基于多模型联合的身份证人脸验证应用研究

胡　强,张太红,赵昀杰,迪力夏提·多力昆
(新疆农业大学 计算机与信息工程学院,新疆 乌鲁木齐 830000)

摘　要:随着互联网技术的发展与普及,各应用软件以“井喷”之势进入人们的生活,而网络实名认证制度伴随着电子商务的发展也出现在大众视野中,在传统实名认证模块中,由于用户上传证件照片不完整、模糊和类别错误等问题,导致人脸验证准确率下降。该文通过对公民身份证的特点进行分析,基于 Mask R-CNN 算法对身份证及其关键信息进行像素级别的目标检测,对用户上传的证件进行质量判别,并根据判别结果联合 MTCNN 模型与 FaceNet 模型完成人脸的验证。最终在自建测试数据集上进行对比实验。实验结果表明,相比于原系统,使用多模型联合的身份证人脸验证方法,在不影响人脸识别精度的前提下,保证了上传证件的有效性和安全性,其中证件的查全率达到了 99.24%,整个系统的准确率为 95.20%。

关键词:目标检测;人脸验证;Mask R-CNN;身份证;多模型联合
中图分类号:TP391　　**文献标识码**:A　　**文章编号**:1673-629X(2021)08-0209-06
doi:10.3969/j.issn.1673-629X.2021.08.036

Research on Application of ID Face Authentication Based on Multi-model Union

HU Qiang,ZHANG Tai-hong,ZHAO Yun-jie,Dilixiati DUOLIKUN
(School of Computer and Information Engineering,Xinjiang Agricultural University,Urumqi 830000,China)

Abstract:With the development and popularization of Internet technology,various application software has entered people's life with the trend of "blowout". With the development of e-commerce,the network real-name authentication system also appears in the public view. In the traditional real-name authentication module,the accuracy of face verification is reduced due to the incomplete,fuzzy and wrong category of photos uploaded by users. Based on the analysis of the characteristics of citizen ID card,we use Mask R-CNN algorithm to detect the identity card and its key information on pixel level,and judge the quality of the user's uploaded ID card. According to the result of the discrimination,the MTCNN model and FaceNet model are combined to complete the face verification. Finally,a comparative experiment was carried out on the self built test data set. The experiment shows that compared with the original system,the ID card face verification method based on multi-model combination ensures the validity and security of the uploaded ID card without affecting the accuracy of face recognition. The recall rate of the ID card reaches 99.24%,and the accuracy rate of the whole system is 95.20%.

Key words:object detection;face verification;Mask R-CNN;ID card;multi-model association

0　引　言

近年来,由于机器学习的快速发展,同时伴随着人脸验证精度的提高,基于该项技术而衍生的产品在大众生活中呈现“井喷”之势,如日常生活中应用较广泛的手机人脸解锁、小区门禁、支付宝刷脸支付等。人脸的验证相对于声音、指纹、掌纹、虹膜等生物特征,具有以下优势:数据采集简单;普遍的可接受性;应用的便捷性;较好的安全性。并且人脸特征具有自然性、方便性和非接触性等优点,使其在人机交互、身份信息验证、地方安全监控等方面具有巨大的应用前景。现阶段大量 Web 应用中的证件人脸验证模块,主要通过简单的人脸检测和人脸识别来完成,忽视了证件的质量与相关性,如要求用户上传居民身份证及本人照片完成人脸实名认证,但部分用户并未按要求上传证件,其通过上传驾驶证、社保卡等其他证件来变相完成人脸验证。而融入基于目标检测的证件判别方法可有效解决该问题。

人脸验证最关键的两个步骤是人脸检测和人脸识

收稿日期:2020-09-14　　**修回日期**:2021-01-16
基金项目:自治区研究生科研创新项目(XJ2020G169)
作者简介:胡　强(1992-),男,硕士研究生,研究方向为农业信息化;通讯作者:张太红,教授,研究方向为人工智能。

