

改进的基于YOLOv3的人脸检测算法

胡一帆, 秦岭, 杨小健

(南京工业大学计算机科学与技术学院, 南京 211816)

摘要: 针对因背景与人脸高度相似和人脸目标尺度过小而导致的人脸检测精度较低的问题, 提出了一种改进的基于YOLOv3的人脸检测算法。首先使用遗传算法改进原算法中随机初始化的影响, 生成更符合目标大小的预测框, 其次用轻量级网络改进原特征提取网络, 提高人脸检测速度, 最后使用边框回归损失代替YOLOv3坐标损失函数并改进置信度损失函数以提升训练收敛速度和结果精度。所设计的算法模型在Wider Face数据集上的检测精度和速度得到了提升。

关键词: 人脸检测; 聚类算法; YOLOv3; 轻量级网络; 损失函数

中图分类号: TP391.4 **文献标志码:** A

Improved Face Detection Algorithm Based on YOLOv3

HU Yifan, QIN Ling, YANG Xiaojian

(College of Computer Science and Technology, Nanjing TECH University, Nanjing 211816, China)

Abstract: Aiming at the low accuracy of face detection caused by the high similarity between background and face and the small scale of face target, an improved face detection algorithm based on YOLOv3 is proposed. Firstly, the K-means clustering algorithm based on genetic algorithm is used to improve the influence of random initialization in the original algorithm and generate a prediction frame more in line with the target size. Secondly, the lightweight network is used to improve the original feature extraction network and improve the face detection speed. Finally, the frame regression loss is used to replace the YOLOv3 coordinate loss function and the confidence loss function is improved to improve the training convergence speed and result accuracy. The accuracy and speed of the designed face algorithm are improved on Wider Face dataset.

Key words: face detection; clustering algorithm; YOLOv3; lightweight network; loss function

引 言

随着时代的发展, 图像和视频是互联网中最常见的数据展现形式。计算机视觉技术可以通过这些数据获取有效信息, 使得人与机器友好交互, 该技术已经应用在人们生活的各个方面。其中, 人脸检测是计算机视觉领域中的一个重点方向, 受到国内外学者的关注。

人脸检测是指通过计算机处理输入的图像、视频, 输出人脸的位置、大小的过程^[1]。人脸检测的研究源于人脸识别, 于20世纪60年代人脸识别就开始被学者关注, 60多年来取得了巨大的进步。而人脸

检测是人脸识别的第一个步骤,是在确定了人脸位置的基础上进行的,因此并没有单独受到重视。直至21世纪,随着人脸识别技术的成熟、人脸识别需求的不断扩大以及对复杂环境下的人脸识别的鲁棒性需求,人脸检测不再仅仅是人脸识别的关键部分,还是视觉监视、三维人脸模型等方面重要的一环。现有的人脸检测技术已经能高精度地检测出特定环境中的人脸,但在自然环境中,由于背景和人脸角度、大小的不同,人脸检测仍然还有很多问题需要面对。最容易受到影响的因素主要有:(1)人脸的表情是多变的,导致眼睛、嘴巴可能会呈现出不同的姿态;(2)图像或视频中人脸可能会被不同的物体遮挡,如头发、胡子、饰品或其他人,使其不具有完整性;(3)图片或视频呈旋转状态;(4)受拍摄光源的影响,亮度、阴影等也对检测结果有很大的影响。

目前各国的研究人员都对人脸检测投入了大量研究,每年的国内外会议以及期刊上都有许多文献发表。近年来,随着卷积神经网络技术的发展,深度人脸检测也取得了很大的进展。主要的研究方向分为两类:两步人脸检测和单步人脸检测。

YOLO是当前最先进的单步目标检测算法之一,常常被应用于计算机视觉领域,改进后的YOLOv3算法相较于其他优秀的目标检测算法单步多框目标检测(Single shot multibox detector, SSD)^[2]、RetinaNet^[3]、R-FCN^[4]有着较高的检测精度和检测速度。但对于现实中复杂的人脸检测环境,检测精度、召回率等依然不够理想。因此本文将针对YOLOv3算法,从聚类算法、特征提取网络、损失函数3方面进行改进,提升训练收敛速度和结果精度。

1 相关工作

1.1 传统人脸检测

传统人脸检测方法主要有两种类型。第一种是基于特征的检测方法,该方法主要通过肤色、灰度、模板等特征进行检测。文献[5]中根据肤色获取候选区域,使得模型检测效果得到很大的提升。文献[6]首先对图像进行分割,计算分割后各部分的像素平均值,最后得到灰度值,通过分析不同器官的灰度值来检测人脸。

第二种是基于统计的检测方法,该方法首先训练人脸样本,提取人脸特征,最后通过提取到的特征进行检测。文献[7]通过支持向量机将特征空间的间隔扩大以扩大正、负样本间的间隔。文献[8]通过梯度统计直方图对人脸特征提取,并通过latentSVM分类的方式检测。

