

基于 YOLO 的手机外观缺陷视觉检测算法*

杨戈^{1,2}, 刘思瀚¹

(1. 北京师范大学珠海分校 智能多媒体技术重点实验室, 广东 珠海 519087;

2. 北京师范大学 自然科学高等研究院, 广东 珠海 519087)

摘要: 针对在小规模手机屏幕缺陷数据检测效率不高的问题, 提出了一个基于 YOLO v3 的手机外观缺陷视觉检测算法, 即 YOLO-q-v3 算法。YOLO-q-v3 算法通过改进 DarkNet-53 网络结构, 减少原网络的层数和相应的模型参数, 有效地提高算法的检测效率。将 YOLO v3 算法和本文提出的 YOLO-q-v3 算法对同一数据集进行对比实验, 实验结果表明, YOLO-q-v3 算法能够正确识别手机屏幕裂痕, YOLO-q-v3 在检测速率上比 YOLO v3 算法提高了 24%。

关键词: YOLO; 屏幕缺陷检测; 深度学习

中图分类号: TN183

文献标识码: A

DOI: 10.16157/j.issn.0258-7998.211888

中文引用格式: 杨戈, 刘思瀚. 基于 YOLO 的手机外观缺陷视觉检测算法[J]. 电子技术应用, 2022, 48(7): 81-85.

英文引用格式: Yang Ge, Liu Sihan. A visual detection algorithm for mobile phone appearance defect based on YOLO[J]. Application of Electronic Technique, 2022, 48(7): 81-85.

A visual detection algorithm for mobile phone appearance defect based on YOLO

Yang Ge^{1,2}, Liu Sihan¹

(1. Key Laboratory of Intelligent Multimedia Technology, Beijing Normal University, Zhuhai 519087, China;

2. Advanced Institute of Natural Sciences, Beijing Normal University, Zhuhai 519087, China)

Abstract: Aiming at the problem that some algorithms are not efficient in the detection of small-scale mobile phone screen defect data, this paper proposes a visual detection algorithm for mobile phone appearance defect based on YOLO v3, namely YOLO-q-v3 algorithm. The YOLO-q-v3 algorithm improves the DarkNet-53 network structure, reduces the number of layers of the original network and the corresponding model parameters, and effectively improves the detection efficiency of the algorithm. The YOLO v3 algorithm and the YOLO-q-v3 algorithm proposed in this paper are compared and tested on the same data set. The experimental results show that the YOLO-q-v3 algorithm can correctly identify the cracks on the mobile phone screen, and the detection rate of YOLO-q-v3 is 24% higher than that of the YOLO v3 algorithm.

Key words: YOLO; screen defect detection; deep learning

0 引言

手机是当前人们生活中的必需品, 如何高效并且尽可能低成本地制造手机成为了热点之一。而手机的屏幕缺陷检测则是属于手机制造过程中重要的一个环节, 如何在这个环节上化繁为简并且提高效率至关重要。传统的人工检测不仅效率低下, 并且会增加手机制造成本。还有一些传统的视觉处理方法, 通过提取图像特征再通过提取的特征去选择、识别。这种传统的办法虽然降低了人工成本, 但是受限于特征提取时的环境以及方式, 使得特征提取环节变得复杂, 所以效率也不会太高。因此, 寻求一种高效、可靠的基于深度学习的智能化的手

机屏幕缺陷检测算法是很有必要的。

1 机器视觉缺陷检测方法

1.1 传统的机器视觉缺陷检测方法

传统的机器视觉缺陷检测方法主要是研究图像的特征选取、特征的选择以及特征的识别这三个方面。在图像特征选取方面, 文献[1]的主要研究内容是图像特征的提取、特征的选择和图像的识别。在特征提取时有根据统计量来提取的, 也有根据信号域来提取的。文献[2]的方法是由获取图像、图像处理和图像反馈三部分组成的, 先用特殊的专业相机去获取清晰度高的图像, 再将获取的图像进行降噪的处理, 目的是降低图像的环

