ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 37,No. 1,Jan. 2022,pp. 82－93
DOI:10. 16337/j. 1004-9037. 2022. 01. 007

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

MEL-YOLO:多任务人眼属性识别及关键点定位网络

吴东亮，沈文忠，刘林嵩

（上海电力大学电子与信息工程学院，上海 201306）

摘　要：针对当前人眼定位相关算法任务单一、且在多种干扰因素影响下（如光照、眼镜、遮挡）性能下降的问题，提出了可同时检测人眼感兴趣区域、识别人眼多种属性及定位关键点的轻量型神经网络MEL-YOLO。将YOLOV3算法与改进的DS-sandglass模块结合，在关键点回归分支应用去归一化的编解码方法提高网络定位宽度，并且在损失函数引入完全交并比（Complete intersection-over-union，CIoU）和均方误差（Mean square error，MSE），使得网络整体性能提升。MEL-YOLO算法在近红外虹膜数据集上人眼检测准确率为100%；属性识别和关键点定位准确率分别为98.7%和96.5%，在可见光数据集UBIRIS上分别达到92%和91%。实验结果证明：MEL-YOLO能同时实现人眼检测、属性识别及关键点定位，且准确率高、模型较小、泛化能力强，能够适用于低性能的边缘计算设备。

关键词：人眼定位；属性识别；卷积神经网络；关键点定位；轻量型算法

中图分类号：TP301.6　　**文献标志码：**A

MEL-YOLO:Multi-task Human Eye Attribute Recognition and Key Point Location Network

WU Dongliang，SHEN Wenzhong，LIU Linsong

(College of Electronics and Information Engineering，Shanghai University of Electric Power，Shanghai 201306，China)

Abstract: The existing eye location algorithms have some disadvantages of single task and performance degrade in complex environment such as illumination，glasses and occlusion，so a multi- efficient，light-YOLO and lightweight neural network，MEL-YOLO，is designed for obtaining eye multi-attributes and landmarks. Based on the YOLOV3 network，combining with the enhanced DS-sandglass block，a denormalized coding and encoding method is used in the regression branch of key points to promote the network positioning depth，and the complete intersection-over-union（CIoU）and the mean square error（MSE）are introduced into the loss function，so promoting the overall performance of the network. On the near-infrared dataset，the MEL-YOLO network achieves the position accuracy of 100%，and achieves the attribute recognition rate and the landmark accuracy rate of 98.7% and 96.5%，while reaches 92% and 91% on the UBIRS dataset. The experimental results demonstrate that the MEL-YOLO network can accurately obtain eye multi-attributes and key point information. Also，it is proved that MEL-YOLO is small and robust，and has the firm generalization ability，thus applying to low-performance edge computing devices.

Key words: human eye location; attribute recognition; convolutional neural network; eye landmark location; lightweight algorithm

收稿日期：2021-03-25；**修订日期：**2021-09-10

引　言

眼睛是人脸上最显著的特征,可以反映人的状态、情绪、视线方向及年龄等一系列重要信息,和虹膜识别、眼周识别和表情识别等应用有着密切联系[1]。准确识别眼睛属性(如位置信息、左右眼类别信息以及是否佩戴眼镜等)及定位关键点是实现上述应用的重要前提。图1展示了虹膜或眼周识别系统完整流程:图片预处理阶段是整体系统的基础阶段;定位人眼区域可以缩小虹膜精确定位范围;左右眼分类可以减少虹膜或眼周匹配次数(左、右眼分开匹配);判断是否佩戴眼镜可避免用户注册虹膜或眼周模板时因佩戴眼镜而影响采集图像质量的情况;关键点定位可实现虹膜轮廓初步定位,为精确定位等其他操作奠定基础。不同位置的关键点可分别用来拟合上、下眼睑、虹膜等轮廓,内外眼角与瞳孔中心点可用来计算当前视线角度。由此可见,人眼属性的准确识别及关键点的准确定位将直接影响虹膜或眼周识别体验。

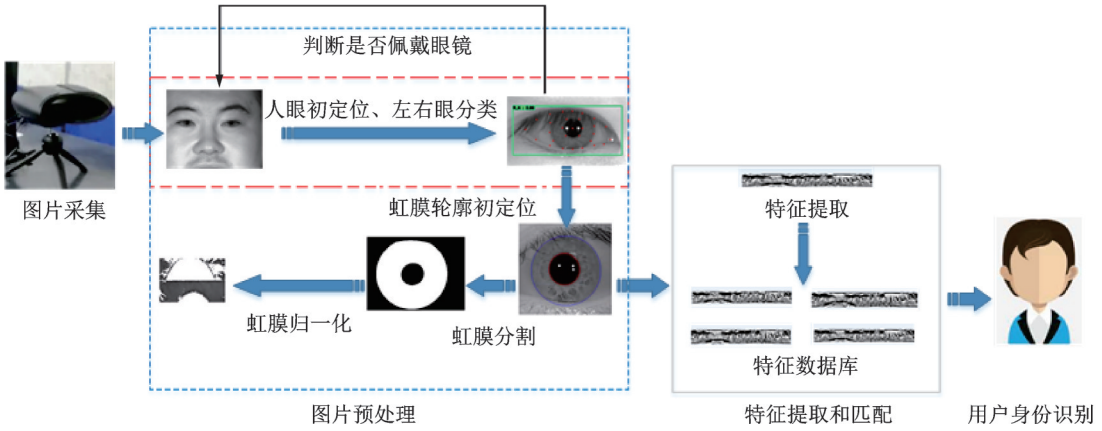


图1　虹膜(眼周)识别系统

Fig.1　System of iris (periocular) recognition

传统算法一般利用人眼区域中某一或某些特征信息(如固有特征、外观特征及结构信息等)实现单一任务,如人眼定位、关键点检测等。虽然在理想情况下能够取得不错的效果,但在实际应用场景中,由于脸部表情、姿态变化及光照等因素的干扰,鲁棒性差,测试效果不佳。文献[2]提出改进的SDM算法进行精确的人眼定位,该方法在每层回归提取不同尺度的尺度不变特征变换(Scale-invariant feature transform, SIFT)进行形状增量的学习,但是在多姿态下定位精确率不够理想。晁静静等[3]基于方向梯度直方图(Histogram of oriented gradient, HOG)和支持向量机(Support vector machine, SVM)进行人眼定位。该算法依靠灰度梯度变化进行人眼定位,而且不是端到端的方法,存在计算量大、复杂度高的问题,且算法中没有对虹膜图像进行左右眼分类,存在误检和准确率不高的情况。Xu等[4]使用主动形状模型(Active shape model,ASM)首先定位出79个面部关键点,然后根据其关键点位置信息定位出人眼区域,该方法是二阶段算法,实时性较差。

