◎热点与综述◎

改进的多目标回归实时人脸检测算法

吴志洋,卓 勇,廖生辉
WU Zhiyang, ZHUO Yong, LIAO Shenghui

厦门大学 航空航天学院,福建 厦门 361005
College of Aerospace Engineering, Xiamen University, Xiamen, Fujian 361005, China

WU Zhiyang, ZHUO Yong, LIAO Shenghui. Improved multi-objective regressive real-time face detection algorithm. Computer Engineering and Applications, 2018, 54(11):1-7.

Abstract: Real-time multi-objective regression algorithm optimizes four location parameters respectively and separates four location variables. It makes localization less accurate and training hard to converge. Based on the problems, this paper proposes a real-time multi-object regression face detection algorithm, combining detection evaluation function Intersection over Union(IoU) as loss function. Based on multi-objective regression algorithm model proposed by Redmond etc., it adopts its real-time detection mechanism. Then IoU is introduced as loss function of location parameter. Four independent location parameters in real-time multi-object regression model are integrated into one unit and optimized, avoiding the defects of the base model. The algorithm is tested in face detection benchmark FDDB. The results indicate that this algorithm is superior to a mainstream traditional one in terms of face detection's effectiveness, and it outperforms other classical deep learning methods in terms of detection speed. The algorithm achieves a balance between face detection's effectiveness and detection speed. It provides some reference value to construct practical face applications.

Key words: multi-objective regression; face detection; Intersection over Union(IoU); convolutional neural network

摘 要:针对物体检测实时多目标回归算法中分别优化各四个位置参数,割裂了四个位置变量之间的关系,造成对物体的边框回归不够准确且训练不易收敛的问题,提出一种带检测评价函数(Intersection over Union,IoU)作为损失函数的实时多目标回归人脸检测算法。首先基于Redmond等提出实时多目标回归模型,采用该模型检测实时性的机制,然后融合了IoU函数作为位置参数的损失函数,将实时多目标回归模型中的四个独立位置参数整合成一个单元进行优化,避免了基础模型的缺陷。算法在人脸检测基准库FDDB上进行测试,实验结果表明:在人脸检测的有效性上优于主流的传统人脸检测算法,检测速度上领先于其他经典深度学习方法。提出的算法在检测人脸的有效性和检测速度两者之间取得了一个较好的平衡,为构建实用的人脸相关应用系统提供了参考价值。

关键词:多目标回归;人脸检测;检测评价函数;卷积神经网络
文献标志码:A **中图分类号:**TP391 doi:10.3778/j.issn.1002-8331.1803-0029

1 引言

人脸检测是计算机视觉和模式识别中一个重要而又基础的研究,同时也是众多跟人脸相关应用的关键环节,比如人脸识别、人证比对等。传统计算机领域的研究者对人脸检测的研究主要集中在人工设计特征提取器,如SIFT[1]、HOG[2]用传统的机器学习算法训练有效的分类器来进行图像中的人脸检测和识别任务。这样的方法要求研究人员必须手工提取到有效的特征,然后对每个部分分别进行优化,这导致了在检测过程中得到的往往是局部最优而不是全局最优。

基金项目:2016年工信部智能制造综合标准化与新模式应用项目(No.2016-213)。

作者简介:吴志洋(1989—),男,硕士研究生,主要研究方向为计算机视觉、深度学习,E-mail:wuzhiyang18@163.com;卓勇(1970—),男,博士,教授,主要研究方向为计算机视觉、机电一体化;廖生辉(1994—),男,硕士研究生,主要研究方向为计算机视觉、机电一体化。

收稿日期:2018-03-02 **修回日期:**2018-04-17 **文章编号:**1002-8331(2018)11-0001-07

基于Adaboost[3]的传统人脸检测算法现阶段在速度上仍然具有明显优势,而深度学习方法在检测的准确率上则可以取得更好的性能表现,比如,在人脸测评数据集FDDB[4]上,传统方法只有85%的准确率,而深度学习方法已超过95%,包括人脸识别的深度学习方法[5]也取得了极大进展。因此,基于深度学习的人脸检测方法已经成为当前的研究主流。

目前,基于深度学习的主流方法可以总结为三个步骤:首先,从一张图片中提取目标候选区,常用的方法有Selective Search[6]等;然后,把这些提取到的候选区送入一个卷积神经网络中进行识别或者分类;最后,对某些分类结果的候选区进行边框微调。对于以上的三个环节,其中的瓶颈在于候选区的提取,即第一个环节,这个环节同时制约着物体检测的准确率与检测速度。一方面,由于候选区提取环节是基于低级的语义特征的,且传统的区域推荐算法对局部外观变化敏感,导致了算法在许多情况下会失效,比如物体遮挡等情况;另一方面,大量的区域推荐算法基于图像分割[6]或者是稠密的滑动窗口形式[7],这带来了庞大的计算量,使得算法无法在实时的物体检测系统中得到应用。