1.2 基于卷积神经网络的人脸检测方法

基于卷积神经网络的目标检测算法主要分为单步检测和两步检测。两步检测算法中首先被人们熟知的是R-CNN^[9]算法,随后学者们又对其做出了改进。文献[10]提出通过分类、边框回归,统一两个阶段。文献[11]将文献[10]中的选择性搜索方法^[12]替换成区域生成网络,使得目标检测的各个步骤都能共享卷积特征,目标检测速度得到了很大的提升。特征丢失对人脸检测有着重要影响,文献[13]为此改进目标检测网络,选取位置敏感的卷积层。文献[14]通过对多个区域进行特征提取以达到丰富目标特征的目的,实际检测效果也得到了提高。单步检测算法最经典的主要有SSD^[2]和YOLO^[15]两类。前者将文献[11]提出的思想进行归纳,预先针对不同情况生成先验框。后者通过减少计算复杂度,使得检测速度得到明显的提升,但检测精度却有一定的下降。YOLOv2^[16]引入锚框(Anchor box)、高分辨率分类器等,在保持其原先速度的同时,检测精度也得到了明显的提升。YOLOv3^[17]在YOLOv2的基础上进一步改进,同时融入特征金字塔、残差网络等方法,在小目标检测精度方面也得到了提升。YOLOv4^[18]在不改变YOLOv3的骨干网络情况下,使用空间金字塔池化并将参数聚合方式更改为路径聚合网络。YOLOv5^[19]则将原先的DarkNet53网络改为CSPDarkNet和Focus网络结构,同时结合路径聚

合网络和特征金字塔聚合特征,并设计出自适应的描框计算。

由于人脸检测与目标检测有一定的相似性,因此很多人脸检测算法是在目标检测算法上进行改进,其中最为常见的人脸检测算法主要分为3类:级联卷积、单步、两步。对于级联卷积方式文献[20]首先提出在网络中设置多个分类器并通过级联结构组织,虽然相较于传统的人脸检测精度有很大的提升但模型过于复杂。文献[21]对其做出改进,用联合损失函数使其合并,同时合并各级的分类回归任务。文献[22]合并人脸区域性检测和关键点检测,该网络中候选区网络受限获取大致区域,接着使用微调网络对人脸区域进行优化,最后使用输出网络输出最终的检测结果。两步人脸检测主要是基于文献[11]提出的Faster R-CNN的改进,文献[23]提出了Face R-CNN,但在检测速度方面还有很大进步空间。单步人脸检测主要是基于SSD的改进,文献[24]减少了网络参数,设计出尺度不变性网络。文献[25]重新设计了SSD中的Anchor,针对小尺度人脸检测的问题,降低了阈值,使得小尺度人脸检测精度得到了很大的提升。

通过以上分析可以看出在各类目标检测和人脸检测算法中,各种YOLO算法都有着最佳的事件性能,在后面改进的YOLO算法中精确度也得到了很大的提升。相较于YOLOv4和YOLOv5,YOLOv3在当前形势下有着最广阔的应用领域,对于人脸检测方面,时间性能和精确度都有不错的表现,因此本文将对YOLOv3进行改进,提出一种改进的基于YOLOv3的人脸检测算法,从聚类算法、特征提取网络、损失函数3方面出发针对性地进行改进以提升训练收敛速度和结果精度。

2 YOLOv3算法

2.1 边界框预测

YOLOv3中使用K-means算法进行聚类,从而获得最优的Anchor box尺寸。预测结果会对每个网格生成3个边界框,并用其中心坐标、宽、高、置信度和条件类别概率来表示。网格计算出每个预测框和真实框的交并比,并选取交并比最大的预测框负责预测,交并比计算公式为

$$\text{IoU} = \frac{\text{box}_p \cap \text{box}_t}{\text{box}_p \cup \text{box}_t} \quad (1)$$

式中: box_p 为预测框的面积大小, box_t 为真实框的面积大小。

用 t_x, t_y, t_w, t_h 表示预测框的输出,用 b_x, b_y, b_w, b_h 表示预测结果。 b 和 t 之间的转换公式为

$$\begin{aligned} b_x &= \sigma(t_x) + c_x \\ b_y &= \sigma(t_y) + c_y \\ b_w &= p_w e^{t_w} \\ b_h &= p_h e^{t_h} \end{aligned} \quad (2)$$