目前,基于卷积神经网络的目标检测算法经历了从双阶段到单阶段、从目标识别定位[5]到关键点检测[6]的发展过程,具有检测速度快、鲁棒性强等优点,并广泛应用于各个领域。越来越多的研究人员将该类算法应用到人眼定位和人脸关键点定位等任务中,并取得了显著的效果。这说明利用卷积神经网络提取人眼特征,进行人眼属性识别和关键点定位的方法是可行的。

陈金鑫等[7]提出了基于EL-YOLO的虹膜图像人眼定位及分类算法。该算法基于目标检测算法

YOLO,将轻量级网络MobileNetV3[8]引入模型,在不损失准确率的前提下,大大降低了模型的参数量和计算量,最终的准确率能达到99%左右。而且对于正负样本,网络都可以拥有很好的区分能力以及定位效果,但任务缺少关键点定位和是否戴眼镜检测结果,需要采用二阶段网络进行其他任务。文献[9]提出了一种弱监督的眼睛关键点检测算法,尝试将目标检测算法Faster R-CNN应用到眼睛关键点检测上,能够同时完成人眼定位和7个关键点定位;但该算法需要基于先前的预测对结果进行微调,影响测试速度。

本文结合了目标检测和关键点检测两种深度学习算法,提出了一种可以同时检测人眼多个属性及关键点位置信息的单阶段卷积神经网络模型MEL-YOLO。该算法不仅可以做到对多尺度图像进行多属性识别(人眼感兴趣区域、左右眼类别以及是否佩戴眼镜)和关键点定位,并在跨光谱图像上能完成同样任务,同时还解决了传统算法在不同数据集上表现不稳健的缺陷,展现出计算量低、检测速度快及泛化力强的优势。本文创新点包括:

(1)利用深度可分离(Depthwise separation, DS)卷积进行下采样,替换原有的激活函数并进行跨层连接,得到改进的DS-sandglass模块[10],并提出了单阶段多任务轻量型网络MEL-YOLO,在保证准确率的前提下降低参数量与计算量,提升在边缘设备的运行速度。

(2)针对关键点定位任务提出了全新的去归一化的编解码方法,新增关键点回归分支,可同时检测人眼区域、识别左右眼类别、判别是否佩戴眼镜及定位人眼中29个关键点。

(3)在边框定位损失函数和关键点损失函数中分别引入完全交并比(Complete intersection-over-union, CIoU)[11]和均方误差(Mean square error, MSE)损失函数,将均方根平均检测误差和定位成功率作为人眼关键点的评价指标,能有效判断关键点定位的优劣。

1 MEL-YOLO模型

YOLO[12]通过输出物体的中心点坐标、宽和高来确定物体的具体位置,即YOLO可以得到输入图片中某点的具体坐标,同理利用YOLO直接回归关键点位置信息的思路是可行的。另外,为更好地适用于嵌入式设备,通过实验本文最终确定输入图像为384×288×3,并对回归分支设计、编解码方法、网络结构和损失函数进行了深入研究,设计了可同时完成人眼多属性识别(人眼感兴趣区域检测、左右眼分类、是否戴眼镜)和关键点定位的单阶段多任务轻量级模型MEL-YOLO。

1.1 回归分支设计

相比一般的目标检测任务,本文在单阶段网络中增加了多属性识别和关键点定位任务。不同于直接在原来的回归分支上增加关键点信息,本文新增关键点分支来定位29个关键点,回归分支结构如图2所示。特征图大小不同,包含的语义信息的丰富程度也不相同,所以本文采用多尺度的特征图做检测(这里使用了2种尺度特征图)。另外为了提高小目标的检测精度,使用了特征金字塔网络(Feature pyramid network, FPN)[13]。经过特征融合的大尺寸特征图分支f_1和小尺寸特征图分支f_2分别进行不同的卷积操作可生成对应的边框(Box)分支和关键点(Landm)分支。在边框分支中,每一格点由4个边框信息、1个置信度信息和左眼戴(不戴)眼镜、右眼戴(不戴)眼镜4类类别信息组成;在关键点分支中,每一格点由58个关键点位置信息(对应29个关键点的横、纵坐标信息)组成。

此外,特征图f_1、f_2上的每一个格点会反映此区域是否存在检测目标。在目标格点上需要训练所有信息,而在背景格点上只需训练置信度,降低训练难度并减少了候选框生成。同时,在测试时引入非极大值抑制(Non-maximum suppression, NMS)算法[14],可以有效去掉重叠候选框,最终得到定位结果。

