CelebA 的性能指标

人脸检测模型	平均精确率 (MAP)	召回率 (Recall)	Average IOU
YOLOV2	90.63%	95.00%	75.27%
基于 DenseNet-201 的 YOLOV2	90.90%	100%	90.03 %

实验四:在 Fddb 人脸数据集上进行算法性能测试。

实验四中选择 Fddb 人脸数据集作为算法测试数

据集,对人脸检测模型进行性能评测,得出 ROC 曲线图,如图 6、图 7 所示。

实验三:在 CelebA 人脸数据集上进行算法性能测试。

实验三选取 CelebA 的测试数据集对人脸检测模型进行测试,算法测试指标如表 2 所示。

由表 2 可知,DenseNet-201 的 YOLOV2 人脸检测算法的各项性能指标相较于原模型,性能更优。

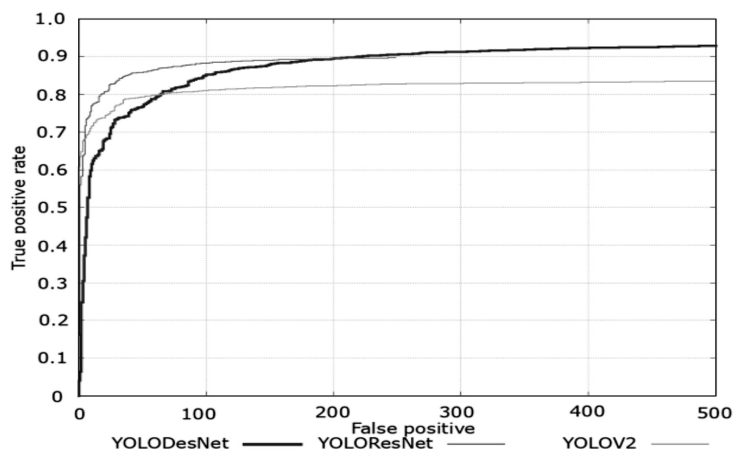


图 6 YOLOV2 人脸检测模型和改进的模型 ROC 曲线

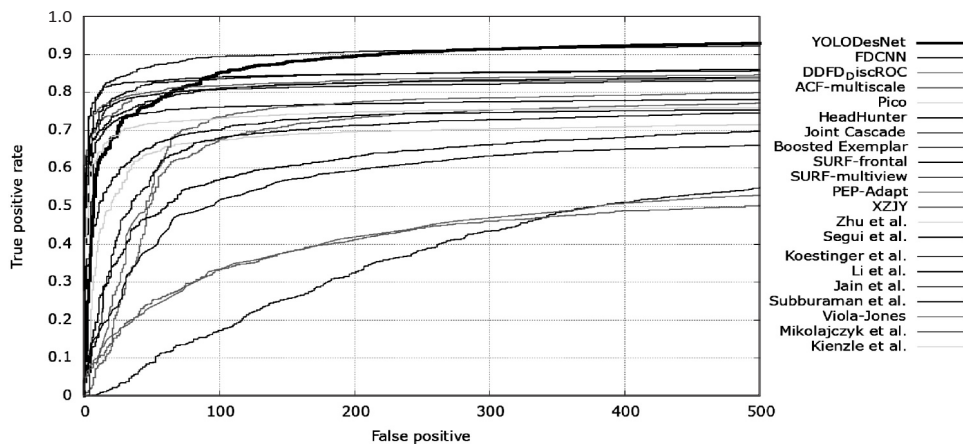


图 7 基于 DenseNet-201 的 YOLOV2 模型与其他模型对比

由上图可得,在YOLOV2系列的人脸检测算法中,DenseNet-201模型在Fddb数据集上的效果最好,并且与其他人脸算法进行比较时,基于DenseNet-201的YOLOV2人脸检测算法的表现性能最好。