别,随着深度学习在计算机视觉领域的快速发展,越来越多基于卷积神经网络的人脸验证方法被相继提出。其中效果显著的人脸检测算法有 Joint Cascade[1],Cascade CNN[2],MTCNN[3]。主流的人脸识别算法有:DeepID[4-5]系列、FaceNet[6];DeepID 利用卷积神经网络提取人脸特征,在 LFW 人脸数据集上识别率可以达到 97.45%,但受限于训练样本数量;在 DeepID 的基础上引入了验证信号的 DeepID2 表现更好,识别率达到了 99.15%;FaceNet 采用三元组损失函数对网络模型进行训练,其在 LFW 人脸数据集上的识别率可以高达 99.63%。典型的目标检测算法有:SPP-Net[7]、Fast-RCNN[8]、YOLO[9]、Mask R-CNN[10] 算法等,YOLO 算法的检测速度很快但精度不是很好,而 Mask R-CNN 检测精度相比于 YOLO 精度好了很多,也是由于 Mask R-CNN 良好的表现,众多学者将它应用在自己的研究中,如李琦结合双目视觉来测量牛体尺寸[11],喻立春用它识别火焰图像[12],姜红花应用该算法来检测玉米间的杂草[13],张泽堃用它完成服装识别与分割[14]等。该文使用 Mask R-CNN 算法对身份证及其关键信息进行检测,并判别证件质量,联合 MTCNN 模型与 FaceNet 模型完成人脸的验证,在不降低人脸识别精度的情况下,保证了上传证件的有效性和安全性。

1 多模型联合的人脸验证

新疆马产业科技创新平台共有马匹记、马场管理、马匹竞拍、天马赛事等几大模块,用户想要使用该平台均需要完成实名认证。该平台落地于新疆伊犁昭苏县,据统计昭苏县主要少数民族为哈萨克族、维吾尔族、蒙古族、柯尔克孜族、回族。占当地人口最大比重的哈萨克族牧民是养马业的主力军[15]。该文主要研究该平台的用户实名认证模块,将人脸验证流程分为:证件分类、证件质量检测、人脸检测、人脸验证。

首先使用 Mask R-CNN 网络进行身份证的判别和检测,训练及测试数据集为真实环境下拍摄的照片,通过交叉验证法[16]反复训练与测试,选择误差最小的模型。其次将 Mask R-CNN 网络勾画出的人像送入 MTCNN 中进行人脸检测及对齐,最终在检测到人脸的基础上使用 FaceNet 计算人脸的嵌入向量,通过计算人脸向量间的欧氏距离对人脸进行验证。整个验证流程如图 1 所示。