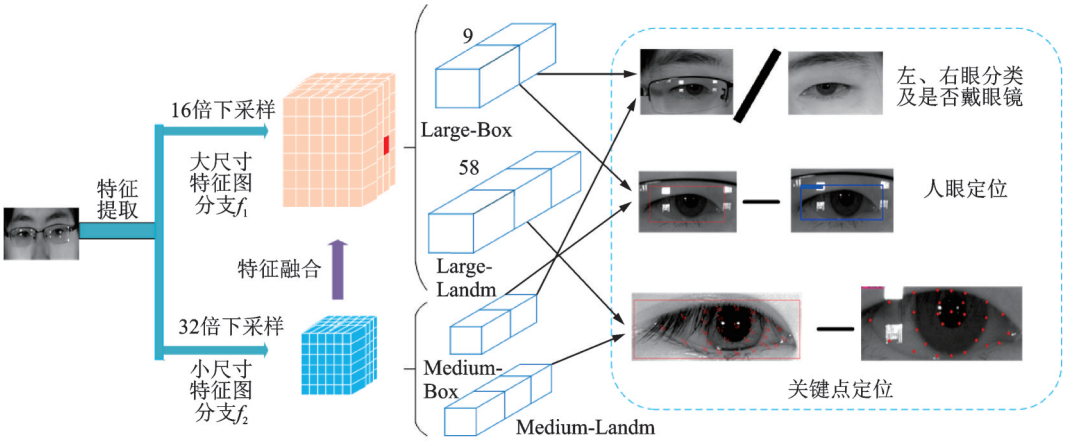


图2 MEL-YOLO网络输出结构

Fig.2 Output structure of MEL-YOLO Network

1.2 边界框和关键点预测

YOLO在训练中使用了锚点框,可以提高目标定位的速度,加快网络收敛。其中网络的真实预测值是边界框相对于先验框的转换值,在预测过程中使用以下解码方式:采用Sigmoid函数将物体中心距离网格左上角的偏移量t_x、t_y限制在[0,1)之间,预测的中心点因此落在对应网格中,预设锚点框的宽、高和以e为底数的边框宽高缩放比例相乘得到边框实际的宽、高。

在实际训练过程中,因无法保证所有的关键点和物体中心点落在同一格子里,故关键点的解码不能采用类似YOLO目标中心点的解码方式[12],所以本文针对关键点任务提出去归一化的编解码方法:去掉Sigmoid归一化,直接将网络预测值(l_x, l_y)作为关键点相对于锚点框中心的偏移,锚点框中心点加上对应的偏移量即预测关键点真正位置。如图3所示,其中虚线框是预设锚点框,以(c_x, c_y)为矩形框中心坐标,P_w和P_h分别为预设锚点框的宽与高,实线框是实际目标的边框,(t_x, t_y)是实际边框相对于锚点框的偏移。所有的关键点解码公式都是相同的,这里选用其中一点解释解码过程。

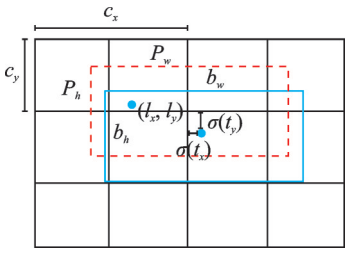


图3 先验框示意图

Fig.3 Diagram of bounding box

预测的边框信息和关键点信息表达式为

$$b_x = \sigma(t_x) + c_x,\ b_y = \sigma(t_y) + c_y \tag{1}$$

$$b_w = p_w e^{t_w},\ b_h = p_h e^{t_h} \tag{2}$$

$$p_x = P_w \times (l_x \times \partial) + c_x + 0.5,\ p_y = P_h \times (l_y \times \partial) + c_y + 0.5 \tag{3}$$

式中:σ为Sigmoid函数;(b_x, b_y, b_w, b_h)为实际目标边框中心点和宽、高;(t_w, t_h)为实际目标边框相对于预设锚点的宽高缩放比例;(p_x, p_y)为预测的关键点真实坐标;∂为超参数,这里取0.1。

1.3 网络结构

MobileNetXt[10]首次使用了Sandglass模块,并通过实验证明了其优越的性能,而MEL-YOLO采用包含下采样操作的改进残差块DS-sandglass,它们的结构分别如图4(a, b)所示。不同于MobileNetXt[10]通过设置Sandglass第2个3×3深度(Depth wise, DW)卷积的步长实现下采样操作,改进的DS-sandglass模块利用第1个3×3深度可分离卷积(步长为2)实现下采样操作,并将第2个逐点(Point

wise, PW)卷积的步长设为1,从而减少特征提取过程中信息丢失;另外DS-sandglass中同样进行跨层连接,从而增强残差结构中梯度跨层传播的能力。

与Sandglass模块相比,DS-sandglass模块将激活函数替换为Mish[15]函数。Mish函数比ReLU系函数曲线更加平滑,对负值的容许度更高,允许更好的信息进入神经网络,从而提高网络的准确性和泛化性。

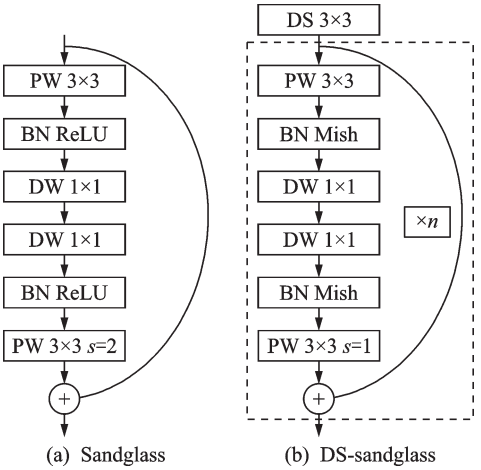


图4 Sandglass模块和DS-sandglass模块

Fig.4 Sandglass block and DS-sandglass block

MEL-YOLO采用和DarkNet53类似的特征提取+下采样和残差结构的网络设计,整体结构如图5所示,其中虚线框内为主干特征提取网络,具体细节如表1所示。为保证不丢失过多的眼睛纹理信息,MEL-YOLO第1个模块(图5中Conv-sandglass)使用了具有小感受野的3×3卷积核,移动步长从1个像素增加到2个像素,即DS-sandglass的DS替换为3×3普通卷积。同时为尽可能保证24×18特征图信息不丢失,对网络最后输出特征图(尺寸为12×9)采用普通卷积和上采样的方法,将调整后的特征图和24×18特征图直接相加;网络末端采用普通卷积操作,得到最终的回归分支。其中网络末端2个Box分支输出的通道数都为9,其中前1~4通道表征人眼边框信息,第5通道表征定位人眼的置信度,第6~9通道表征属性信息(左右眼及是否戴眼镜),2个Landm分支输出的通道数为58,表征单眼29个关键点的位置信息。

