中图法分类号：TP391　　**文献标识码**：A　　**文章编号**：
论文引用格式：

多模态深度神经网络的固废对象分割

张剑华,陈嘉伟,张少波,郭建双,刘盛

浙江工业大学计算机科学与技术学院, 杭州, 310023

摘　要：目的 对城市进展过程中产生的建筑固废进行处理，并将之转换为资源和能源，是极佳的保护环境的经济发展模式。然而人工分拣处理存在着效率慢、污染严重、对人身危害大等问题。目前工业界在探索一种有效的基于机械臂自动抓取的建筑固废自动分拣系统，其中图像分割技术是非常必要的一个环节。但是工业现场的环境因素造成固废对象的颜色严重退化，会影响最终的固废对象分割。本文针对建筑固废图像分割对象难度大的现状，提出一种基于多模态深度神经网络的方法来解决固废对象分割问题。**方法** 首先, 在颜色退化严重的场景下，把RGB图像和深度图一起作为深度卷积神经网络的输入，利用深度卷积神经网络进行高维特征学习，通过softmax分类器获得每个像素的标签分配概率。其次，基于新的能量函数建立全连接条件随机场，通过最小化能量函数寻找全局最优解来分割图像，从而为每一类固废对象产生一个独立的分割块。最后，利用局部轮廓信息计算深度梯度，实现同一类别的不同实例的固废对象精确分割。**结果** 在固废图像测试集上,该方法取得了90.02%均像素精度和89.03%均交并比（MIOU）。此外，与目前一些优秀的语义分割算法相比，也表现出了优越性。**结论** 本文提出的方法能够对每一个固废对象同时进行有效的分割和分类，为建筑垃圾自动分拣系统提供准确的固废对象轮廓和类别信息，从而方便实现机械臂的自动抓取。

关键词 ：　多模态信息；固废对象分割；卷积神经网络；条件随机；深度梯度

Multimodal Deep Neural Network for Construction Waste Object Segmentation

Zhang Jianhua, Chen Jiawei, Zhang Shaobo, Guo Jianshuang, Liu Sheng

College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

Abstract: Objective Construction waste is no longer useless nowadays. It is an excellent economic development mode to protect the environment by recycling waste generated during construction and converting it into resources and energy. The current situation of construction waste in China has become more and more severe. With the development of urbanization, the old buildings are demolished and rebuilt and especially are replaced by skyscrapers. And once inhabited areas have been gradually transformed into cities with ever expanding sizes of this cities. These cities have high speed development along with serious hidden dangers. Construction waste generated by a large number of construction work sites have become more and more difficult to ignore. Urban construction waste refers to all kinds of construction waste generated during the construction, transformation, decoration, demolition and laying of various buildings and structures and their auxiliary facilities. It mainly includes muck, waste concrete, waste brick, waste pipe, waste wood, etc. According to statistics, the amount of construction building waste in China now accounts for 30% to 40% of the total amount of municipal waste. In the next ten years, China will produce more than 1.5 billion tons of construction waste per year on average. It is estimated that by 2020, the construction waste will reach 2.6 billion tons; by 2030, it will reach 7.3 billion tons. Resource utilization and recycling is are inevitable choices for dealing with construction waste in buildings. To effectively deal with construction waste in buildings, you can start from its characteristics. The construction waste is a mixture of various

基金项目:国家自然科学基金（No.61876167，U1509207）
收稿日期:　　；　**修回日期**:
Suppoted by: National Natural Science Foundation of China (No. 61876167, U1509207)

building materials wastes, which are actually a unutilized resources. In the 1990s, several communities in California first launched a single-stream recycling project, which referred to the mixture of all paper products, plastics, glass, metal, and other waste. It was separated into single item by a sorting system. In the sorting system, waste was mainly processed by a combination of hardware equipment and manpower. The system was not fully automated and relied mainly on human recycling, so it was not efficient. This was a meaningful attempt to let people understand the feasibility of recycling waste. For construction waste, there are many construction wastes, such as waste bricks, waste rock, scrap steel, etc., which can be recycled after being sorted, rejected or crushed. However a system like a single-stream recycling project is not capable to handle a large amount of construction waste. With the development of artificial intelligence technology, the use of intelligent robotic equipment in the field of construction waste recycling can greatly improve the capability, efficiency and safety of recycling. Among them, the robotic arm is the most widely used automated mechanical device in the industrial field. It can quickly grasp objects and can work continuously. The emergence of robotic arms provides a new and efficient solution for the automatic sorting of construction waste in buildings. The use of robotic arms to sort construction waste is a revolutionary innovation for the construction waste treatment industry. For the robotic arm grabbing task, the position information and contour information of the object are indispensable. The application of computer image segmentation algorithms in this scene is undoubtedly very suitable. Through the image segmentation algorithms, the construction waste image can be accurately segmented to obtain the position and contour of each object. Combining robotic arms and image segmentation algorithms to achieve efficient construction waste recovery is worth looking forward to. However, due to the characteristics of industrial sites and construction waste objects, it is very difficult to segment construction waste objects from the obtained construction waste images by the segmentation algorithms. In terms of the difficulty of object segmenting in construction waste image, this paper proposes a construction waste object segmentation method based on multimodal information deep neural network to solve the image segmentation problem, and provides accurate construction waste object contour and category information for the construction waste automatic sorting system. Therefore, it is capable to realize automatic grabbing using the robot arm. **Method** First of all, in the scenes with severe color degradation, feature learning with RGB images alone does not meet the actual needs. Therefore, it is necessary to train the salient features with depth information. We treat the RGB image and the corresponding depth image as the input of the deep convolutional neural network. The deep convolutional neural network is used to perform high-dimensional feature learning, and the feature maps obtained from the convolutional layers of the last layer are weighted and summed, and then feed as input data of the Softmax classifier, finally we obtain the label allocation probability of each pixel. Based on the probability that each pixel belongs to a category, we construct a multi-label full connected conditional random field. The unary energy term treat each pixel as an independent item, without considering the relationship between the pixels. The binary energy term represents the relationship among pixels. So that similar pixels are divided into the same category, and pixels with large differences between each other are assigned to different categories, which makes the segmented edges smoother. We able to obtain more accurate segmentation results. Therefore, according to the actual situation, We propose an energy function suitable for construction waste objects. The global optimal solution is obtained by minimizing the energy function to segment the object in the image, thereby generating an independent segmentation block for each type of construction waste object. Finally, Fine segmentation of local ambiguous regions is done based on the depth gradient information. The ambiguous area refers to an adhesion areas between construction waste objects that are difficult to distinguish due to degradation of visual characteristics. The depth gradient information is used to obtain the local depth edge map, from which the local ambiguity area is extracted. For the local ambiguity area, the algorithm extracts the effective internal edge to segment the adhesion objects belonging to the same class. **Result** On the construction waste image test set, our method achieves 90.02% Mean Pixel Accuracy(MPA) and 89.03% Mean Intersection Over Union(MIOU). In addition, compared with some excellent semantic segmentation algorithms, the experimental results show that the proposed method obtain better performance and improve the segmentation accuracy. **Conclusion** The algorithm proposed in this paper can segment and classify most construction waste object effectively at the same time, and provide accurate contour and classification

information of the construction waste object to a construction waste automatic sorting system, so as to facilitate the automatic grasping construction waste by the robotic arm.