* 基金项目: 广东高校省级重大科研项目(2019KZDXM015, 2020ZDZX3058); 广东省学科建设专项资金(2013WYXM0122); 智能多媒体技术重点实验室(201762005); 校级教学团队(202012); 校级课程思政(201932); 2020年广东省教改项目(655); 2021年广东省课程教研室(104)

境噪声,尽可能地降低环境干扰,使得图像的信息得到增强。文献[3]提出基于图像处理技术的手划痕检测算法,该方法需要将图片进行预处理,降低图像的噪声,剔除环境噪声,增强图像的有用信息,将灰度进行变换从而提高图片的质量。

文献[4]设计了一套包含光学成像模块、图像采集模块以及软件算法的整体方案。为了提高图像增强的效果,改进了引导滤波,把改进的中值滤波后的图像作为引导图像,将引导滤波的高效边缘保持效果和改进的中值滤波的平滑作用相结合,去除低照度图像的噪声。文献[5]进行二值化和边缘检测确定手机屏幕区域的位置,其次采用几何校正使目标区域保持水平,利于进行目标提取操作,最后进行颜色空间转换提高屏幕缺陷与周围的对比度,以便于之后进行缺陷检测。

传统的机器视觉缺陷检测方法虽然可以对一些具有重复性的缺陷进行高效率的检测,但由于这种方法的检测效率对图像的特征选取和特征的选择这两方面有较高的依赖性,并且在实现中有很多的缺陷特征和类型都不同,使得该方法的检测效率并不高,无法作为生产中实时缺陷检测的技术^[6]。

1.2 基于深度学习的缺陷检测方法

基于深度学习的缺陷检测方法通过卷积计算,将图片的特征参数化,提高了算法的实用性和有效性,可以再通过优化参数从而来提高算法的精准度。文献[7]将现有的FCN(Fully Convolutional Network)进行改进,构建了一套适合裂纹检测的全卷积神经网络Crack FCN,该网络保留FCN网络的优点,输入的图片没有限制,任意尺寸的图片都可以作为网络的输入,该网络通过反卷积操作来还原最后一层卷积层的图片特征尺寸,提高了预测的精度。文献[8]构建了一种能够模拟视觉感知的深度学习网络结构,结合了神经网络非线性映射的能力,提出了一种网络结构自我生长的和参数自我学习的方法。通过模拟人类大脑深度学习的过程,可以使得可疑区域直接通过深度学习网络结构,再进行卷积网络的逐层学习提取可疑区域的特征,最后利用径向基网络实现图像的识别功能^[8]。但是该算法也存在相应的缺点,即在背景噪声太大和有效区域太过模糊的情况下会导致算法结果不稳定。

文献[9]提出CNN(Convolutional Neural Network)算法,CNN属于前馈神经网络,是一种包含了卷积计算的深度学习网络。CNN网络一般包括了卷积层、池化层和全连接层。卷积神经网络的主要模型包括LeNet模型、AlexNet模型、VGGNet模型和ResNet模型等^[10-12]。

文献[13]提出YOLO(You Only Look Once)系列算法,该系列算法属于端到端的深度学习神经网络,它将网络结构应用在整个图像上,将图像分成若干个网格,对网格的边界进行预测并预测网格中内容的所属类别,它最终将目标检测问题转化为了回归问题。不同于R-CNN

系列算法,它直接采用预测目标对象的边界框的方法,将区域识别和对象识别这两个阶段融合成一个阶段,可以在预测边界坐标的同时对目标进行分类。这就使得YOLO系列算法网络结构更加简洁,速度更快,准确率更高。YOLO算法的网络结构图包含了24个卷积层和2个全连接层,用卷积层网络来提取特征,用全连接层来得到预测值^[14-15]。

