计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.191200127

全局双边网络的语义分割算法

任天赐[1] 黄向生[2] 丁伟利[1] 安重阳[1] 翟鹏博[3]
1 燕山大学电气工程学院 河北 秦皇岛 066000
2 中国科学院自动化研究所 北京 100190
3 中国科学院微电子所 北京 100029
(tianci.ren@stumail.ysu.edu.cn)

摘 要 语义分割任务是对图像中的物体按照类别进行像素级别的预测,其难点在于在保留足够空间信息的同时获取足够的上下文信息。为解决这一问题,文中提出了全局双边网络语义分割算法。该算法将大尺度卷积核融入BiSeNet网络中,在BiSeNet网络原有的空间路径和上下文路径两条分支的基础上增加全局路径分支,使网络能够捕获更多的上下文信息,同时提出将BiSeNet网络中的注意力优化模块和特征融合模块中的全局池化模块替换为全局卷积模块,进一步提高了网络获取上下文信息的能力,从而使预测结果更加准确。实验结果表明,该算法在Cityscapes数据集上将交并比(MIoU)指标提高了0.84%,获得了优于BiSeNet网络的表现。

关键词:语义分割;双边分割网络;全局卷积网络
中图法分类号 TP389.1

Global Bilateral Segmentation Network for Segmantic Segmentation

REN Tian-ci[1], HUANG Xiang-sheng[2], DING Wei-li[1], AN Chong-yang[1] and ZHAI Peng-bo[3]
1 Institute of Electrical Engineering, Yanshan University, Qinhuangdao, Hebei 066000, China
2 Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
3 Institute of Microelectronics, Chinese Academy of Sciences, Beijing 100029, China

Abstract The task of semantic segmentation is to predict the objects according to the category at the pixel level. The difficulty lies in retaining enough spatial information and obtaining enough context information. In order to solve this problem, this paper proposes a global bilateral network semantic segmentation algorithm. In this algorithm, the large-scale convolution kernel is integrated into the BiSeNet Network, and the global path branches are added to the original spatial path and context path of the BiSeNet Network, so that the network can capture more context information. At the same time, the global pooling module in the attention optimization module and feature fusion module is replaced by the global convolution module to further improve the network acquisition. The experimental results show that the algorithm improves the MIoU index by 0.84% on Cityscaps dataset, and achieves better performance than BiSeNet Network.

Keywords Semantic segmentation, Bilateral segmentation network, Global convolutional network

1 引言

图像语义分割是计算机视觉领域的一个基本问题。在视频监控、自动驾驶、场景理解等领域有着重要的应用价值。语义分割的目标是将图像中相同类别的像素分成一类,将不同类别的物体分割出来。与分类任务不同的是,语义分割不仅需要预测出物体的类别及粗略的位置信息,还需要将物体沿着自身的轮廓边界分割出来,与其他物体分开。近年来,卷积神经网络由于对复杂问题强大的拟合能力,能够从大量样本中自动学习提取图像中的隐含特征,并避免了人工显式的特征提取,使其在语义分割任务上取得了巨大成功。在语义分割任务中有两个相互制约的问题:1)多层卷积堆叠,逐层缩小特征图能够增大感受野,有利于提取语义信息,使得分类更加准确,然而,低分辨率的特征图不利于细小物体和包含大量细节物体边缘的类别预测;2)高分辨率的特征图有利于物体边缘区域的类别预测,但容易导致细节较少的区域类别预测错误,如质地均匀物体的中间部分。因此,如何平衡这两个问题成为语义分割网络研究的热点。

2 相关研究

自FCN[1]将CNN结构应用到图像语义分割领域以来,许多优秀的语义分割方法陆续被提出。一类方法采用

本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家重点研发计划(2018YFB1308302,2018YFB1308300);复杂物体表面纹理获取和三维重建的关键技术研究项目(61573356)
This work was supported by the National Key R&D Program of China (2018YFB1308302,2018YFB1308300) and Research on Key Technologies of texture acquisition and 3D reconstruction of complex object surface (61573356).
通信作者:丁伟利(weiye51@ysu.edu.cn)