Key words: Multimodal information； construction waste object segmentation； convolutional neural network； conditional random； depth gradient

0 引 言

随着我国基础建设的不断加快，城市化的不断扩展，伴随着产生大量建筑固废。城市建筑固废主要是指在对各类建筑物设施进行建设、改造、装修、拆除、铺设等过程中产生的各类固体废物[1]，主要包括渣土、废旧混凝土、废砖石、废旧管材、废旧木材等。针对这些固废，存在着难处理，效率低，环境污染大等一系列问题。

传统的方法大多采用填埋，极少部分用作焚烧，占用土地资源，造成严重的污染。对固废有效的资源回收与利用，是目前建筑固废处理的热点与难点。

作为循环经济[2]的一项重要内容，建筑固废的资源化利用成为了工业上的一个难题。随着人工智能技术的发展，在建筑固废资源化领域使用智能机器人设备可以极大的提升回收的效率和保障安全。其中机械臂是机器人技术在工业领域中得到最广泛实际应用的自动化机械装置，它能够快速的抓取物体，并且可以持续的工作。机械臂的出现，为建筑固废的自动分拣提供了一种新的、高效的解决方案。为了完成固废对象的自动抓取，对象的轮廓信息是不可或缺的。所以，图像分割技术在其中的应用是非常必要的。然而由于工业现场的环境因素，摄像头往往处于比较恶劣的环境下，如：震动、高灰、物体的快速移动，在这些情况下获取到的固废对象图像，对其进行分割是非常困难的。

在实际生产线上，到达机械臂自动分拣这条皮带线上需要分拣出来的固废主要包含废旧木材（木头）、废旧混凝土（石头）、废砖石（砖头）这 3 类。针对这 3 类固废，本文提出一种基于多模态深度神经网络的图像分割算法解决固废对象分割问题。首先，利用深度相机采集固废对象的三维信息和颜色信息，用卷积神经网络提取高维多模态类别特征；然后用获得的特征信息训练 softmax 分类器，产生的每个像素标签分配概率作为全连接条件随机场(DenseCRF)[3]的输入，进行全局的固废类别分割，输出每个类别的分割信息，完成固废分类的第一步任务；最后基于深度边缘轮廓完成固废对象分割的最终任务，从而在根本上解决了固废对象分类分割问题。

1 图像分割技术现状

1.1 经典的图像分割方法

传统的图像分割方法主要包括阈值法[4]，边界检测法[5]，区域法[6]等。2000 年左右，开始出现基于超像素的图像分割[7]方法，这种处理方式将具有相似特征的像素分组。根据算法实现原理不同，超像素方法可以分为基于图论的方法[8]和基于聚类的方法[9]。前者的主要思想是将图像映射为带权无向图，图像的像素对应图的顶点，像素信息对应顶点属性，像素之间的相似性（或差异性）对应边的权值，将图像分割问题转换为图的顶点标注问题。后者的主要思想是根据图像中的单个像素及像素之间的相互信息，如颜色、亮度、纹理等，利用数据挖掘中聚类算法，将具有相近特征的相邻像素聚到同一个像素块。

1.2 图像分割中深度学习方法的应用

近年来深度学习方法在物体检测、图像分类、目标识别等计算机视觉领域得到了广泛的应用，在图像分割方面也进行了一些有益的尝试。基于学习的语义分割方法将图像分割问题看做是图像单个像素的分类问题，先以有标注的图像作为训练样本，训练支持向量机(SVM) [10]、逻辑回归(L R) [11]等分类器，再以训练好的分类器对输入图像进行逐像素分类，根据像素分类情况得到图像分割结果。语义分割的输出是二维分割图像，保留空间信息，从而使得分割结果更加准确。UC Berkeley 的 Long 等人提出 FCN[12]全卷积神经网络用于图像的分割。该网络试图从抽象的特征中恢复出每个像素所属的类别，即从图像级别的分类进一步延伸到像素级别的分类。Zhao 等人提出 PSPNet[13]金字塔场景解析网络。该网络提出金字塔池化模块来聚合背景信息，利用全局场景信息提升分割效果。

1.3 基于 RGB-D 的图像分割

对于颜色退化或者场景复杂的图像，基于可见光图像的分割算法，得到的分割结果并不能够满足需求。因此结合 3D 传感器，引入深度数据作为额外的一维信息，可以有效辅助分割。Qiu 等人[14]提

出一种基于 RGB-D 传感器的新型空间自适应投影方法。通过 RGB-D 传感器获得点云之后，使用 3D 背景差异来去除背景，并且使用欧几里德聚类提取来获得一组点云。对于每个点云，通过一些预处理来处理它，然后将其投影到不同的平面中以获得易于分割的优化图像。Holz 等人[15]提出一种快速计算表面法线的方法，并把像素在法向量空间中进行聚类，合并成不同的候选平面。候选平面在法向量空间和球面坐标中进行再一次聚类，候选平面被分割和分类成不同的对象。Richtsfeld 等人[16]所提的算法，在基于表面法向量将点云聚类得到平面集合的同时，也通过 NURBS（非均匀有理 B 样条曲线）同样拟合点云得到曲面集合，并通过模型选择找到给定点云的最佳表示形式。然后，计算表面之间的关系特征向量，并通过 SVM 训练得到成对的能量项，最后执行基于图论的分割，通过能量函数的全局最优化来得到对象。由于固废形状的不规则，基于平面的方法可能会得到过分割的结果。而在几个固废对象堆放在一起的情况时，基于平面的方法无法将固废对象分割开。

2 分割算法设计

2.1 算法的完整流程

在工业环境下，固废表面常常被粉尘覆盖，颜色信息严重退化。在固废分拣线上，固废通过皮带传输，需要经过多级破碎。在经过破碎后，固废的形状也变得不规则。固废在皮带上会出现堆叠的情况，导致黏连的同类固废对象难以区分。在这种情况下，将固废对象分割出来是非常困难的。传统的二维分割算法，主要依赖于颜色和边缘信息进行分割，而固废对象存在严重的颜色退化，所以很难得到高精度的分割效果。

