Computer Science and Application 计算机科学与应用, 2020, 10(7), 1327-1334
Published Online July 2020 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2020.107137

Hans汉斯

RGB-D Object Recognition Based on ResNet

Ximin Wang, Bin Ding*, Yu Zhang, Xunguang Ju

Xuzhou University of Technology, Xuzhou Jiangsu
Email: *872167234@qq.com

Received: Jun. 24th, 2020; accepted: Jul. 8th, 2020; published: Jul. 15th, 2020

Abstract

Traditional object recognition research is generally based on RGB image and gray image. RGB image and gray image have their own limitations. Due to the lack of shape information of the object surface, it is easy to make mistakes when recognizing objects with similar color but different shapes. Using RGB-D cameras, we can obtain the depth value of each pixel while obtaining high-resolution RGB images. The depth data contains the information about the shape of the object, which can provide new features for object recognition. This paper mainly studies the RGB-D object recognition algorithm based on deep learning, selects the V1B type of ResNet residual network, and sets the convolution neural network model suitable for extracting RGB-D image features through manual parameter adjustment. Two ResNet V1b network models are used to extract effective RGB features and depth features respectively. In order to extract more diverse features, dilated convolution is added to the residual network. A full connection method is used to fuse RGB features and depth features. Experiments show that the neural network structure used in this paper is effective for RGB-D object recognition, and the contrast experiments show that the introduction of dilated convolution effectively improves the recognition rate of RGB-D images.

Keywords

RGB-D Camera, Object Recognition, Convolutional Neural Network, ResNet, Dilated Convolution

基于ResNet的RGB-D物体识别

王熙敏，丁　宾*，张　宇，鞠训光

徐州工程学院，江苏　徐州
Email: *872167234@qq.com

收稿日期：2020年6月24日；录用日期：2020年7月8日；发布日期：2020年7月15日

*通讯作者。

文章引用：王熙敏，丁宾，张宇，鞠训光. 基于 ResNet 的 RGB-D 物体识别[J]. 计算机科学与应用，2020，10(7): 1327-1334. DOI: 10.12677/csa.2020.107137

摘　要

传统的物体识别研究一般是基于RGB图像和灰度图像，RGB图像和灰度图像自身具有一定的局限性，由于缺少物体表面的形状信息，对颜色相近但形状不同的物体进行识别时容易造成错误。使用RGB-D相机可以在获取高分辨的RGB图像的同时，获取每个像素的深度值，深度数据中蕴含着有关物体形状的信息，可以为物体识别提供新的特征。本文主要研究了基于深度学习的RGB-D物体识别算法，选择了ResNet残差网络的V1b型，通过手动调参设置适合于提取RGB-D图像特征的卷积神经网络模型。通过两个ResNetV1b网络模型分别提取有效的RGB特征和深度特征，为了提取到更多样性的特征，在残差网络子模块中加入了空洞卷积；采用了全连接的方法对RGB特征和深度特征进行融合。实验证明了本文所采用的神经网络结构对于RGB-D物体识别是有效的，对比实验表明空洞卷积的引入有效提高了RGB-D图像的识别率。

关键词

RGB-D相机，物体识别，卷积神经网络，ResNet，空洞卷积

1. 引言

在传统的基于 RGB 图像或者灰度图像的物体识别研究中，通常对图像提取各种人工定义的特征，例如 HOG 特征[1] [2]、LBP 特征[3]、SIFT 特征[4]、Haar-like 特征[5] [6]等。这些特征提取的方法只能在特定的识别任务下才能取得一定的效果，鲁棒性并不高。人工定义的特征只代表了图像特征的一部分，更多有效的、鉴别能力强的特征被忽略了。比如 SIFT 特征对特征点构造 128 维的向量，如果图像没有足够多的纹理将会造成误匹配或者是没有办法匹配，此外，SIFT 特征完全忽略了色彩信息。对于光照变化不均匀的图像，LBP 特征无法稳定提取，基于 LBP 特征的物体识别方法表现不佳。基于神经网络与深度学习的特征提取避免了人工定义的特征提取，通过神经网络可以提取到更有鉴别能力的特征。