### 3.3 结果分析

实验一到实验四的结果均显示,改进的DenseNet-201的YOLOV2人脸检测模型相较于YOLOV2模型,性能得到较大的提升,并且在多角度、不同光照条件下,以及不同数据集的评测中,基于DenseNet-201的YOLOV2人脸检测模型的鲁棒性以及准确性等性能表现最优。结果表明,DenseNet的密集连接的网络结构相比于Darknet-19网络,不仅能够提取到更为细粒度、更为抽象的人脸特征,而且使得网络提取到的人脸特征在整个任务中最大化地被网络利用,并学习到新的人脸特征,提升人脸检测算法的性能。

## 4 结束语

基于DenseNet-201对YOLOV2算法进行了改进,提出了一种新的多角度人脸检测算法。该算法相较于之前YOLOV2中的DarkNet网络,可以提取更为丰富的人脸特征,并且使提取到的特征更具准确性和鲁棒性。在CelebA和Fddb两个人脸数据集上对YOLOV2和改进后的YOLOV2方法进行测试,实验结果表明,改进后的YOLOV2算法对多角度人脸检测的准确性更高,且具有更强的鲁棒性。经过对YOLOV2算法的改进,虽然在多角度人脸检测的性能上有所提升,但是在算法的准确性上还存在一定的差距,尚需进一步完善。

### 参考文献:

- [1] 刘晓波. 浅谈基于深度学习的人脸识别技术[J]. 信息通信, 2019(6): 18-20.
- [2] 常思远, 李有乘, 孙培岩, 等. 一种基于MTCNN的视频人脸检测及识别方法[J]. 许昌学院学报, 2019, 38(2): 149-152.
- [3] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu: IEEE, 2017: 6517-6525.
- [4] 刘 博, 王胜正, 赵建森, 等. 基于Darknet网络和YOLOv3算法的船舶跟踪识别[J]. 计算机应用, 2019, 39(6): 1663-1668.
- [5] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE, 2016: 779-788.
- [6] 郭瑞琴, 陈雄杰, 骆 炜, 等. 基于优化的Inception ResNet A模块与Gradient Boosting的人群计数方法[J]. 同济大学学报: 自然科学版, 2019, 47(8): 1216-1224.
- [7] 黄继鹏, 史颖欢, 高 阳. 面向小目标的多尺度Faster-RCNN检测算法[J]. 计算机研究与发展, 2019, 56(2): 319-327.
- [8] 钟 熙, 孙祥娥. 基于Kmeans++聚类的朴素贝叶斯集成方法研究[J]. 计算机科学, 2019, 46(S1): 439-441.
- [9] RAHMAN M A, WANG Y. Optimizing intersection-over-union in deep neural networks for image segmentation [C]//International symposium on visual computing. Las Vegas: Springer International Publishing, 2016: 234-244.
- [10] 田 雄, 吴 薇, 刘晓尚, 等. 多人脸跟踪与最佳人脸提取[J]. 电子科技, 2019, 32(9): 32-36.
- [11] JUNG E, CHIKONTWE P, ZONG X, et al. Enhancement of perivascular spaces using densely connected deep convolutional neural network[J]. Access IEEE, 2019, 7: 18382-18391.
- [12] TONG N, CAO M, SHENG K. Shape constrained fully convolutional DenseNet with adversarial training for multi-organ segmentation on head and neck low field MR images[J]. Medical Physics, 2019, 46(6): 2669-2682.
- [13] KRASIN I, DUERIG T, ALLDRIN N, et al. OpenImages: a public dataset for large-scale multi-label and multi-class image classification [DB/OL]. 2017. <https://github.com/open-images>.
- [14] 刘 晨. 基于机器人视觉系统的人脸检测技术研究[J]. 电子设计工程, 2019, 27(5): 109-112.
- [15] JAIN V, LEARNED-MILLER E. Fddb: a benchmark for face detection in unconstrained settings [R]. Amherst: University of Massachusetts, 2010.