Computer Vision based Polyethylene Terephthalate (PET) Sorting for Waste Recycling

Ouiem Bchir, Shahad Alghannam, Norah Alsadhan, Raghad Alsumairy, Reema Albelahid, Monairh Almotlaq
Department of Computer Science, College of Computer and Information Sciences
King Saud University, Riyadh, Saudi Arabia

***Abstract*—Recycling plays a vital role in saving the planet for future generations as it allows keeping a clean environment, reducing energy consumption, and saving materials. Of special interest is the plastic material which may take centuries to decompose. In particular, the Polyethylene Terephthalate (PET) is a widely used plastic for packaging various products that can be recycled. Sorting PET can be performed, either manually or automatically, at recycling facilities where the post-consumed objects are moving on the conveyor belt. In particular, automated sorting can process a large amount of PET objects without human intervention. In this paper, we propose a computer vision system for recognizing PET objects placed on a conveyor belt. Specifically, DeepLabv3+ is deployed to segment PET objects semantically. Such system can be exploited using an autonomous robot to compensate for human intervention and supervision. The conducted experiments showed that the proposed system outperforms the state of the art semantic segmentation approaches with weighted IoU equals to 97% and Mean BFscore equals to 89%.**

***Keywords*—PET; recycling; computer vision; machine learning**

I. Introduction

Over the last decade, people around the world have a rising concern about efficient waste management due to the yearly waste increase. In fact, according to the World Bank Group 2020 statistics, 2.01 billion tons of solid municipal waste are engendered every year worldwide [1]. Furthermore, according to the same source, it is predicted that this amount would increase to 3.4 billion tons by 2050. One way of processing this huge amount of waste is by incineration. However, it can be harmful to the environment because of greenhouse gas emissions. Another commonly used way to process waste is landfill. Nevertheless, it is not appropriate for certain materials that need a very long time to biodegrade. In particular, plastic material, which constitutes 14% of the total waste amount [2], takes over 100 years to biodegrade. Therefore, recycling, which consists of processing and reusing the waste, emerged as an alternative method suitable for waste processing. Since the way of processing the waste depends on its type, the waste needs to be sorted. To make the sorting process easier, there are sometimes specific waste bins for the most common waste types such as plastic, glass, and paper. Even though different types of plastic require specific methods of treatment. Therefore, plastic materials must be sorted according to their type since the quality of waste separation highly affects the quality of the recycled plastic. One of the most valuable types of plastic for recycling is Polyethylene Terephthalate (PET). It is widely used for plastic bottles. It is recognized by the symbol “PET” or “PETE” imprinted in the container.

The wide use of PET is due to the fact that it is environmentally friendly and inexpensive. For that reason, recycling centers sort plastic waste into PET and non-PET plastics. It is even further sorted into transparent, blue and green, and mixed color PET since they do not have the same sale price. In fact, transparent PET is the most valuable one and the mixed color is the least valuable [3]. Manual waste sorting is exhaustive and time consuming. Moreover, it may be affected by the worker’s condition. On the other hand, the PET chemical sorting process is very delicate, dangerous, and generates chemical residue [4]. Electrostatic systems that disperse plastics according to their types are alternative solutions for plastic sorting [5]. Nevertheless, they are not cost effective. Thus, mechanical approaches have been used instead, as they are safe and less costly [6]. They mainly use visual sensors to localize the PET materials that would be moved to the appropriate waste bin. Typically, mechanical sorting systems use a conveyor belt to carry the waste. When the waste reaches the camera position, an image of the waste scene is captured. Then, a computer vision system localizes the PET object in the captured image and categorizes it using image processing and machine learning techniques. More specifically, the image is segmented into objects. Then, these objects are conveyed as input to a recognition system in order to categorize it as PET or non-PET. Nevertheless, suitable visual descriptors need to be extracted from the image in order to discriminate PET objects effectively. In this respect, considering the diversity of PET waste and the background clutter, the determination of such features is arduous and constitutes a hindrance for computer vision systems [7]. One way of alleviating the problem of choosing the appropriate visual features is through the use of deep learning techniques ability to semantically segment the waste image.