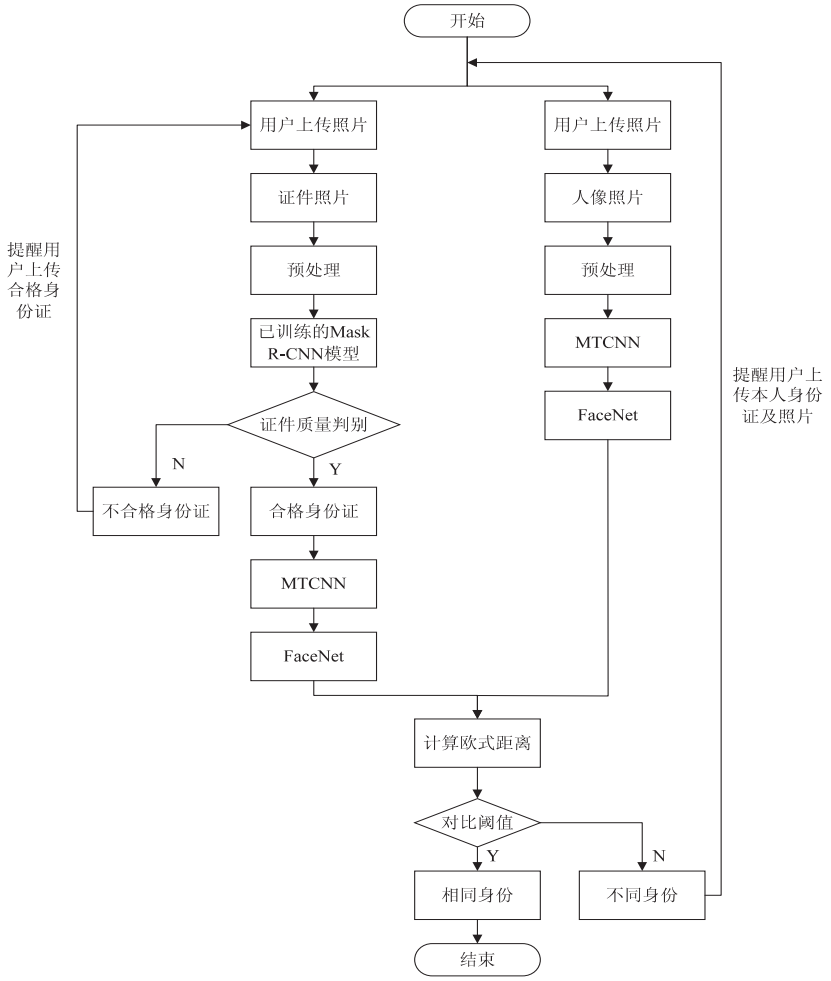

图 1 系统整体实现模型

1.1　基于目标检测的证件辨别方法

1.1.1　Mask R-CNN 框架

Mask R-CNN 是何凯明在 2017 年基于以往的 Faster-RCNN 构架提出的新卷积网络,该算法在 2016 年荣获微软大型图像数据集处理比赛的冠军,在有效的检测目标的同时完成了高质量的语义分割。图片输入后,首先通过骨干网络提取多尺度特征图,然后根据区域建议网络(region proposal network,RPN)选择出候选区域目标,接着使用 softmax 分类器区分前景目标和背景目标,同时使用边框回归器修正候选框位置,生成最终候选框。

1.1.2　证件质量验证

中国于 2004 年开始换发第二代身份证,到 2013 年 1 月 1 日二代身份证全部换发完毕。中国第二代身份证有壮文、维文、彝文、藏文、蒙文、朝鲜文六种民族文字与汉文字并列的双文版身份证。新疆马匹养殖户大多以少数民族为主,因考虑到不同民族的身份证存在差异,通过对大量不同民族身份证图像的观察,发现身份证上内容排版样例分为两种,一是汉文在下民族文在上,二是汉文在右蒙文在左。

标准身份证长 85.6 mm,宽 54 mm,由姓名、性别、民族、出生日期、住址、公民身份号码及本人相片 7 个登记项目组成,七个登记项位置固定,本实验对其中三个登记项(姓名、公民身份号码、本人相片)进行了标注,图 2 为三个登记项位置分布,结合图 2 所示与实际测量得到表 1 数据,W 为该长度在 AD 上的占比,H 为该长度 AB 在上的占比,测量误差小于 1 mm。由于用户拍照的角度或设备不同,提交上来的身份证照片大小均可能不同,所以采用长度占比进行判断。