式中: c_x, c_y 表示每个网格左上角点的坐标, p_w 和 p_h 分别表示预测框的宽的高。

置信度的值取决于该网格中是否有对象,如果有,置信度为1,反之则为0。对象类别概率表示对检测目标类别的预测。

2.2 特征提取网络

YOLOv3提出了DarkNet-53的特征提取网络,网络共53层卷积,最后一层为全连接层,其余52层为主体结构,包括:1个32 filters的卷积核,5组残差模块先执行 1×1 再执行 3×3 的卷积层,分别执行1、2、8、8、4次,其中每个残差模块的第一层使用步长为2的卷积核。网络输入为 416×416 ,通过步长为2的卷积,经历5次下采样后得到5个大小的特征图,并在 $13 \times 13, 26 \times 26, 52 \times 52$ 的特征图上预测。通过32倍、16倍、8倍下采样得到3种特征图以融入更多的浅层语义和细节信息。

2.3 损失函数

YOLOv3的损失函数主要有3个部分:定位坐标损失、置信度损失、目标类别损失。整体损失函数公式为

$$\text{Loss} = \text{Loss}_{\text{coord}} + \text{Loss}_{\text{IoU}} + \text{Loss}_{\text{class}} \quad (3)$$

定位坐标的损失函数通过方差来计算,由式(2)可知,边框最终输出为 b_x, b_y, b_w, b_h ,映射到 416×416 图上后输出值为 x, y, w, h ,原图中实际的真实框中心坐标和宽高为 $\hat{x}, \hat{y}, \hat{w}, \hat{h}$,定位坐标损失函数为

$$\begin{aligned} \text{Loss}_{\text{coord}} = & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} [(x_{ij} - \hat{x}_{ij})^2 + (y_{ij} - \hat{y}_{ij})^2] + \\ & \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} [(\sqrt{w_{ij}} - \sqrt{\hat{w}_{ij}})^2 + (\sqrt{h_{ij}} - \sqrt{\hat{h}_{ij}})^2] \end{aligned} \quad (4)$$

式中: S^2 表示特征图有 S^2 个网格,每个网格产生 B 个预测框, I_{ij}^{obj} 表示是否检测这个目标,如果是则为1,反之为0。

置信度损失函数通过交叉熵损失函数计算,整个损失函数分为两部分,当有目标物体时损失函数为

$$\text{Loss}_{\text{IoU1}} = \lambda_{\text{obj}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} [\hat{C}_i^j \lg C_i^j + (1 - \hat{C}_i^j) \lg (1 - C_i^j)] \quad (5)$$

当没有目标物体时损失函数为

$$\text{Loss}_{\text{IoU2}} = \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{noobj}} [\hat{C}_i^j \lg C_i^j + (1 - \hat{C}_i^j) \lg (1 - C_i^j)] \quad (6)$$

式中: C_i^j 为预测的置信度, \hat{C}_i^j 为真实置信度。

目标类别损失也是用相同的交叉熵损失函数计算,公式为

$$\text{Loss}_{\text{class}} = \sum_{i=0}^{S^2} I_{ij}^{\text{obj}} \sum_{c \in \text{classes}} [\hat{P}_i^j \lg P_i^j + (1 - \hat{P}_i^j) \lg (1 - P_i^j)] \quad (7)$$

式中: P_i^j 为在第 i 个网格的第 j 个预测框预测为 c 类别的概率, \hat{P}_i^j 表示 c 类真实的概率。

3 针对YOLOv3的算法改进

为解决复杂背景下的人脸检测问题,本文在YOLOv3的基础上进行改进,对于Anchor的预测使用遗传算法对原预测方式进行改进,特征提取网络方面使用GhostNet代替DarkNet-53,最后使用GIOU边框回归损失进行定位坐标的损失计算以及使用Focal loss进行置信度损失计算。

3.1 基于遗传算法的K-means聚类

如上文所述,YOLOv3要获取最优尺寸的Anchor box,则需要使用K-means算法对数据集集中的相关参数进行聚类。思想是首先随机设置 k 个中心,通过欧式距离公式筛选出与中心距离较近的目标分为一类,接着根据分类结果重新分配 k 个中心并根据中心重新划分聚类,不断重复直至聚类中心不变。

K-means算法有着速度快、消耗资源少的优点,但初始点的选取过于随机,对于分布不均的离散数据很难找到可靠的初始中心点,因此本文的改进方向是获取更加可靠的初始中心点,接着使用K-means聚类。

遗传算法是指将对象看作染色体,在其中随机选取若干形成聚类,通过适应度函数算出每个的适应值,并根据适应值选取出领头或是淘汰。接着将范围内的目标与其他聚类交叉,并产生子代,而适应度高的则直接进入下一代,然后使子代与适应度高的目标重新计算适应值,重复上述步骤,直到选取出最高适应值的目标。从算法过程可以看出遗传算法可以自适应地从全局搜索出最优的初始点。整个改进后的聚类算法流程如图1所示,具体算法如下:

(1)染色体编码。由于聚类算法的目的是获取最优尺寸的 Anchor box,因此算法设计的染色体为锚框的坐标。本文设置锚框的总数据为 N ,聚类数为 k ,并采用二进制编码。

