doi:10.3969/j.issn.1672-1292.2018.03.005

# 基于深度学习监控场景下的多尺度目标检测算法研究

程显毅[1,2],胡海涛[2],季国华[1],孙丽丽[1]

(1.硅湖职业技术学院计算机系,江苏 昆山 215323)
(2.南通大学南通先进通信技术研究院,江苏 南通 226019)

[摘要] 针对监控环境下的视频图像处理存在漏检这一问题,分析现有目标检测算法中普遍使用的深度学习方法—Faster R-CNN,在 VGG16 卷积神经网络基础上,对深度卷积神经网络进行改进,在第一层卷积层中加入空洞卷积核,扩展神经网络的宽度,使得目标检测模型具有尺度不变性. 在深度学习平台 PyTorch 下对 Cifar-10 数据集进行了实验,实验结果显示,改进的目标检测算法具有较好的尺度不变性,在监控场景下更具优势.

[关键词] 深度学习,目标检测,空洞卷积核,监控场景
[中图分类号]TP181 [文献标志码]A [文章编号]1672-1292(2018)03-0033-06

# Research on Algorithm of Multi-Scale Target Detection Based on Deep Learning in Monitoring Scenario

Cheng Xianyi[1,2],Hu Haitao[2],Ji Guohua[1],Sun Lili[1]

(1.Department of Computer,Silicon Lake Vocational and Technical College,Kunshan 215323,China)
(2.Nantong Research Institute for Advanced Communication Technologies,Nantong University,Nantong 226019,China)

**Abstract**:In view of a problem of missed inspection in the video image processing under the monitoring environment,we analyze the deep learning method commonly used in existing target detection algorithms-Faster R-CNN,and improve a deep convolution neural network based on VGG16 convolution neural network. Expanding the width of the neural network,by adding an empty core to the first volume layer,makes the target detection model have scale invariance. The Cifar-10 dataset is tested on the in-depth learning platform PyTorch. The experimental results show that the improved target detection algorithm has a better scale invariance and has more advantages in the monitoring scene.

**Key words**:deep learning,target detection,dilated kernelof convolution,monitoring scenarios

目标检测是计算机视觉领域热门的研究方向之一,是视频监控、无人驾驶、辅助诊断等应用的基础. 随着越来越多的研究人员投入到目标检测的研究中来,目标检测算法近几年取得了跨越式的发展. 目标检测算法分为 3 类[1]:非神经网络结构的机器学习方法、后卷积神经网络方法和前卷积神经网络方法.

非神经网络的机器学习方法试图通过对待检测目标的局部或全局特征进行统计和分析. 例如,Viola[2]提出基于 AdaBoost 的算法框架,使用 Haar-like 小波特征进行分类,然后采用滑动窗口搜索策略实现准确有效地定位. 由于 Haar-like 特征对边缘比较敏感,可以区分人的眼眶,因而适用于人脸检测. Dalal[3]提出使用 HOG 作为特征,利用 SVM 作为分类器进行行人检测. HOG 特征是描述行人的最好的特征,对于行人的形变、遮挡具有较好的鲁棒性,HOG+SVM 在当时是行人检测最成功的方法.

后卷积神经网络方法[4]把卷积神经网络放在整体算法结构的后面,总体上是先通过处理得到可能包含目标的区域,然后再使用卷积神经网络提取区域的特征,完成对目标的识别. 该方法的代表就是区域卷

收稿日期:2018-04-18.
基金项目:国家自然科学基金(61771265)、江苏省现代教育技术研究课题(2017-R-54131)、南通大学-南通智能信息技术联合研究中心开放课题(KFKT2016B06).
通讯联系人:程显毅,博士,教授,研究方向:计算机视觉. E-mail:xycheng@ ntu.edu.cn

积神经网络(regions of CNN,R-CNN)[5]. 区域卷积神经网络首先使用 Selective Search(SS)算法[6],从图片中提取出 2 000 个可能包含有目标的区域,再将这 2 000 个候选区(region of interest,ROI)压缩到统一大小(227 * 227)送入卷积神经网络中进行特征提取,在最后一层将特征向量输入 SVM 分类器,得到该候选区域的种类. 为了得到更精确的候选区,减少卷积神经网络重复提取候选区特征,万维[7]融合了 LUV 色彩、梯度幅值和梯度方向直方图这 3 类特征,使用决策树和 Adaboost 算法根据邻近尺度特征相似性的原理级联构成多尺度 Adaboost 分类器,采用滑窗法遍历图像,最终得到疑是行人窗口. 针对传统滑动窗口法产生的候选物体框质量不高、数量多和冗余性大的问题,吴慧[8]提出了一种遥感目标高质量候选框选择算法,进一步过滤掉大量不可能包含物体的候选框,最终产生数量较少、质量好的候选框,最后使用卷积神经网络提取候选框特征. 这些方法利用了卷积神经网络抽取目标特征方面的优势,但是却存在两个问题:

