



## 基于 Mask R-CNN 的人脸检测与分割方法

林凯瀚, 赵慧民, 吕巨建, 詹 瑾, 刘晓勇, 陈荣军

(广东技术师范大学 计算机科学学院, 广州 510665)

**摘 要:** 针对现有主流的人脸检测算法不具备像素级分割, 从而存在人脸特征具有噪声及检测精度不理想的问题, 提出了一种基于 Mask R-CNN 的人脸检测及分割方法。通过 ResNet-101 结合 RPN 网络生成候选区域, 再利用 RoIAlign 算法实现像素级的特征点定位, 旨在提高定位精度。根据全卷积网络生成相应的人脸二值掩码, 实现图像中人脸信息与背景的分割。此外, 构建了一个具有分割标注信息的人脸数据集用于训练相应模型。在通用人脸检测数据集的实验结果表明, 该方法具有较好的人脸检测效果, 并能在准确检测的同时实现像素级的人脸信息分割。

**关键词:** 人脸检测; Mask R-CNN 算法; 实例分割; RoIAlign 算法; 全卷积网络

开放科学(资源服务)标志码(OSID):



**中文引用格式:** 林凯瀚, 赵慧民, 吕巨建, 等. 基于 Mask R-CNN 的人脸检测与分割方法[J]. 计算机工程, 2020, 46(6): 274-280.

**英文引用格式:** LIN Kaihan, ZHAO Huimin, LÜ Jujian, et al. Face detection and segmentation method based on Mask R-CNN[J]. Computer Engineering, 2020, 46(6): 274-280.

## Face Detection and Segmentation Method Based on Mask R-CNN

LIN Kaihan, ZHAO Huimin, LÜ Jujian, ZHAN Jin, LIU Xiaoyong, CHEN Rongjun

(School of Computer Science, Guangdong Polytechnic Normal University, Guangzhou 510665, China)

**[Abstract]** Face detection is an important research direction in computer vision and information security, which has been widely studied over the past few decades. In the traditional face detection method, there is no pixel-level segmentation process, which leads to the problem of face features with noise and unsatisfactory detection accuracy. In order to overcome this shortcoming, a face detection and segmentation method based on Mask R-CNN is proposed in this paper. In this method, ResNet-101 and RPN is used to generate RoIs, and RoIAlign faithfully retains the exact spatial locations to generate binary mask through Fully Convolution Network. In order to train the model, this paper constructs a face dataset with segmentation annotation information. The experimental results of well-known face detection dataset show that the proposed method has better face detection effect and can achieve pixel-level face information segmentation at the same time.

**[Key words]** face detection; Mask R-CNN algorithm; instance segmentation; RoIAlign algorithm; Full Convolutional Network(FCN)

**DOI:** 10.19678/j.issn.1000-3428.0054566

### 0 概述

人脸检测是计算机视觉和信息安全领域的一个重要研究方向,也是目标检测技术的关键分支,具有重要的研究意义及应用价值。人脸检测一般包括人脸的识别和定位两个过程,通过利用图像处理以及机器学习等技术从图像或视频中检测定位出人脸,从而获取人脸信息。近年来,人脸检测在模式识别

领域得到了广泛的研究与发展,现有人脸检测研究主要包括基于传统机器学习的方法和基于深度学习的方法。

在传统机器学习方法中,主要是通过手工特征结合分类器实现检测过程。Viola-Jones<sup>[1]</sup>算法是最早实时有效人脸检测的算法,标志着人脸检测开始进入实用阶段。但是,该方法存在特征维度较大、对复杂的背景情况识别率较低等问题。为解决以上问

**基金项目:** 国家自然科学基金(61772144); 广东省教育厅创新团队项目(2017KCXTD021); 广东省自然科学基金-博士启动基金(2016A030310335); 广东省普通高校青年创新人才类项目(2018KQNCX139); 广州市对外科技合作计划项目(201807010059)。