In this paper, we propose to localize and categorize PET plastics on the conveyor belt. The proposed approach will semantically segment the PET material. To achieve this, DeepLabv3+ deep neural network architecture [8] will be trained to learn PET containers’ visual characteristics.

II. Semantic Segmentation

In the field of computer vision, image segmentation is the task of dividing the image into sets of pixels called segments. It is considered as a one of the most difficult and challenging problems in the computer vision field [9]. Image segmentation aims to represent the image at a higher level in a way that facilitates its analysis by localizing objects and edges. Over the last decades, image segmentation has been used in several applications such as medical image analysis, scene understanding, robotic vision, and self-driving cars [10]. The image segmentation process can be supervised or unsupervised. Unsupervised image segmentation does not require a training phase, and thus previous knowledge of the object is not needed [11]. Alternatively, supervised segmentation requires a training phase that uses a set of labeled pixels. It can be perceived as a classification of the pixels that constitute the image [12]. While instance segmentation treats multiple instances of the same object as distinct objects [13], semantic segmentation treats multiple instances of the same object as a single one. It does not differentiate between two or more instances referring to the same object in the same image. In the last decade, semantic segmentation approaches were mainly based on the extraction of suitable engineered features fed to a classifier. However, the efficiency of these approaches depends heavily on the extracted features. This is considered a critical factor for the progress of semantic segmentation [14]. Recently, the boost of Deep Learning in the context of computer vision has also affected semantic segmentation [14].

The development of Deep Convolutional Neural Net (DCNN) led to a significant improvement in semantic segmentation [15]. One of the main characteristics that led to the success of DCNN is its ability to learn abstract data representation [16]. However, while the special abstraction is recommended for classification tasks, it impedes semantic segmentation. In fact, semantic segmentation approaches based on DCNN face three main problems. The first one is related to the repeated combination of the max-pooling layer and striding that yields a feature map with decreased feature resolution [17]. The second obstacle concerns the multi-scale challenge, where the objects may have different scales [18]. This induces increasing the number of computations since it requires training the network with different scale versions of the image. The third hindrance is due to the discard of the location information [19]. DCNN, designed for image classification and object detection, is invariant to special transformation. This results in inconsistent segmentation outcomes. Furthermore, as semantic segmentation implicates segmentation and classification processes, the key point is then how to adjoin the two processes. We distinguish three types of deep learning approaches for semantic segmentation. The first type starts by learning the object regions [20] [21]. These regions, integrating the shape information, are then conveyed to a DCNN classifier [22] [23]. This type of approach depends on the results of the segmentation phase which in its turn depends on the engineered features. Alternatively, the second type of approach uses the convolution layers of DCNN to extract the features to use them for the segmentation phase [19] [24] [25]. However, segmentation and classification tasks are still performed in cascade. Therefore, classification still depends on the segmentation phase and consequently, any segmentation error cannot be recovered by the classification task. The third type employs DCNN directly on the images to learn the pixels' categories [17] [26]. This eliminates the segmentation phase. In order to enhance the segmentation performance along the edges, the Conditional Random Fields (CRFs) approach [23] has been integrated into the DCNN based approaches [24] [27]. In fact, by taking into consideration the neighboring pixels, the object boundaries are better localized. CRF has been used as a post-processing step [8]. It has also been integrated to the DCNN architecture in [23], [24], [25], [26] and [27].

III. Related Work

In the literature, several PET sorting approaches have been reported. Among the reported works, some works designed a handcrafted feature suitable for PET categorization [28] [29] [30]. Other works used available generic handcrafted features [31] [32].