针对固废对象分割，本文提出一种融合多模态信息的深度卷积神经网络的固废对象分割算法。本文算法利用深度卷积神经网络融合 RGB 信息和深度信息进行特征学习，提升了特征的丰富性，解决了由于固废颜色严重退化导致的特征缺失问题，生成了高质量的像素类别预测分数图。结合全连接条件随机场迭代优化，解决了固废形状不规则导致的细节处分割不精细的问题，为每一类固废对象产生一个精确的分割块。针对同类固废对象难以区分的情况，利用局部轮廓的深度梯度信息，实现同一类别的不同实例的固废对象精确分割，解决了黏连导致同类固废对象难以区分的情况。

综上，本文算法的主要步骤如下：

1)据预处理：用深度卷积神经网络对原始 RGB-D 数据集进行学习颜色、纹理、形状、边缘信息等特征；

2)构造能量项，应用全局 CRF 模型将原始图像不同类别对象分割开，获得完整的分类信息，减小下一步局部对象分割的难度；

3)基于深度梯度的局部对象精细分割。利用深度梯度信息得到局部深度边缘图，从中提取局部歧义区。针对局部歧义区，提取有效的内部边缘分割同类粘连对象。

2.2 原始数据的采集以及预处理

由于工业环境下大量的灰尘颗粒导致了固废物体颜色的严重退化，仅仅依靠颜色信息和轮廓信息，并不能对固废物体进行很好的分割。因此，我们采用卷积神经网络学习特征，避免显式的特征抽取，最后用训练出来的 softmax 分类器产生每个像素标签分配概率。由于固废对象颜色的严重退化，单单用 RGB 图进行特征学习并不符合实际需求，因此需要把深度这一显著特征也进行训练。

因此本文在 VGG16[17]网络结构的基础上融合了深度信息，并把 VGG16[17]最后的全连接层改造成 3 层卷积层，形成一个全卷积网络，进行多模态信息的特征学习。本文提出的深度卷积神经网络包含 2 个输入层 $data_1$ 和 $data_2$ 层，$data_1$ 层输入 RGB 图像和对应的标签文件，$data_2$ 层输入对应的深度图。将两者最后一层卷积层输出的特征图进行加权求和后作为输入训练 softmax 分类器，以获得每个像素点的标签分配概率。神经网络结构如下图 1 所示，其中卷积块里各包含 2 个或 3 个卷积层，激活层和一个最大池化层。

具体来讲，本文提出的网络结构主要改进如下：

1)修改输入。增加一个并列分支 $data_2$ 层，用来训练对应的深度图。在颜色退化的场景下，固废物体颜色信息和轮廓信息在图像上显示失真，仅仅用 RGB 图像进行特征学习，会导致训练出来的模型准确率下降。因此需要融合 RGB 信息和深度信息训练 softmax 分类器。

2)避免特征图空间分辨率下降。因为普通的池化会缩小图片的尺寸，比如 VGG16[17]五次池化后图片被缩小了 32 倍，从而损失很多精细结构信息。为了获得更大尺寸的特征图，综合内存和计算量限制，令第 4 个第 5 个池化层的步长为 1，再加上 1 填充，池化核尺寸都为 3。这样经过池化后特征图大小不变，由于步长比池化核的尺寸小，输出之间会有重

叠和覆盖，提升了特征的丰富性。但后层的感受野发生了变化。为使感受野不变，后面的卷积层使用空洞卷积[18]，其作用是在不增加参数的前提下，增加感受野。

3)修改卷积层滤波器核数。滤波器的核数是每一层能获取到的特征的种类数, 也是下一层可用于组合的特征数量。在特征学习中, 特征的多样性能够提高学习效果。由于进行学习的图片场景较为复杂, 其包含的信息、拥有的特征数相对较多, 需要更高维数上的特征学习, 因此全连接层改造的 3 层卷积层, 滤波核分别为 1 024, 1 024 和 4。

4)修改输出数。输出数要根据实际分类类别数进行设置。根据实际情况，网络输出结果需要对木头、石头、砖头加上背景一共 4 个类别做出预测，为每个像素点产生标签分配概率。为了方便起见用 0, 1, 2, 3 来分别表示 4 种不同的类别。

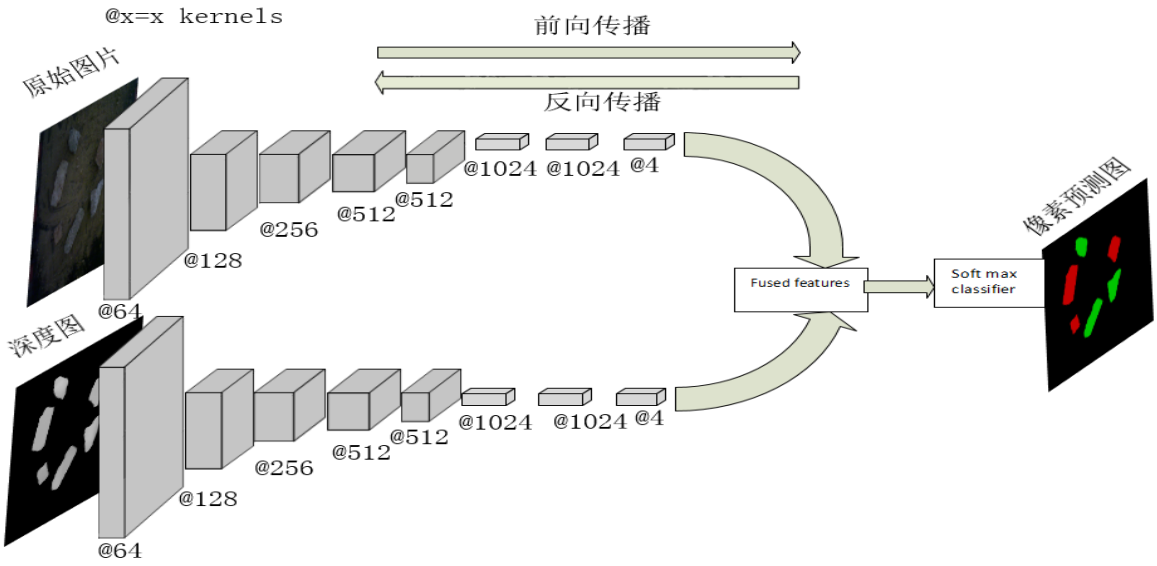


图 1 网络结构图

Fig.1 Network structure diagram

2.3 全连接条件随机场全局分割

2.3.1 能量函数构建

用本文提出的网络结构训练出来的模型能够为每个像素点产生分别属于每个类别的概率，生成的概率图如图 2 所示。

(a)真值 （b）背景概率图(c)木头概率图(d)石头概率图(e)砖头概率图

图2 概率图

Fig.2 Probability map（(a) truth； (b) Background probability map；(c) Wood probability map；(d) Stone probability map；(e) Brick probability map)