2 YOLO-q-v3 算法

YOLO-q-v3算法采用53层卷积层,其中卷积层由filters过滤器和激活函数构成,通过设定filter的数值,按照设置的步长不断地提取图片的特征,从局部特征再到总体的特征,从而完成图像的识别。卷积层的参数包括了卷积核的大小、步长和填充。卷积层里面有多个卷积核,卷积核的大小影响了选择区域的大小即感受野的大小,卷积核越大提取的图片特征越复杂。全连接层的作用是将经过卷积层和池化池处理后的数据串联起来,再对串联出的结果进行识别和分类。全连接层接受卷积层或者池化层的输入,并输出一个多维向量,这个多维向量代表了每一个类别的概率,即输出各个结果的预测值,然后取最大的概率作为全连接层识别分类的结果。

2.1 YOLO-q-v3 算法设计

本文提出YOLO-q-v3算法中用于提取特征的主要网络是卷积神经网络DarkNet-53。DarkNet-53提取了ResNet网络的优点,它在某些卷积层之间添加了一条快捷路径,这条路径和卷积层形成了残差组件。残差组件可以使得网络层数相应减少,参数减少,从而减少计算量,使得DarkNet-53的性能得到了较大提升。但是在本文运用的数据集中,图片缺陷数量不多,不需要检测多个对象,因此本文提出将原DarkNet-53网络在保持原本的检测精度的情况下进行删减网络层数,通过减少参数来提高算法的检测效率,算法相关参数如表1所示。

表1 算法参数对比

算法	卷积层	特征图	残差组件
YOLO v3	53	3	23
YOLO-q-v3	53	3	15

YOLO-q-v3算法还采用了FPN(Feature Pyramid Networks)的结构改善对较小物体的检测结果,每一个特征图被拿来做了边界框回归,其中尺寸最小的顶层特征用来检测较大的物体,而尺寸较大的特征图则用来检测较小的物体。如果是单一尺度检测器,尺度太大而数据集上特征过小就会导致生成的锚框对小特征检测效果不好。采用多尺度检测器就能避免这个缺点,不论在检测特征过小还是特征过大的图片时,都有适当尺寸的检测器去进行检测。该网络结构没有池化层和全连接层,是由卷积层构成了全卷积神经网络,这样使得算法减少了卷积层的运算量,从而提升了DarkNet-53网络的计算效

率。并且在预测对象类别时不再使用 softmax,而是改用 logistic 来进行预测,能够支持多标签的对象。

如图 1 所示,改进的 DarkNet-53 网络的主体卷积层与原网络相同,一共有 53 层卷积层,除去最后一个 1×1 的卷积层共有 52 个卷积作为 DarkNet-53 网络的网络卷积主体。DarkNet-53 网络提取了 ResNet 网络的优点,它在某些卷积层之间添加了一条快捷路径,这条路径和卷积层形成了残差组件,使得 DarkNet-53 网络结构层次

	类型	输出通道数	卷积核	输出特征图大小
	卷积	32	3×3	256×256
	卷积	64	3×3/2	128×128
1×残差块	卷积	32	1×1	128×128
	卷积	64	3×3	
	残差			
2×残差块	卷积	128	3×3/2	64×64
	卷积	64	1×1	
	卷积	128	3×3	
2×残差块	卷积	256	3×3/2	32×32
	卷积	128	1×1	
	卷积	256	3×3	
6×残差块	卷积	512	3×3/2	16×16
	卷积	256	1×1	
	卷积	512	3×3	
4×残差块	卷积	1024	3×3/2	8×8
	卷积	256	1×1	
	卷积	1024	3×3	
	平均池化		全局池化	
	全连接		1 000	
	Softmax			