**作者简介:** 林凯瀚(1994—),男,硕士研究生,主研方向为计算机视觉、图像处理; 赵慧民,教授、博士; 吕巨建,讲师、博士; 詹瑾,副教授、博士; 刘晓勇,教授、博士; 陈荣军,副教授、博士。

**收稿日期:** 2019-04-10 **修回日期:** 2019-06-17 **E-mail:** jujianlv@gpnu.edu.cn

题,研究人员提出了更为复杂的手工设计特征,如HOG<sup>[2]</sup>、SIFT<sup>[3]</sup>、SURF<sup>[4]</sup>、LBP<sup>[5]</sup>等。其中一个重要的进展是文献[6]提出的DPM(Deformable Part Model)<sup>[6]</sup>,DPM在HOG和SVM的基础上进行性能扩展,充分利用两者的优势,在人脸识别、行人检测等任务上取得重要突破。然而传统的机器学习方法仍然存在两个主要缺陷,即基于滑动窗口的区域选择策略针对性较弱,设计较为复杂,手工设计特征对复杂情景检测稳定性较差。

自从AlexNet<sup>[7]</sup>在ImageNet中使用深度卷积神经网络,并且在图像分类的准确率上有了大幅提高以来,人们尝试将深度学习应用于人脸检测领域,取得了很好的应用效果。文献[8]利用深度卷积神经网络进行特征提取,对没有被Adaboost过滤过的人脸和非人脸的图像提取出更有价值的特征,取得了较好的实现效果。文献[9]提出一种基于深度卷积神经网络提取的归一化特征的形变部件模型。近年来,随着基于候选区域(region-based)的R-CNN系列目标检测算法的快速发展,R-CNN系列算法在人脸检测领域的应用研究逐渐兴起。文献[10]在WIDER数据集训练了Faster R-CNN模型,在Fddb和IJB-A数据集实现了人脸检测。文献[11]针对小尺度人脸检测精度不高的问题,在Faster R-CNN的基础上提出了DSFD算法。文献[12]通过特征拼接、困难样本挖掘、多尺度训练等策略改进了Faster R-CNN模型,提高了模型性能,取得较好的检测效果。基于深度学习的人脸检测方法主要采用卷积神经网络进行特征提取,在准确性和多目标检测方面具有良好的实现效果,并能够以较少的时间花费换取大幅的准确率提升,因此,基于深度学习的人脸检测算法已成为人脸检测的主流研究方向。

上述的研究工作都取得较好的人脸检测效果,但是缺乏对人脸图像分割的关注。如需在人脸图像上获取更精准的人脸信息,则现有的方法存在提取的人脸目标特征维度大、空间量化较为粗糙、具有背景噪声等问题,导致一些实用的图像处理技术(如多姿态人脸矫正<sup>[13]</sup>、人脸图像的超分辨率重构<sup>[14]</sup>、遮挡人脸姿态识别<sup>[15]</sup>等)在安防监控所抓拍的视频图像上难以应用。因此,需要一种适用于图像的人脸检测及分割方法。在检测分割方面,Mask R-CNN是文献[16]基于Faster R-CNN提出的一种改进算法,它增加了对实例分割的关注。除

了分类和边界框回归之外,Mask R-CNN还为每个RoI添加并行分支以进行实例分割,并对上述3种损失进行联合训练,在目标检测数据集中取得了良好的效果。

在以上研究工作的基础上,本文提出了一种基于Mask R-CNN的人脸检测及分割方法,该方法在现有的人脸检测仅实现人脸定位的基础上增加了分割分支,能够在人脸检测的同时实现像素级的人脸信息分割。同时,为了训练相应模型,本文从Fddb<sup>[17]</sup>和ChokePoint<sup>[18]</sup>数据集中随机选择了5115张图像进行标注,构建了一个具有分割标注的新数据集。

## 1 相关工作

基于候选区域的CNN检测方法是目标检测领域的主流研究方向,例如R-CNN<sup>[20]</sup>、Fast R-CNN<sup>[21]</sup>、Faster R-CNN<sup>[22]</sup>以及Mask R-CNN<sup>[16]</sup>等,这些方法不仅检测精度高,且具有较快的检测速度。