为了克服这些缺陷,近几年出现了一系列改进的深度学习算法,大大加速了区域推荐环节。Zhang Xiang[8]等为检测的目标训练一个分类器,用一种高效的滑动窗口的方式遍历多张不同尺寸的图像,实现了物体的分类与定位,由于检测器需要在图像金字塔上面遍历所有层级的图像,当层级过多时,将耗费大量的计算时间,而当层级太少时,检测效果则会明显下降;Girshick R等[9]提出了R-CNN方法,该方法首先在一张图像上产生2 000个候选区域,然后把这些区域送入SVM分类器,最后把含有物体的区域传入下一个网络进行边框的回归,这种复杂的方式导致检测速度慢,且每个部分是独立训练的,造成优化十分困难;为了克服这些问题,Ren S等[10]提出了Faster R-CNN方法,在生成候选框的部分采用一个浅层的全卷积网络RPN在每张图像上生成约300个候选框,但是由于这些候选框的尺度和比例是提前设计好的,且是固定的,这就造成了当图像中物体尺寸范围波动较大时,RPN网络表现不理想;Redmon J等[11]把物体检测看作是一个简单的回归问题,把图片划分成7×7的网格,直接回归出每个物体的种类与边框,且不需要图像金字塔,因而在检测速度方面具有十分明显的优势。然而,该方法对物体边框信息的四个变量用L2损失函数(平方误差)分别进行回归,这种过于简单的方式割裂了四个位置变量之间的关系,导致在物体定位上效果不够理想且网络训练不易收敛,而Jiang Y等[12]提出了IoU Loss避免了L2损失函数的缺陷,但在检测速度上却达不到实时性,很难应用于实际的工程项目。

本文受回归思想与IoU Loss的启发,创造性地提出了结合回归思想与检测评价函数IoU作为损失函数的人脸检测算法,该算法与传统算法以及经典深度学习算法相比具有如下3个优点:

(1)应用卷积神经网络能够自动学习到数据中的特征,比传统的人工设计特征更加有效。

(2)融合IoU函数克服了实时多目标回归算法变量分离的缺陷,使得模型的代价函数更加合理,不仅使原有的多目标回归算法在检测不同尺度的人脸时更加鲁棒,而且使得深度网络的训练更加容易收敛。

(3)不需要采用图片金字塔的方式,只需处理一个层级的图片,较好地权衡了算法的检测速度与检测精度。

2　相关算法与分析

2.1　实时多目标回归算法YOLO

YOLO是Redmon J等[11]提出的一种通用物体检测深度卷积神经网络模型,它主要由24个卷积层、4个最大池化层、2个全连接层、L2损失函数层组成,如图1所示。图中省略了激活函数层、Batch Normalization(BN)层[13],其中C代表卷积层,P代表最大池化层,FC代表全连接层,L2 Loss代表平方损失层,且在所有的卷积层、倒数第二个全连接层后附加Leaky[11]激活函数层,所有卷积层之前带有BN层。

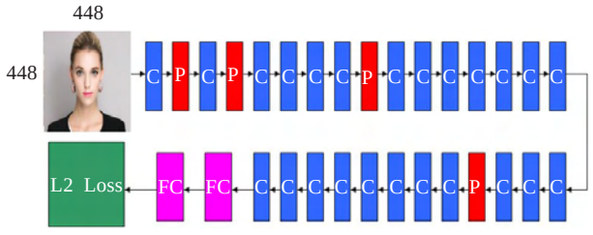
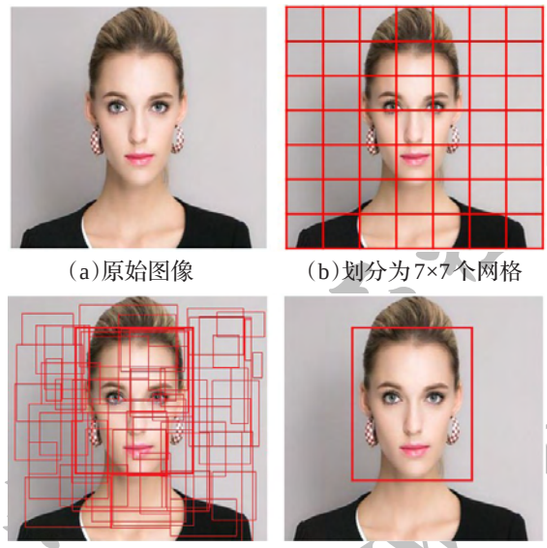


图1　网络结构示意图(省略了激活函数层、BN层)

YOLO把物体检测分开的几个部分整合到一个深度卷积神经网络中,整体流程如图2所示,该方法把一张图片分割成7×7个小网格,然后每个网格回归出两个包围框,最后用NMS[14]算法合并多余的人脸框,得到最终的人脸区域。