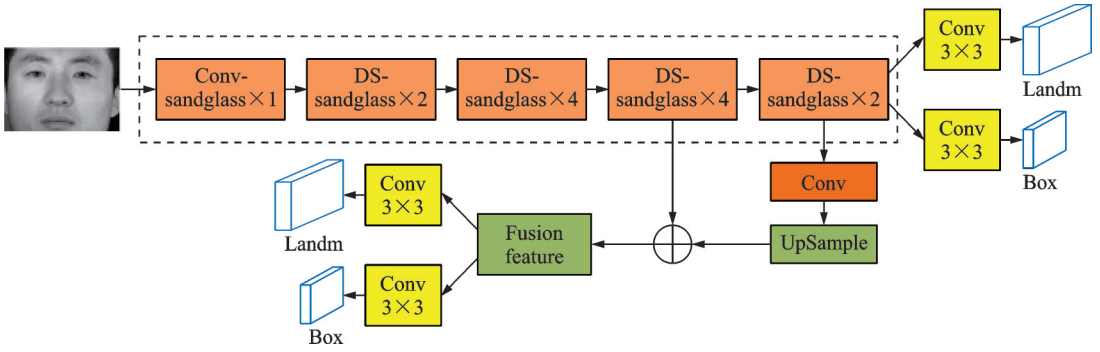


图5 MEL-YOLO网络整体结构图

Fig.5 Overall structure diagram of MEL-YOLO network

表1 特征提取网络结构

Table 1 Network structure of feature extraction

输入尺寸	模块×t	通道降维倍数	输出尺寸
384×288×3	Conv-sandglass ×1	4	192×144×16
192×144×16	DS-sandglass ×2	6	96×72×32
96×72×32	DS-sandglass ×4	4	48×36×64
48×36×64	DS-sandglass ×4	4	24×18×128
24×18×128	DS-sandglass ×2	4	12×9×256

注:表中"×t"是指DS-sandglass块重复次数。

1.4 损失函数设计

整体的损失函数的计算公式主要由4部分组成,即

$$L=\lambda_{\text{scale}}\sum_{i=0}^{w\times h}\sum_{j=0}^{n}I_{ij}^{\text{obj}}(1-\text{CIoU})+\lambda_{\text{allobj}}\sum_{i=0}^{w\times h}\sum_{j=0}^{n}[I_{ij}^{\text{obj}}\text{cross\_entropy}(\hat{c}_i,c_i)+I_{ij}^{\text{noobj}}\text{cross\_entropy}(\hat{c}_i,c_i)]+$$
$$\lambda_{\text{allobj}}\sum_{i=0}^{w\times h}\sum_{j=0}^{n}I_{ij}^{\text{obj}}\text{cross\_entropy}(\hat{c}_i,c_i)+\alpha\times\lambda_{\text{allobj}}\sum_{i=0}^{w\times h}\sum_{j=0}^{n}I_{ij}^{\text{obj}}\text{MSE}(\hat{c}_i,c_i) \tag{4}$$

式中:λ_{scale}取值在(1,2)之间,可弱化边界框尺寸对损失的影响,且和检测目标的大小成反比;$w\times h$表示预测特征图的尺寸;n表示每个尺寸的锚点数;I_{ij}^{obj}、I_{ij}^{noobj}分别代表特征图此处存在和不存在检测目标;λ_{allobj}为衡量预测整体结构与标签之间的距离情况,本文选用L_2距离;$\text{cross_entropy}(c_i,\hat{c}_i)$和$\text{MSE}(c_i,\hat{c}_i)$分别代表交叉熵损失函数和均方误差损失函数,$c_i$和$\hat{c}_i$分别代表目标预测值和目标真实值;$\alpha$为超参数,用来调整关键点损失函数所占比例,本文经过实验,最终确定为0.01。另外除了置信度损失需要考虑无检测目标时的情况,其他损失函数都只需要考虑检测目标存在的情况。

式(4)第1部分是边框交并比(Intersection-over-union,IoU)损失。为了解决IoU无法优化无重叠的边框问题,文献[7]在定位人眼上首次采用广义交并比(Generalized IoU, GIoU),取得不错效果。而CIoU[11]在GIoU的基础上还考虑了目标的重叠面积、中心点距离及长宽比等因素,使得目标框回归变得更加稳定,避免了IoU在训练过程中出现发散等问题。CIoU的表达式为

$$\begin{aligned}\text{CIoU}&=1-\text{IoU}+\frac{\rho^2(b,b^{\text{gt}})}{C^2}+\alpha\nu\\ \alpha&=\frac{\nu}{(1-\text{IoU})+\nu}\\ \nu&=\frac{4}{\pi^2}\left(\arctan\frac{w^{\text{gt}}}{h^{\text{gt}}}-\arctan\frac{w}{h}\right)^2\end{aligned} \tag{5}$$

式中:b和b^{gt}分别表示预测框和真实框的中心;ρ为欧式距离;C^2指刚好能包含预测框和真实框的最小矩形框的对角线长度平方;α代表权重函数;ν用来度量长宽比的相似性。

式(4)第2部分和第3部分分别表示置信度损失和分类损失,其中交叉熵公式为

$$\text{cross\_entropy}=\hat{c}_i\log(c_i)+(1-\hat{c}_i)\log(1-c_i) \tag{6}$$

式中c_i和$\hat{c}_i$分别代表目标预测值和目标真实值,在不同的任务中指代不同的对象。

式(4)第4部分是关键点损失函数。一般关键点损失函数采用L_1、L_2函数。本文经过实验,最终选择了MSE损失函数,表达式为

$$\text{MSE}=\frac{1}{29}\sum_{i=0}^{28}[(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2] \tag{7}$$

式中(x_i,y_i)和$(\hat{x}_i,\hat{y}_i)$分别代表关键点预测值和关键点真实值。

2 实验评估

本节从实验细节(包括数据集准备、锚点框设置、数据增强等)、实验结果、消融实验、在可见光数据集上的表现以及和其他算法对比5个方面证明MEL-YOLO网络模型的性能优势。