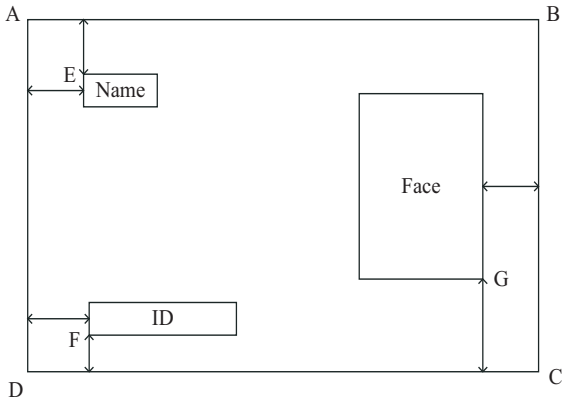

图 2　身份证结构示意图

表 1　身份证登记信息相对位置占比

身份证	名称	距离/mm	$W = x / AB$ /%	$H = y / AD$ /%
身份证的长	AD	54	—	—
身份证的宽	AB	85.6	—	—
非蒙古族身份证	E 到 AB	8.7	—	16.1
	E 到 AD	6	7	—
	F 到 AD	6	7	—
	F 到 CD	5.7	—	10.6
	G 到 BC	6	7	—
	G 到 CD	14	—	25.9
蒙古族身份证	E 到 AB	8.5	—	15.7
	E 到 AD	7	8.2	—
	F 到 AD	16	29.6	—
	F 到 CD	7	—	13
	G 到 BC	4	4.6	—
	G 到 CD	14.5	—	26.8

由 A、B、C、D 四点坐标根据 $y = kx + b$ 可以求得,AD、AB、CD、BC 四条直线,结合以下公式:

$$d = \frac{|Ax_0 + By_0 + C|}{\sqrt{A^2 + B^2}}$$

点到直线的距离为 d,分别可以求出表 1 中各点到直线的距离,以及它们在长、宽上的占比。从表 1 中可以看出,无论是非蒙古族身份证还是蒙古族身份证除居民身份号码的相对位置差异较大,其他关键信息位置基本没有什么变化。由此可以推断,若三个登记项存在且距离在阈值内,则认为该身份证照片质量合格,不在阈值内则认为证件质量不合格,系统便要求用户重新提交本人身份证照片。多次调整阈值区间并不断测试,筛选后的照片再次经人工验证基本合格,详细数据见表 2。

表 2 证件质量检测

证件	名称	W/%	H/%	测试数量	筛选后照片数	筛选通过率/%
身份证	E 到 AB	—	13-18	263	255	96.96
	E 到 AD	5-12	—			
	D 到 AD	5-32	—			
	F 到 CD	—	8-15			
	G 到 BC	3-10	—			
	G 到 CD	—	24-29			

1.2 人脸验证

1.2.1 人脸检测

该文选择 MTCNN 网络来进行人脸的检测和对齐,MTCNN 三个阶段使用的卷积神经网络结构分别是:P-Net,R-Net 和 O-Net。网络将送入的图片生成图像金字塔[3],第一阶段由 P-Net 获得候选窗体和边界回归向量,候选窗体通过边界框进行校正,并用非极大抑制算法筛选重叠窗体。第二阶段任务由 R-Net 完成,将经过 P-Net 确定的包含候选窗体的图片拿来训练,使用全连接网络进行分类,利用边界框向量微调候选窗体,再用非极大抑制算法去除重叠窗体。第三阶段使用 O-Net 卷积神经网络进行操作,O-Net 是 MTCNN 里最精细的网络结构,功能与 R-Net 类似,在去重后同时标定 5 个人脸关键点的位置,MTCNN 网络通过交叉熵损失函数来判断该区域是否包含人脸,函数如下:

$$L_i^{\det} = -(y_i^{\det}\log(p_i)) + (1 - y_i^{\det})(1 - \log(p_i))$$
$$y_i^{\det} \in \{0,1\}$$

其中,p_i 为人脸出现的概率,$y_i^{\det}$ 为该区域的真实标签。采用欧氏距离作为距离度量的回归损失函数,函数如下:

$$L_i^{\text{box}} = \| \hat{y}_i^{\text{box}} - y_i^{\text{box}} \|_2^2 \quad y_i^{\text{landmark}} \in R^4$$

其中,$\hat{y}$ 为通过网络预测得到的边框坐标,y 为实际的边框坐标。关键点定位选用的依旧是欧氏距离,函数如下:

$$L_i^{\text{landmark}} = \| \hat{y}_i^{\text{landmark}} - y_i^{\text{landmark}} \|_2^2 \quad y_i^{\text{landmark}} \in R^{10}$$