其中概率图用单通道灰度图像表示，属于哪个的类别的概率越大（即概率值趋于 1）在类别概率图像上显示越趋于白色，属于哪个的类别的概率越小（即概率值趋于 0）在类别概率图像上显示越接近黑色。

本小节在得到每个像素点属于每个类别的概率的基础上，构建了一个多标签的全连接条件随机场，提出了适用于固废对象的能量函数，然后通过最小化能量函数，来实现对不同类别区域的标记。定义 $\mathcal{S}=\{1,2,3,\cdots n\}$ 表示像素索引集合，像素与随机变量一一对应。对于随机变量 $\boldsymbol{X}=\{x_1,x_2,x_3,\cdots x_n\}$，每一个 x_i 都有 $x_i \hat{\mathrm{I}}\ \boldsymbol{C}$，$\boldsymbol{C}=\{0,1,2\cdots l\}$。$\boldsymbol{C}$ 表示标签集合，l 由得到固废对象的类别数量决定。全连接的条件随机场的能量函数可以定义为：

$$E(x)=\sum_{i}^{n}\alpha_u(x_i)+\sum_{i<j}^{n}\beta_p(x_i,x_j) \quad (1)$$

其中 i，j 是像素的索引，α_u 表示一元能量项，β_p 表示二元能量项。其中二元能量项符合高斯核的线性组合，可以通过高效近似推理算法，来降低计算复杂度，实现高效的像素标记。

2.3.2 一元能量项

一元能量项 $\alpha_u(x_i)$ 表示的是，分配像素 i 标签

x_i 的代价。定义如公式所示

$$\alpha_u(x_i) = -\log p(x_i) \tag{2}$$

其中 $p(x_i)$ 表示每个像素点属于每个类别的概率。一元能量项是把每个像素点都作为一个独立体，没有考虑像素之间的联系。因此仅仅通过一元能量项并不能精确的分割图像。

2.3.3 二元能量项

二元能量项 $\beta_p(x_i, x_j)$ 表示的是像素点与像素点之间的关系，让两者相似的像素点分到同一个类别，两者相差较大的像素点分到不同类别，使分割出来的边缘更加平滑，得到更加精确的分割结果。

本节根据 RGB-D 信息，将二元能量项函数定义如公式所示

$$\beta_p(x_i, x_j) = \begin{cases} \sum_{n=1}^{4} w_n g_n(i,j) & if \quad x_i \neq x_j \\ 0 & if \quad x_i = x_j \end{cases} \tag{3}$$

$$g_1(i,j) = \exp\left(\frac{|p_i - p_j|^2}{\theta_\alpha^2} - \frac{|I_i - I_j|^2}{\theta_\beta^2}\right) \tag{4}$$

$$g_2(i,j) = \exp\left(\frac{|p_i - p_j|^2}{\theta_\gamma^2}\right) \tag{5}$$

$$g_3(i,j) = \exp\left(\frac{|d_i - d_j|^2}{\theta_\eta^2} - \frac{|I_i - I_j|^2}{\theta_\nu^2}\right) \tag{6}$$

$$g_4(i,j) = \exp\left(\frac{|p_i - p_j|^2}{\theta_\rho^2} - \frac{|I_i - I_j|^2}{\theta_\tau^2} - \frac{|d_i - d_j|^2}{\theta_\varsigma^2}\right) \tag{7}$$

其中 $g_1(i,j)$、$g_2(i,j)$、$g_3(i,j)$、$g_4(i,j)$ 为四个对比度敏感函数，w_1、w_2、w_3、w_4 分别是它们对应的权重。用 i，j 表示像素，I_i，I_j 表示它们的 RGB 颜色信息，d_i，d_j 表示它们的深度信息，而 p_i，p_j 表示它们的位置信息。

$g_1(i,j)$ 控制颜色相似且位置相近的像素被标记成同一标签。$g_2(i,j)$ 控制相邻的像素被尽可能分配同一标签，保证得到的分割结果平滑，减少孤立的像素或者区域。$g_3(i,j)$ 控制深度值信息相似且位置相近的像素被标记成同一标签。由于在固废图像中颜色退化严重，所以增加 $g_4(i,j)$ 控制颜色和深度信息相似且位置相近的像素被标记成同一标签。

2.4 局部精确分割

在 2.3 小节，已经实现对不同类别固废的精确分割，完成了对固废对象的类别标记。接下来需要在相同类别里对不同对象进行精确分割。

对于非黏连的同类固废对象，经过 CRF 分割后，可以很容易的把固废对象分割出来。

而对于黏连的同类固废对象，因为视觉特征退化，造成粘连的同类固废对象间存在难以区分的歧义区域。针对这种模糊区域，采用深度梯度算法提取深度边缘分割模糊区。

深度梯度信息能够较好地反应相邻像素间连接紧密程度和相对变化程度， 同时反映了图像的水平方向特征与垂直方向特征。

令 $\boldsymbol{G}$ 为图像 $\boldsymbol{I}$ 的梯度模值图像，任一像素点 P 的梯度模值为

$$G(p) = \sqrt{G_x(p)^2 + G_y(p)^2}, \quad p \in \boldsymbol{I} \tag{8}$$

其中 G_x、G_y 为水平方向梯度与垂直方向梯度，利用 Sobel 算子垂直方向模板 $\begin{bmatrix} 1 & 0 & 1 \\ 2 & 0 & 2 \\ 1 & 0 & 1 \end{bmatrix}$ 和水平方向模板 $\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$ 与图像 $\boldsymbol{I}$ 进行卷积计算所得。

根据公式(9)可得深度边缘图 $\boldsymbol{E_g}$。

$$E_g(x,y) = \begin{cases} 255 & if \quad G(p) > Threshold \\ 0 & if \quad G(p) \leq Threshold \end{cases} \tag{9}$$

其中 $p(x,y)$ 为图像 $\boldsymbol{I}$ 上的像素点，$E_g(x,y)$ 为 $\boldsymbol{E_g}$ 图上第 y 行，第 x 列的像素值，$G(p)$ 为图像 $\boldsymbol{I}$ 上像素点 P 的梯度模值，*Threshold* 为阈值，本实验中设置为 0.8。

在得到每一类的分割结果基础上，令集合 $\boldsymbol{F_w}\{f_{w1}, f_{w2}, f_{w3}, \cdots f_{wn}\}$、$\boldsymbol{F_s}\{f_{s1}, f_{s2}, f_{s3}, \cdots f_{sn}\}$ $\boldsymbol{F_b}\{f_{b1}, f_{b2}, f_{b3}, \cdots f_{bn}\}$ 分别表示图像 $\boldsymbol{I}$ 中木头石头砖头所有的闭合轮廓集合。$\boldsymbol{S_w}\{s_{w1}, s_{w2}, s_{w3}, \cdots s_{wn}\}$、$\boldsymbol{S_s}\{s_{s1}, s_{s2}, s_{s3}, \cdots s_{sn}\}$、$\boldsymbol{S_b}\{s_{b1}, s_{b2}, s_{b3}, \cdots s_{bn}\}$ 为与每个类别轮廓集合一一对应的每个闭合轮廓的大小（即每个闭合轮廓内所含像素点的个数）。