传统的物体识别通常基于 RGB 图像，随着深度学习的深入研究，基于 RGB 图像的物体识别已经取得重要进展。在一些特定的应用场合，比如室内机器人或工业流水线上的物体识别任务中，使用 RGB-D 相机采集图像，除了在每个像素点获取 RGB 值之外，还能获得该像素点的深度值。深度值含有新的信息，有望为物体识别提供更多有用的特征。Kevin Lai 等来自美国华盛顿大学的研究学者通过对于 RGB-D 图像的多年的研究，建立了一个具有多分类、多角度的 RGB-D 物体数据库[7]。在国内，来自宁波大学的卢良峰等人提出的基于原始图像层[8]进行融合得到 RGB-D 物体的特征进行 RGB-D 物体识别。在 2015 年由来自合肥工业大学的胡良梅，杨慧等人[9]提出的通过多核学习以及自适应加权的方法进行 3D 目标识别方法。2018 年同样来自合肥工业大学的张治安等人提出基于稀疏联结卷积递归神经网络的 RGB-D 图像识别算法[10]。随着对于 RGB-D 物体的识别的研究，基于多模态深度学习的 RGB-D 物体识别等方法相继被提出。尽管近年来深度学习算法得到了广泛的研究，但是它们仍是主要基于 RGB 图像和灰度图像的研究，RGB 颜色和深度数据属于多模态数据，RGB 特征和深度特征的提取以及特征融合方法还没有得到充分的研究。本文主要研究了基于深度学习的 RGB-D 识别算法框架，以及 RGB 特征和深度特征的提取和融合方法。

2. RGB-D 物体识别中的深度神经网络

ResNet 卷积神经网络网络结构

ResNet 网络[11]是在 2015 年由何凯明、张翔宇和任少卿共同提出的，该网络是基于 VGG-19 网络进行的改进，因为随着 VGG 网络深度的增加，会出现梯度爆炸或者梯度消失的问题，而 ResNet 网络的出现使得这个问题迎刃而解，并还减少了参数的数量。ResNet 提出了一个全新的思想，假设需要涉及到一个网络层且存在最优化的网络层次结构，那么大多时候所设计出的深层次网络的很多层实际上是冗余层。如果冗余层的网络层可以完成恒等映射，能够保证所有经过这些恒等层的输入输出完全相同。ResNet 网络由 Building Block 结构组成。Building Block 结构使用了一个残差支路和一个 Short Connection 支路。

大量实验证明，残差网络的确解决了退化的问题，在训练集和测试集上都是网络层数越多，错误率越小。图 1 为两种残差网络层数不同的对比，其中横坐标指迭代的次数，纵坐标指错误率。在具有 18 层和 34 层残差网络的 ResNet 上，随着迭代次数的逐次增加，可以看到具有残差网络层数越多的 ResNet 发生错误率越小。

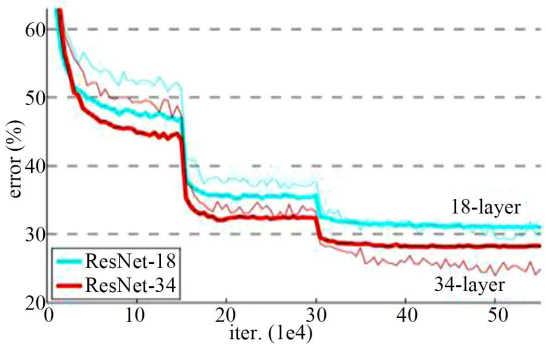

Figure 1. ResNet error rate trend graph [11]
图 1. ResNet 错误率趋势图[11]

3. 基于 ResNet 的 RGB-D 物体识别算法

3.1. 算法流程

本文所提出的基于深度学习的 RGB-D 图像识别算法的基本思路是，分别使用两个 ResNet 网络实现对 RGB 特征和深度特征的提取，通过特征融合算法进行 RGB 特征和深度特征的融合，最终通过 softmax 分类器进行分类。算法的总体流程图如图 2 所示。