R-CNN算法由Ross Girshick等人于2014年提出,该算法在VOC 2007数据集中取得了66%(mAP)的成绩,掀起了基于候选区域(region-based)CNN的研究热潮。首先由于R-CNN特征提取环节存在重复计算的问题,导致检测速度偏慢,因此Fast R-CNN基于R-CNN,在卷积层对待检测图像整图进行一次特征提取;其次引入RoI Pooling层进行特征尺度的统一化,生成固定长度的特征向量;然后用softmax(归一化指数函数)代替了SVM,并将分类和边框回归进行了合并,在提升准确率的同时,提高了检测速度。Faster R-CNN是在Fast R-CNN基础上提出的改进算法,该算法引入了区域建议网络(Region Proposal Network,RPN)模型,通过在特征图上划窗,使用不同尺寸、不同长宽比的9种锚点框映射到原图上,得到候选区域。该算法将候选框提取合并到深度网络中,大幅提升了检测的速度与精度。

为了能够在目标检测的同时,实现像素级的实例分割,满足计算机视觉任务中更为精准的检测定位需求,文献[16]提出了Mask R-CNN算法。该算法在Faster R-CNN的基础上,添加实例分割分支,通过RoIAlign获得更精准的像素信息,利用全卷积网络生成相应的二值掩码,实现了目标检测及实例分割技术的结合应用,在目标检测领域的COCO数据集和自动驾驶领域的Cityscapes数据集上的实验结果表明,该算法均取得了较好的效果。

## 2 人脸检测及分割方法

Mask R-CNN 算法主要由两大分支组成,即检测分支和分割分支。检测分支实现了对图像中目标的定位及分类,分割分支则通过全卷积网络(Fully Convolutional Network, FCN)<sup>[23]</sup>生成二值掩码实现实例分割,达到像素级的区分效果。

本文模型的整体框架如图 1 所示,首先将待检测

图像传入该模型,通过卷积神经网络对整张图像进行特征的提取,得到对应的特征图,利用区域建议网络(RPN)在特征图上迅速生成候选区域,再通过候选区域匹配(RoIAlign),得到固定尺寸的特征图输出,然后在分类分支做出目标框的定位及分类,在分割分支通过全卷积网络对人脸图像绘出相应的二值掩码实现实例分割,最后输出系统预测的图像。

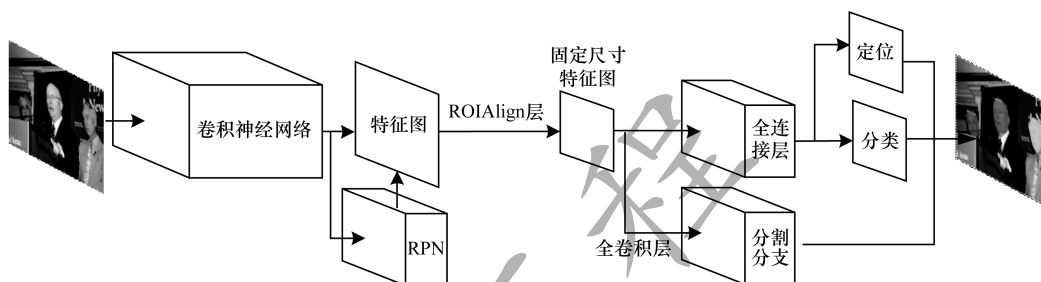


图 1 本文算法模型框架

Fig. 1 The model framework of the proposed algorithm

### 2.1 RPN 网络

区域建议网络(RPN)通过倍数和长宽比例不同的窗口或锚点(Anchors)在特征图上进行滑窗从而迅速生成候选区域。RPN 算法示意图如图 2 所示,图中背景图像表示经卷积神经网络提取特征后的特征图,虚线表示窗口为基准窗口,假设基准窗口大小为 16 个像素点,其包含 3 个分别表示长宽比例为 1:1、1:2、2:1 的窗口,则点划线及实线分别表示 8 像素和 32 像素大小的窗口,同理,其各有 3 个长宽比例为 1:1、1:2、2:1 的窗口。RPN 利用上述 3 种倍数和 3 种比例的共 9 种尺度窗口的方法对特征图进行滑窗,当  $\text{IoU} \geq 0.5$  时,认为其为正例,并对其进行回归。