由于在实际生产线中，机器人自动分拣只是其中的一个环节，其前面的破碎环节基本上能保证要分割的相同类别的固废对象尺寸大小基本相近，不会出现过大尺寸的固废对象。因此，在保证分割精度的前提下，根据前面环节破碎设备能够输出的固废对象的最大尺寸为阈值，从 $\boldsymbol{F_w}$、$\boldsymbol{F_s}$、$\boldsymbol{F_b}$ 中提取可能存在歧义区的固废轮廓，从而减小计算量。对

$\boldsymbol{F}_w$、$\boldsymbol{F}_s$、$\boldsymbol{F}_b$ 中的每个轮廓处理如下。

$$\boldsymbol{B}_w = \{ f_{wi} \mid s_{wi} > T_w, i ? [1,n] \} \quad (10)$$

$$\boldsymbol{B}_s = \{ f_{si} \mid s_{si} > T_s, i ? [1,n] \} \quad (11)$$

$$\boldsymbol{B}_b = \{ f_{bi} \mid s_{bi} > T_b, i ? [1,n] \} \quad (12)$$

其中 $\boldsymbol{B}_w$、$\boldsymbol{B}_s$、$\boldsymbol{B}_b$ 分别为提取出来的可能存在模糊区的木头、石头、砖头闭合轮廓集合。T_w、T_s、T_b 根据前面环节破碎设备能够输出最大尺寸的固废对象轮廓区域面积给出。

令 f_{si} 为 $\boldsymbol{B}_s$ 中任意一个闭合轮廓，外轮廓图 $\boldsymbol{F}_c$ 如图 3 所示。

图 3 局部边缘轮廓图

Fig.3 Local edge contour map

以该轮廓的最小外接矩形为边界，$p(x,y)$ 为图像 $\boldsymbol{I}$ 上最小外接矩形内的像素点，根据公式提取局部深度边缘轮廓图 $\boldsymbol{E}_m$。

$$E_m(x,y) = \begin{cases} 255 & if \ E_g(x,y) == 255 \\ 0 & if \ E_g(x,y) \neq 255 \end{cases} \quad (13)$$

其中 $E_m(x,y)$ 为 $\boldsymbol{E}_m$ 图上第 y 行，第 x 列的像素值，深度边缘轮廓图 $\boldsymbol{E}_m$ 如图 4 所示。

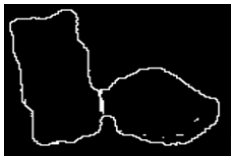


图 4 深度边缘轮廓图

Fig.4 Depth edge contour map

在外部轮廓已知的情况下，内部边缘成为分割粘连同类对象的关键。通过外轮廓图 $\boldsymbol{F}_c$ 和深度边缘图 $\boldsymbol{E}_m$，根据公式得到局部的内部边缘图 $\boldsymbol{E}_{in}$。

$$\boldsymbol{E}_{in} = \boldsymbol{E}_m - \boldsymbol{F}_c \quad (14)$$

由于固废对象表面是不规则的，这样得到的内部边缘 $\boldsymbol{E}_{in}$ 会存在很多干扰的无效边缘。如何判断 $\boldsymbol{E}_{in}$ 是否存在有效的内部边缘成为分割模糊区域的关键。

由于固废对象外部形状大多是凸多边形，因此其凸包区域与轮廓区域的差值会比较小。但如果存在黏连区域，其外轮廓上不可避免的会存在凹的形状，导致其凸包区域与轮廓区域的差值会较大。本小节利用这点判断内部边缘 $\boldsymbol{E}_{in}$ 是否存在有效的内部边缘。

通过计算得到外轮廓 $\boldsymbol{F}_c$ 的凸包区域 $\boldsymbol{F}_t$，利用外轮廓图 $\boldsymbol{F}_c$ 得到对应的掩码区域 $\boldsymbol{F}_y$,如图 5 所示。

(a)外轮廓掩码 (b)凸包区域

图 5 外轮廓与对应的凸包

Fig.5 Outer contour and corresponding convex hull ((a) Outer contour mask ； (b) Convex area)

根据公式可以得到轮廓凹处区域图 $\boldsymbol{E}_o$。

$$\boldsymbol{E}_o = (\boldsymbol{F}_t - \boldsymbol{F}_y) ? \ C_{2*k+1} \quad (15)$$

其中 Å 表示在 $(\boldsymbol{F}_t - \boldsymbol{F}_y)$ 的结果上执行核大小为 C_{2*k+1} 的腐蚀操作，k=1。

得到的轮廓凹处区域图 $\boldsymbol{E}_o$ 如图 6 所示。

图 6 轮廓凹处区域图

Fig.6 Concave area map

计算轮廓凹处区域图 $\boldsymbol{E}_o$ 掩码面积如果小于阈值，则认为该外轮廓是一个独立的固废对象轮廓，如果大于阈值，则认为该外轮廓可能存在固废对象黏连情况。

在轮廓凹处区域图 $\boldsymbol{E}_o$ 上计算所有闭合轮廓点集, 计算每一个轮廓点集像素点坐标 x 均值 $\bar{x}$，y 均值 $\bar{y}$。根据如果凹处区域是上下相对，则 $\bar{x}$ 相近，$\bar{y}$ 相差较大；如果凹处区域是左右相对，则 $\bar{x}$ 相差较大，$\bar{y}$ 相近的规律将闭合轮廓点集两两匹配。

令闭合轮廓点集 $\boldsymbol{P}_u$、$\boldsymbol{P}_d$ 为上下匹配的点集，遍历 $\boldsymbol{P}_u$、$\boldsymbol{P}_d$ 分别得到 $\boldsymbol{P}_u$ 中 y 值最小的像素点 $p_1(x_1,y_1)$ 和 $\boldsymbol{P}_d$ 中 y 值最大的像素点 $p_2(x_2,y_2)$。令 $E_{in}(x,y)$ 为 $\boldsymbol{E}_{in}$ 图上第 y 行，第 x 列的像素值，通过公式提取两者之间的边缘点集 $\boldsymbol{E}_u$。

$$\boldsymbol{E}_u = \{ p(x,y) \mid x \text{常} x_1 \ \&\& x \quad x_2 \} \quad (16)$$