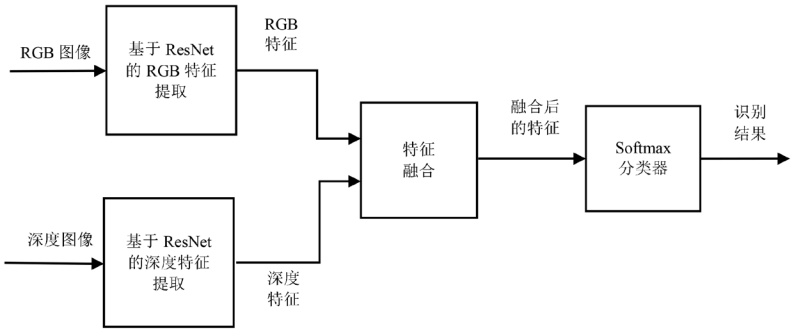

Figure 2. Overall flow chart of RGB-D object recognition algorithm based on ResNet
图 2. 基于 ResNet 的 RGB-D 物体识别算法总体流程图

3.2. RGB 特征提取

基于 ResNet 的 RGB-D 物体识别算法中提取 RGB 特征和深度特征的流程图如图 3 所示。

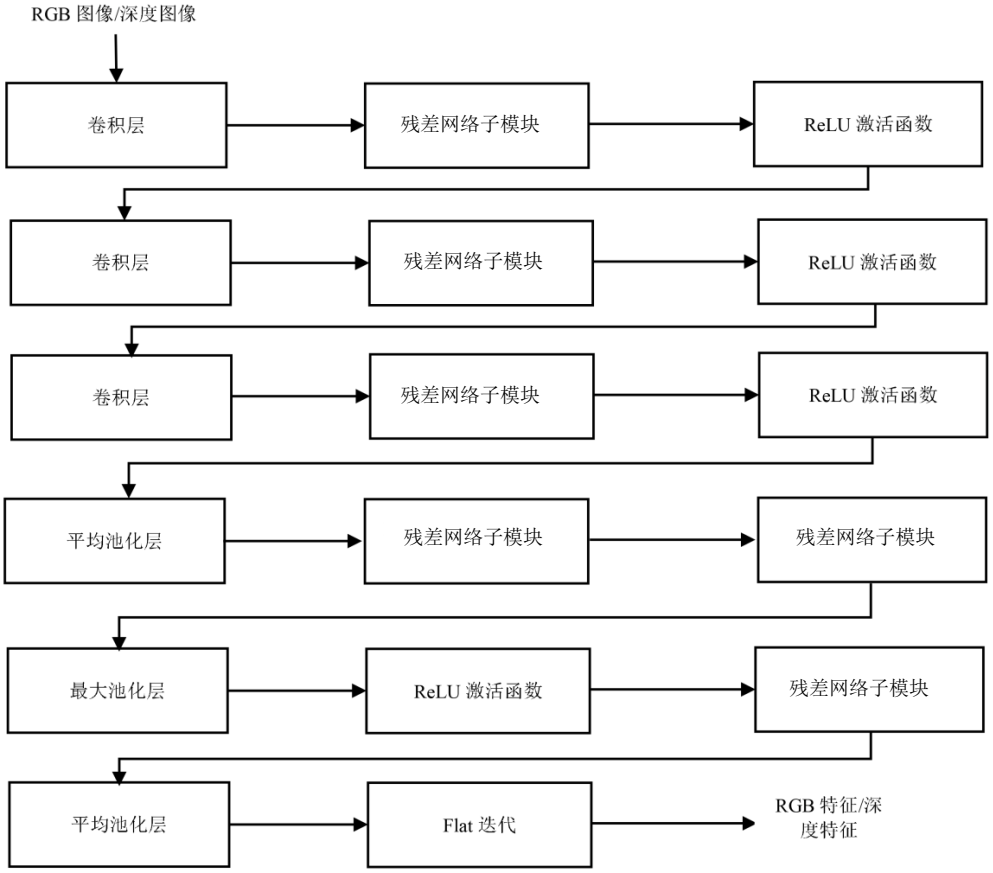

Figure 3. RGB features/depth features extraction flow chart
图 3. RGB 特征/深度特征提取流程图