(1)不管是 SS 还是使用图像特征来确定目标候选区域,候选区域提取的过程都是在 CPU 内计算完成,候选框定位占用了 CPU 的大量资源;

(2)因为共有 2 000 个候选区域,在对候选区域进行卷积操作提取特征时,会有大量重复的计算,增加了算法的复杂性.

针对第 2 个问题,Ross[9]提出了快速区域卷积神经网络(Fast R-CNN),Fast R-CNN 借鉴了 SPP-Net[10]思想. 潘广贞[11]借助 Fast R-CNN 的思想,采用 SS 法对视频车辆图像提取多个车辆候选目标矩形区域,采用 Hessenberg 分解法将运动车辆和其阴影区域分开,结合 PCA 分析法检测阴影,最终识别移动阴影中包含的车辆区域,实现快速去除阴影的效果.

虽然 Fast R-CNN 在目标检测上又一次实现了飞跃,但在使用 SS 进行候选区域选择时速度仍然很慢. 目标检测出现了第 3 种方法——前卷积神经网络方法,其代表是 He[12]提出的 Faster R-CNN. 何凯明发明了区域推荐网络(region proposal net,RPN)替代 SS 算法进行候选区域推荐,在特征图之后使用 ROI 池化层使得区域卷积网络和区域推荐网络共享卷积层,实现了真正的端到端的计算,目标识别的速度和精度得到了大幅提升. 桑军等[13]使用 ZF、VGG-16 及 ResNet-101 3 种卷积神经网络分别与 Faster R-CNN 结合,在 BIT-Vehicle 数据集上进行实验,实验结果显示,不同卷积神经网络在 Faster R-CNN 下的检测效果各不相同.

# 1　多尺度目标检测算法设计

针对 Faster R-CNN 目标场景下目标漏检问题,本文设计了多尺度目标检测算法,在第一层卷积层中加入空洞卷积核,扩展了神经网络的宽度,使得目标检测算法具有尺度不变性.

## 1.1　Faster R-CNN

Faster R-CNN 从 Faster R-CNN 基础上改进而来,是目前综合性能最好的目标检测算法,其使用区域推荐网络(RPN)替代 Fast R-CNN 的 SS 方法为网络提供目标区域,在 VOC 和 COCO 数据集上都有非常卓越的表现. Faster R-CNN 的算法流程示意图如图 1 所示.

图 1 中,Input 为网络输入的图像. roi_data_layer 主要为网络提供 data、im_info、gt_bbox. VGG 是卷积神经网络,在 Faster R-CNN 中卷积神经网络可使用 VGG16,故图中用 VGG 来代替卷积神经网络. RPN 主