其中 $p(x,y)$ 为内部边缘 $\boldsymbol{E}_{in}$ 上的像素点，然后计算点集 $\boldsymbol{E}_u$ 的大小 S_u，根据公式判断 $\boldsymbol{E}_u$ 是否是有效的内部边缘点集，得到结果 r 。

$$r=\begin{cases}1 & if\ \ S_u \geq t\\ 0 & if\ \ S_u \leq t\end{cases} \tag{17}$$

其中 t 为给定阈值。r 为 1 代表点集 $\boldsymbol{E}_u$ 为有效内部边缘，为 0 代表点集 $\boldsymbol{E}_u$ 为干扰边缘。

结合内部边缘 $\boldsymbol{E}_u$ 和外轮廓信息，根据公式就可以把粘连的同类固废对象分割开，得到分割后的局部轮廓图 $\boldsymbol{F}_n$。

$$\boldsymbol{F}_n=\boldsymbol{F}_y-\boldsymbol{E}_u \oplus C_{2*k+1} \tag{18}$$

其中 $\oplus$ 表示在 $\boldsymbol{E}_u$ 的结果上执行核大小为 C_{2*k+1} 的膨胀操作，k=1。

3 实验与结果分析

由于目前没有公开可用的固废数据集，本节首先从工业现场中通过 ASUS XtionPRO 摄像头采集 1 006 张固废 RGB 图像和深度图。其中这些图像包含了实际生产线中的皮带上堆放的石头，砖块，木头，混泥土等固废物体，它们形状不规则且表面被粉尘或者碎屑覆盖。

我们对数据集进行人工标注，主要分成木头，砖头，石头加上背景一共 4 类，按照 4:1 的比例划分训练集与测试集。用 VGG16[17] 模型作为预训练模型进行超参微调。训练参数如下：原始图像 640×480 像素，学习率 0.001，batch_size：4, 迭代次数：40 000。

深度卷积神经网络各个层输入输出具体如表 1 所示。

表 1 各层输入输出

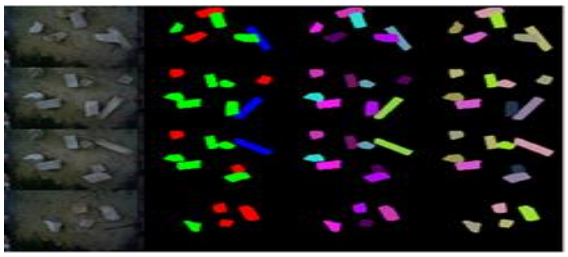
Table 1 Input and output of each layer

参数层	$data_1$ 通道	$data_2$ 通道
C1 卷积块	(1×3×640×480)	(1×1×640×480)
	(1×64×640×480)	(1×64×640×480)
Pool1 层	(1×64×640×480)	(1×64×640×480)
	(1×64×321×241)	(1×64×321×241)
C2 卷积块	(1×64×321×241)	(1×64×321×241)
	(1×128×321×241)	(1×128×321×241)
Pool2 层	(1×128×321×241)	(1×128×321×241)
	(1×128×161×121)	(1×128×161×121)
C3 卷积块	(1×256×161×121)	(1×256×161×121)
	(1×256×161×121)	(1×256×161×121)
Pool3 层	(1×256×161×121)	(1×256×161×121)
	(1×256×81×61)	(1×256×81×61)
C4 卷积块	(1×256×161×121)	(1×256×161×121)
	(1×512×81×61)	(1×512×81×61)
Pool4 层	(1×512×81×61)	(1×512×81×61)
	(1×512×81×61)	(1×512×81×61)
C5 卷积块	(1×512×81×61)	(1×512×81×61)
	(1×512×81×61)	(1×512×81×61)
Pool5 层	(1×512×81×61)	(1×512×81×61)
	(1×512×81×61)	(1×512×81×61)
Conv6 层	(1×512×81×61)	(1×512×81×61)
	(1×1 024×81×61)	(1×1 024×81×61)
Conv7 层	(1×1 024×81×61)	(1×1 024×81×61)
	(1×1 024×81×61)	(1×1 024×81×61)
Conv8 层	(1×1 024×81×61)	(1×1 024×81×61)
	(1×4×81×61)	(1×4×81×61)

其中第二第三列中 2 到 13 行表格中第一行是输入参数，第二行是输出参数，对应（图像张数×图像通道数×图像宽度×图像高度）。

全连接条件随机场二元项公式中的参数设置如下 w_1：6，w_2：8，w_3：2，w_4：4，迭代次数 1 次。

二元能量项 $\beta_p\left(x_i,x_j\right)$ 保证特征相似且位置相近的像素分配相同的标签，而特征差异大或者相距较远的像素分配不同的标签。因此得到的分割结果更加精确，如图 7 所示。



(a)原图 (b)CRF 结果 (c)分割结果 (d)真值

图7 分割结果

Fig.7 Segmentation results（(a) Original image ； (b) Crf result ； (c) Segmentation result ； (d) Truth)

其中(b) CRF 结果是本文算法第二步之后的结果（即每种颜色代表某一类固废对象独立的分割块），红色代表砖头，绿色代表石头，蓝色代表木头，(c)分割结果是本文算法第三步之后的结果（不同的颜色代表不同分割对象），(d) 代表真值。

全连接条件随机场可以通过多次迭代对分割结果进行优化，但是随之而来的是耗时的增加。本文通过实验来分析迭代次数对分割精度和分割耗时的影响。从表 2 可以看出，随着迭代次数的增加，分割精度也得到提升，但是分割的精度提升不是很明显。因为迭代一次的分割精度已经比较高，多次迭代对歧义区域的分割更加精细，只会有小部分像素改变标签。所以从分割精度上看，提升的比较小。迭代次数增加，随之而来的不仅仅是分割精度的提升，还有算法耗时的增加。从表 2 可以看到，迭代一次的耗时需要 6.7 秒，而迭代 4 次的耗时已经超过了 9 秒。在实际应用中，算法速度越快越有益于它的应用。因此，综合分割精度和耗时，本文选择迭代 1 次。

表 2 迭代次数对算法分割精度的影响

Table 2 Influence of the number of iterations on the segmentation precision of the algorithm

迭代次数	MPA/%	MIOU/%	Timing/秒/张
1	90.02	89.03	6.7
2	90.07	89.07	7.4
3	90.14	89.15	8.2
4	90.29	89.27	9.0
5	90.67	89.61	9.8

其中 MPA（均像素精度）和 MIOU（均交并比）是常用像素标记的精度标准，PA 为标记正确的像素占总像素的比例，MPA 表示均像素精度，计算每个类内被正确分类像素数的比例，之后求平均。MIOU 表示均交并比，计算两个集合的交集和并集之比，在语义分割的问题中，这两个集合为真实值和预测值。这个比例可以变形为正真数比上真正、假负、假正（并集）之和，在每个类上计算 IOU（交并比），之后求平均。