图 1 改进 DarkNet-53 网络结构

变得更深。而本文提出的改进算法将原网络的残差组件数量从 1、2、8、8、4 个分别删减到 1、2、2、6、4 个。

2.2 算法实现

YOLO-q-v3 算法实现步骤如图 2 所示,首先通过特征提取网络 DarkNet-53 对输入的图像进行特征提取,得到不同尺度大小的特征图。然后对图像的特征图进行网格划分,划分出来的网格可以负责来预测落在该网格中的真实边界框中的目标。每个目标有固定数量的边界框,在 YOLO-q-v3 中有 3 个边界框,最后使用逻辑回归来确定预测的回归框。

2.2.1 检测框预测

如图 3 改进算法直接预测相对位置,预测出 b-box 中心点相对于网格单元左上角的相对坐标。

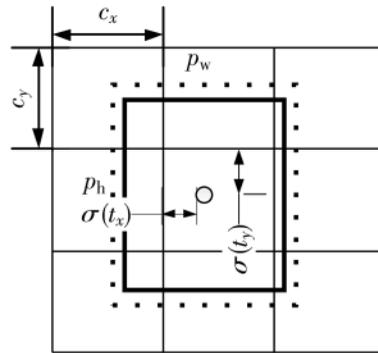


图 3 检测框预测

根据 anchor 的预测值来预测宽和高的公式如下:

$$b_x = \sigma(t_x) + c_x \quad (1)$$

$$b_y = \sigma(t_y) + c_y \quad (2)$$

$$b_w = p_w e^{t_w} \quad (3)$$

$$b_h = p_h e^{t_h} \quad (4)$$

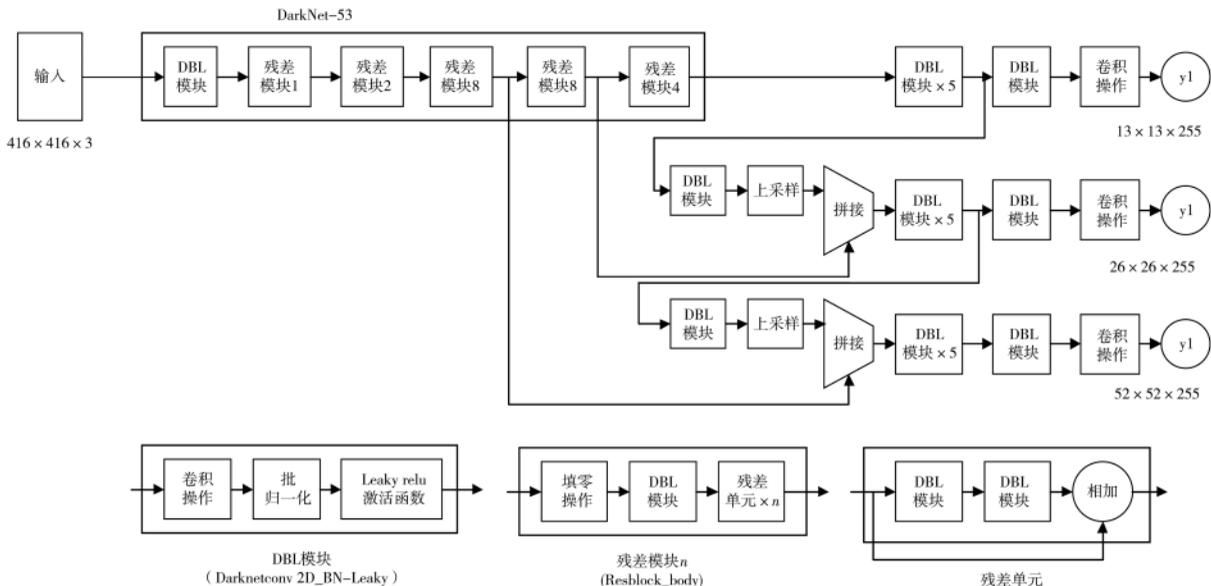


图 2 YOLO-q-v3 框架

其中, b_x, b_y, b_w, b_h 分别为预测框的中心坐标 x, y 、宽度、高度, t_x, t_y, t_w, t_h 分别为网络输出的预测框的中心坐标 x, y 、宽度、高度, c_x, c_y 分别是网格左上角的坐标 x, y, p_w, p_h 分别是锚框的宽和高。