要提取到 RGB 特征需要将 RGB 图像数据输入到 ResNet 中得到，首先在卷积层中对输入的图像做 2D 卷积，其中要将图像数据进行升维，为了能够最大程度上的提取到特征数目并且做一个线性变换，降低参数的数量，其数学公式如表达式(1)所示。

$$out\left(N_i, C_{out_j}\right) = bias\left(C_{out_j}, k\right) + \sum_{k=0}^{C_{in}-1} weight\left(C_{out_j}, k\right) * input\left(N_i, k\right) \tag{1}$$

其中，N 意为 Batch Size，C 意为 Channels。

RGB 图像数据通过卷积层经过升维后，提取到的特征数目是众多的，之后进入到残差网络子模块中，残差网络子模块流程图如图 3 所示，即输入到三层残差网络，分别为 1 × 1、3 × 3、1 × 1 卷积，经过了先降维在升维的操作，相当于对特征值进行了二次筛选，去掉了部分没有区别力的 RGB 特征，再次减少了参数的数量而且不会出现退化的现象，另外错误率亦大大的降低了，再通过 ReLU 函数的激活，因为线性模型的表达能力是远远不够的，其原因是如果输入和输出都是线性组合的话，其效果和没有隐藏层是一样的结果，无论网络叠加多少层数，其结果不过是矩阵相乘，故经过 ReLU 函数的激活从线性变为非线性可以将特征保留并能够映射出来，即将每个节点的输入映射到输出端。

ReLU 函数的表达式为：

$$f(x)=\max(0,x) \tag{2}$$

由表达式(2)所示，激活函数 ReLU 的本质是取最大值，当输入的值是负数的时候，节点不会被激活，故输出的值为 0，这说明了在同一时间内只有部分的节点被激活，所以网络会变得很稀疏，进而使得计算变得更加有效率。RGB 图像数据在经过重复多次的卷积层和残差网络子模块以及 ReLU 的激活后，其中提取到的 RGB 特征值进入到 MaxPooling 进行最大池化作用于 RGB 图像中的不重合区域，使 RGB 特征能够保持平移、旋转以及尺度的不变性和增大感受野的范围，MaxPooling 的作用如图 4 所示。

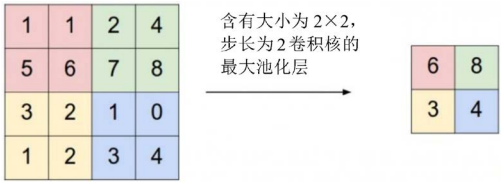

Figure 4. The role of MaxPooling
图 4. MaxPooling 的作用

通过相同的多个特征提取和 ReLU 函数以及最大池化层，图像进入到平均池化层进行特征提取，可以减小邻域大小受限后直接造成的估计值的方差值，从而能够更多的保留图像的重要信息，随后进行 flat 迭代，是为对 RGB 图像的特征的扁平化排序的处理，使得其 RGB 特征能够便于更快地索引随后输出，至此完成 RGB 特征提取。

3.3. 深度特征提取

要提取到深度特征需要将深度图像数据输入到ResNet中得到，其具体的流程与RGB特征通过将RGB图像数据输入到 ResNet 中提取相似，流程图如图 3 所示。其不同点在于将深度图像输入到卷积层的时候，有 N 个卷积层可以去进行循环的卷积，进行循环的卷积可以充分地提取到深度特征，有利于之后的物体识别的准确率的提升。从 N 个卷积层输出后，依次输入到与 RGB 特征提取的相同的多个残差网络子模块、ReLU 函数、最大池化层以及平均池化层和 flat 迭代进行特征提取。