U-shape 网络结构，先通过逐层下采样获取上下文语义信息，再逐层恢复空间信息。如：SegNet 等[2-3]通过逐层反卷积来恢复空间信息；U-Net 等[4]通过引入高分辨率特征图信息来恢复空间信息；RefineNet[5]通过在网络结构中引入多条路径来恢复空间信息。U-shape 结构虽然能一定程度上恢复空间信息，但是丢失的空间信息很难完全恢复。另一类方法采用融合不同层级特征的思想。如：DeepLabV2[6]、DeepLabV3[7]、DeepLabV3+[8]通过多尺度孔洞卷积融合多尺度上下文信息；PSPNet[9]融合不同尺度的平均池化层；Zhang 等[10]通过增加一个尺度自适应卷积层来自适应地获取上下文信息。这些方法捕获了不同尺度的特征，但不能利用全局图像中对象之间的关系。

在上述方法中，BiSeNet[11]网络在语义分割任务上获得了瞩目的成绩，原因在于 BiSeNet[11]网络不仅设计了 Context Path 来获取上下文信息，还设计了 Spatial Path 来保留空信息，并在二者中做了较好的平衡。本文是对 BiSeNet[11]的改进，主要思想是通过增加一条 Global Path[12]来获取更大的感受野，并将原有 ARM 和 FFM 模块中的全局平均池化替换为 GlobalConvBlock[12]。

本文第 3 节详细描述了提出的算法；第 4 节给出了具体的实验结果；最后总结全文。

3 全局双边分割网络

本文提出的全局双边分割网络是对 BiSeNet[11]网络结构的改进。BiSeNet[11]网络结构包含两条分支，Spatial Path 分支输出的特征图尺寸为原始特征图的 1/8，保留了丰富的空间信息；Context Path 分支提取上下文信息，受 Global Convolution Network[12]启发，本文提出在 BiSeNet[11]原有两条分支的基础上增加一条 Global Path 分支，以增大感受野并提高分类准确性（所增加的全局分支如图 1 中左侧红色实线框所示）。全局分支采用“大卷积核”对输入数据进行卷积处理，增大了感受野，将全局分支输出的特征图与 Spatial Path 分支和 Context Path 分支所输出的特征图通过特征融合模块进行融合。使得网络中包含了更加丰富的语义信息，从而提高了网络分类的准确性。除此之外，本文将 BiSeNet[11]的原有注意力优化模块（ARM）和特征融合模块（FFM）的全局平均池化都替换为全局卷积网络（GlobalConvBlock）模块[12]，进一步提高了网络分类的准确性。整体的网络结构如图 1 所示。中间绿色虚线框为 Spatial Path，该分支堆叠三层 stride 为 2 的卷积层以保留较大尺寸的特征图，从而保留了丰富的空间信息，每个卷积层由卷积（图 1 中 Conv）、批归一化（图 1 中 Bn）和激活函数（图 1 中 ReLU）组成；右侧蓝色虚线框为 Context Path，该分支使用轻量级网络迅速下采样，使得网络能够获取上下文信息；将 Context Path 所输出的特征图输入注意力优化模块，并通过上采样处理增大特征图尺寸，使特征图尺寸与 Spatial Path 输出的特征图尺寸一致。，通过特征融合模块对这两种不同级别的特征进行融合，最后经过 8 倍上采样后与本文所添加的 Global Path 输入特征融合模块再进行融合得到最终的输出结果。Global Path 由卷积模块（图 1 中的 Conv 模块）、ResNet 第二层特征图（图 1 中的 Res-2 模块）、ResNet 第三层特征图（图 1 中的 Res-3 模块）、全局卷积网络模块（图 1 中的 GCN 模块）、边界优化模块（图 1 中的 BR 模块）及反卷积模块（图 1 中 Deconv 模块）组成。