(2)适应度函数。本文采用误差平方和的方式,设锚框 N_n 的宽高分别为 w_n, h_n ,样本集为 $N = \{N_1, N_2, \dots, N_n\}$,则公式为

$$G_{\text{inner}} = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_j^{(i)} - m_i)^2 \quad (8)$$

$$G_{\text{inter}} = \sum_{i=1}^k (m_i - m)^2 \quad (9)$$

式中: G_{inner} 为类内距离和, G_{inter} 为类间距类和, m_i 为聚类后每个聚类的均值, m 为整个样本的均值, $x_j^{(i)}$ 为第 i 个样本中的第 j 个对象。由于 G_{inter} 与适应度成正比, G_{inner} 与适应度成反比,因此适应度公式为

$$F(N_n) = \frac{G_{\text{inter}}}{D + G_{\text{inner}}} \quad (10)$$

式中 D 为防止分母为 0 设置的常数。

(3)选择算子。遗传算法中使用的是优胜劣汰的方式进行淘汰,本文依照其法则将适应度值高的目标大概率进入下一代,适应度值低的大概率进入交叉,其中适应度值最低的 10% 对象则被淘汰。因此选择算子的计算公式为

$$P(N_n) = \frac{F(N_n)}{\sum_{j=1}^N F(N_j)} \quad (11)$$

(4)自适应交叉算子。本文的设计为当适应值高于平均值时进行交叉,低于平均值时根据概率公式决定是否交叉,概率公式为

$$P_c = \begin{cases} \frac{F_{\text{max}} - F'}{F_{\text{max}} - F_{\text{avg}}} & F' \geq F_{\text{avg}} \\ 1 & F' < F_{\text{avg}} \end{cases} \quad (12)$$

式中: F_{max} 为适应度的最大值, F_{avg} 为适应度的平均值。

具体的交叉方式为:将输入值转化为 10 位的二进制数,并为二进制数设置分段,以每三位分为一段,例如 128 的二进制数为 0010000000,划分后则为 0|010|000|000。随机选取一个分段,将两个目标的分段的部分进行交叉互换。

(5)自适应变异算子。方式与交叉相同,概率公式为

$$P_m = \begin{cases} \frac{F_{\text{max}} - F'}{F_{\text{max}} - F_{\text{avg}}} & F' \geq F_{\text{avg}} \\ 1 & F' < F_{\text{avg}} \end{cases} \quad (13)$$

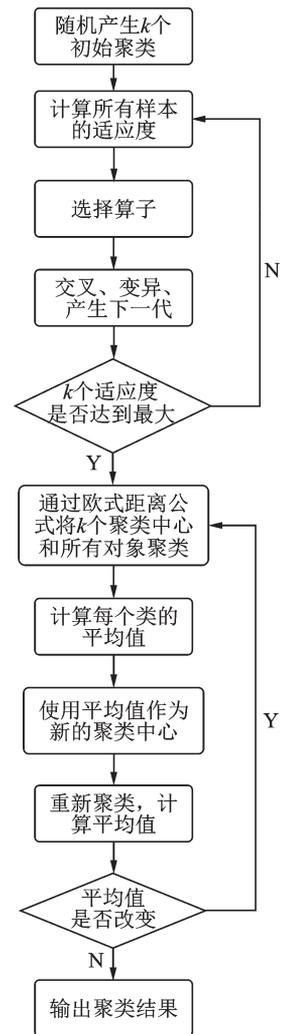


图1 基于遗传算法的 K-means 聚类算法流程图

Fig.1 Flow chart of K-means clustering algorithm based on genetic algorithm

具体的变异规则为:与交叉方式相同地设置为二进制并分段,接着随机选取其中一段并将该段中的3个数替换为随机生成的3位二进制数。

3.2 网络架构的改进

如上文所述,在YOLOv3模型中使用的是DarkNet-53的网络结构,相较于以往的网络,尽管该网络在性能方面更加优异,但较大的计算量以及较大的内存空间使得检测速度还有一定的提升空间。轻量级网络模型通过压缩模型大小,依然有着很好的检测效果,同时也具备一定的泛化能力,因此本文通过轻量化的网络结构代替DarkNet-53对网络架构进行改进。

GhostNet是最优秀的轻量级网络之一。GhostNet目标是通过廉价操作生成更多的特征图,在特征层间有很多是相似的,将这些相似的特征层看作彼此间的Ghost。卷积层的特征输出会包含很多相似的冗余,因此不需要全部卷积,并直接使一部分作为冗余信息。对比与目前的主流卷积,GhostNet模块的原始卷积可以自定义卷积核数量代替大量的1×1卷积。大多数方法是先卷积降维,再特征提取,而Ghost是先原始卷积,再通过线性变换获取特征。