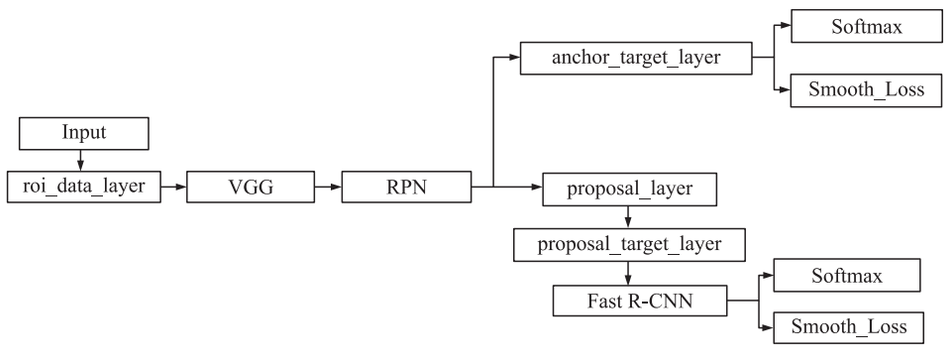


**图 1　Faster R-CNN 算法流程示意图**

**Fig. 1　The flow diagram of Faster R-CNN algorithm**

要的作用是生成建议窗口(proposals),并将建议窗口输出到卷积神经网络最后一层卷积的特征图(feature map)上. anchor_target_layer 的作用类似 SS,对于 roi_data_layer 中 data 的图片和对应的 gt_bbox 生成锚点(anchors),计算出所有可能的候选框(rois). proposal_layer 的作用是过滤掉比较小的候选框,然后使用非极大值抑制(non-maximal suppresion,NMS),按照一定的规则,选取较少的候选框(如 256 个),并利用 rpn_cls_prob 将 bbox 从[0,1]映射回实际大小,用作 propasal_target_layer 的输入. proposal_target_layer 在训练时从候选框选择一部分用于训练,同时给定训练目标. Fast R-CNN 用来实现候选框的分类和位置的回归,如 Cifar-10 数据集共 10 个种类,则经过 Fast R-CNN 就分成了 11 类(背景也是一类).

文献[12]使用 Faster R-CNN 对监控场景的检测结果如图 2 所示. 在图 2 中,只检测到了 2 辆车,漏检了大多数的目标.

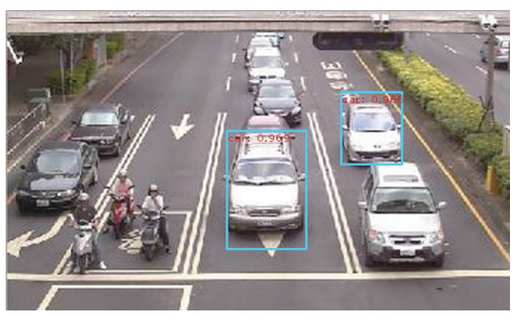

**图 2　Faster R-CNN 检测结果**

**Fig. 2　The detection result of Faster R-CNN**

## 1.2　空洞卷积核

卷积神经网络使用卷积核提取图像特征. 卷积核的大小代表局部感知野的大小,不同大小的局部感知野可以感受到不同尺度的特征,在卷积过程中使用不同大小的卷积核可以提取到不同尺度的特征,使卷积神经网络具有尺度不变性.

图 3 给出了 Inception 结构,这种结构既达到了多尺度的目的,又扩展了网络的宽度. 但这种做法增加了需要训练的参数. 为此,有学者设计出了如图 4 所示的 Bottleneck 结构[14]. Google 以 Bottleneck 结构为基础训练出了 GoogleNet,性能十分优越.

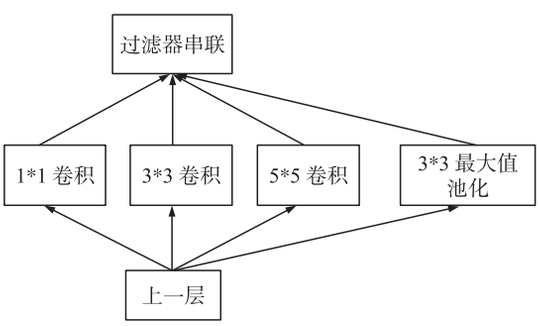


**图 3　Inception 结构**

**Fig. 3　The structure of Inception**

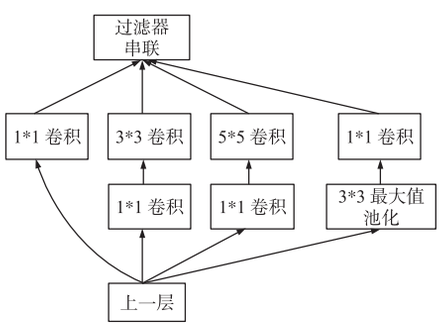


**图 4　Bottleneck 结构**

**Fig. 4　The structure of Bottleneck**

图 5 所示为空洞卷积核,该卷积核大小是 5 * 5,但中间的一些参数被设置为 0,在卷积核中有一部分“空洞”,故而实际上只有 3 * 3 个参数,而其却能感受到 5 * 5 范围的特征. 可以发现,空洞卷积核就是 Laplace 算子的变形,这不仅让其易于应用,也符合对于卷积神经网络多尺度的需求.

| | | | | |
|---|---|---|---|---|
| $W_{11}$ | 0 | $W_{13}$ | 0 | $W_{15}$ |
| 0 | 0 | 0 | 0 | 0 |
| $W_{31}$ | 0 | $W_{33}$ | 0 | $W_{35}$ |
| 0 | 0 | 0 | 0 | 0 |
| $W_{51}$ | 0 | $W_{53}$ | 0 | $W_{55}$ |

**图 5　空洞卷积核**

**Fig. 5　Dilated kernel of convolution**

## 1.3　多尺度卷积神经网络

根据 Faster R-CNN 算法的流程,其中的卷积神经网络就是 VGG16 网络,需要训练出一个多尺度卷积神经网络来代替 VGG16 抽取出图像的多尺度 feature maps.