其损失函数定义如下:

$$\begin{aligned}Loss=&\lambda_{\text{pos}}\sum_{i=1}^{49}\sum_{j=1}^{2}\theta_{ij}^{\text{obj}}[(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2]+\\&\lambda_{\text{pos}}\sum_{i=1}^{49}\sum_{j=1}^{2}\theta_{ij}^{\text{obj}}[(\sqrt{w_i}-\sqrt{\hat{w}_i})^2+(\sqrt{h_i}-\sqrt{\hat{h}_i})^2]+\\&\sum_{i=1}^{49}\sum_{j=1}^{2}\theta_{ij}^{\text{obj}}(Conf_{ij}-\hat{Conf}_i)^2+\\&\lambda_{\text{noobj}}\sum_{i=1}^{49}\sum_{j=1}^{2}\theta_{ij}^{\text{obj}}(Conf_{ij}-\hat{Conf}_i)^2+\\&\sum_{i=1}^{49}\theta_i^{\text{obj}}\sum_{c=1}^{Class}(p_i(c)-\hat{p}_i(c))^2\end{aligned}\tag{1}$$



(a)原始图像　(b)划分为7×7个网格

(c)每个网格预测2个bboxes　(d)NMS得出最后的bbox

图2　多目标回归算法示意图

其中，$Loss$ 表示网络的损失值；$\lambda_{\text{pos}}=5$；$\lambda_{\text{noobj}}=1$；θ_{ij}^{obj} 表示第 i 个网格中的第 j 个包围框含有物体中心，当有物体中心时，$\theta_{ij}^{\text{obj}}=1$，否则 $\theta_{ij}^{\text{obj}}=0$；$\theta_{i}^{\text{obj}}$ 表示第 i 个网格是否含有物体中心，如果有物体中心时，$\theta_{i}^{\text{obj}}=1$，否则为0；$x_i$、$y_i$ 表示预测出来的包围框中心点坐标相对于网格的大小；$\hat{x}_i$、$\hat{y}_i$ 表示训练图片中标记的物体边框的中心点坐标相对于网格的大小；w_i、h_i 表示预测出来的包围框的宽和高相对于整张图片的大小；$\hat{w}_i$、$\hat{h}_i$ 表示训练图片中标记的物体边框的宽和高相对于图片的大小；$Conf_{ij}$、$\hat{Conf}_{ij}$ 分别表示第 i 个网格中预测的第 j 个包围框的置信度与第 i 个网格训练图片标注的置信度；$p_i(c)$，$\hat{p}_i(c)$ 分别表示第 i 个网格预测出的类别概率和训练数据标注的类别概率，其中位置参数如图3示。

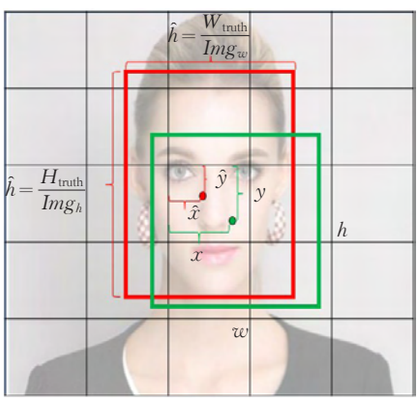


图3　位置参数示意图

2.2　YOLO算法缺陷分析

检测评价函数IoU(Intersection over Union)是被用来评价模型检测效果好坏的一个标准，表示的是两个框的交集 I 和并集 U 的比例，如图4所示：重叠程度越高，IoU值越大。其中红色框是标注框，绿色框是预测框，IoU函数表达式为 $IoU=I/U$ 。

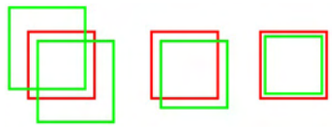


图4　IoU示意图

通过式(1)所示：模型的预测框参数 x,y,w,h 通过L2损失函数 $L_2(x_i,\hat{x}_i)=\sum_i(x_i-\hat{x}_i)^2$ 独立进行优化，这样的方式割裂了四个位置参数之间的强相关性。可以得出：

(1)在相同IoU的条件下，理论上网络损失函数的贡献应该是均等的，但当用L2损失函数时，如图5所示，尺度大的人脸对损失函数所产生的误差将大大超过小尺度人脸所产生的误差，导致深度网络在训练时，更加偏向于尺度较大的人脸，而容易忽略尺度较小的人脸，这对于网络的收敛以及模型的检测效果都将带来负面影响。