图2是Ghost bottleneck结构图,主要由Ghost模块堆叠两次,第一个模块用于增加特征维度,第二个减少特征维度。当Stride=2时,shortcut采用下采样层,并增加Stride=2的卷积。

改进的网络结构图即用GhostNet的结构代替YOLOv3中的标准卷积,以减少卷积中的计算量。如表1所示为改进后的特征提取层部分,输入尺寸归一化为416×416×3,表中的S为步长,为了检测出更多的人脸目标,本文使用4个不同的储存进行预测,分别是13×13×960、26×26×112、52×52×40和104×104×24。

3.3 损失函数的改进

YOLOv3模型中为了保持正负样本平衡采用阈值划分、采样设置等措施,但由于人脸检测中人脸目标较小、背景复杂等问题,正负样本不平衡的问题依然存在。针对此问题,本文引入GIoU^[26]边框回归损失对坐标损失函数进行改进,通过Focal loss对置信度损失函数进行改进。

如式(4)所示,由于YOLOv3坐标损失函数的设计中当预测框和真实框未重合时,该方式无法衡量两框之间的距离远近,模型无法得到优化,因此本文引入GIoU进行坐标误差计算

$$GIoU = IoU - \frac{|C - (A \cup B)|}{|C|} \tag{14}$$

$$LossGIoU = 1 - GIoU \tag{15}$$

GIoU示意如图3所示,其中A、B为边界框,C为A、B的最小闭包区域。当A、B部分重合,则IoU=1,GIoU=1-0=1,当A、B部分相距很远时C趋于很大的值,则GIoU=0-1=-1,因此GIoU的取值范围为[-1,1],LossGIoU的取值范围则为[0,2]。

因此针对式(4)改进后的坐标误差计算为

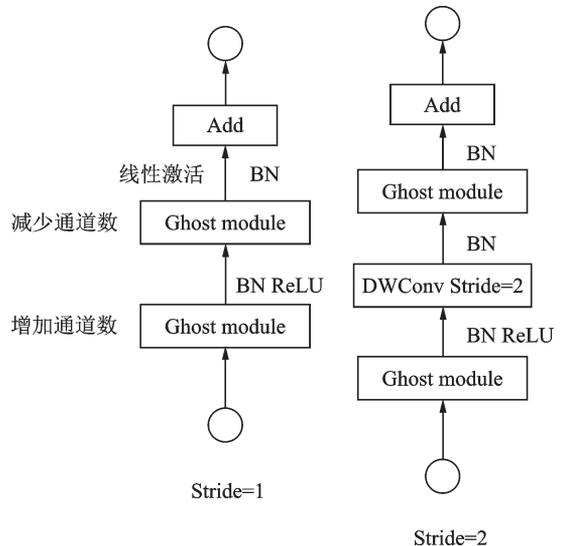


图2 Ghost bottleneck 结构图
Fig.2 Ghost bottleneck structure diagram

表 1 改进后的网络层次结构
Table 1 Improved network structure

Name	Input size	Operator	S	Output
第一特征提取层	416×416×3	Conv2d 3×3	2	208×208×16
	208×208×16	G-bneck	1	208×208×16
	208×208×16	G-bneck	2	104×104×24
	104×104×24	G-bneck	1	104×104×24
第二特征提取层	104×104×24	G-bneck	2	52×52×40
	52×52×40	G-bneck	1	52×52×40
第三特征提取层	52×52×40	G-bneck	2	26×26×80
	26×26×80	G-bneck	1	26×26×80
	26×26×80	G-bneck	1	26×26×80
	26×26×80	G-bneck	1	26×26×80
	26×26×80	G-bneck	1	26×26×112
	26×26×80	G-bneck	1	26×26×112
第四特征提取层	26×26×80	G-bneck	2	13×13×160
	13×13×160	G-bneck	1	13×13×160
	13×13×160	G-bneck	1	13×13×160
	13×13×160	G-bneck	1	13×13×160
	13×13×160	Conv2d 1×1	1	13×13×960

$$\text{Loss}_{\text{coord}} = \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} (1 - \text{GIoU}) \quad (16)$$

对于人脸检测,训练样本中,大多数预测框不包含人脸,直接使用二分类解决方式会出现很严重的不平衡问题。为解决以上问题,Focal loss 对标准交叉熵函数进行了改进

$$\text{FL} = \begin{cases} -(1-p)^\gamma \log_a p & y=1 \\ -p^\gamma \log_a (1-p) & y=0 \end{cases} \quad (17)$$