VGG16 通过使用双 3 * 3 的卷积核代替一个 5 * 5 卷积核,达到增加网络深度、减少参数的目的. 本文设计的多尺度卷积神经网络在不增加参数量的同时,让卷积神经网络对特征尺度具有鲁棒性. 受 GoogleNet 启发,为了减少算法复杂度,本文算法在第一层卷积中使用 3 * 3 的空洞卷积核,算法描述如图 6 所示.

# 2　实验分析

## 2.1　PyTorch 深度学习平台

PyTorch[15] 由 Torch7 团队开发,支持单一或多个 CPU 或 GPU 运算,且可自由切换,提高了深度神经网络的训练速度. 与现有的主流架构相比,PyTorch 具有两个突出特点:不仅能够实现 GPU 加速,还能支持动态神经网络.

实验数据集选用 Cifar-10,有 10 类物体,最终的图片分类结果为 10 类.

## 2.2　多尺度目标检测算法性能比较评估

按照图 6 的连接方式更改神经网络结构,在卷积层 1-1 和卷积层 1-2 中加入空洞卷积核,即在定义多尺度卷积层时设置膨胀比为 1,设置填充为 2,步长保持不变. 具体代码如下:

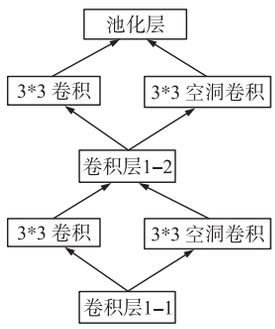


**图 6　多尺度卷积层**

**Fig. 6　Multiscale layer of convolution**

self.conv1 = nn.Sequential( Conv2d( filter = 32, filter_size = 3, strides = 1, dilation_rate = 1, padding = 2), Conv2d( filter = 32, filter_size = 3, strides = 1, dilation_rate = 1, padding = 2), nn. MaxPool2d( filter_size = 2, strides = 2, padding = 1).

self.conv1 为变量名,该段代码建立了第 1 层卷积神经网络. 代码的第 1 个 Conv2d 定义了 1-1 层卷积神经网络,第 2 个 Conv2d 定义了 1-2 层卷积神经网络,nn.MaxPool2d 定义了第 1 层池化网络. 代码中的 filter 代表卷积核个数,filter_size 代表卷积核尺寸,strides 代表步长,dilation_rate 代表扩张单位,padding 是图像填充的行列数. 1-1 和 1-2 卷积层使用的是 3 * 3 的空洞卷积核,图像经过这两层卷积变为 224 * 224 * 32 的张量(tensor). 第 1 层池化层是卷积核为 2 * 2、步长为 2 的池化操作.

因为 Cifar-10 数据集共有 10 类物体,所以最终的图片分类结果是 10 类,这就需要在最后一层全连接层(fc layer)的输出由 4 096 改为 10.

具体代码如下:

self.fc6 = FC(512 * 7 * 7,4 096)

self.fc7 = FC(4 096,4 096)

self.fc8 = FC(4 096,10)

图 7 所示为本文算法和 Faster R-CNN 算法在 Cifar-10 数据集上评估结果对比,虚线是本文算法,实线是 Faster R-CNN 算法结果.