当上一层的输出作为输入进入残差网络子模块时，首先进行下采样，采用下采样的作用主要是经过处理之后让图像变得符合一定的大小并且可以成为物体所对应的图像的缩略图，而且还可以相应的减少计算量。下采样的原理是对于一幅像素大小为 A*B 的图像，对该图像进行 C 倍的下采样，下采样过后得到的图像的像素大小为(A/C)*(B/C)，该步骤中最重要的一点是 C 必须要为 A 和 B 的公约数。另外当图像以矩阵的形式呈现时，即为把最初的图像 C*C 大小的窗口内的图像变成一个像素点，且该像素点的值的大小与 C*C 窗口内所有像素的平均值相等。所以，经过下采样后，应该输出的为 RGB 图像或者深度图像的一个小的缩略图。该输出进而经过平均池化层进行 RGB 图像或者深度图像的平均特征值的提取和卷积层的对于 RGB 图像或者深度图像的特征值的再卷积之后输入到残差网络的 1 × 1、3 × 3、1 × 1 三层卷积，先降维再升维对特征值进行二次筛选，去掉部分没有区别力的 RGB 特征，再减少了参数的数量而且不会出现退化的现象之后再进行空洞卷积。空洞卷积的作用是在不做池化的情况下保持信息的完整不再损失和扩大了其感受野的面积，使得每次卷积之后的输出都可以含有大范围的信息，从而可以简洁的提高 RGB-D 物体的图像的识别的准确率。另外通过查阅文献中的理论知识可以复现出一个 7 × 7 的卷积层，其的正则等效于 3 个 3 × 3 卷积层还可以大幅度的减少参数的设置。

经过残差网络子模块时内部详细流程图如图 5 所示。

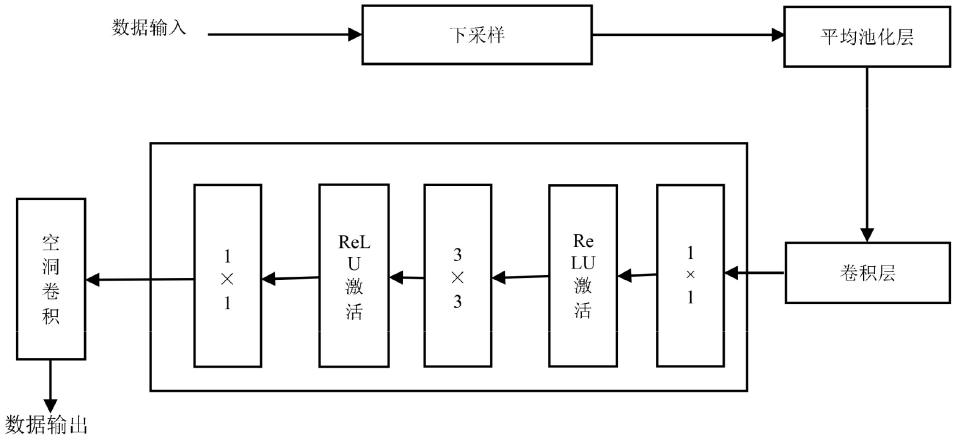

Figure 5. Detailed internal flowchart when passing through the residual network
图 5. 经过残差网络子模块时内部详细流程图

3.4. RGB 特征和深度特征融合

RGB 特征和深度特征融合流程图如图 6 所示。在对 RGB 特征和深度特征进行融合时，首先通过决策树算法将融合网络参数初始化，对提取的 RGB 特征和深度特征所输入的网络赋予不同的参数，这些参数确定了 RGB 特征和深度特征在融合时的比例系数，反映了两种特征对分类结果的影响权重。然后选择最大池化法提取更有效的 RGB-D 特征，再经过 softmax 分类器进行分类，获得 RGB-D 物体的识别结果。

采用决策树算法的原因是，决策树算法具有较高的分类精度，而且生成的模式简单，其数学表达公式如表达式(3)所示。

$$H(X,Y)=-\sum_{i=1}^{n}p(x_i,y_i)\log p(x_i\mid y_i)=\sum_{j=1}^{n}p(y_j)H(X\mid y_i) \tag{3}$$

由表达式(3)条件熵表达式可知，在知道 X 值时可以有效的减少 Y 值的不确定性，从而有效地提高分类的准确率。

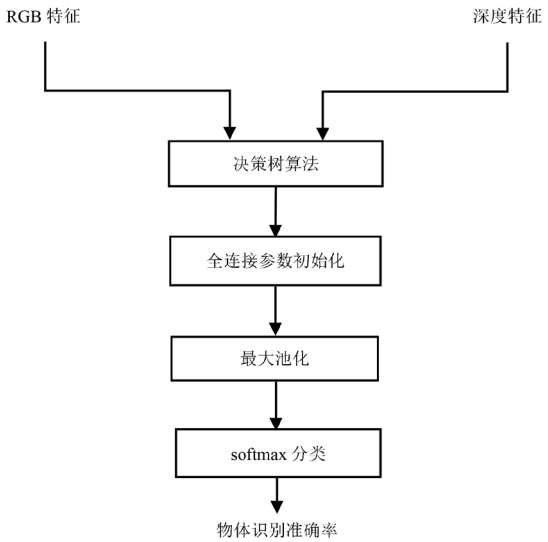

Figure 6. Flowchart of feature fusion of RGB and depth features in the image layer
图 6. RGB 特征和深度特征在图像层进行特征融合流程图