因此本文改进的置信度损失函数为

$$\begin{aligned} \text{Loss}_{\text{IoU}} = & \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{\text{obj}} [a(1-C_i^\gamma) \hat{C}_i^\gamma \ln C_i^\gamma + (1-a)(C_i^\gamma)^\gamma (1-C_i^\gamma) \ln(1-C_i^\gamma)] + \\ & \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^B [a(1-C_i^\gamma) \hat{C}_i^\gamma \ln C_i^\gamma + (1-a)(C_i^\gamma)^\gamma (1-C_i^\gamma) \ln(1-C_i^\gamma)] \end{aligned} \quad (18)$$

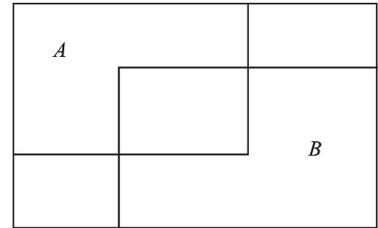


图3 GIoU示意图
Fig.3 GIoU sketch map

4 实验与分析

4.1 数据集

本文实验采用 Wider Face^[27]人脸检测数据集,该数据集由 62 个不同类型场景下的 32 203 个图像和 393 703 个人脸图像构成,每张图像都可以呈现出人脸的不同表情、姿势、光照、遮挡方式等。该数据集分为训练集、验证集和测试集 3 部分,每个部分中还定义了 3 种尺寸的人脸大小,可以很好地反映出不同

同算法针对不同大小尺寸人脸的检测效果。

4.2 实验设置

在模型训练时,针对人脸检测,实验对原配置信息进行了部分修改,部分参数设置如表2所示。

4.3 评价指标

人脸检测效果主要通过精确率(Precision)、召回率(Recall)、每秒处理图像数(Frames per second, FPS)、平均精确度(mean average precision, mAP)来评价。

精确率是指检测结果中检测正确的概率,表示为正确预测的个数与被预测个数之比。召回率指对于所有含有人脸的图像中真正为人脸的概率,表示为预测人脸数与图像中所含人脸个数之比

$$\text{Precision} = \frac{\text{TP}}{\text{FP} + \text{TP}} \quad (19)$$

$$\text{Re call} = \frac{\text{TP}}{\text{FN} + \text{TP}} \quad (20)$$

式中:TP指预测正确的样本数,FP指将非人脸预测为人脸的样本数,FN为将真实人脸预测为非人脸的样本数。

根据不同阈值下的精度值和召回率,分别以精度值和召回率为横、纵坐标,画出一条曲线,该曲线可以综合考虑精度值和召回率的影响,AP(Average precision)值就指该曲线与横纵坐标轴所围成的面积,而mAP则是在各个类别下AP值的平均值,AP和mAP的计算公式分别为

$$\text{AP} = \int_0^1 \text{Precision}(r) dr \quad (21)$$

$$\text{mAP} = \frac{\text{AP}_1 + \text{AP}_2 + \dots + \text{AP}_k}{k} \quad (22)$$

4.4 消融实验

本文对YOLOv3的改进主要从K-means聚类算法、特征提取网络、损失函数3方面改进,为了验证各个方面改进的有效性,本文设置了以下实验进行验证:

- (1) 对YOLOv3进行实验。
- (2) 使用改进后的聚类算法的Anchor进行实验。
- (3) 使用新的特征提取网络进行实验。
- (4) 使用改进后的损失函数进行实验。
- (5) 使用改进后的聚类算法的Anchor和新的特征提取网络进行实验。
- (6) 使用改进后的聚类算法的Anchor、新的特征提取网络和改进后的损失函数进行实验。

实验结果分别如表3、4所示。从表3、4可以看出,本文提出的算法在3个不同数据集上的mAP值可以分别达到0.938、0.923、0.920,相较于原始的YOLOv3算法分别提升了5.2%、4.3%、6.3%,速度方面也接近提升了1倍。其中,对于Anchor、损失函数的改进都使得mAP值有一定的提升,特征提取网络由于选取了轻量级网络,虽然精度无明显提升但速度大幅提升,综合以上证明了算法改进的有效性。

表2 实验参数设置

Table 2 Experimental parameter setting table

参数描述	参数值
Width	416
Height	416
Channels	3
Batch	16
Decay	0.000 5
Angle	5
Saturation	1.5
Exposure	1.5
Hue	0.1
Learning_rate	0.001
Max_batches	500 200
Scales	0.1, 0.1

表3 针对mAP值的效果对比

Table 3 Effect comparison for mAP value

实验序号	实验描述	mAP		
		Easy	Medium	Hard
(1)	YOLOv3	0.886	0.880	0.857
(2)	改进 Anchor	0.910	0.899	0.891
(3)	新特征网络	0.890	0.882	0.858
(4)	改进损失函数	0.898	0.886	0.878
(5)	改进 Anchor+新特征网络	0.927	0.913	0.907
(6)	改进 Anchor+新特征网络+改进损失函数	0.938	0.923	0.920