2.2.2 损失函数

本文算法在计算损失函数时,根据 IOU(Intersection over Union)对预测框结果进行分类,IOU 的计算公式为^[13]:

$$W_c = \max((\min(X_A + W_A, X_B + W_B) - \max(X_A, X_B)), 0) \quad (5)$$

$$H_c = \max((\min(Y_A + H_A, Y_B + H_B) - \max(Y_A, Y_B)), 0) \quad (6)$$

$$IOU = \frac{A \cap B}{A \cup B} = \frac{W_c H_c}{W_A H_A + W_B H_B - W_c H_c} \quad (7)$$

其中, (X_A, Y_A) 为边框 A 的左上角坐标, W_A 为 A 的宽, H_A 为 A 的高; (X_B, Y_B) 为边框 B 的左上角坐标, W_B 为 B 的宽, H_B 为 B 的高; C 为边框 A 与 B 相交的矩形区域, W_c 为 C 的宽, H_c 为 C 的高^[13]。

YOLO v3 算法的损失函数与每个特征图的相关损失有关,其中每个特征图的损失 $loss_M$ 与预测框大小和位置相关 $loss_{box}$ 、预测结果置信度相关 $loss_{con}$ 、预测分类结果相关 $loss_{class}$ 有关,公式如下:

$$loss_{box} = \lambda_{box} \sum_{i=0}^{N1 \times N1} \sum_{j=0}^3 1_{ij}^{obj} [(t_x - t'_x)^2 + (t_y - t'_y)^2 + (t_w - t'_w)^2 + (t_h - t'_h)^2] \quad (8)$$

$$loss_{con} = -\lambda_{con} \sum_{i=0}^{N1 \times N1} \sum_{j=0}^3 [1_{ij}^{obj} \log(c_{ij}) + 1_{ij}^{noobj} \log(1 - c_{ij})] \quad (9)$$

$$loss_{class} = -\lambda_{class} \sum_{i=0}^{N1 \times N1} \sum_{j=0}^3 1_{ij}^{obj} \sum_{c \in classes} [p_{ij}(c) \log(p_{ij}(c)) + (1 - p_{ij}(c)) \log(1 - p_{ij}(c))] \quad (10)$$

$$loss_M = loss_{box} + loss_{con} + loss_{class} \quad (11)$$

$$loss = loss_M + loss_{M2} + loss_{M3} \quad (12)$$

3 实验与结果

3.1 数据集

手机屏幕缺陷包括裂痕、油脂、破裂等,本文主要采用裂痕数据集,数据集中标注裂痕缺陷目标,采用 LabelImg 软件对数据集进行标记,包括标注出数据集中的目标位置和缺陷的种类,保存为 xml 格式。本文用于实验的数据集来源于学生自己拍摄的手机屏幕缺陷图集,一共有 2 071 张裂痕缺陷图片,分辨率为 1 432×1 432,由于数据集量偏少,因此将图片通过旋转和翻折等操作对数据集进行扩充,可以降低模型过拟合的机率,提高实验的检测效果。将数据集以 8:2 的比例分为训练集和测试集,其中随机选取 1 657 张图片作为训练集样本,选取 414 张图片作为测试集样本。

3.2 实验环境

本文将以检测手机屏幕裂痕缺陷为目的,通过改进原 YOLO v3 算法提出 YOLO-q-v3 算法,并将两个算法在同一实验环境下进行实验,再通过对比指标进行比较,实验环境如表 2 所示。

表 2 实验环境

实验环境	配置参数
操作系统	Windows 10 家庭版
GPU	NVIDIA GeForce GTX 960M
CPU 处理器	i7-6700HQ 8 核
内存/GB	16
编程语言	Python3.6
深度学习框架	Torch1.2
GPU 加速库	CUDA10.0