2.1 实验细节

(1)数据集

本实验选用的数据集(包括公开数据集和实验室自采数据集)详细信息及训练集、测试集分配情况如表2所示,其中右上角标*表示是公开数据集。这4个数据集图片都包含上半部分人脸或整张人脸,包含完整的人眼信息,可以作为本实验的数据集。

表2 数据集构成

Table 2 Dataset composition

数据集	分辨率/(像素×像素)	训练集人脸数	测试集人脸数
MIR2016*	1 968×1 024	400	60
CASIA-IrisV4-Distance*	2 352×1 728	2 000	500
CASIA-IrisV4单眼*与SEPAD_V1	640×480	1 600	1 032
SEPAD_V2	800×600	600	270
合计		7 000	2 362

本文重新标注数据集,标注需遵循以下准则:边界框必须包含完整的眼睛区域,包括内外眼角、眼睑;文献[16]选择眼角、上、下眼睑、虹膜轮廓和瞳孔轮廓在单眼图片上标注了28个关键点。本文在此基础上调整了标注顺序,并增加了瞳孔中心这一新关键点,编号规则如下:编号01、13、07分别对应内、外眼角点及瞳孔中心点;编号14、02、04、06、09、11、20对应上眼睑关键点;编号15、03、05、08、10、12、29对应下眼睑关键点;编号16、17、18、26、25、27对应瞳孔边缘关键点;编号20、19、21、22、23、24对应虹膜边缘关键点。本文选择了CASIA-IrisV4单眼数据库中图片编号为S5997L08的图片来展示标注细节,如图6所示,其中边框标签g代表检测对象佩戴眼镜。

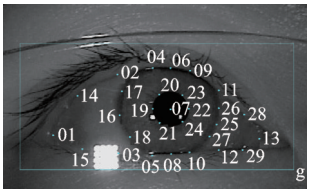


图6 眼睛标注细节信息

Fig.6 Eye label details

(2)锚点框参数设置

利用k-means聚类算法[17]确定边界框先验值,随机选择2个聚类和1个比例,聚类结果分别为(7.704×3.303)与(3.312×1.422)。

(3)数据增强

数据集图片在采集过程中容易受到光照、拍摄距离及被拍摄者眼部活动等因素的影响,图片质量参差不齐。为了提高模型的泛化能力和鲁棒性,本文选择了以下3种数据增强方法:随机亮度、对比度增强、随机裁剪和随机平移,且每种数据增强方法是否使用都是随机的(每种概率都设为0.5)。

(4)评价指标

为准确评价MEL-YOLO的性能,本文针对不同的任务选取了不同的评价指标。在属性识别任务上,采用了准确率(Accuracy)、精确率(Precision)与召回率(Recall)作为评价指标;在关键点定位任务上,文献[9]用均方根误差(Root mean square error, RMSE)和失败率(Failure rate)来评价预测值和真实值之间的关系。当RMSE大于一定阈值,说明该图上的关键点定位失败,否则定位成功,所以失败率是指关键点定位失败的图片数占总图片数的比例。参照文献[9]设置,本文将阈值同样设为8%。由于关键点数量过多,本文选取所有关键点和每类关键点的平均检测误差、RMSE、整体和每类的成功率(Success rate)作为评价指标,其中每类关键点成功率是指关键点定位成功的图片数占总图片数的比例,平均检测误差d_i和RMSE计算公式如下

$$d_i=\frac{1}{N}\sqrt{(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2} \tag{8}$$

$$\mathrm{RMSE}=\frac{1}{29}\sum_{i=1}^{29}d_i \tag{9}$$

式中:(x_i,y_i)和$(\hat{x}_i,\hat{y}_i)$分别代表关键点预测值和关键点真实值;N为归一化因子,一般$N=\sqrt{g_w*g_h}$,g_w

和g_h分别代表真实框的宽和高。通常在目标检测任务中,预测值和真实值的IoU达到0.5即可认为是真正例[18]。

(5)训练部署及训练策略

本实验所使用的硬件环境为:Inter(R) i7-8700K CPU,32 GB内存、Nvidia GTX 2080Ti GPU。使用Tensorflow深度学习框架对本文提出的MEL-YOLO网络进行训练和测试。网络输入分辨率为384×288×3,优化器选用Adam[19],使用余弦退火衰减法[20]调整学习率,初始和截止学习率分别为1×10^{-4}、1×10^{-6},每个阶段训练epoch都为50。

2.2 实验结果

2.2.1 数据集测试结果

根据2.1节可知,目标检测任务中一般认为当IoU大于0.5时定位成功,所以统计了IoU阈值为0.5时[18]定位及分类准确性,平均定位准确率达到98.7%。图7是MEL-YOLO测试的精确率-召回率曲线图,图例中L、R表示左、右眼,G、N表示有和没有戴眼镜,数值表示置信度。可以看出左、右眼不戴眼镜属性识别准确率分别达到99.8%及100%,戴眼镜的准确率分别为96.4%及96.5%。

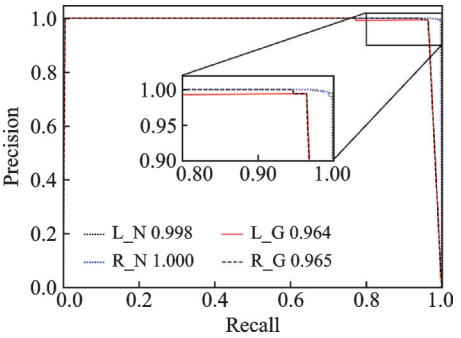


图7 精确率-召回率曲线

Fig.7 Precison-recall curve

为了与EL-YOLO进行对比(EL-YOLO模型不区分检测对象是否佩戴眼镜),本文单独统计了MEL-YOLO只进行人眼定位和分类时的性能,如表3第1,2行所示。可以发现:在相同任务和同一IoU时,MEL-YOLO的准确率均高于EL-YOLO。且MEL-YOLO在IoU为0.5~0.8时预测准确率都保持在99%以上,说明MEL-YOLO在人眼定位任务上效果更好、更稳定。表3第3行统计了MEL-YOLO网络模型区分是否佩戴眼镜的性能,当IoU为0.5~0.7时,准确率都能达到98%以上,在IoU=0.8时,准确率略微下降,为94.8%。但用来检测是否佩戴眼镜,IoU为0.5以上的检测结果已能满足实际应用。最后还统计了IoU分别为0.5~0.8的所有属性识别准确率,如表3第4行所示。由于MEL-YOLO分类要求更高(同时检测出左右眼状态和是否佩戴眼镜),MEL-YOLO在IOU=0.5时能达到98.7%,在0.8时预测准确率在96.1%。总体来说,MEL-YOLO多属性识别任务达到理想效果。