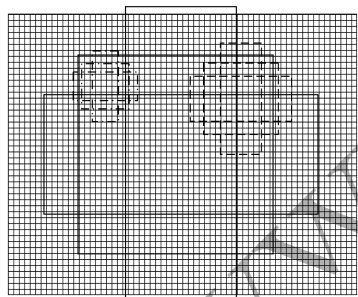


图 2 RPN 算法示意图

Fig. 2 Schematic diagram of RPN algorithm

交并比(Intersection over Union, IoU)计算公式如下:

$$\text{IoU} = \frac{S_{A \cap B}}{S_{A \cup B}} \quad (1)$$

其中,  $A$ 、 $B$  分别为 RPN 网络生成的候选框及训练集中正确的目标框,  $S_{A \cap B}$  为  $A$ 、 $B$  的相重叠处面积,  $S_{A \cup B}$  为  $A$ 、 $B$  并集面积。

### 2.2 RoIAlign 层

RoIAlign 层实现对产生的候选区域(Region of Interest, RoI)进行池化,从而将不同尺度的特征图通过 RoIAlign 层池化为固定的尺度特征图的效果, RoIAlign 算法流程如图 3 所示。在 Faster R-CNN 模型中候选区域池化(Region of Interest Pooling, RoI Pooling)存在两次取整操作,从而产生了量化误差,导致了图像的像素点定位精度度较低。

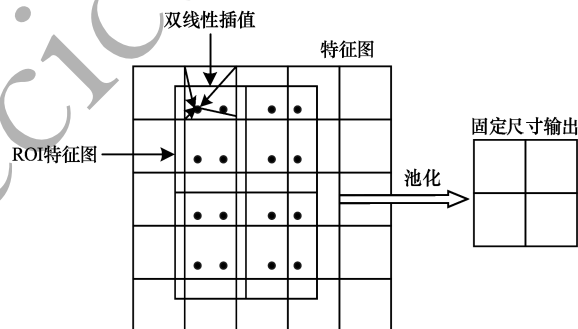


图 3 RoIAlign 算法流程

Fig. 3 RoIAlign algorithm procedure

在面对小目标检测及实例分割的任务时, RoI Pooling 达不到精准的特征点定位要求。针对上述问题, RoIAlign 在生成的 RoI 特征图上取消量化过程,采用双线性插值算法,保留浮点型坐标,避免了量化误差,可以使原图像像素与特征图像素相匹配。双线性插值算法公式如下:

1) 对  $x$  方向进行线性插值:

$$f(R_1) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}), R_1 = (x, y_1) \quad (2)$$

(2)

$$f(R_2) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}), R_2 = (x, y_2) \quad (3)$$

2) 对  $y$  方向进行线性插值:

$$f(P) = f(x, y) \approx \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y - y_1}{y_2 - y_1} f(R_2) \quad (4)$$

如图 4 所示,  $f(x, y)$  为待求解点  $P$  的像素值,  $f(Q_{11})$ 、 $f(Q_{12})$ 、 $f(Q_{21})$ 、 $f(Q_{22})$  分别为已知四点  $Q_{11} = (x_1, y_1)$ 、 $Q_{12} = (x_1, y_2)$ 、 $Q_{21} = (x_2, y_1)$  及  $Q_{22} = (x_2, y_2)$  的像素值,  $f(R_1)$ 、 $f(R_2)$  为  $x$  方向插值得到的像素值。