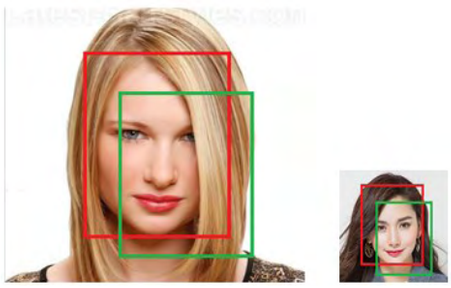


图5　相同IoU下的大小人脸

(红色为标注框，绿色为回归出的人脸区域)

(2)人脸数据集中人脸尺寸的跨度较大，且小人脸占了一定的比例，在FDDB[4]中，分辨率为40×40的小人脸占到10%左右，在Wider Face[14]中则占到33%左右，也就意味着(1)中所分析的情况，如果采用L2 Loss，则带来的缺陷是不可避免的。

(3)当单独优化各个位置参数时，容易导致仅有部分变量回归正确，如图6所示，回归出的人脸区域(绿色)仅有人脸区域的左上角坐标回归正确，而无法完全正确回归出整个人脸位置。

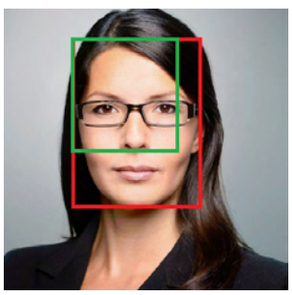


图6　位置参数单独优化的缺陷

(红色为标注框，绿色为回归出的人脸区域)

3 融合多目标回归与IoU损失的人脸检测算法MIFD(Multi-objective regression with IoU loss Face Detection)

3.1 算法的实时性分析

基于卷积神经网络的人脸检测算法在检测的准确程度上比传统人脸检测算法有较大的优势,而在检测速度上,大多数深度学习算法却达不到实时性。

如基于Faster RCNN[14]的人脸检测,其不能达到实时的核心问题在于人脸推荐区域的环节上,其设计相当于是用一个滑动窗口对最后的特征图上的每一个位置都进行了估计,每个位置上预测9种不同尺度的候选区域,一张图片约推荐2 000个候选区,代价在于采用滑动窗口的方式十分耗时,且推荐了过多的候选区,而这些候选区还要进入下一级网络再次进行特征提取;DenseBox[15]与Overfeat[8]则需要通过构建多个层级的图像金字塔来保证检测效果,由于要处理多张图像,将带来巨大的计算量;Unibox[12]提出了一种不需要图像金字塔的人脸检测策略,通过对图像中的每一个像素进行分类(人脸与非人脸),以及在每一个像素预测出距离人脸边界的距离,通过获取人脸置信度大于阈值的像素以及该像素对应的边界距离来实现人脸检测,但由于其对像素的分类提取的是网络的浅层特征,往往会得到过多被预测为人脸的像素,遍历这些像素需要较多的计算时间。

在多目标回归算法中,直接将输入图像划分为7×7个网格,每个网格预测2个人脸区域,共有98个候选区,相比于Faster RCNN,这种网格划分的方式在获取候选区方面,速度上有着巨大的优势;相比于DenseBox与Overfeat,多目标回归算法则不需要构建图像金字塔;相比于Unibox,多目标回归算法则直接对这98个候选框进行位置参数的调整,并不会产生不可预估的大量候选区。

为了实现实时检测的目的,本文选择多目标回归的机制进行人脸检测。

3.2 改进思路

基于2.2节的算法缺陷分析,本文拟作如下改进:

首先,发现IoU函数在面对任意尺度的人脸时,当人脸预测框与标注框(ground truth)具有相同的重合效果时,其IoU值是一致的,如果用IoU函数来作为位置参数的损失函数,将能够避免2.2节中(1)、(2)所分析的缺陷,即克服了不同人脸尺度带来误差不均衡的问题。

其次,从2.1节式(1)中截取了部分关于位置参数的L2损失函数,令$J_1=(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2$,则$\frac{\partial J_1}{\partial x_i}=2(x_i-\hat{x}_i)$、$\frac{\partial J_1}{\partial y_i}=2(y_i-\hat{y}_i)$,可以看出$x_i,y_i$的梯度并无牵连。当用IoU作为损失函数时,如果一个网格中包含物体中心,其IoU值应为1,当不包含物体中心时,其IoU值应为0,因此,将该输出情况看作0~1分布,引入交叉熵损失函数来对IoU进行优化,约束模型的输出分布与训练数据标签分布的一致性。设期望的输出分布为p,则在含有物体中心的网格,$p=1$,则交叉熵损失函数为:$J_2=-p\ln(IoU)-(1-p)\ln(1-IoU)=-\ln(IoU)$,对$J_2$求导数得:

$$\begin{cases}\frac{\partial J_2}{\partial x_i}=\frac{-1}{IoU(x_i,y_i,\hat{x}_i,\hat{y}_i)}\frac{\partial IoU(x_i,y_i,\hat{x}_i,\hat{y}_i)}{\partial x_i}\\ \frac{\partial J_2}{\partial y_i}=\frac{-1}{IoU(x_i,y_i,\hat{x}_i,\hat{y}_i)}\frac{\partial IoU(x_i,y_i,\hat{x}_i,\hat{y}_i)}{\partial y_i}\\ \frac{\partial J_2}{\partial \widehat{x_i}}=\frac{-1}{IoU(x_i,y_i,\hat{x}_i,\hat{y}_i)}\frac{\partial IoU(x_i,y_i,\hat{x}_i,\hat{y}_i)}{\partial \widehat{x_i}}\\ \frac{\partial J_2}{\partial \widehat{y_i}}=\frac{-1}{IoU(x_i,y_i,\hat{x}_i,\hat{y}_i)}\frac{\partial IoU(x_i,y_i,\hat{x}_i,\hat{y}_i)}{\partial \hat{y}_i}\end{cases}\tag{2}$$

从式(2)可以看出,各个变量的梯度都是关于x_i,y_i,$\hat{x}_i$,$\hat{y}_i$的函数,即网络在更新参数时,是进行联动更新,而非独立优化四个位置参数,更具体的求导公式参考3.2节。

此外,由于IoU的取值范围为[0,1],自动地将任意尺度的输入数据标签进行了归一化处理。

3.2.1 IoU函数前向传播算法

前向传播算法如下所示。

前向传播算法(Forward)步骤如下:

输入:G表示训练样本中的标注框

P表示模型预测出的包围框

输出:L表示位置参数的损失

步骤1:对含有物体中心的网格进行如下计算:

$X=P_w\times P_h$

$\hat{X}=G_w\times G_h$

$I_h=\max(G_y-\frac{1}{2}\times G_h,P_y-\frac{1}{2}\times P_h)-\min(G_y+\frac{1}{2}\times G_h,P_y+\frac{1}{2}\times P_h)$

$I_w=\max(G_x-\frac{1}{2}\times G_w,P_x-\frac{1}{2}\times P_w)-\min(G_x+\frac{1}{2}\times G_w,P_x+\frac{1}{2}\times P_w)$

$I=I_h\times I_w$

$U=X+\hat{X}-I$

$IOU=\frac{I}{U}$

$L=-\ln(IOU)$

步骤2:对不含物体中心的网格:

$L=0$

其中,X表示预测出的包围框的面积;$\hat{X}$表示训练标注框的面积;G_x、G_y、P_x、P_y分别表示预测出的包围框和训练标注框的中心点坐标值;G_w、G_h、P_w、P_h分别表示预测出的包围框和训练标注框的宽和高;I表

示预测框与标注框的交集；U 表示预测框与标注框的并集；I_w、I_h 表示预测框和标注框交集部分的宽和高，参考图3。

3.2.2　IoU函数反向传播算法

为了更简洁地描述反向传播算法的计算公式，本文进行了相应的符号规定：$\nabla_p I$ 表示 I 对 P 中任意一个参数的偏导数，即 $\nabla_p I$ 为 $\nabla_{p_x} I$、$\nabla_{p_y} I$、$\nabla_{p_w} I$、$\nabla_{p_h} I$ 中任意一个；$\nabla_p X$ 表示 X 对 P 中任意一个参数的偏导数；且令：

$$G_y-\frac{1}{2}\times G_h=a_1,\ P_y-\frac{1}{2}\times P_h=b_1$$

$$G_y+\frac{1}{2}\times G_h=c_1,\ P_y+\frac{1}{2}\times P_h=d_1$$

$$G_x-\frac{1}{2}\times G_w=a_2,\ P_x-\frac{1}{2}\times P_w=b_2$$

$$G_x+\frac{1}{2}\times G_w=c_2,\ P_x+\frac{1}{2}\times P_w=d_2$$

则位置信息损失函数 L 对预测框的梯度为：

$$\frac{\partial L}{\partial P}=\frac{\partial(-\ln(IOU))}{\partial P}=\frac{-1}{IOU}*\frac{\partial(IOU)}{\partial P}=\frac{-1}{IOU}*\frac{U\nabla_P I-I(\frac{\partial(X+\hat{X}-I)}{\partial P})}{U^2}=\frac{I\nabla_P X-I\nabla_P I-U\nabla_P I}{U^2*IOU}\tag{3}$$