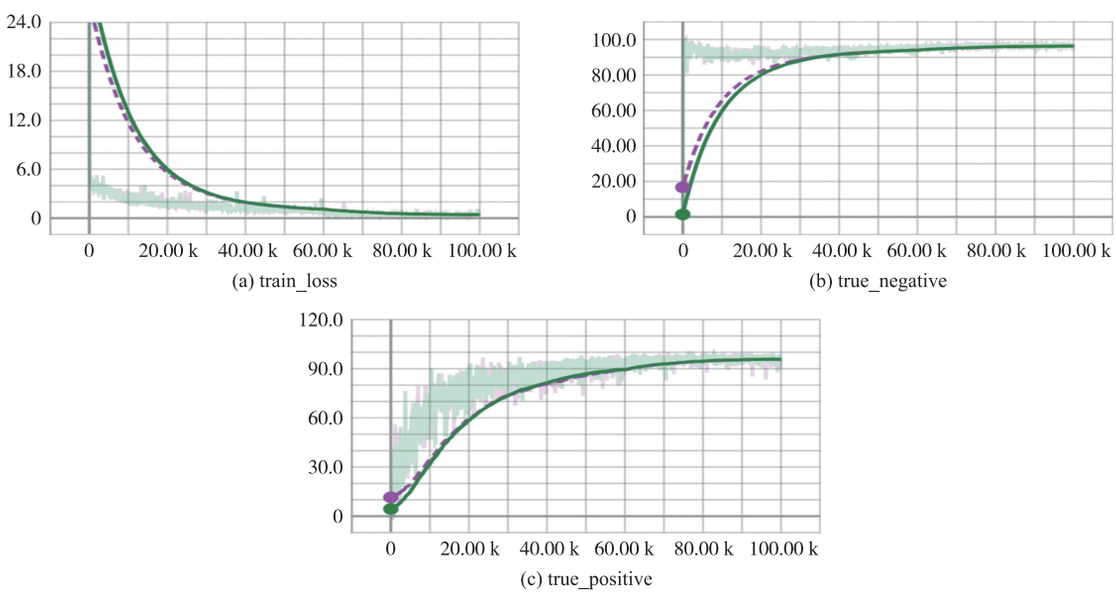


**图 7　Faster R-CNN 和本文算法的性能比较**

**Fig. 7　The performance comparison of Faster R-CNN and the algorithm in this paper**

从图 7 可以看出,在学习率保持相同情况下,本文算法的损失率曲线下降较快(见图 7(a)),精确率和召回率曲线上升较快(见图 7(b)和(c)),这说明本文方法在 Cifar-10 数据集上检测性能较好.

### 2.3 实验结果

将 Faster R-CNN 应用于非监控场景的检测结果如图 8 所示. 共检测出 4 个目标,包括人和汽车,检测效果较好. 再次将 Faster R-CNN 模型应用于监控场景,检测结果如图 9 所示,目标对象数量超过 10 个,但 Faster R-CNN 只检测出 7,漏检了一部分汽车目标和大部分行人目标.

图 10 为本文方法在监控场景下的检测结果,虽未将目标全部检测出来,但相对图 8 的检测效果更好,多了两个行人和右下角的目标.

**图 8 Faster R-CNN 非监控场景检测结果**

**Fig. 8 The detection result of Faster R-CNN in non-monitoring scene**

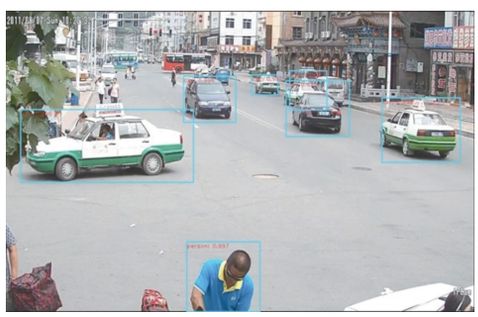

**图 9 Faster R-CNN 监控场景下的检测结果**

**Fig. 9 The detection result of Faster R-CNN in monitoring scene**

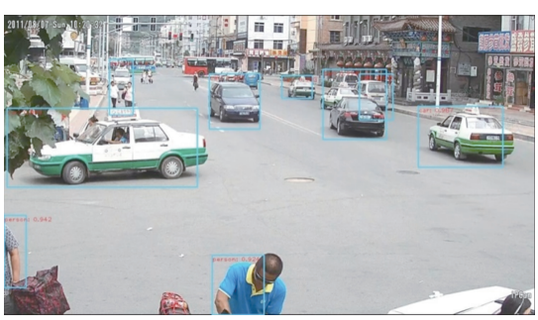

**图 10 多尺度目标检测模型检测结果**

**Fig. 10 The detection result of multiscale target detection**

## 3 结语

基于深度学习目标检测的方法中,Faster R-CNN 的方法特别优秀. 但在监控场景下,Faster R-CNN 的检测结果并非特别理想. 本文根据监控场景的实际需要,改进了 Faster R-CNN 的卷积神经网络结构,使得改进后的目标检测模型具有一定的尺度不变性,从而在监控场景下可以检测到更多的目标.

### [参考文献](References)


[1] 赵玉吉. 基于视频序列的运动目标检测与跟踪算法研究[D]. 扬州:扬州大学,2017.
ZHAO Y J. Research on motion target detection and tracking algorithm based on video sequence[D]. Yangzhou:Yangzhou University,2017.(in Chinese).

[2] VIOLA P,JONES M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Kauai,USA,2001.