A. PET Sorting Approaches based on Application Dedicated Handcrafted Features

The approach in [28] extracts the foreground object (the waste object) by employing background subtraction. After connecting the obtained objects and enhancing the border using morphological operations, small blobs are discarded according to a pre-defined size threshold. For the remaining blobs called "white strips", a contour box is determined along its eight surrounding boxes of the same size called "grey strips". After the detection of the plastic blobs, a visual descriptor is extracted. The authors in [28] designed a new handcrafted feature. It is based on modeling the color distribution of the "grey strips". Alternatively, the authors in [29] propose the "white pixel" approach. They first start by preprocessing the image by performing noise removal, background subtraction, and grey level transformation. Then, they employ the MATLAB function "regionprops" [33] to split the image into a set of disconnected objects. This results in reducing the problem to a classification problem where only one object is present in the image. From the obtained grey level image, the authors designed two handcrafted features. The first one is extracted from the whole image by computing the average of the last 106 entries of the 256-bin normalized color histogram. Assuming that the bottom of the container is not covered by a label and is transparent showing the black color of the conveyer belt, the second proposed feature divides the image into five parts and extracts the center of the fifth one. From the extracted Region of Interest, ROI, the mean and standard deviation of the first 100 entries of the normalized 256-bin color histogram are computed. The resulting two features are then fed to the Linear Discriminant Analysis (LDA) classifier [34]. On the other hand, the reported approach in [30] assumes that only one object is present in the scene. It starts by converting the RGB image to a greyscale one. Then, the Canny edge detector [35] is employed to detect the object in the image. The 256-bin histogram is computed from the detected object based on which the authors in [30] designed a new handcrafted feature. The proposed feature consists of two values. The first one is the sum of the first one hundred entries of the 256-bin histogram, v_1, and the second one is the sum of the last one hundred entries, v_2. Similar to the proposed approach in [29], the authors in [30] assume that PET objects are transparent. Thus, they will be perceived as black, like the color of the conveyer belt. Considering this assumption, they design a rule to classify PET and non-PET objects. More specifically, an object is considered PET if v_1 is greater than v_2.

B. PET Sorting Approaches based on Generic Handcrafted Features

The authors in [31] proposed a PET sorting system. They assume that there is only one object in the captured image and propose to classify plastic bottles carried on a conveyor belt as PET or non-PET. Moreover, they propose to further classify non-PET plastic bottles as High Density Polyethylene (HPDE) or Polypropylene (PP). The preprocessing step starts by segmenting the image using Otsu's thresholding method [36] in order to locate the object. It is followed by background subtraction and segmentation enhancement using morphological operators. The authors suggest working directly on the pixels of the considered object. However, due to the image's size, the obtained feature has a high dimension and the system would then be prone to the curse of dimensionality. That is why they propose to reduce the dimensionality using five techniques. Namely, they used Principal Component Analysis (PCA) [[37] [38], Kernel PCA [39], Fisher's Linear Discriminant Analysis (FLDA) [40], Singular Value Decomposition (SVD) [41], and Laplacian Eigenmaps (LEMAP) [42]. The resulting feature vectors are fed separately to the Support Vector Machine (SVM) classifier [43]. Then, the classification results obtained using each feature are combined using the majority vote approach. Alternatively, the system proposed in [32] treats each object present in the image separately. First, edges are detected. Then, standard shape features are extracted. These are the length, the width, the area, the aspect ratio, and the filling fraction. The Cartesian and polar coordinates of the 90 equally spaced points of the perimeter are also considered. The authors in [32] considered three classifiers. Namely, the K-Nearest Neighbor (KNN) [34] with K=1, Kohonen map [44], and Artificial Neural Net (ANN) [45]. For the KNN classifier they used the geometric feature, the Cartesian coordinates of the perimeter, and its polar coordinates separately. On the other hand, the geometric feature is used with Kohonen map, and the polar coordinate of the perimeter is used with ANN. Moreover, the authors designed a "factor-of-merit" measure to decide on the final category of the object. In fact, the "factor-of-merit" is computed for the different considered system results in order to combine them. The system is assessed in a 50-instance dataset.

C. Convolutional Neural Net based Approaches

In [46], the authors developed a system that sorts four kinds of waste: glass, paper, plastic, and metal, based on a pre-trained ResNet-50 architecture [47]. ResNet was pre-trained using ImageNet dataset [48]. It is used to extract the feature automatically from the whole image. In fact, a single object is considered per image. A multi-class soft kernel SVM [43] is used instead of softmax for the classification task. In [49], the authors proposed a waste management system using ResNet-34 deep learning architecture. The system assumes the presence of a single object in the captured image. The work in [49] classifies the waste into six categories which are cardboard, glass, metal, plastic, paper, and trash. Nevertheless, the proposed system aims to classify the waste as digestible and indigestible. In fact, cardboard, glass, metal, plastic, and paper categories are considered indigestible while the remaining waste is considered digestible. A computer vision waste sorting approach is proposed in [50]. It considers a single object per image. The authors adopted AlexNet [51] deep learning architecture to categorize various types of waste material. However, this system performed poorly compared to the system based on extracting Scale Invariant Feature (SIFT) [52] and feeding it to the SVM classifier [43]. The authors in [53] proposed a waste sorting system for all types of materials. It is based on DenseNet-121 [54] deep learning architecture. The choice of DenseNet was motivated by the small size of the dataset [53]. In an attempt to improve the performance, data augmentation is employed by considering vertical, horizontal, and random 25° rotations. To further improve the system's performance, a genetic algorithm is utilized to optimize the hyper-parameters of the fully connected layers.