其中，

$$\nabla_p X=\begin{cases}\dfrac{\partial X}{\partial P_x}=0,\ \dfrac{\partial X}{\partial P_y}=0\\ \dfrac{\partial X}{\partial P_w}=P_h,\ \dfrac{\partial X}{\partial P_h}=P_w\end{cases}\tag{4}$$

$$\frac{\partial I}{\partial P_x}=\begin{cases}-1,\ a_2>b_2\text{且}c_2>d_2\\ 0,\ \text{其他}\\ 1,\ a_2<b_2\text{且}c_2<d_2\end{cases}\tag{5}$$

$$\frac{\partial I}{\partial P_y}=\begin{cases}-1,\ a_1>b_1\text{且}c_1>d_1\\ 0,\ \text{其他}\\ 1,\ a_1<b_1\text{且}c_1<d_1\end{cases}\tag{6}$$

$$\frac{\partial I}{\partial P_w}=\begin{cases}-1,\ a_2<b_2\text{且}c_2>d_2\\ -\dfrac{1}{2},\ a_2>b_2\text{且}c_2>d_2\text{或者}a_2<b_2\text{且}c_2<d_2\\ 0,\ \text{其他}\end{cases}\tag{7}$$

$$\frac{\partial I}{\partial P_h}=\begin{cases}-1,\ a_1<b_1\text{且}c_1>d_1\\ -\dfrac{1}{2},\ a_1>b_1\text{且}c_1>d_1\text{或者}a_1<b_1\text{且}c_1<d_1\\ 0,\ \text{其他}\end{cases}\tag{8}$$

式(3)~(8)即为随机梯度下降算法的深度网络位置参数的学习算法。

3.3　深度网络结构MIFD及整体损失函数IoU Loss

本文基于实时多目标回归模型YOLO，融合IoU函数，构建了本文的模型结构，如图7所示。

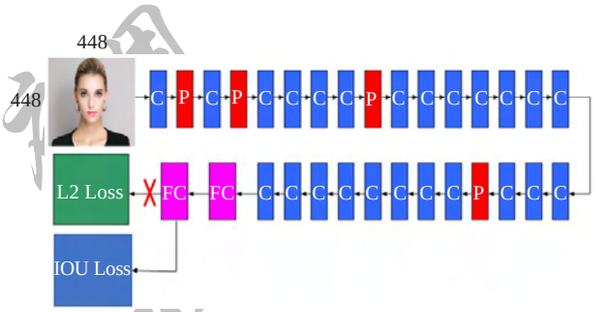


图7　本文提出的网络结构MIFD示意图
（省略了激活函数层、BN层）

IoU Loss定义如下：

$$IoU\ Loss=-\lambda_{\text{pos}}\sum_{i=1}^{49}\sum_{j=1}^{2}\theta_{ij}^{\text{obj}}\ln(IOU_{ij})+\sum_{i=1}^{49}\sum_{j=1}^{2}\theta_{ij}^{\text{obj}}(Conf_{ij}-\hat{Conf}_i)^2+\lambda_{\text{noobj}}\sum_{i=1}^{49}\sum_{j=1}^{2}\theta_{ij}^{\text{obj}}(Conf_{ij}-\hat{Conf}_i)^2+\sum_{i=1}^{49}\theta_i^{\text{obj}}\sum_{c=1}^{Class}(p_i(c)-\hat{p}_i(c))^2\tag{9}$$

其中，$\ln(IOU_{ij})$ 表示取第 i 个网格的第 j 个包围框与标注信息IoU的对数值，其余参数的含义与取值参考公式(1)中的相关说明。

4　实验结果与分析

4.1　模型训练策略

实验建立在64位的Linux操作系统和NVIDIA GTX Geforce 1080 GPU的服务器上，采用的深度学习框为caffe，下载地址为：https://github.com/BVLC/caffe，相关软件有Python2.7版本、Matlab2014b版本。

为验证本文提出的算法MIFD在图像中人脸检测的有效性，采用的数据集为香港中文大学公开的人脸检测基准数据集Wider Face[16]，有32 203张图片，共包含393 703张人脸，全部手工标注，标注的人脸有较大程度的尺寸、姿态和遮挡等变化。另一个数据集为马萨诸塞大学计算机系维护的一套公开数据库FDDB[4]，共有2 845张图片包含5 171张人脸，涵盖了在自然环境下的各种姿态的人脸。Wider Face分为2个部分，分别用于训练集、验证集，FDDB为测试集。

为了方便本文提出的算法MIFD与YOLO的算法对比，训练两个模型时，采取了相同的训练数据与训练策略，图片均划分为11×11个网格。将每张训练图片随机截取面积不小于图片面积70%，舍弃残缺的标注框，对保留下来的框的坐标进行相应的变换，然后截取的区域缩放到448×448，作为数据增强的手段，来减小过拟合。初始学习率(learning rate)设置为1×10^{-5}，每个批次的图片数量(batch size)为32，网络从YOLO的原始模型获得初始权重，采用随机优化算法Adam[17]进行网