表3 不同IoU下区域测试准确率

Table 3 Testing accuracy under different IoUs

模型	任务			准确率/%			
	人眼定位	左右眼分类	区分是否佩戴眼镜	IoU=0.5	IoU=0.6	IoU=0.7	IoU=0.8
EL-YOLO[13]	√	√		99.9	99.9	98.6	84.1
MEL-YOLO(loc)	√	√		100.0	100.0	99.9	99.1
MEL-YOLO(glass)	√		√	98.5	98.5	98.4	94.8
MEL-YOLO(all)	√	√	√	98.7	98.7	97.2	96.1

整体和每类关键点定位情况如表4所示。由表4可见,关键点整体的平均检测误差为2.7%,定位成功率达到了96.5%。另外可以发现瞳孔中心的平均检测误差最小,定位成功率达到了99%;内、外眼角的误差最大,成功率只有90%左右。经过多次实验发现,所有关键点类定位准确度(误差越小,准确度越高)依次是:瞳孔中心>虹膜边缘>下眼睑边缘>上眼睑边缘>瞳孔边缘>外眼角>内眼角。

表4 整体和每类关键点定位情况

Table 4 Position about key points of overall and each category

类别	整体	外眼角	内眼角	上眼睑	下眼睑	瞳孔边缘	虹膜边缘	瞳孔中心
d_i/%	2.7	4.0	3.9	2.8	2.5	3.1	2.3	1.8
定位准确率/%	96.5	91.5	90.0	98.0	98.3	96.5	99.0	99.5

测试集中部分图片的眼睛定位及分类识别置信率如图8所示,分图题中L、R表示左、右眼,G、N表示有和没有戴眼镜,数值表示置信度。图8(a,b)是同一对象不佩戴眼镜和佩戴眼镜的2张双眼虹膜图像,可以看出MEL-YOLO可以准确定位出人眼区域、区分左右眼和是否佩戴眼镜,以及定位出29个关键点。图8(c,e,f)展示了单眼图片中待测人眼区域受到头发干扰、遮挡以及光斑影响的测试情况,可以发现定位分类的效果不受影响,关键点定位准确,表明MEL-YOLO不受图片大小、光照、遮挡等外界因素的限制,不依赖双眼图像的眼睛相对空间位置关系。图8(d)是数据集中关键点定位稍微不准的情况,从图中可以发现瞳孔区域过小,手工标注时不能准确标注,导致网络无法准确学习瞳孔特征。图8(g,h)是分类失败的2个例子,网络能够正确完成左右眼分类、但未能正确识别是否佩戴眼镜。这2张图片包含眼镜特征信息较少,增加了网络分类的难度,但其分类置信度较低,说明网络对这幅图像的判别结果并不肯定。

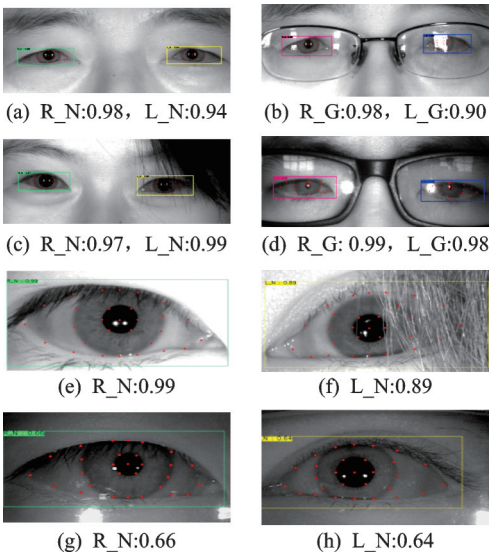


图8 MEL-YOLO网络测试效果

Fig.8 MEL-YOLO network test results

综上所述,MEL-YOLO可以克服眼镜、光照以及遮挡等干扰的影响,能同时完成人眼属性识别和关键点定位多种任务,效果已达到实际应用要求。

2.2.2 消融实验

为了验证MEL-YOLO模型的优势,本文在输入图片尺寸、编解码方法、网络结构、边框损失函数和关键点损失函数进行了相关实验。由式(4)可知,损失函数是一个整体,修改其中1个损失函数可能会影响其他的任务,因此本文还统计了关键点的定位成功率,实验结果如表5、6所示。

为了探究网络输入分辨率对性能的影响,在MEL-YOLO上分别测试了3种输入尺寸:384×288×1、384×288×3和416×416×3。实验结果如表5所示,可以发现输入大小为384×288×1效果最差,因为输入为灰度图像,损失了大量的信息,降低了关键点定位精度。还可以发现增大输入分辨率之后在增加参数量的基础上定位性能提升并不明显,甚至关键点定位成功率还下降了1.1%,所以本文最终选择384×288×3作为网络输入。

表5 不同输入尺寸的实验结果

Table 5 Results of different input sizes

序号	输入图像尺寸	准确率/%				成功率/%	参数量/MB
		IoU=0.5	IoU=0.6	IoU=0.7	IoU=0.8		
1	384×288×1	98.6	97.6	96.5	93.2	94.2	0.23
2	384×288×3	98.7	98.7	97.2	96.1	96.5	0.24
3	416×416×3	98.9	98.9	97.5	96.2	95.4	0.38

表6 不同编码方式的实验结果结果

Table 6 Results of different encoding methods

方法	准确率%				成功率/%	参数量/MB
	IoU=0.5	IoU=0.6	IoU=0.7	IoU=0.8		
YOLO	98.8	98.6	96.5	95.2	93.2	0.24
去归一化	98.7	98.7	97.2	96.1	96.5	0.24