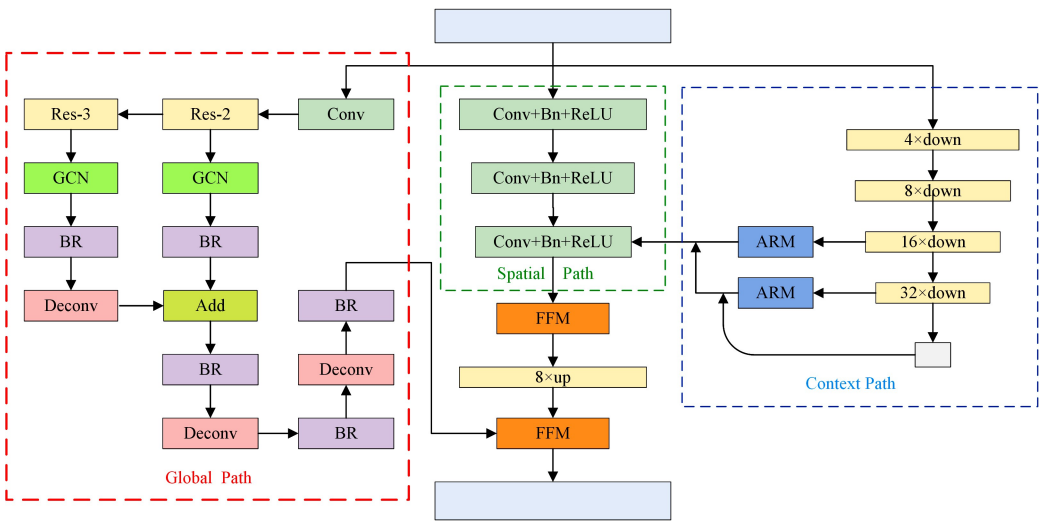


图 1 整体网络结构图（电子版为彩色）

Fig. 1 Network architecture

3.1 全局卷积网络

由于计算更加高效，堆叠多个小尺寸卷积核[13-14]代替大尺寸卷积核成为近年来深度学习中卷积核尺寸的趋势。针对语义分割任务，GCN（Global Convolutional Network）[12]提出大的卷积核更有利与获取足够大的感受野。从分类角度来看，如果卷积核大小增加到与特征图空间大小一致（全局卷积），则网络能够更好地利用全局信息，因此有利于更好地完成分类任务。而小尺寸的感受野在原图上对应很小的像素块，但单独的小像素块并不能准确地预测物体的类别。例如一个单独的蓝色像素块，其既可能是天空，也可能是大海，也可能是蓝色衣服上的某一部分，因此其周围的像素对该像素类别的预测起到了很大的作用。也就是说，语义分割任务需要获取较大的感受野，才能使像素点类别预测融合更多全局信息，从而使该像素点类别预测更加准确。然而，使用大尺寸卷积核大大增加了计算量，为了减小参数量，本文采用 $1*k+k*1$ 和 $k*1+1*k$ 的组合卷积替代 $k*k$ 的卷积，对于一个 $k*k$ 的卷积操作，使得参数量缩小到 $O(2/k)$。全局卷积模块如图 2 所示。

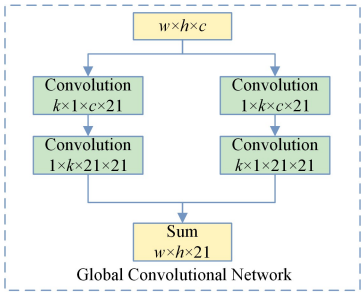


图 2 全局卷积网络

Fig. 2 Global convolutional network

图 2 中，w，h，c 分别代表全局卷积模块输入数据的宽、高和通道数；k 为卷积核的尺寸；Convolution 和 Sum 分别为对应卷积和加操作。

3.2 边界优化模块

为了在物体边缘得到更加精细的分割结果，GCN[12] 设计了边界优化模块。边界优化模块堆叠两层卷积构成残差结构作为网络结构的一部分，相比 CRF 之类[15] 的后处理操作，该结构的特点在于能够端到端地进行训练。具体来说，定义 $S^{\sim}$ 作为优化后的特征图：

$$S^{\sim}=S+R(S) \tag{1}$$

其中，S 是原始特征图，$R(\cdot)$ 是残差分支。边界优化模块的结构如图 3 所示。图 3 中，w 和 h 分别代表全局卷积模块输入数据的宽和高；Conv 和 Convolution 代表卷积操作；Sum 代表加操作；激活函数使用 ReLU。

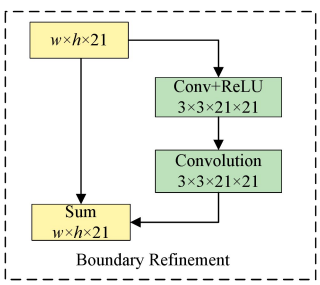


图 3 边界优化模块

Fig. 3 Boundary refinement

3.3 注意力优化模块

BiseNet[11] 注意力优化模块通过计算一个注意力向量来指导特征学习，能够利用全局上下文信息且不需要消耗太多计算量。与 BiSeNet[11] 中的注意力优化模块不同，我们将原本的全局池化替换为全局卷积模块（GlobalConvBlock）[12]，使注意力向量包含更多的全局信息（所替换的模块如图 4 中红色虚线框所示）。注意力优化模块的结构如图 4 所示。