4. 实验结果及分析

4.1. 基于 2D3D 数据库的实验结果及分析

为了验证本文算法的有效性，采用与文献[7]相同的实验过程，在 2D3D 数据库[12]上进行了实验。在每一次实验中，随机的按照 7:3 的比例将 2D3D 数据库分为训练集和测试集两部分。最终确定的训练集为 110 个实例，约为 4200 张图片，测试集为 46 个实例，约为 1800 张图片。此次实验共计重复 50 次，2D3D 数据库在每一次实验中都会被随机切割，最终的实验结果为 50 次实验结果的平均值。

表 1 将本文算法和其它典型算法在 2D3D 数据库进行物体识别的准确率进行对比。本文提出的基于 ResNet 的 RGB-D 物体识别算法取得了较好的识别率，对比实验证明，本文算法能较好地实现 RGB 特征和深度特征的提取与融合。

Table 1. Accuracy results and comparison of various algorithms for object recognition (%)
表 1. 多种算法物体识别的准确率结果和对比(%)

算法	RGB	Depth	RGB-D
SP + HMP [13]	86.3	87.6	91.0
KSAE-SPMP [8]	87.5	87.6	92.4
CNN-SPM-RNN [15]	89.4	88.5	92.9
本文算法(未加入空洞卷积)	89.3	91.6	92.5
本文算法(加入空洞卷积)	90.5	92.1	93.2

4.2. 基于 RGB-D 数据库的实验结果及分析

采用与 4.1 节相同的实验过程，对本文算法在华盛顿大学 RGB-D 数据集上进行了物体识别实验。最终确定的训练集为 210 个实例，约为 3000 张图片，测试集为 90 个实例，约为 12,000 张图片。此次实验共计重复 50 次，RGB-D 数据库在每一次实验中都会被随机切割，最终的实验结果为 50 次实验结果的平均值。

表 2 将本文中提出的基于 ResNet 的 RGB-D 物体识别算法和其它典型算法进行了对比。本文的算法取得了较好的分类效果，证明了基于 ResNet 的 RGB 特征与深度特征提取和融合方法对于 RGB-D 物体识别是有效的。

Table 2. Various algorithms based on RGB-D dataset object recognition accuracy results (%)
表 2. 多种算法基于 RGB-D 数据库的物体识别的准确率结果(%)

Method	Both
SP + HMP [13]	87.5 ± 2.9
CNN-RNN [14]	86.8 ± 3.3
CNN-SPM-RNN + CT [15]	90.7 ± 1.1
Fus-CNN [16]	91.3 ± 1.4
本文算法(未加入空洞卷积)	90.2 ± 1.8
本文算法(加入空洞卷积)	91.4 ± 1.2

以上通过分别将 RGB-D 数据集和 2D3D 数据集输入到基于 ResNet 的 RGB-D 物体识别算法来测试得到算法的准确率，经实验证明，空洞卷积的引入使得 RGB-D 物体识别率提高了 0.7%~1.2%。经分析，空

洞卷积之所有能有效提高识别率，在于它可以使神经元的感受野变大，从而在特征提取时能提取到更全面、有效的特征。

5. 结论

本文给出了一种基于 ResNet 的 RGB-D 物体识别算法，利用 ResNet 从 RGB 图像和深度图像中提取有效的特征，并采用了决策树算法和全连接方式将 RGB 特征和深度特征进行融合；为了提高特征提取的有效性，在残差网络子模块中引入了空洞卷积。实验结果表明，该算法在 RGB-D 物体识别的准确率方面取得了较好的效果。将来需要对网络结构和特征融合方法进一步优化，以进一步提高识别正确率。