络训练。

4.2　实验结果与分析

4.2.1　MIFD与YOLO的性能对比

图8为本文算法MIFD和YOLO算法的人脸检测效果图,可以看到:YOLO采用了L2 Loss,在面对不同尺度的人脸时,本文提出的MIFD更具鲁棒性;L2 Loss不采用位置参数联合优化的方式,虽然能够将某些尺度下的人脸框住,但是框住人脸的准确程度却不如IOU Loss。

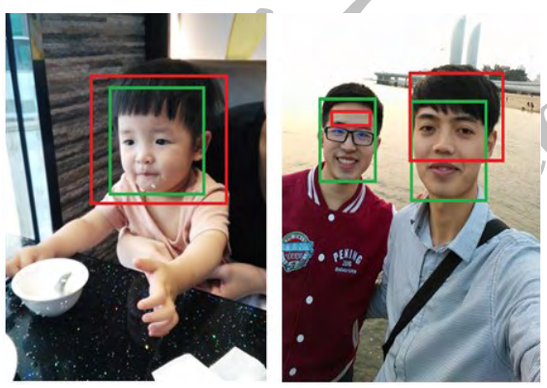


图8　检测结果图

(红色为YOLO,绿色为本文提出的MIFD)

为进一步比较本文算法与基础算法的性能,本文将两个模型在人脸数据库FDDB上进行测试,绘制ROC曲线(图9),并给出模型训练时Loss的收敛情况(图10)。

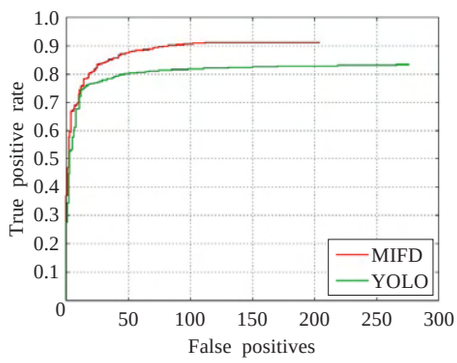


图9　MIFD与YOLO的ROC曲线对比

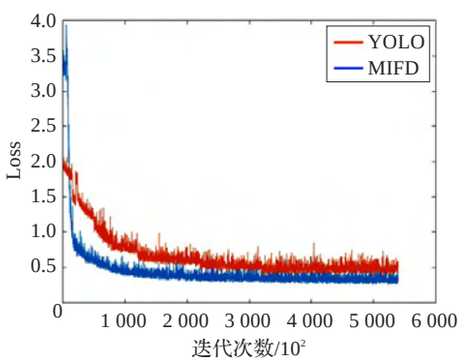


图10　训练情况对比

如图9所示,横轴表示图像中非人脸区域被误检为人脸的数量,YOLO的误检数量为275,MIFD的误检数量为205,误检率降低了(275 – 205)/275=24.5%;纵轴表示人脸区域被正确检出的比例,YOLO为82.5%,MIFD达到91.2%,准确率提高了91.2% – 82.5%=8.7%;如图10所示,横坐标代表训练迭代次数,纵坐标代表训练过程中的Loss值,可以看出,MIFD最终的loss比YOLO模型的更小,且更为稳定,充分说明了本文算法在加快训练收敛的有效性。

4.2.2　MIFD与传统主流人脸检测算法Adaboost的对比

基于Opencv库的Adaboost分类器,在CPU模式下进行对比分析。从FDDB人脸库中随机抽取1 000张图片,共含有1 605张人脸,进行算法性能比较。

从表1可以得出,本文提出的方法相比于传统的人脸检测算法Adaboost,检测精度上:准确率提升了7.9%,漏检率减少了7.88%,误检率减少了30.5%;检测速度上:Adaboost则具有明显的优势,MIFD可通过GPU加速来弥补检测速度上的不足。

表1　MIFD与Adaboost算法性能比较

算法	正确检测率/%	漏检率/%	误检率/%	检测速度/(s·张$^{-1}$)
Adaboost	84.1	15.90	34.1	0.95
MIFD	92.0	8.02	3.6	1.50

4.2.3　MIFD与其他深度学习方法对比

这部分对比包括检测精度与检测速度,选取了其他四种经典的人脸检测深度学习算法进行检测精度上的比较,分别是CascadeCNN[18]、Boosted Exemplar[19]、PEP-Adapt[20]、Faster-RCNN[14],将这些算法在FDDB数据集上进行评估,绘制ROC曲线,在1080 GPU的服务器上进行测试,结果如图11所示。

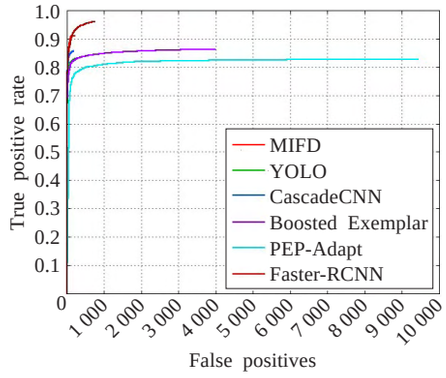


图11　各算法检测效果比较

选取了MIFD、YOLO、Faster RCNN进行了算法检测速度的比较,结果如表2所示。

表2　模型检测速度对比

模型	检测速度/(f·s^{-1})	最大准确率/%
MIFD	38.0	91.2
YOLO	33.8	82.5
Faster RCNN+VGG16	9.2	96.1[14]