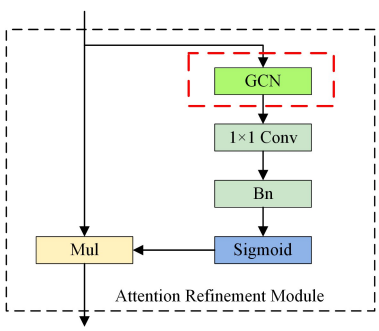


图 4 注意力优化模块（电子版为彩色）

Fig. 4 Attention refinement module

图 4 中，GCN 代表全局网络模块，1×1 Conv 代表 1×1 卷积操作，Bn 代表批归一化操作，激活函数为 Sigmoid 函数，Mul 代表乘操作。

3.4 特征融合模块

Spatial Path 输出的特征主要编码包含丰富细节的低层次信息，Context Path 主要编码包含上下文信息的高层次信息，为了将不同级别的特征表示融合到一起，BiSeNet[11] 设计了特征融合模块。与 SENet[16] 类似，通过级联操作计算特征向量来调整特征的权重以达到特征选择和融合的作用。与 BiSeNet[12] 中的特征融合模块不同，我们将原本的全局池化替换为全局卷积模块（GlobalConvBlock）[12]，使其计算的特征向量能更好地融合全局信息（所替换的模块如图 5 中红色虚线框所示）。特征融合模块的结构如图 5 所示。图 5 中，Concatenate 代表拼接操作；Conv 代表卷积操作；Bn 代表批归一化操作；ReLU 代表 ReLU 激活函数；GCN 代表全局卷积网络模块；Sigmoid 代表 Sigmoid 激活函数；Mul 和 Add 分别代表乘操作和加操作。

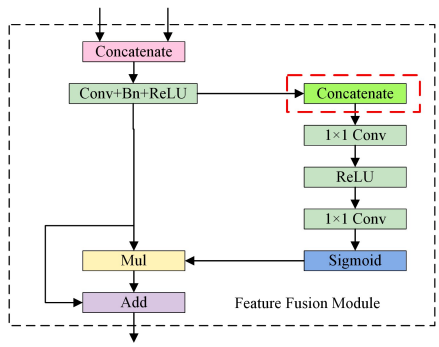


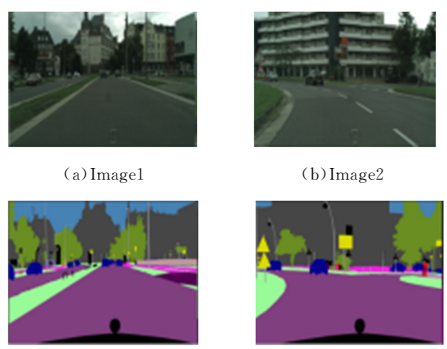
图 5 特征融合模块（电子版为彩色）

Fig. 5 Feature fusion module

4 实验结果与分析

4.1 数据集

在道路场景语义分割任务中，Cityscapes[17] 是当前较为权威的数据集，其在覆盖面积、复杂度及语义注释像素的数量方面远大于 CamVid[18] 数据集。Cityscapes[17] 覆盖了 50 个城市街道的春季、夏季和秋季的视频序列，共包含 5000 张高质量像素级标签的精细图像和 20000 张具有粗略标签的附加图像。其中，精细图像中包含训练样本 2975 张，验证样本 500 张，测试样本 1525 张。本实验利用 Cityscapes[17] 中的 5000 高质量像素级标签的精细图像进行实验，数据集样例如图 6 所示。



(a) Image1 (b) Image2

(c) Label1 (d) Label2

图 6 Cityscape 样例

Fig. 6 Cityscape examples

4.2 实验环境

本文实验采用 Resnet-18[19] 作为骨干网络。选择小批量随机梯度下降进行训练，具体参数为：batch size=8，momentum=0.9，weight decay=5×10^{-4}。采用“poly”学习率，power=0.9，初始学习率设置为 1×10^{-5}。模型性能采用平均交并比(mIoU)来衡量，所有实验都是在深度学习框架 PyTorch 上进行的，计算机配置如表 1 所列。

表 1 计算机软件/硬件配置表

Table 1 Result on Cityscape dataset

Hardware/Software Name	Model/Version
Processors	Intel(R) Xeon(R) Gold 6150
Graphics cards	NVIDIA TITAN Xp(12GB)
Ubuntu System	16.04
CUDA	10.1
Python	3.6.3
PyTorch	1.0.1