基金项目

徐州市科技计划项目(KC17078)。

参考文献

[1] Behera, S.K., Rath, A.K. and Sethy, P.K. (2020) Maturity Status Classification of Papaya Fruits Based on Machine Learning and Transfer Learning Approach. *Information Processing in Agriculture*. https://doi.org/10.1016/j.inpa.2020.05.003

[2] Wang, Y., Li, M., Zhang, C., Chen, H. and Lu, Y.M. (2020) Weighted-Fusion Feature of MB-LBPUH and HOG for Facial Expression Recognition. *Soft Computing*, **24**, 5859-5875. https://doi.org/10.1007/s00500-019-04380-x

[3] Kaplan, K., Yılmaz, K., Melih, K. and Metin Ertunç, H. (2020) Brain Tumor Classification Using Modified Local Binary Patterns (LBP) Feature Extraction Methods. *Medical Hypotheses*. https://doi.org/10.1016/j.mehy.2020.109696

[4] Lowe, D.G. (2004) Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, **60**, 91-110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

[5] 张彩丽, 刘广文, 詹旭, 才华, 刘智. 基于新增haar特征和改进AdaBoost的人脸检测算法[J]. 长春理工大学学报(自然科学版), 2020, 43(2): 89-93.

[6] 邢益铭, 野莹莹, 程立英, 裴金鹏, 林月, 许翔宇. 基于 Haar-AdaBoost 人脸检测算法的研究[J]. 装备制造技术, 2020(3): 67-70 + 75.

[7] Lai, K., Bo, L., Ren, X., *et al*. (2011) A Large-Scale Hierarchical Multi-View rgb-d Object Dataset. 2011 *IEEE International Conference on Robotics and Automation*, Shanghai, 9-13 May 2011, 1817-1824. https://doi.org/10.1109/ICRA.2011.5980382

[8] 卢良锋. 基于 RGB-D 物体识别的深度学习算法研究[D]: [硕士学位论文]. 宁波: 宁波大学, 2015.

[9] 胡良梅, 杨慧, 张旭东, 董文菁, 陈仲海. 融合 RGB 特征和 Depth 特征的 3D 目标识别方法[J]. 电子测量与仪器学报, 2015, 29(10): 1431-1439.

[10] 张治安, 张旭东, 张骏. 基于稀疏联结卷积递归神经网络的RGB-D图像识别算法[J]. 合肥工业大学学报(自然科学版), 2018, 41(5): 582-588.

[11] He, K., Zhang, X., Ren, S., *et al*. (2016) Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision & Pattern Recognition*, Las Vegas, 26 June-1 July 2016, 770-778. https://doi.org/10.1109/CVPR.2016.90

[12] Browatzki, B., Fischer, J., Graf, B., *et al*. (2011) Going into Depth: Evaluating 2D and 3D Cues for Object Classification on a New, Large-Scale Object Dataset. *IEEE International Conference on Computer Vision Workshops*, *ICCV* 2011 *Workshops*, Barcelona, 6-13 November 2011, 1189-1195. https://doi.org/10.1109/ICCVW.2011.6130385

[13] Bo, L., Ren, X. and Fox, D. (2013) Unsupervised Feature Learning for RGB-D Based Object Recognition. In: *Experimental Robotics*, Springer International Publishing, Berlin, 387-402. https://doi.org/10.1007/978-3-319-00065-7_27

[14] Socher, R., Huval, B., Bath, B.P., *et al*. (2012) Convolutional-Recursive Deep Learning for 3D Object Classification. *NIPS* 2012, Lake Tahoe, 3-6 December 2012, 665-673.

[15] Cheng, Y.H., Zhao, X., Huang, K.Q., *et al*. (2015) Semi-Supervised Learning and Feature Evaluation for RGB-D Object Recognition. *Computer Vision and Image Understanding*, **139**, 149-160. https://doi.org/10.1016/j.cviu.2015.05.007

[16] Eitel, A., Springenberg, J.T., Spinello, L., *et al*. (2015) Multimodal Deep Learning for Robust RGB-D Object Recognition. *Proceedings of IROS*, Hamburg, 28 September-2 October 2015, 681-687. https://doi.org/10.1109/IROS.2015.7353446