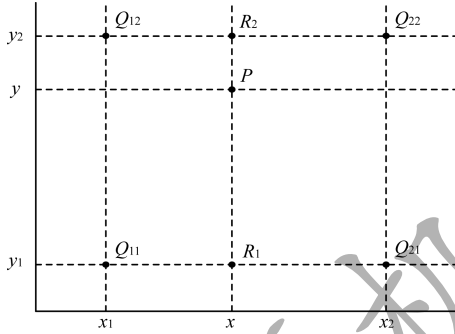


图 4 双线性插值算法流程

Fig. 4 Procedure of bilinear interpolation algorithm

### 2.3 全卷积网络

对于传统的 CNN 网络架构, 为了得到固定维度的特征向量, 其卷积层后大多连接若干全连接层, 最后输出为针对输入的一个数值描述, 因而普遍适用于图像的识别分类、目标的检测定位等任务。全卷积网络 (FCN) 与传统的 CNN 网络相似, 网络同样包含了卷积层和池化层, 不同之处在于全卷积网络对末端一个卷积层的特征图使用了反卷积进行上采样, 使输出图像尺寸恢复到原图像尺寸, 然后通过 Softmax 分类器进行逐像素预测, 从而预测出每个像素点的所属类别。

### 2.4 损失函数定义

本文的 Mask R-CNN 模型完成了 3 个任务, 即人脸框的检测定位、人脸与背景的分类、人脸与背景的分割。因此, 损失函数的定义包括了定位损失、分类损失及分割损失三部分。因此, 损失函数定义如下:

$$L = L_{\text{cls}}^* + L_{\text{box}}^* + L_{\text{mask}}^* \quad (5)$$

其中,  $L_{\text{cls}}^*$  为分类损失,  $L_{\text{box}}^*$  为定位损失,  $L_{\text{mask}}^*$  为分割损失。

分类损失公式如下:

$$L_{\text{cls}}^* = L(\{p_i\}) = \frac{1}{N_{\text{cls}}} \sum_i L_{\text{cls}}(p_i, p_i^*) \quad (6)$$

其中,  $i$  对应第  $i$  个锚点,  $N_{\text{cls}}$  为分类样本数,  $p_i$  为锚点预测为目标的概率值,  $p_i^*$  为标记的标签值,  $p_i^* = 1$  为正例,  $p_i^* = 0$  为负例,  $L_{\text{cls}}$  为两个类别的对数损失:

$$L_{\text{cls}}(p_i, p_i^*) = -\lg[p_i^* p_i + (1 - p_i^*)(1 - p_i)] \quad (7)$$

定位损失公式如下:

$$L_{\text{box}}^* = L(\{t_i\}) = \frac{1}{N_{\text{reg}}} \sum_i p_i^* L_{\text{reg}}(t_i, t_i^*) \quad (8)$$

其中,  $i$  对应第  $i$  个锚点,  $N_{\text{reg}}$  为回归的样本锚点个数,  $t_i$  为预测框位置参数,  $t_i^*$  为标记的真实框位置参数,  $L_{\text{reg}}$  为定位的回归损失,  $p_i^* L_{\text{reg}}$  则表示当锚点检测为正例时, 才计算其回归损失, 其公式为:

$$L_{\text{reg}}(t_i, t_i^*) = \text{smooth}_{L1}(t_i - t_i^*) \quad (9)$$

分割损失即二值掩码的损失, 如果候选框检测出为某一类别, 则使用该类别的交叉熵作为误差值进行计算, 其他类别损失值不计入, 从而避免了类间的竞争, 其公式为:

$$L_{\text{mask}}^* = -\frac{1}{m^2} \sum_{1 \leq i, j \leq m} [y_{ij} \lg \hat{y}_{ij}^k + (1 - y_{ij}) \lg (1 - \hat{y}_{ij}^k)] \quad (10)$$

其中,  $y_{ij}$  为  $m \times m$  区域坐标点  $(i, j)$  的标签值,  $\hat{y}_{ij}^k$  为该点第  $k$  类的预测值。

## 3 实验与结果分析

### 3.1 实验环境及数据集构建

实验环境及数据集构建主要有以下 3 个方面:

#### 1) 实验环境配置

本文实验所使用的环境配置如下: 操作系统为 Ubuntu 16.04; CPU 为 Xeon E5-2690 v4 @ 2.20 GHz; GPU 为 NVIDIA GeForce GTX 1080Ti; 深度学习框架为 Tensorflow 1.9.0; 编程语言为 Python 3.6。