3.3 评估指标

本文模型性能评估指标为精度 (Precision)、召回率 (Recall)、平均精度均值 (mean Average Precision, mAP) 和 F1 值。精度 P 代表预测正确的正例数据占预测为正例数据的比例,计算公式为:

$$P = \frac{TP}{TP + FP} \quad (13)$$

其中,TP 为标签正样本输出正样本实例的数量,FP 为标签负样本输出正样本实例的数量。

召回率 R 表示预测正确的正例数据占实际正例数据的比例,计算公式为:

$$R = \frac{TP}{TP + FN} \quad (14)$$

其中,FN 为标签正样本输出负样本实例的数量。精度 P 与召回率 R 越大说明检测效果越好,然而两者为负相关关系,需要量化指标 mAP 在两者间进行权衡。mAP 值为两类目标 AP 值的平均值,而 F1 值与 mAP 值作用相似,也需要同时考虑 P 和 R 两者的量化指标。F1 值计算公式为:

$$F1 = \frac{2PR}{P + R} \quad (15)$$

3.4 实验结果

3.4.1 训练模型

图 4 给出了 YOLO-q-v3 模型在训练过程的 loss 曲线,可以看出该模型的各项损失值在训练轮数达到一定次数时都处于逐渐收敛的状态,说明 YOLO-q-v3 网络模型训练的各个阶段已经达到了一个稳定的状态,也

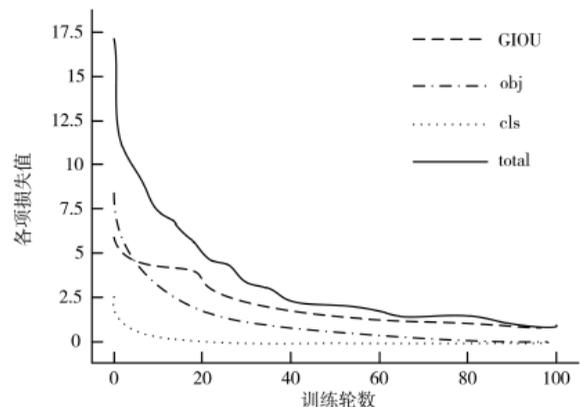


图 4 loss 曲线

明 YOLO-q-v3 模型完成训练。

3.4.2 实验结果对比及分析

本文改进的 YOLO-q-v3 算法首先能够准确地将数据集中的裂痕缺陷识别出来,并与原算法在同一实验环境下对同一数据集进行比较试验。

从表 3 的各项评估指标可以看出,本文 YOLO-q-v3 算法的精度 P 比原 YOLO v3 算法的精度减少了 0.4%,两个算法的召回率 R 相同, YOLO-q-v3 算法的 $F1$ 值比原 YOLO v3 算法下降 0.1%。

表 3 不同算法指标对比

算法	P	R	$F1$	检测速率/(f/s)
YOLO v3	0.921	0.915 3	0.947	41.17
YOLO-q-v3	0.917	0.915 3	0.937	51.28

实验结果表明,本文 YOLO-q-v3 算法能够准确地识别数据集中的缺陷,虽然在检测精度上对比原算法有所下降,但是下降幅度不大,并且在识别效率上比原算法有所提高,比原算法检测速率提高了 24%。

4 结论

本文对 YOLO v3 算法的 DarkNet-53 网络结构进行了相应的删减改进,使用 FPN,提出了 YOLO-q-v3 算法进行手机屏幕缺陷检测,减少了计算量,使得算法更适合需要实时检测的要求,如手机生产线对缺陷检测速度有着严格要求。本文 YOLO-q-v3 能够正确识别出手机屏幕裂痕,该算法能够满足手机裂痕的非人工检测,检测效率比 YOLO v3 算法有所提高。

参考文献

- [1] 刘恒.基于全卷积网络的手机液晶面板缺陷检测算法研究[D].重庆:重庆邮电大学,2019.
- [2] 李晨.基于机器视觉的不同属性表面中微弱缺陷的检测技术研究[D].杭州:浙江大学,2018.
- [3] 王思宇,郭阳宽,郭会梁,等.手机屏幕表面划痕检测系统研究[J].数字技术与应用,2018,36(4):57-58.