[3] DALAL N,TRIGGS B. Histograms of oriented gradients for human detection[C]//Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego,USA,2005.

[4] SHIGO A. Support vector machines for pattern classification[M]. New York:Springer,2012.

[5] ROSS G. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago,Chile,2015.

[6] HINTERSTOISSER S,LEPETIT V,ILIC S,et al. Dominant orientation templates for real-time detection of textureless objects[C]//2010 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). San Francisco,USA,2010.

[7] 万维. 基于深度学习的目标检测算法研究及应用[D]. 成都:电子科技大学,2015.
WAN W. Research and application of target detection algorithm based on in-depth learning[D]. Chengdu:University of Electronic Technology,2015.(in Chinese)

[8] 吴慧. 基于深度学习的遥感影像目标检测[D]. 哈尔滨:哈尔滨工业大学,2016.
WU H. Target detection of remote sensing imaging based on in-depth learning[D]. Harbin:Harbin Institute of Technology, 2016.(in Chinese)

[9] ROSS G,JEFF D. Region-based convolutional networks for accurate object detection and segmentation[J]. IEEE translations on pattern analysis and machine intelligence,2016,38(1):142-158.

[10] HE K,ZHANG X,REN S,et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence,2015,37(9):1904-1916.

[11] 潘广贞,孙艳青,王凤. 基于 Fast RCNN 模型的车辆阴影去除[J]. 计算机工程与设计,2018(3):819-823.
PAN G Z,SUN Y Q,WANG F. Removal of vehicle shadow based on fast RCNN model[J]. Computer engineering and design, 2018(3):819-823.(in Chinese).

[12] REN S,HE K,GIRSHICK R,et al. Faster R-CNN:towards real-time object detection with region proposal networks[C]// Advances in Neural Information Processing Systems. Montreal,Canada,2015:91-99.

[13] 桑军,郭沛,项志立,等. Faster R-CNN 的车型识别分析[J]. 重庆大学学报,2017,40(7):32-36.
SANG J,GUO P,XIANG Z L,et al. Vehicle detection based on faster-RCNN[J]. Journal of Chongqing university,2017, 40(7):32-38.(in Chinese).

[14] ABDELGHAFFAR A A. Influence of sinusoidal and square voltages on partial discharge inception in geometries with point-like termination[J]. High voltage,2018,15(3):31-37.

[15] 廖星宇.深度学习入门之 PyTorch[M]. 北京:电子工业出版社,2017.
LIAO X Y. PyTorch of deep learning[M]. Beijing:Electronic Industry Press,2017.(in Chinese)

[责任编辑:严海琳]

(上接第 32 页)

[11] 秦绪佳,王慧玲,杜轶诚,等. HSV 色彩空间的 Retinex 结构光图像增强算法[J]. 计算机辅助设计与图形学学报, 2013,25(4):488-493.
QIN X J,WANG H L,DU Y C,et al. Structured light image enhancement algorithm based on Retinex in HSV color space[J]. Journal of computer-aided design and computer graphics,2013,25(4):488-493.(in Chinese)

[12] ZHOU Z G,SANG N,HU X R. Global brightness and local contrast adaptive enhancement for low illumination color image[J]. Optik-international journal for light and electron optics,2014,125(6):1795-1799.

[13] 王守觉,丁兴号,廖英豪,等. 一种新的仿生彩色图像增强方法[J]. 电子学报,2008,36(10):1970-1973.
WANG S J,DING X H,LIAO Y H,et al. A novel bio-inspired algorithm for color image enhancement[J]. Acta electronica sinica,2008,36(10):1970-1973.(in Chinese)

[14] 郑江云,江巨浪,黄忠. 基于 RGB 灰度值缩放的彩色图像增强[J]. 计算机工程,2012,38(2):226-228.
ZHENG J Y,JIANG J L,HUANG Z. Color image enhancement based on RGB gray value scaling[J]. Computer engineering, 2012,38(2):226-228.(in Chinese)

[15] GUO P,YANG P X,LIU Y,et al. An adaptive enhancement algorithm for low illumination image based on hue reserving[C]// Proc of Cross Strait Quad-Regional Radio Science and Wireless Technology Conference(CSQRWC). Harbin,China,2011:1247-1250.

[16] RAJU G,NAIR M S. A fast and efficient color image enhancement method based on fuzzy-logic and histogram[J]. Internation journal of electronics and communications,2014,68:237-243.

[责任编辑:严海琳]