#### 2) 数据集

由于公开的人脸数据集中大多数只有目标框的定位信息, 不具备分割标注信息, 因此本文从 FDDB<sup>[17]</sup> 原始图像数据集和 ChokePoint<sup>[18]</sup> 数据集中随机选择了 5 115 张图片并对其进行分割标注作为新数据集。FDDB (Face Detection Data Set and Benchmark) 数据集是马萨诸塞大学公开的人脸检测数据集, 该数据集由原始图像数据集和相应的部分标签数据组成, 标签数据包含了 2 845 张图片 (其中包含了 5 171 张人脸)。ChokePoint 数据集是由澳大利亚信息与通信技术研究 (NICTA) 公开的人脸数据集, 该数据集通过在不同场景出入口处安装不同角度监控摄像头, 从而采集到的人脸数据集, 该数据集包括 48 个视频序列, 共计约 64 204 张人脸图像。

### 3) 参数设置

本文采用 ResNet-101 网络作为特征提取器,首先使用 COCO 预训练模型<sup>[24]</sup>进行参数的泛化,使模型具有一定特征提取能力,从而减少训练时间。然后利用已标注的数据集进行模型训练,参考文献[16]并通过实验调试,设置迭代次数为 50,迭代步数为 3 000,学习率为 0.001,权值衰减率为 0.000 1。

## 3.2 结果分析

### 3.2.1 实验效果与性能分析

在训练出的模型上利用现有的公开数据集进行测试实验,测试图像检测结果如图 5 所示。可以看出,相较于其他人脸检测算法,本文算法在完成人脸目标框定位的同时,通过彩色的二值掩码将人脸信息与背景分割开来,实现了对人脸图像检测定位和分割的效果。



图 5 测试图像检测结果

Fig. 5 Test image detection results

为有效检测模型的性能,对比实验采用本文所构建数据集作为训练集,再进行模型性能的评测。实验在 Fddb 数据集上对比了本文算法与 Faster R-CNN<sup>[10]</sup>、Pico<sup>[25]</sup>、Viola-Jones<sup>[26]</sup>、koestinger<sup>[27]</sup>等算法,性能曲线如图 6 所示。

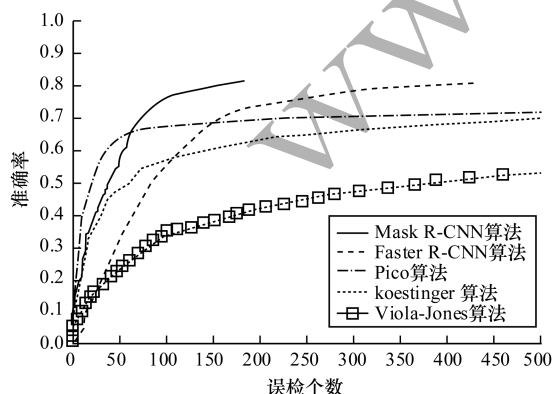


图 6 Fddb 数据集检测性能曲线

Fig. 6 Fddb datasets detection performance curve

从图 6 可以看出,本文算法误检个数明显减少,且准确率高于其他算法。

### 3.2.2 不同数据集的有效性分析

为了验证模型在不同数据集的有效性,本文在 AFW<sup>[28]</sup>数据集上进行了模型性能测试实验,实验对比了文献[29]中部分算法,实验结果如图 7 所示。虽然受限于训练数据集数据规模相对较小,以及标注规则存在差异,但本文模型仍取得了 94.08% AP 值(平均精度)的成绩,优于大部分对比算法,对比表现最优的 SquaresChnFtrs-5AP 值仅低了 1.16,证明了本文算法的有效性。