- [4] 高如新,任晓朵,吴献,等.基于图像处理的手持屏幕缺陷检测系统研究[J].测控技术,2017,36(4):26-30.
- [5] 彭赶,张平,潘奕创.基于机器视觉的手持屏幕缺陷检测系统研究[J].自动化技术与应用,2018,37(9):104-107,127.
- [6] 应冬梅,郭阳宽,张晓青.基于手机屏幕缺陷检测系统设计[J].现代工业经济和信息化,2020,10(2):51-52.
- [7] 王森,伍星,张印辉,等.基于深度学习的全卷积网络图像裂纹检测[J].计算机辅助设计与图形学学报,2018,30(5):859-867.
- [8] 余永维,殷国富,殷鹰,等.基于深度学习网络的射线图像缺陷识别方法[J].仪器仪表学报,2014,35(9):2012-2019.
- [9] 李梦园.深度学习算法在表面缺陷识别中的应用研究[D].杭州:浙江工业大学,2015.
- [10] 宋威.基于深度卷积神经网络的手持屏幕缺陷检测[D].成都:电子科技大学,2019.
- [11] 史磊,荆明娥,范益波,等.基于 R-CNN 算法的分割检测算法[J].复旦学报(自然科学版),2020,59(4):412-418.
- [12] 袁鹏泰.多人姿态估计算法研究及应用[D].南京:南京航空航天大学,2019.
- [13] 黄家才,邹俊,丁凌,等.基于改进 YOLO v3 算法的零件目标快速检测方法[J].南京工程学院学报(自然科学版),2020,18(3):6-11.
- [14] 岳晓新,贾君霞,陈喜东,等.改进 YOLO V3 的道路小目标检测[J].计算机工程与应用,2020,56(21):218-223.
- [15] 周志锋,万旺根,王旭智.基于 YOLO V3 框架改进的目标检测[J].电子测量技术,2020,43(18):102-106.

(收稿日期:2021-06-23)

作者简介:

杨戈(1974-),男,博士,副教授,主要研究方向:人工智能技术、计算机视觉技术。

刘思瀚(1999-),男,本科,主要研究方向:计算机视觉、双目立体视觉技术。



扫码下载电子文档

(上接第 80 页)

- [12] 朱荷欢,武文,孙玉婷.三维建模不同技术方法的特点研究及应用思考[C]//南京市国土资源信息中心 30 周年学术交流论文集,2020:31-34.
- [13] 郭猛.三维建模的技术方法特点研究及应用[J].智慧城市,2021,7(14):58-59.
- [14] 李德仁.基于数字孪生的智慧城市[J].互联网天地,2021(7):12.
- [15] 许旭,鲁金萍.“十四五”时期我国新型智慧城市建设的七大趋势[J].网络安全和信息化,2021(7):4-7.

(收稿日期:2021-07-16)

作者简介:

于海洋(1987-),通信作者,男,硕士,工程师,主要研究方向:数字孪生、5G 行业应用、网联无人机, E-mail: yuhy8@chinatelecom.cn.

封顺天(1980-),男,硕士,高级工程师,主要研究方向:数字孪生城市、数据治理、数据交易。

崔立鹏(1990-),男,硕士,工程师,主要研究方向:智慧交通、数字孪生应用。



扫码下载电子文档

版权声明

经作者授权，本论文版权和信息网络传播权归属于《电子技术应用》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、DOAJ、美国《乌利希期刊指南》、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《电子技术应用》编辑部

中国电子信息产业集团有限公司第六研究所