其中,$\hat{y}_i^{\text{landmark}}$ 为预测结果,y_i^{landmark} 为实际关键点位置。

1.2.2 人脸识别

该文选用 FaceNet 完成人脸的识别工作,其利用海量的人脸数据学习人脸嵌入层,使同一身份的 L_2 嵌入层距离小,不同身份的 L_2 嵌入层距离大。嵌入层可以表示为 $f(x) \in R^d$,它将图片 x 投影到 d 维的欧氏空间,FaceNet 的投影在 d 维超球面上,即 $\| f(x) \|_2 = 1$。FaceNet 将确保同一个人的 x_i^a 与 x_i^p 相近,与他人 x_i^n 相离,公式如下:

$$\| f(x_i^a) - f(x_i^p) \|_2^2 + \alpha < \| f(x_i^a - f(x_i^n)) \|_2^2$$

其中,$\forall (f(x_i^a), f(x_i^p), f(x_i^n)) \in T$,$\alpha$ 强制正样本和负样本存在间隔,上述等价于最小化下面的损失:$\sum_i^N [\| f(x_i^a) - f(x_i^p) \|_2^2 - \| f(x_i^a) - f(x_i^n) \|_2^2 + \alpha]_+$。这样的嵌入层通过大量的数据克服了光照、表情等影响。

网络架构选择 Inception[6] 类型网络,Inception 主要在没有高效的稀疏矩阵乘积设备的情况下,利用现有的密集计算组件来近似视觉卷积网络中的滤波级别上的局部稀疏结构。

2 数据材料

2.1 数据采集

本实验共采集 2 476 幅图像,为模拟马业平台真实需求,人像照片与身份证照片 1∶1 对应采集,878 组图像由用户各自的手机拍照所得。由于每个人所属环境、光照、拍照时间、拍照设备均可能不同,所以 878 组图像复杂多样。学生卡图像 240 张,采集方式与身份证一样,社保卡、驾驶证、银行卡、其他卡图像各 120 张且均是网络下载。将收集好的数据按每个类别 7∶3 的比例随机分配,作为身份证关键信息检测模型的训练集和测试集。

2.2 数据集标注

该文使用图像标注工具 LabelMe[17] 对数据进行标记,标注完成生成 json 文件,标注后的数据由 LabelMe 内的 json_to_dataset. py 批量生成 Mask R-CNN 需要的数据集。实验共对 1 118 张图片进行了标注,并送入模型进行学习。用测试集对模型进行检验,发现该模型未对学生卡、社保卡、驾驶证和其他卡的姓名、人脸信息进行标注,但将身份证上的关键信息进行了标注,并准确地勾画出目标框。

3 实验环境及结果

3.1 实验环境

由于马业信息平台使用 Django 框架进行开发,所以本实验选用 Python 语言进行编程,所用深度学习框架为谷歌 TensorFlow-gpu 1.14.0 版本,Keras2.1.5,CUDA10.0,操作系统为 Ubuntu 18.04.4。计算机

CPU 为 Inter® Xern(R) CPU E5-2620,GPU 为 TITAN V,内存 64 G。

3.2　实验结果

在训练 Mask R-CNN 模型时,采用 0.000 1 的初始学习率,批处理大小(Batch size)设置为 4,模型共计训练 50 个 epoch。在模型训练过程中记录各类别的 AP(average precision)值并计算 mAP(mean average precision),统计结果如图 3 所示。训练集与验证集损失函数值的变化曲线如图 4 所示。