4.3 评价标准

本文采用平均交并比(MIoU)指标作为衡量标准，其是衡量语义分割任务中分割精度的重要指标。在语义分割任务中，MIoU 计算预测值集合与真实值集合的交集与并集之比，在每个类别上计算 IoU，取其平均值即为 MIoU。

$$MIoU=\frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{i=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}} \tag{2}$$

其中，$k+1$ 表示数据集包含 k 个物体类别和 1 个背景类别，p_{ii} 表示真正的像素数量，p_{ij} 和 p_{ji} 分别表示假正和假负的像素数量。

4.4 实验结果及分析

如表 2 所列，本文方法在 Cityscapes[13] 数据集上进行验证，将平均交并比(MIoU)提高到 78.28%。相比 BiSeNet[11] 网络其提高了 0.84%，说明本文提出的网络结构优于前者。

表 2 算法在 Cityscape 数据集上的表现

Table 2 Comparison of BiseNet algorithm with our algorithm on Cityscape dataset

Method	Mean IoU/%
DeepLabV3	72.33
PSPNet	74.53
DeepLabV3+	76.64
BiSeNet	77.44
Ours	78.28

为了说明本文提出的全局双边网络结构能够有效利用图像的上下文信息，本文设计如下实验。

为了说明所添加的全局分支结构的有效性，我们对比了未添加全局分支的网络结构与添加全局分支后的网络结构在 Cityscape[13] 上的表现，添加全局分支后将 MIoU 结果提升了 0.49%，具体结果如表 3 所列。

表 3 添加全局分支的结果

Table 3 GCN branch added results

Method	Mean IoU/%	
	val	test
BiSeNet	77.44	77.12
BiSeNet+GCN 分支	77.93	77.71

为了说明将 ARM 模块中的全局平均池化替换为全局卷积网络模块的有效性，我们对比了替换前与替换后的网络结构在 Cityscape[13] 上的表现，将注意力优化模块中的全局平均池化替换为全局卷积网络模块后，将 MIoU 结果提升了 0.14%，具体结果如表 4 所列。

表 4 替换 ARM 中的全局平均池化为全局卷积网络的结果

Table 4 Replace global average pooling in ARM module with global convolution network

Method	Mean IoU/%	
	val	test
BiSeNet+GCN 分支	77.93	77.71
BiSeNet+GCN 分支+ARM(GCN)	78.07	77.87

为了说明将 FFM 模块中的全局平均池化替换为全局卷积网络模块的有效性，我们对比了替换前与替换后的网络结构在 Cityscape[13] 上的表现，将特征融合模块中的全局平均池化替换为全局卷积网络模块后，将 MIoU 结果提升了 0.21%，具体结果如表 5 所列。

表 5 替换 FFM 中的平均池化操作为全局卷积网络的结果

Table 5 Replace global average pooling in FFM module with global convolution network

Method	Mean IoU/%	
	val	test
BiSeNet+GCN 分支+ARM(GCN)	78.07	77.87
BiSeNet+GCN 分支+ARM(GCN)+FFM(GCN)	78.28	78.03

结束语 本文提出了全局双边网络，将 Global Convolution Network[12] 中的采用大卷积核的思想应用到 BiSeNet[11] 网络结构中，获得了更好的效果，采用大的卷积核从一定程度上增大了感受野，使得像素点在预测类别时能够更好的结合周边像素的信息，提升了分割准确率。在以后的工作中，将会进一步探索语义分割网络结构并探究其中的因果关系。

致谢 感谢中国科学院自动化研究所为本研究提供 GPU 计算支持。

参 考 文 献

[1] LONG J, SHELHAMER E, DARRELL T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 39(4): 640-651.

[2] BADRINARAYANAN V, HANDA A, CIPOLLA R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.

[3] NOH H, HONG S, HAN B. Learning deconvolution network for semantic segmentation[C]// Proceedings of the IEEE International Conference on Computer Vision. 2015: 1520-1528.

[4] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]// International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

[5] Lin G, Milan A, Shen C, et al. RefineNet: Multi-path refinement networks with identity mappings for high-resolution semantic segmentation[J]. arXiv: 1611.06612.

[6] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 40(4): 834-848.

[7] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[J]. arXiv: 1706.05587, 2017.

[8] CHEN L C，ZHU Y，PAPANDREOU G，et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]// Proceedings of the European Conference on Computer vision (ECCV). 2018：801-818.

[9] ZHAO H S，SHI J P，QI X J，et al. Pyramid Scene Parsing Network[J]. arXiv：1612. 01105v2，2017.