为了评估本文算法的有效性，在实验过程中选取 FCN[12]、SegNet[19]、DeepLabV1[20]、DeepLabV2[18]这四种语义分割算法作为对比方法。训练参数一致如下：原始图像 640×480 像素，学习率 0.001，迭代次数：40 000，batch_size：4。统一用迭代 40 000 次的模型进行对比，结果如表 3 所示。

表 3 算法精度

Table 3 Algorithm accuracy

算法	Mpa/%	MIOU/%	Timing/秒/张
Fcn8s[12]	67.73	58.15	--
SegNet[19]	62.03	53.08	--
Deeplabv1[20]	80.65	76.95	6.7
Deeplabv2[18]	85.26	81.47	6.8
ours	90.02	89.03	7.0

从表中可以看到 SegNet[19]算法的精度最低，SegNet[19]是 Cambridge 提出旨在解决自动驾驶或者智能机器人的图像语义分割深度网络,可能不适用于对小目标的分割。Fcn8s[12]算法精度优于 SegNet[19]算法，但是进行 8 倍上采样的结果还是比较粗糙，图像中的细节不敏感，导致其对小目标的分割精度还是比较低。后端有 CRF 优化的 deeplabv1[20]和 deeplabv2[18]分割算法在精度上明显高于 SegNet[19]和 Fcn8s[12]算法。针对多尺度目标使用不同采样率进行平行空洞卷积的 deeplabv2[18]算法精度比 deeplabv1[20]算法精度有明显提高。我们的算法融合深度信息提取特征，分割精度比以上四种算法都要高。

为了进一步评估本文算法的先进性，选取 Fcn32s-Color-D[12]、Fusenet-Sf1[21]、DeepLabV2-D[18]三种针对 RGB-D 数据的语义分割算法进行对比实验。训练参数一致如下：原始图像 640×480 像素，学习率 0.001，迭代次数 40 000，batch_size：4。统一用迭代 40 000 次的模型对比，结果如表 4 所示。

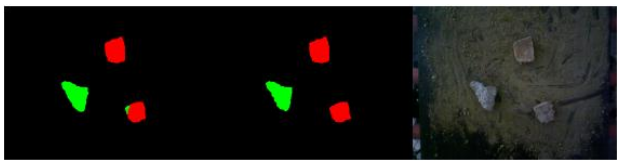
表 4 RGB-D 数据的语义分割算法精度

Table 4 Semantic segmentation algorithm accuracy of RGB-D data

算法	Mpa/%	MIOU/%
Fcn32s-Color-D[12]	72.33	68.15
FuseNet-Sf1[21]	78.03	72.08
DeepLabV2-D[18]	86.01	82.15
ours	90.02	89.03

从表中可以看到 Fcn32s-Color-D[12]算法的精度最低，Fcn32s-Color-D[12]算法是把深度信息和 RGB 信息整合成一个 4 通道图像进行训练，真正能学习到深度信息的有效特征只有第一个卷积层，对深度信息的特征提取过于粗糙，导致分割精度较低。FuseNet-Sf1[21]算法采用双分支的网络同时从 RGB 图像和深度图提取特征进行融合作为第一个池化层的输入，并采用多尺度信息作为反池化层的输入，因此分割精度较高。DeepLabV2-D[18]算法是在 DeepLabV2[18]的基础上针对第一个卷积层加入深度信息融合成 4 通道图像进行训练，对比表三表四的结果可以发现，精度提高并不明显。我们的算法采用双分支的网络同时从 RGB 图像和深度图提取特征，加入空洞卷积扩大感受野，增强语义信息，精度比以上三种算法要高。

综合表三表四，我们的算法比不用深度信息的语义分割算法精度有明显提升，耗时上跟同样后端有 CRF 优化的 deeplab 分割算法接近；跟用深度信息的语义分割算法相比也处于先进水平。从理论上分析，在颜色退化的环境下，仅仅依靠颜色信息和轮廓信息，并不能对固废物体进行很好的分割，加入深度信息显得更加合理，在一些细节处分割效果更好，如图 8 所示。

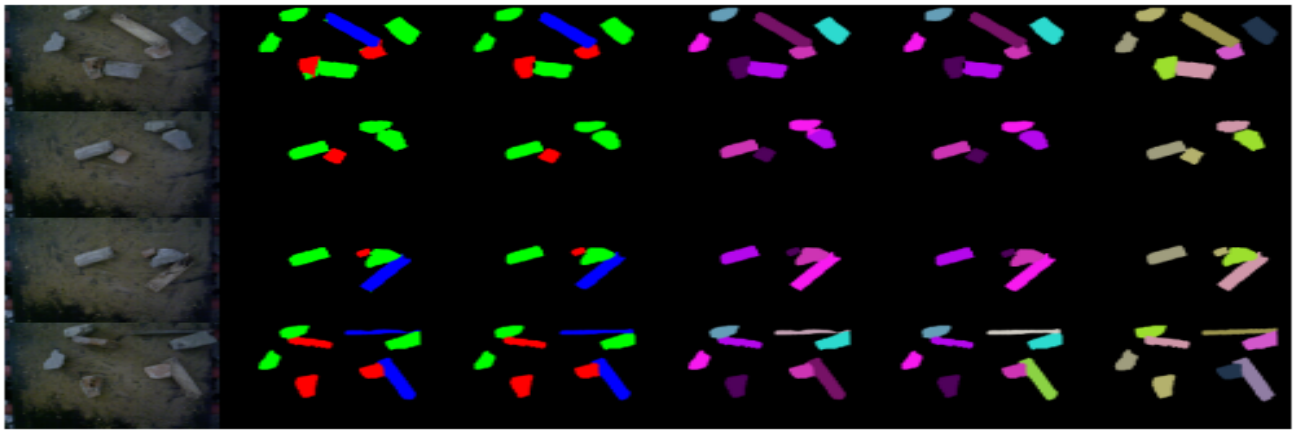


(a) Deeplabv2-D 结果 (b)CRF 结果 (c)真值

图8 分割对比效果图

Fig.8 Adhesive segmentation contrast effect chart（(a) Deeplabv2-D results ;(b) Crf Result; (c) truth)

其中(a)是 deeplabv2-D[18]的结果图，(b)是本文第二步算法结果图，绿色代表石头，红色代表砖头，蓝色代表木头。(a)图右下角一块砖头出现类内不一致的情况（即砖头一部分误识别为石头），(b)图中则是正确分割，显然本文 CRF 结果比 deeplabv2-D[18]结果图更符合实际情况。



(a)原图 (b)deeplabv2-D 结果 (c)本文 CRF 结果 (d) deeplabv2-D 最终 (e)本文结果 (f)真值

图 9 对比效果图

Fig.9 Contrast rendering（(a) Original image ； (b) Deeplabv2-D results； (c) Ours crf result； (d) Deeplabv2-D eventually； (e) ours result； (f) truth)