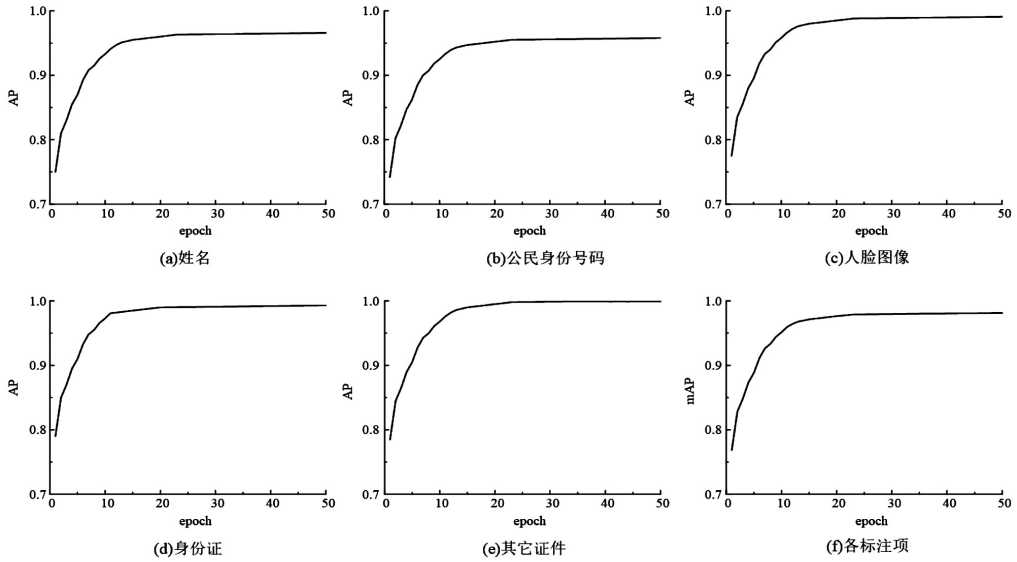

图 3　各标注项的 AP 与 mAP 值随 epoch 的变化曲线

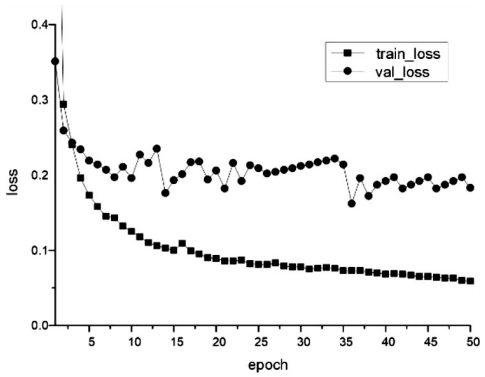

图 4　Mask R-CNN 训练损失变化曲线

观察图 3 可以看出,随着 epoch 的不断增长,各类别的 AP 值逐步提高并达到较高的稳定值,并且在图 4 中模型稳定收敛于较低的 Loss 值,说明了模型在训练集与验证集上有着较好的身份证关键信息检测能力。

基于关键信息检测模型对测试集中共计 479 张图片进行二分类,其中身份证 263 张,非身份证 216 张,分类结果如表 3 所示。

表 3　证件分类

类别	总数	判定为正	判定为负	查准率/%	查全率/%
身份证	263	261	2	100	99.24
非身份证	216	0	216		

最后将本实验方法与原系统方法进行对比,取 263 组人像与身份证对应的照片,对 263 张身份证照片中 50% 做随机的图像处理(模糊、随机遮挡),其中 72 人各增加一张学生卡照片,系统准确率=(清晰身份证通过数+其他证件未通过数)/总数量,将测试数据送入系统中得到表 4 实验结果。

表 4　实验结果对比

方法	数据名称	测试数量	系统通过数量	系统准确率/%
原系统方法	清晰身份证	132	132	74.32
	模糊身份证	131	51	
	学生卡	72	70	
	无关证件	144	2	
多模型联合方法	清晰身份证	132	132	95.20
	模糊身份证	131	23	
	学生卡	72	0	
	无关证件	144	0	