通过图11可以得出，本文提出的算法MIFD在误检数量上低于其他深度学习方法；在检测准确率上不如Faster RCNN方法，但与其他深度学习方法相比，仍然具有明显优势。根据表2可以得出，MIFD的检测速度达到38 f/s，能达到实时检测人脸的目的，速度是Faster RCNN的4.13倍，在检测速度上具有十分明显的优势。因此，本文算法MIFD在检测精度与检测速度上取得了一个很好的权衡。

5　结束语

构建实用的人脸检测相关的应用系统，需要解决自然环境下的各种姿态、不同尺度人脸的检测准确性、鲁棒性问题，同时在检测速度上必须达到一定的要求。本文基于深度卷积神经网络自动学习特征的优越特性，且基于实时多目标回归思想，使得提出的算法满足了检测实时性的要求；同时，分析了实时多目标回归算法割裂了位置参数之间的关系，造成了模型的检测效果不够理想的问题，针对存在的缺陷，引入了IOU函数，把位置参数变量融合为一个整体进行优化，克服了该缺陷，提升了检测效果。实验结果表明：提出的算法在人脸检测的精度以及检测速度上取得了一个较好的平衡，检测精度上优于传统主流的Adaboost算法，检测速度也能够达到实时性，该算法可用于出入口人证比对、视频监控分析等人脸相关的视觉系统。

参考文献：

[1] Lowe D G.Distinctive image features from scale invariant keypoints[J].International Journal of Computer Vision，2004，60(2)：91-110.

[2] Dalal N，Triggs B.Histograms of oriented gradients forhuman detection[C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition.Los Alamitos：IEEE Computer SocietyPress，2005：886-893.

[3] 曾鸿军，沈燕飞，王毅.基于感兴趣区域的头像视频前处理方法[J].计算机工程与应用，2017，53(6)：188-192.

[4] Jain V，Learned-Miller E.FDDB：A benchmark for face-detection in unconstrained settings[R].UMass Amherst-Technical Report，2010：222-231.

[5] 张国云，向灿群，罗百通，等.一种改进的人脸识别CNN结构研究[J].计算机工程与应用，2017，53(17)：180-185.

[6] Uijlings J R R，Sande K E A V D，Gevers T，et al.Selective search for object recognition[J].International Journal of Computer Vision，2013，104(2)：154-171.

[7] Zitnick C L，Dollar P.Edge boxes：Locating object proposals from edges[C]//European Conference on Computer Vision，2014：162-172.

[8] Zhang Xiang，Sermanet P.Overfeat：Integrated recognition，localization and detection using convolution networks[C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition.Columbus：IEEE Computer Society Press，2014：651-667.

[9] Girshick R，Donahue J，Darrell T，et al.Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition，2014：580-587.

[10] 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，2015，39(6)：1137-1148.

[11] Redmon J，Divvala S，Girshick R，et al.You only look once：Unified，real-time object detection[C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition.Las Vegas：IEEE Computer Society Press，2016：779-788.

[12] Jiang Y，Jiang Y，Cao Z，et al.UnitBox：An advanced object detection network[C]//ACM on Multimedia Conference，2016：516-520.

[13] Ioffe S，Szegedy C.Batch normalization：Accelerating deep network training by reducing internal covariate shift[J].Computer Science，2015，70(2)：23-35.

[14] Jiang H，Learned-Miller E.Face detection with the Faster RCNN[EB/OL].[2016-07-10].http://arxiv.org/abs/1606.03473.

[15] Huang L，Yang Y，Deng Y，et al.DenseBox：Unifying-landmark localization with end to end object detection[J].Computer Science，2015，26(3)：254-267.

[16] Yang S，Luo P，Chen C L，et al.WIDER FACE：A face detection benchmark[C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition，2016：5525-5533.

[17] Kingma D P，Ba J.Adam：A method for stochastic optimization[J].Computer Science，2014，32(3)：111-125.

[18] Li H，Lin Z，Shen X，et al.A convolutional neural network cascade for face detection[C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition，2015：5325-5334.

[19] Li H，Lin Z，Brandt J，et al.Efficient boosted exemplar based face detection[C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition，2014：1843-1850.

[20] Li H，Hua G，Lin Z，et al.Probabilistic elastic part model for unsupervised face detector adaptation[C]//Proceedings of 2014 IEEE International Conference on Computer Vision，2014：793-800.