为了证明本文提出的关键点编解码方法的有效性,图片输入大小固定为384×288×3,在MEL-YOLO网络的基础上分别使用YOLO中心点编解码方法和去归一化的编解码方法进行实验。实验结果如表6所示,可以发现使用本文方法在人眼定位和关键点定位上的性能优于使用YOLO中心点编解码方式,原因是使用去归一化的编解码方法保证网络偏移量范围为$(-\infty,+\infty)$,而不是$[0,1)$,即网络定位宽度进一步提高,可以回归任一关键点的位置。

为了验证MEL-YOLO网络结构的性能,本文在3种网络结构上进行了训练和测试,结果如表7所示,其中DarkNet 0.25[12]是指将DarkNet通道数缩小为原来的1/4。

表7 消融实验结果

Table 7 Results of ablation study

序号	网络结构	边框损失函数		关键点损失函数		准确率/%				成功率/%	参数量/MB	计算量/GB
		GIoU	CIoU	L_2	MSE	IoU=0.5	IoU=0.6	IoU=0.7	IoU=0.8			
1	DarkNet0.25	√			√	98.7	98.7	97.6	95.8	95.2	1.53	1.31
2	DarkNet0.25		√		√	99.2	98.9	98.1	96.8	96.4	1.53	1.31
3	MobileNetXt	√			√	99.3	99.3	97.6	96.7	95.3	2.16	1.32
4	MobileNetXt		√		√	99.6	99.5	98.7	97.2	97.8	2.16	1.32
5	MEL-YOLO	√			√	98.5	98.5	96.8	95.3	94.3	0.25	0.24
6	MEL-YOLO		√		√	98.7	98.7	97.2	96.1	96.5	0.25	0.24
7	MEL-YOLO		√	√		98.4	98.4	97.1	95.8	94.7	0.25	0.24

表7中实验1、3、5和实验2、4、6为2组对比实验,保证其他条件(边框损失分别是GIoU和CIoU、关键点损失是MSE)相同的情况下只替换了网络结构。从表7结果可以发现,MEL-YOLO在参数量和计算量上有明显优势,在大大降低模型参数的情况下,仍然可以达到、甚至超越其他2个模型的性能,说明MEL-YOLO是兼顾性能和准确性的轻量级模型,更适用于边缘设备。

表7中实验1和2,实验3和4,实验5和6为3组对照实验,在保证其他实验设置相同(关键点损失函数是MSE损失)的情况下只修改了边框损失函数,用来验证CIoU和GIoU损失函数在本任务上的性能。可以看出CIoU和GIoU在IoU取不同阈值的情况下边框准确率均能保持在99%以上,但使用CIoU时关键点定位上成功率更高,所以本文最终选择了CIoU损失函数。

最后,本文在MEL-YOLO和CIoU的基础上分别将MSE损失函数和L_2损失函数作为关键点损失函数,并对比实验结果。从表7中实验6、7可以发现,使用MSE损失函数时关键点定位成功率达到97%,而使用L_2损失函数之后成功率只有95%,所以本文最终选择了MSE损失函数。

2.3 算法效果及性能对比

本文选取的训练数据集MIR2016和CASIA-IrisV4等都是近红外图像,为验证MEL-YOLO模型在

跨光谱图片的性能，本文直接在UBIRIS.V1数据集[21]上进行属性识别及关键点定位，选择其中1 200张图片进行标注并测试，标注和部分测试结果分别如图9和图10(a,b)所示。可以看出虽然MEL-YOLO没有在可见光图片上进行训练，但MEL-YOLO依然能够定位人眼区域，进行左右眼识别以及关键点定位多项任务。最终MEL-YOLO在UBIRIS数据集上定位准确率达到了92%，关键点成功率达到了91%，即MEL-YOLO可以克服光谱影响，在训练中能够准确提取出人眼特征。

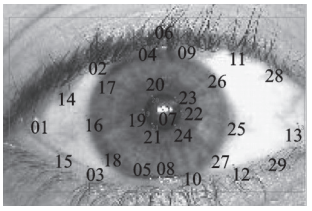
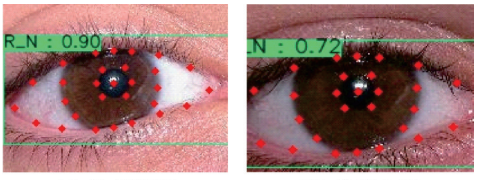


图9 UBIRIS标注细节信息

Fig.9 UBIRIS label details

MEL-YOLO网络和其他人眼相关算法性能如表8所示，可以发现本算法在人眼算法上优于其他算法，而且还可以准确区分是否佩戴眼镜以及关键点定位。其次，MEL-YOLO网络模型的浮点运算次数(FLOPs)为0.244 GB，参数量仅为0.246 MB，相比EL-YOLO算法，本文网络在增加了一定参数量的基础上，能够同时兼顾准确性和速度，且完成任务更多。总体来说，MEL-YOLO属轻量型的多任务模型，适合移植到嵌入式边缘计算设备上运行。



(a) R_N:0.90 (b) R_N:0.72

图10 UBIRIS数据集测试效果

Fig.10 UBIRIS dataset test results

表8 不同方法的计算量和性能

Table 8 FLOPs and peformance of different methods

方法	人眼定位准确率/%	左右眼分类准确率	戴眼镜分类准确性/%	关键点定位准确性/%	参数量/MB	计算量/GB
HOG和SVM[3]	99.9	—	—	—	—	—
主动外观模型[4]	99.0	—	—	—	—	—
EL-YOLO[13]	99.9	99.9	—	—	0.119 9	0.140 7
MEL-YOLO	100.0	100.0	98.5	96.5	0.246 3	0.244 4