表4 针对检测速度的效果对比

Table 4 Effect comparison for detection speed

实验序号	实验描述	Speed	
		FPS	ms
(1)	YOLOv3	22.2	45
(2)	改进 Anchor	20.9	44
(5)	改进 Anchor+新特征网络	39.0	26
(6)	改进 Anchor+新特征网络+改进损失函数	40.2	25

4.5 精度对比实验

为了验证本文提出算法与其他各种人脸检测算法检测精度的提升效果,选取了以下算法进行对比: CMS-RCNN^[28]、HR^[29]、SSH^[24]、Face-RFCN^[30]、SRN^[31]、YOLOv4、YOLOXs。CMS-RCNN的主要检测目标是小尺度目标,HR是基于多尺度融合的人脸目标检测方式,SSH是针对VGG网络的改进,融入多尺度特征分析,Face-RFCN是为了应对复杂场景的优化,SRN算法通过STC模块过滤掉浅层特征图上的负样本,通过STR精确回归边界框位置。具体精度对比结果如表5所示。从表5可以看出,本文算法相较于CMS-RCNN、HR、SSH、Face-RFCN在3个子数据集上都有显著提升,由于SRN模型中引入了二阶分类和回归任务,提升了定位的精确度,但该

表5 不同模型mAP值对比实验结果

Table 5 Comparison of experimental results of mAP value for different models

网络模型	mAP		
	Easy	Medium	Hard
CMS-RCNN	0.899	0.874	0.624
HR	0.925	0.897	0.756
SSH	0.931	0.899	0.745
Face-RFCN	0.937	0.901	0.831
SRN	0.944	0.938	0.901
YOLOv4	0.940	0.924	0.918
YOLOXs	0.932	0.918	0.916
本文模型	0.938	0.923	0.920

模型主要针对于小尺度人脸检测,虽然在Easy和Medium数据集上略优于本文算法,但在Hard数据集上本文算法mAP值明显高于SRN。相较于YOLOv4,本文模型在提高速度的同时仍能 and YOLOv4有相近的精确度。由于本文模型选取轻量级网络注重于速度,所以类似于YOLOv5缩放规则将YOLOX缩放为YOLOvXs进行对比,结果可以看出本文模型略优于YOLOvXs。

4.6 检测效果对比

本文选取一张包含约1000张人脸的人脸检测热门图片进行效果对比,检测效果分别如图4~6所示。通过实验,原始YOLOv3模型可检测出315张人脸,YOLOv4模型可检测出537张人脸,而本文模型可检测出751张人脸。本文选取图片右上部分呈现小尺度特征能够有效判断检测效果,图7为局部细节对比。对比图7可以发现本文模型在复杂环境、以及小尺度特征的情况下依然有着较好的检测效果,优于其他模型。



图4 YOLOv3检测效果图

Fig.4 YOLOv3 detection effect diagram



图5 YOLOv4检测效果图

Fig.5 YOLOv4 detection effect diagram



图6 本文模型检测效果图

Fig.6 Detection effect diagram of the proposed model



图7 局部细节对比图

Fig.7 Partial detail comparison diagram

5 结束语

为了改善图像中复杂背景对人脸检测精度和速度的影响,本文选择YOLOv3模型为基础,从聚类方式、特征提取网络和损失函数3方面进行改进,提出了改进的基于YOLOv3的人脸检测算法。算法首先使用遗传算法对数据集进行聚类获取先验框,接着改进网络架构使用GhostNet进行特征提取,最后提出GIoU损失函数代替原坐标损失函数并使用Focal loss改进置信度函数。实验结果表明,该算法在平均精确度和速度方面都得到了明显的提升。本文仅从聚类方式、特征提取网络和损失函数3方面进行了改进,在接下来的研究中可以从特征融合等方面进一步提升模型性能。

参考文献:

- [1] YANG M H, KRIEGMAN D J, AHUJA N. Detecting faces in images: A survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(1): 34-58.
- [2] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//*Proceedings of European Conference on Computer vision*. [S.l.]: Springer, 2016: 21-37.
- [3] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//*Proceedings of International Conference on Computer Vision*, [S.l.]: IEEE, 2017: 2980-2988.
- [4] DAI J, LI Y, HE K, et al. R-FCN: Object detection via region-based fully convolutional networks[C]//*Proceedings of the 30th International Conference on Neural Information Processing Systems*, [S.l.]: [s.n.], 2016: 379-387.
- [5] CHANDRAPPA D N. Face detection in color images using skin color model algorithm based on skin color information[C]//*Proceedings of 2011 3rd International Conference on Electronics Computer Technology*, [S.l.]: IEEE, 2011: 254-258.
- [6] YANG G, HUANG S. Human face detection in a complex background[J]. *Pattern Recognition*, 1994, 27(1): 53-63.
- [7] OSUNA E. An improved training algorithm for support vector machines[C]//*Proceedings of IEEE Neural Networks for Signal Processing*. [S.l.]: IEEE, 1997: 276-285.
- [8] BENENSON R. Face detection without bells and whistles[C]//*Proceedings of ECCV*. [S.l.]: Springer, 2014: 720-735.
- [9] 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*. Columbus: IEEE, 2014: 580-587.
- [10] GIRSHICK R. Fast R-CNN[C]//*Proceedings of IEEE International Conference on Computer Vision*. Chile: IEEE, 2015: 1440-1448.
- [11] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *IEEE Trans on Pattern Anal Mach Intell*, 2015, 39(6): 1137-1149.
- [12] UIJLINGS J R, SANDE K E, GEVERS T, et al. Selective search for object recognition[C]//*Proceedings of International Journal of Computer Vision*, [S.l.]: Springer, 2013: 154-171.
- [13] DAI J, LI Y, HE K, et al. R-FCN: Object detection via region-based fully convolutional networks[J]. *Advances in Neural Information Processing Systems*, 2016. DOI: 10.48550/arXiv.1605.06409.
- [14] GIDARIS S, KOMODAKIS N. Object detection via a multi-region and semantic segmentation-aware CNN model[C]//

- Proceedings of ICCV. [S.l.]: IEEE, 2015: 1134-1142.
- [15] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [C]//Proceedings of Computer Vision & Pattern Recognition. [S.l.]: IEEE, 2016: 779-788.
- [16] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]//Proceedings of IEEE Conference on Computer Vision & Pattern Recognition. [S.l.]: IEEE, 2017: 6517-6525.
- [17] REDMON J, FARHADI A. YOLOv3: An incremental improvement[EB/OL]. (2018-04-08). <https://arxiv.org/abs/1804.02767>.
- [18] BOCHKOVSKIY A, WANG Y, LIAO H. YOLOv4: Optimal speed and accuracy of object detection [EB/OL]. (2020-04-23). <https://arxiv.org/pdf/2004.10934.pdf>.
- [19] YOLOv5[EB/OL]. (2020-06-09). <https://github.com/ultralytics/yolov5>.
- [20] LI H, ZHE L, SHEN X, et al. A convolutional neural network cascade for face detection[C]//Proceedings of Computer Vision & Pattern Recognition. [S.l.]: IEEE, 2015: 5325-5334.
- [21] QIN H, YAN J, XIU L, et al. Joint training of cascaded CNN for face detection[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S.l.]: IEEE, 2016: 3456-3465.
- [22] 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.
- [23] WANG H, LI Z, JI X, et al. Face R-CNN[J]. 2017. DOI: 10.48550/arXiv.1706.01061.
- [24] NAJIBI M, SAMANGOUEI P, CHELLAPPA R, et al. SSH: Single stage headless face detector[C]//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). [S.l.]: IEEE, 2017: 4875-4884.
- [25] ZHANG S, ZHU X, LEI Z, et al. S3FD: Single shot scale-invariant face detector[C]. Proceedings of ICCV. [S.l.]: IEEE, 2017: 192-201.
- [26] SHRIVASTAVA A, GUPTA A, GIRSHICK R. Training region-based object detectors with online hard example mining[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2016: 761-769.
- [27] 韩兴, 张红英, 张媛媛. 基于高效通道注意力网络的人脸表情识别[J]. 传感器与微系统, 2021, 40(1): 118-121
HAN Xing, ZHANG Hongying, ZHANG Yuanyuan. Facial expression recognition based on high efficient channel attention network[J]. Transducer and Microsystem Technologies, 2021, 40(1): 118-121.
- [28] ZHU C, ZHENG Y, LUU K, et al. Contextual multi-scale region-based CNN for unconstrained face detection[M]. [S.l.]: Springer, 2017: 57-79.
- [29] SUN K, ZHAO Y, JIANG B, et al. High-resolution representations for labeling pixels and regions[J]. arXiv preprint arXiv: 1904.04514, 2019. DOI: 10.48550/arXiv.1904.04514.
- [30] CHEN Q, SHEN F, DING Y, et al. Face detection using R-FCN based deformable convolutional networks[C]//Proceedings of 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC). [S.l.]: IEEE, 2018: 4165-4170.
- [31] ZHANG Z, SHEN W, QIAO S, et al. Robust face detection via learning small faces on hard images[C]//Proceedings of IEEE Winter Conference on Applications of Computer Vision. [S.l.]: IEEE, 2020: 1361-1370.

作者简介:



胡一帆(1996-),通信作者,男,硕士研究生,研究方向:工业信息化,E-mail: 1365880855@qq.com。



秦岭(1980-),男,副研究员,研究方向:工业信息化。



杨小健(1963-),男,教授,研究方向:工业信息化,计算机控制。

(编辑:夏道家)