如图 9 所示，综合来看可以发现加了分支网络深度通道训练出来的模型产生每个像素标签分配概率作为一元项的分割结果明显比 DeepLabV2-D[18]的结果更加精确。

其中(b)(c)中绿色代表石头，红色代表砖头，蓝色代表木头（即每种颜色代表某一类固废对象独立的分割块）。(d)是(b)经过本文算法第三步之后的结果，(e)是(c)经过本文算法第三步之后的结果（不同颜色代表不同的实例固废对象）。(b)中第一行图片，DeepLabV2-D [18]结果无法分割出搭在砖头上木条的自然轮廓，而本文结果可以实现精准分割；(b)中第二行图片，DeepLabV2-D [18]结果无法把右上角两块黏连的石头分割开，而本文结果可以明显分割开；(b)中第四行图片，DeepLabV2-D [18]结果无法把右上角的木条准确的分割出来，而本文结果可以实现正确的分割。

(c)本文 CRF 结果得到的是每一类固废对象的一个独立的分割块，(e)本文结果将黏连的同类固废对象分割开，得到的是同一类别的不同实例的固废对象精确分割。具体如表 5 所示。

表 5 分割结果对比

Table 5 Segmentation result comparison

	图像数量/张	固废对象数量/个	本文 CRF 结果/个	本文结果/个
简单场景	496	1 061	1 061	1 061
复杂场景	510	3 569	3 104	3 569

其中简单场景是指不存在歧义区域的固废图像（即不存在同类固废对象黏连的情况），复杂情况是指存在歧义区域的固废图像（即存在同类固废对象黏连的情况）。从表中可以看到，在简单场景下，本文 CRF 分割结果和本文结果和实际固废对象数量是一样的。但在复杂情况下，本文结果和实际固废对象数量是一样的，而本文 CRF 结果则小于实际固废对象数量。这表明，在复杂情况下，本文 CRF 结果并不能把所有的黏连的同类固废对象分割开，而本文结果则可以做到这一点。这也直接证明了本文算法第三步的必要性和有效性。

4 结 论

本文对固废图像分割技术进行研究与探索，结合深度学习、全连接条件随机场等技术，提出了一种新的分割方法。该方法利用深度卷积神经网络提

取特征，结合全连接条件随机场和图像深度梯度信息实现了复杂情况下的固废对象分割。理论分析和实际实验均表明本文的分割方法精度较高，在本文实验的测试集上取得了90.02%均像素精度和89.03%均交并比，与目前一些先进的语义分割算法相比在精度上表现出一定的优越性。该方法在建筑垃圾智能机器人自动分拣行业具有一定的应用价值。本文通过实验发现黏连同类固废对象的分割精度是整体分割精度的瓶颈。较低的分割精度主要来自于小尺寸固废对象和黏连同类固废对象的干扰。

下一步将针对黏连的同类固废对象探索算法的优化，扩大数据集，提升算法的速度以及提高算法的泛化能力和鲁棒性。

参 考 文 献(References)

[1] Hao Z G, Su X M. Research on the reuse of construction waste[J]. Architecture & Culture, 2017(2):110-111.[郝占国, 苏晓明. 建筑垃圾再利用研究[J]. 建筑与文化, 2017(2):110-111.]

[2] Wang X W. Development guide to building materials[J]. China Building Materials, 2005, 3(1):67-71.[王武祥. 建筑垃圾的循环利用[J]. 中国建材, 2005, 3(1):67-71.]

[3] Krähenbühl P, Koltun V. Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials[J]. 2012:109-117.

[4] Reddi S S, Rudin S F, Keshavan H R. An optimal multiple threshold scheme for image segmentation[J]. Systems Man & Cybernetics IEEE Transactions on, 1984, SMC-14(4):661-665.

[5] Ma W Y, Manjunath B S. EdgeFlow: a technique for boundary detection and image segmentation.[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2000, 9(8):1375-88.

[6] Leung T K, Malik J. Contour Continuity in Region Based Image Segmentation[C]// European Conference on Computer Vision. Springer-Verlag, 1998:544-559.

[7] Achanta R, Shaji A, Smith K, et al. SLIC superpixels compared to state-of-the-art superpixel methods[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012, 34(11):2274-2282.

[8] Felzenszwalb P F, Huttenlocher D P. Efficient graph-based image segmentation[J]. International journal of computer vision, 2004, 59(2): 167-181.

[9] Achanta R, Shaji A, Smith K, et al. SLIC superpixels compared to state-of-the-art superpixel methods[J]. IEEE transactions on pattern analysis and machine intelligence, 2012, 34(11): 2274-2282.

[10] Mingjun S. Road Extraction Using SVM and Image Segmentation[J]. Photogrammetric Engineering & Remote Sensing, 2004, 70(12):1365-1371.

[11] Li J, Bioucas-Dias J M, Plaza A. Spectral–Spatial Hyperspectral Image Segmentation Using Subspace Multinomial Logistic Regression and Markov Random Fields[J]. IEEE Transactions on Geoscience & Remote Sensing, 2012, 50(3):809-823.

[12] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 39(4):1-1.

[13] Zhao H, Shi J, Qi X, et al. Pyramid Scene Parsing Network[J]. 2017:6230-6239.

[14] Qiu Y, Chen J, Guo J, et al. Three Dimensional Object Segmentation Based on Spatial Adaptive Projection for Solid Waste[J]. 2017.

[15] Holz D, Holzer S, Rusu R B, et al. Real-time plane segmentation using RGB-D cameras[C].Robot Soccer World Cup. Springer, Berlin, Heidelberg, 2011, 306-317.

[16] Richtsfeld A, Mörwald T, Prankl J, et al. Learning of perceptual grouping for object segmentation on RGB-D data[J]. Journal of visual communication and image representation, 2014, 25(1): 64-73.

[17] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.

[18] 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, 2018, 40(4):834-848.

[19] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99):2481-2495.

[20] Chen L C, Papandreou G, Kokkinos I, et al. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs[J]. Computer Science, 2015(4):357-361.

[21] Hazirbas C, Ma L, Domokos C, et al. FuseNet: Incorporating Depth into Semantic Segmentation via Fusion-Based CNN Architecture[C]// Asian Conference on Computer Vision. Springer, Cham, 2016:213-228.

作者简介

张剑华(1980-), 男, 副教授, 主要研究方向为计算机视觉、机器学习。E-mail: zjh@zjut.edu.cn

陈嘉伟，男, 硕士研究生, 主要研究方向为计算机视觉、深度学习。

张少波，男, 硕士, 主要研究方向为计算机视觉。

郭建双，男, 硕士研究生, 主要研究方向为计算机视觉。

刘盛, 男, 副教授, 主要研究方向为数字图像处理、计算机视觉。