3 结束语

为了准确识别人眼属性(人眼感兴趣区域、左右眼类别、是否戴眼镜)和定位关键点，本文提出了一种新的网络MEL-YOLO。该网络在YOLOV3的基础上结合了改进的Sandglass模块，针对关键点任务提出去归一化的编解码方法并引入了新的损失函数CIoU和MSE。与其他人眼定位网络相比，MEL-YOLO实现任务更多，且在各种评价指标上都表现突出。该模型的参数量为0.246 3 MB，计算量为0.244 4 GB，使得模型拥有在边缘计算设备上运行的能力。实验结果表明，MEL-YOLO网络在人眼属性识别的平均准确率98.7%，其中不戴眼镜属性识别准确率100%，戴眼镜属性识别准确率为94.6%；关键点定位准确率达到96.5%，同时在可见光图片UBIRIS数据集上属性识别和关键点定位准确率分别达到92%和91%，可以在多光谱环境下准确地识别人眼属性并完成关键点定位任务。本文模型为疲劳检测、虹膜识别、眼周识别和视线估计等研究奠定了较好的基础，具有较高的实用价值。后续工作在继续优化网络模型的基础上增加样本的多样性，提高关键点尤其是内、外眼角点的准确率，同时在网络中增加预测年龄、性别、虹膜纹理类别等更多属性，提升网络的应用范围和价值。

参考文献：

[1] 黄洁媛. 基于CNN的人眼定位与状态分类[D]. 北京:北京交通大学, 2019.
HUANG Jieyuan. Eye location and state classification based on CNN[D].Beijing: Beijing Jiaotong University, 2019.

[2] ZHOU M, WANG X, WANG H, et al. Precise eye localization with improved SDM[C]//Proceedings of Image Processing (ICIP), 2015 IEEE International Conference on. [S.l.]: IEEE, 2015: 4466-4470.

[3] 晁静静,沈文忠,宋天舒.基于HOG和SVM的双眼虹膜图像的人眼定位算法[J].计算机工程与应用,2019,55(9):184-189.
CHAO Jingjing, SHEN Wenzhong, SONG Tianshu. Eye location algorithm of binocular iris image based on HOG and cascade SVM[J]. Computer Engineering and Applications, 2019, 55(9): 184-189.
[4] XU F, SAVVIDES M. Unconstrained periocular biometric acquisition and recognition using COTS PTZ camera for uncooperative and non-cooperative subjects[C]//Proceedings of 2012 IEEE Workshop on the Applications of Computer Vision (WACV). [S.l.]: IEEE, 2012: 201-208.
[5] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2018-09-16) [2021-03-10]. https://arxiv.org/pdf/1409.1556v6.pdf.
[6] ZHANG Z P, LUO P, CHEN C L, et al. Facial landmark detection by deep multi-task learning[C]//Proceedings of European Conference on Computer Vision. Cham: Springer, 2014: 94-108.
[7] 陈金鑫,沈文忠.基于EL-YOLO的虹膜图像人眼定位及分类算法[J/OL].计算机工程与应用.(2020-12-21)[2021-03-10]. http://kns.cnki.net/kcms/detail/11.2127.TP.20200820.1000.008.html.
CHEN Jinxin, SHEN Wenzhong. Human eye location and classification algorithm based on EL-YOLO[J/OL]. Computer Engineering and Application. (2020-12-21) [2021-03-10]. http://kns.cnki.net/kcms/detail/11.2127.TP.20200820.1000.008.html.
[8] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3[C]//Proceedings of the IEEE International Conference on Computer Vision. [S.l.]: IEEE, 2019: 1314-1324.
[9] HUANG Bin, CHEN Renwen, ZHOU Qinbang, et al. Eye landmarks detection via weakly supervised learning[J]. Pattern Recognition, 2020, 98: 107076.
[10] ZHOU Daquan, HOU Qibin, CHEN Yunpeng, et al. Rethinking bottleneck structure for efficient mobile network design[EB/OL]. (2020-11-07) [2021-03-10]. http://arXiv preprint arXiv:2007.02269.
[11] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]//Proceedings of AAAI Conference on Artificial Intelligence, [S.l.]: AAAI, 2020.
[12] REDMON J, FARHADI A. Yolov3: An incremental improvement[EB/OL]. (2018-04-08) [2021-03-10]. http://arXiv preprint arXiv:1804.02767.
[13] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2017: 2117-2125.
[14] NEUBECK A, GOOL L. Efficient non-maximum suppression[C]//Proceedings of 18th International Conference on Pattern Recognition (ICPR'06). [S.l.]: IEEE, 2006: 850-855.
[15] MISRA D. Mish: A self regularized non-monotonic neural activation function[EB/OL]. (2019-08-23) [2021-03-10]. http://arXiv preprint arXiv:1908.08681.
[16] 隋秀娟,薛雷,许翠单.基于约束局部模型的人眼特征点定位[J].工业控制计算机,2020,33(8):105-107.
SUI Xiujuan, XUE Lei, XU Cuidan. Location of eye feature points based on constrained local model[J]. Industrial Control Computer, 2020, 33(8): 105-107.
[17] WAGSTAFF K, CARDIE C, ROGERS S, et al. Constrained *k*-means clustering with background knowledge[C]//Proceedings of the Eighteenth International Conference on Machine Learning. [S.l.]: ACM, 2001: 577-584.
[18] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2014: 580-587.
[19] KINGMA D P, BA J. ADAM: A method for stochastic optimization[EB/OL]. (2017-01-30)[2021-03-10].http://arXiv preprint arXiv:1412.6980.
[20] LOSHCHILOV I, HUTTER F. SGDR: Stochastic gradient descent with warm restarts[EB/OL]. (2017-05-03) [2021-03-10]. http://arXiv preprint arXiv:1608.03983.
[21] PROENÇA H, ALEXANDRE L A. UBIRIS: A noisy iris image database[C]//Proceedings of Image Analysis and Processing. Berlin, Heidelberg: Springer, 2005: 970-977.

作者简介:

吴东亮(1997-),男,硕士研究生,研究方向:虹膜识别、计算机视觉,E-mail:3105223172@qq.com。

沈文忠(1978-),**通信作者**,男,副教授,研究方向:虹膜识别、机器视觉与智能控制,E-mail:7841423@qq.com。

刘林嵩(1997-),男,硕士研究生,研究方向:计算机视觉,E-mail:844678196@qq.com。

(编辑:张黄群)