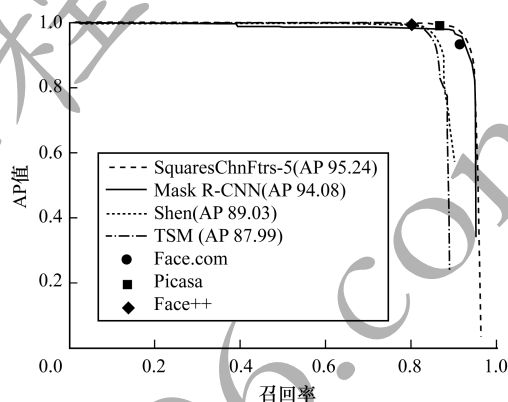


图 7 AFW 数据集检测性能曲线

Fig. 7 AFW datasets detection performance curve

此外,为进一步验证不同数据集的有效性,针对本文模型及在 Fddb 数据集性能表现较好的 Faster R-CNN 模型,本文分别在 ChokePoint 数据集、Caltech 10k Web Faces 数据集<sup>[30]</sup>、WIDER<sup>[31]</sup>数据集进行了测试,实验分别从以上 3 个数据集随机抽取 100 张图像进行测试,通过计算其 mAP(均值平均精度)值进行比较,结果如表 1 所示,其中加黑字体为最优成绩。

表 1 不同数据集性能对比  
Table 1 Performance comparison of different dataset

方法	Caltech	ChokePoint	WIDER
Faster R-CNN 方法	90.46	89.43	90.80
本文方法	<b>98.28</b>	<b>98.39</b>	<b>97.91</b>

通过不同数据集性能对比实验结果可以看出,本文方法比性能表现较好 Faster R-CNN 在 Caltech 10k Web Faces 数据集、ChokePoint 数据集、WIDER 数据集分别提高了 7.82%、8.96%、7.11%,验证了本方法在不同数据集上的有效性。

### 3.2.3 检测时间分析

时间测试实验同样在 ChokePoint 数据集、Caltech 10k Web Faces 数据集、WIDER 数据集进行

验证,从以上数据集分别随机抽取100张图像进行测试,比较其平均耗时,结果如表2所示,其中加黑字体为最少用时。

表2 不同数据集检测时间对比  
Table 2 Comparison of detection time of different datasets

方法	Caltech	ChokePoint	WIDER
Faster R-CNN 方法	<b>0.320</b>	<b>0.285</b>	<b>0.308</b>
本文方法	0.347	0.331	0.320

与Faster R-CNN算法对比,由于本文算法的RoIAlign层计算复杂度较高以及模型添加了分割分支,因此在Caltech 10k Web Faces数据集、ChokePoint数据集、WIDER数据集实验中,检测时间分别多了0.027 s、0.046 s和0.012 s。但本文算法能够以较少的时间花费换取更高的准确率,并能实现在人脸准确定位的同时将人脸信息从背景中分割出来,达到满意的检测与分割效果。

#### 4 结束语

本文提出一种基于Mask R-CNN的人脸检测及分割方法,构建一个具有分割标注信息的人脸数据集,并在建立的数据集上训练模型。该模型利用RoIAlign算法使得人脸图像特征点定位精准度达到像素级的效果,结合ResNet-101网络与RPN网络,提高了检测精度,并通过全卷积网络生成相应的人脸二值掩码,实现了人脸信息与背景图像的分割。在公开数据集上的对比实验结果表明,本文方法能够在不显著增加计算和模型复杂度的情况下,取得较好的检测效果。下一步将对数据集进行扩充,并在现有研究的基础上对模型进行优化,进一步提高模型检测性能。