[10] ZHANG R，TANG S，ZHANG Y，et al. Scale-Adaptive Convolutions for Scene Parsing[C]// 2017 IEEE International Conference on Computer Vision (ICCV). IEEE，2017.

[11] YU C，WANG J，PENG C，et al. BiSeNet：Bilateral Segmentation Network for Real－Time Semantic Segmentation[C]// Computer Vision－ECCV 2018. Lecture Notes in Computer Science，vol 11217.

[12] PENG C，ZHANG X，YU G，et al. Large Kernel Matters－Improve Semantic Segmentation by Global Convolutional Network [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE，2017.

[13] SIMONYAN K，ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science，2014.

[14] SZEGEDY C，LIU W，JIA Y，et al. Going Deeper with Convolutions[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston，MA，2015：1-9.

[15] CHEN L C，PAPANDREOU G，KOKKINOS I，et al. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs[J]. Computer Science，2014(4)：357-361.

[16] HU J，SHEN L，ALBANIE S，et al. Squeeze-and-Excitation Networks[J]. arXiv：1709. 01507v4，2019.

[17] CORDTS M，OMRAN M，RAMOS S，et al. The Cityscapes Dataset for Semantic Urban Scene Understanding[J]. arXiv：1604. 01685，2016.

[18] BROSTOW G J，SHOTTON J，FAUQUEUR J，et al. Segmentation and Recognition Using Structure from Motion Point Clouds [C]// European Conference on Computer Vision. Springer，Berlin，Heidelberg，2008.

[19] HE K，ZHANG X，REN S，et al. Deep Residual Learning for Image Recognition[J]. arXiv：1512. 03385，2015.

REN Tian-ci，born in 1995，postgraduate. His main research interests include computer vision and pattern recognition.

DING Wei-li，born in 1979，Ph.D，professor，Ph.D supervisor，is a member of China Computer Federation. Her main research interests include computer vision，pattern recognition and human-computer interaction.

（上接第 160 页）

Check24 色卡/ColorCheck＋Vectorscope 为仿真对象，结果表明，本文方法取得了很好的色彩校正效果，降低了色彩偏差。然而，由于本文使用的 Gamma 曲线对黑色部分调整过大，使得 ColorCheck24 中易受噪声影响的色块的噪声被进一步放大，造成了部分色块上的 RGB 偏差较大，因此，如何修改 Gamma 曲线，使整个图像的色彩获得更好的表现效果，是下一步的研究方向。

参 考 文 献

[1] CARROLL J. Smart cameras evolve to meet expanding machine vision needs[J]. VisionSystems Design，2019，24(8).

[2] GUO Y，GAO K，ZHU J，et al. A color camecorrection method based on LASSO regression model [J]. Imaging Science and P hotochemistry，2017，35(2)：153-161.

[3] LEVENSON R，BRACKEN C，SHARMA C，et al. Calibration between trigger and color：Neutralization of a genetically encoded coulombic switch and dynamic arrest precisely tune reflection assembly[J]. The Journal of biological chemistry，2019，294(45).

[4] ZHOU Z M. Spatial consistency correction and yarn color measurement of multispectral imaging system [D]. Hangzhou：Zhejiang University，2018.

[5] GONG R. Evaluation method and Application Researchonimage display performance of intelligent mobile terminal [D]. Hangzhou：Zhejiang University，2014.

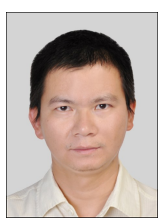
[6] LU Z G，HE P，ZHI L J，et al. Study on trigonometric model fitting based on least square polynomial [J]. Applied Optics，2019，40(5)：853-858.

[7] FANG J，ZHANG Y，CUI W，et al. Outdoor light estimation algorithm based on white balance correction[J]. Computer Science，2019，46 (a)：211-214.

[8] BA T，GUANG D，ROBERT R. A fast white balance algorithm based on pixel greyness[J]. Signal，Image and Video Processing，2017，11(3).

[9] WANG F，WANG W. A fast automatic white balance algorithm with dark channel priority [J]. Optoelectronic Engineering，2018，45 (1)：7.

[10] WU Q Q，SONG X N. A face anti-fraud method based on multi-channel feature and least square decision[J]. Computer Application research，2019，163.



TANG Jia-lin，born in 1982，Ph.D，lecturer. His main research interests include artificial intelligence and so on .

GUO Yan-feng，born in 1978，Ph.D. His main research interests include artificial intelligence and so on.