从表 4 中可知,原系统使用学生卡、社保卡或驾驶证也可以完成系统的人脸验证工作,而使用多模型联合判别的方法可以有效地过滤掉非身份证证件。对于经图像处理过的身份证,该方法提升不是十分明显,原因是随机遮挡没有遮住标注的关键信息该系统会认为身份证合格,总体来说多模型联合的人脸验证应用,在没有降低人脸识别率的基础上,完成了证件质量的筛选,更加符合实际应用的要求。

4　结束语

在新疆维吾尔自治区马产业科技创新平台建设

中,用户实名认证模块仍有部分功能需要完善。根据实名认证要求,用户需上传本人照片及身份证照片,来完成实名认证,但一些用户选择使用生活照或其他证件照来进行认证。原系统方法未对不合格照片进行审核,致使系统存在隐患。故该文选用 Mask R-CNN 网络来训练身份证判别模型,实验表明该模型可以很准确地检测到标注目标,且判别效果良好,训练过程中网络模型能够快速收敛,该模型具有较高的准确性。将模型应用于马产业科技创新平台中可以将非身份证和低质量证件照过滤,并完成人脸验证工作;目前系统完成的是 1 对 1 场景下的人脸验证工作,将来开展赛马赛事等活动,若要运用人脸签到、人脸门禁,要面临的便是 1 对 N 的人脸识别,而该应用可以建立优质人脸资源数据库,为将来的系列活动实现做基础。

参考文献:

[1] QIN H,YAN J,LI X,et al. Joint training of cascaded CNN for face detection[C]//2016 IEEE conference on computer vision and pattern recognition (CVPR). Las Vegas:IEEE,2016:3456-3465.

[2] CHEN D,REN S,WEI Y,et al. Joint cascade face detection and alignment[C]//Computer vision - ECCV 2014. Zurich,Switzerland:Springer,2014:109-122.

[3] 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.

[4] SUN Y,WANG X G,TANG X O. Deep learning face representation from predicting 10,000 classes[C]//IEEE conference on computer vision and pattern recognition. Columbus,OH,USA:IEEE,2014:1891-1898.

[5] SUN Y,WANG X G,TANG X O. Deeply learned face representations are sparse,selective,and robust[C]//IEEE conference on computer vision and pattern recognition. Boston,MA,USA:IEEE,2015:2892-2900.

[6] SCHROFF F,KALENICHENKO D,PHILBIN J. FaceNet:a unified embedding for face recognition and clustering[C]//IEEE conference on computer vision and pattern recognition. Boston,MA,USA:IEEE,2015:815-823.

[7] HE K,ZHANG X,REN S,et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2015,37(9):1904-1916.

[8] 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 & Machine Intelligence,2017,39(6):1137-1149.

[9] 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.

[10] HE K,GKIOXARI G,DOLLAR P,et al. Mask R-CNN[C]//International conference on computer vision. Venice:IEEE,2017:2980-2988.

[11] 李 琦,刘 伟,赵建敏. 基于双目视觉及 Mask RCNN 的牛体尺无接触测量[J]. 黑龙江畜牧兽医,2020(6):46-50;159-160.

[12] 喻丽春,刘金清. 基于改进 Mask R-CNN 的火焰图像识别算法[J]. 计算机工程与应用,2020,56(21):194-198.

[13] 姜红花,张传银,张 昭,等. 基于 Mask R-CNN 的玉米田间杂草检测方法[J]. 农业机械学报,2020,51(6):220-228.

[14] 张泽堃,张海波. 基于 Mask-RCNN 的服装识别与分割[J]. 纺织科技进展,2020(6):20-24.

[15] 崔浩然,杨浩宏,席琳乔,等. 新疆昭苏县定居牧户马匹饲养管理现状调查与分析[J]. 草食家畜,2014(4):21-26.

[16] 范永东. 模型选择中的交叉验证方法综述[D]. 太原:山西大学,2013.

[17] RUSSELL B C,TORRALBA A,MURPHY K P,et al. LabelMe:a database and web-based tool for image annotation[J]. International Journal of Computer Vision,2008,77(1-3):157-173.