#### 参考文献

- [1] VIOLA P, JONES M J. Robust real-time face detection[J]. International Journal of Computer Vision, 2004, 57(2): 137-154.
- [2] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2005: 886-893.
- [3] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [4] BAY H, ESS A, TUYTELAARS T, et al. Speeded-up robust features[J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.
- [5] AHONEN T, HADID A, PIETIRAINEN 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] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part-based models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627-1645.
- [7] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [C]//Proceedings of IEEE NIPS '12. Washington D. C., USA: IEEE Press, 2012: 1097-1105.
- [8] ZHAN Shu, TAO Qinqin, LI Xiaohong. Face detection using representation learning [J]. Neurocomputing, 2016, 187: 19-26.
- [9] RANJAN R, PATEL V M, CHELLAPPA R. A deep pyramid deformable part model for face detection[C]//Proceedings of the 7th IEEE International Conference on Biometrics Theory, Applications and Systems. Washington D. C., USA: IEEE Press, 2015: 1-8.
- [10] JIANG H, LEARNED-MILLER E. Face detection with the faster R-CNN [C]//Proceedings of the 12th IEEE International Conference on Automatic Face and Gesture Recognition. Washington D. C., USA: IEEE Press, 2017: 650-657.
- [11] WU Wenqi, YIN Yingjie, WANG Xingang, et al. Face detection with different scales based on faster R-CNN[J]. IEEE Transactions on Cybernetics, 2018, 49(11): 4017-4028.
- [12] SUN X, WU P, HOI S C H. Face detection using deep learning: an improved faster rcnn approach [J]. Neurocomputing, 2018, 299: 42-50.
- [13] GUO Xiaobo, ZHOU Zhaoyong, LI Songyang. Nasal tip detection and posture correction of 3D face based on effective energy [J]. Computer Engineering, 2018, 44(9): 236-242. (in Chinese)  
郭小波, 周兆永, 李松阳. 基于有效能量的3D人脸鼻尖点检测与姿态矫正[J]. 计算机工程, 2018, 44(9): 236-242.
- [14] ZOU Guofeng, FU Guxia, GAO Mingliang, et al. Pose varied face recognition based on self learning deep convolutional neural network[J]. Journal of Chinese Computer Systems, 2018, 39(6): 1156-1162. (in Chinese)  
邹国锋, 傅桂霞, 高明亮, 等. 基于自学习深度卷积神经网络的姿态变化人脸识别[J]. 小型微型计算机系统, 2018, 39(6): 1156-1162.
- [15] QIU Yiming, LIAO Haibin, CHEN Qinghu. Occluded face pose recognition based on dictionary learning with discrimination performance[J]. Geomatics and Information Science of Wuhan University, 2018, 43(2): 275-281, 288. (in Chinese)

- 邱益鸣,廖海斌,陈庆虎.基于鉴别字典学习的遮挡人脸姿态识别[J].武汉大学学报(信息科学版),2018,43(2):275-281,288.
- [16] HE K, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press,2017:2961-2969.
- [17] JAIN V, LEARNED-MILLER E. Fddb: a benchmark for face detection in unconstrained settings[EB/OL]. [2019-03-04]. <https://www.researchgate.net/publication/>.
- [18] WONG Y, CHEN S, MAU S, et al. Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press,2011:74-81.
- [19] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press,2016:770-778.
- [20] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press,2014:580-587.
- [21] GIRSHICK R. Fast R-CNN [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press,2015:1440-1448.
- [22] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C]//Proceedings of IEEE NIPS '15. Washington D. C., USA: IEEE Press,2015:91-99.
- [23] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of IEEE Conference on Oomputer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press,2015:3431-3440.
- [24] LIN T Y, MAIRE M, BELONGUE S, et al. Microsoft coco: common objects in context [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer,2014:740-755.
- [25] MARKUS N, FRLJAK M, PANDZIC I S, et al. A method for object detection based on pixel intensity comparisons organized in decision trees [EB/OL]. [2019-03-10]. <https://www.researchgate.net/publication/>.
- [26] JENSEN O H. Implementing the Viola-Jones face detection algorithm [D]. Lyngby, Denmark: Technical University of Denmark,2008.
- [27] KOESTINGER M, WOHLHART P, ROTH P M, et al. Robust face detection by simple means [C]//Proceedings of IEEE DAGM '12. Washington D. C., USA: IEEE Press,2012:125-137.
- [28] RAMANAN D, ZHU X. Face detection, pose estimation, and landmark localization in the wild [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press,2012:2879-2886.
- [29] MATHIAS M, BENENSON R, PEDERSOLI M, et al. Face detection without bells and whistles [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer,2014:720-735.
- [30] FINK M, FERGUS R, ANGELOVA A. Caltech 10 000 Web faces [EB/OL]. [2019-03-10]. [http://www.vision.caltech.edu/Image\\_Datasets/Caltech\\_10K\\_WebFaces](http://www.vision.caltech.edu/Image_Datasets/Caltech_10K_WebFaces).
- [31] YANG S, LUO P, LOY C C, et al. Wider face: a face detection benchmark [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press,2016:5525-5533.