As stated above, various vision-based recognition approaches have been proposed in the literature. The extracted features differ between these approaches. Some papers focus on designing handcrafted features suitable for the PET sorting application [28] [29] [30]. However, these approaches assumed that PET materials are transparent. The designed features are based on the fact that PET containers appear black like the conveyer belt color. Nevertheless, this is not the case. PET containers can be transparent, blue and green, and mixed colors. This infers that these approaches addressed only the problem of sorting transparent PET materials. Other approaches used existing generic features [31] [32]. One of them used dimensionality reduction on the image pixels as a feature. The other one employed the shape feature. However, in addition to using only 50 instances as a dataset, the shape feature would not be able to recognize crashed containers. These feature-based approaches face the challenge of feature choice or feature design. Moreover, the images need to be preprocessed and segmented in order to separate the object from the background. This makes the system performance sensitive to the performance of these preprocessing and segmentation techniques. Convolutional Neural Nets, CNN, would alleviate these problems by learning the appropriate feature without the need of preprocessing and segmentation techniques. However, the only approach that used deep learning to classify PET bottles did not classify plastic as PET or non-PET [55]. Rather, only PET bottles are fed to their system, which identifies the state of the PET bottles. Namely, it checks if the PET bottle has a cap, a seal, or content. Thus, to the best of our knowledge, no reported work addressed the problem of PET sorting using CNN. Alternatively, sorting all kinds of waste approaches based on various Deep CNN have been reported [46] [49] [50] [53]. Among these approaches, two are based on ResNet architecture [46] [49]. Another is based on AlexNet [50] and performed poorly. While the other is based on DenseNet [53] and would be practical only for small datasets.

IV. Proposed Approach

We propose to segment the images captured from the conveyer belt semantically. Three categories need to be localized and identified. These are transparent PET, blue and green PET, and mixed color PET. For this purpose, we employ DeepLabv3+ [8]. Fig. 1 displays the architecture of the proposed system. DeepLabv3+ [8] is designed to overcome the limitations of existing semantic segmentation approaches based on DCNN. More specifically, DeepLabv3+ [8] adopts an encoder-decoder architecture and uses Resnet as a backbone for the encoder. Nevertheless, as shown in Fig. 2, it introduces modifications to the Resnet through the use of atrous convolution. Moreover, Atrous Spatial Pyramid Pooling (ASPP) and fully connected Conditional Random Fields (CRF) are incorporated.

Fig. 2 shows a simplified structure of DeepLabv3+ model. As shown, the Resnet model reduces the size of the input image by a factor of 16. Nevertheless, DeepLabv3+ discards the striding of the last convolutional layer and replaces it by atrous convolution with rate equal to 2, and appends it by the Atrous Spatial Pyramid Pooling (ASPP) module. The output of ASPP is then up-sampled by 4 in the decoder module. The obtained feature map is concatenated with a feature map from the encoder module that has the same size, specifically the one down-sampling the input by a factor of 4. Next, it is convoluted using a set of 3X3 filters, and then up-sampled by 4 to engender an output of the same size of the input. Recurrent pooling and convolution layers decrease the resolution of the obtained feature. To remedy that, atrous convolution is introduced. Its idea comes from the wavelet decomposition. It up-samples the filter by inserting holes that are filled with zeroes to enlarge the receptive field. This is called Atrous convolution, or dilated convolution. Moreover, DCNNs are able to handle objects with different scales by using the Atrous Spatial Pyramid Pooling (ASPP) method.

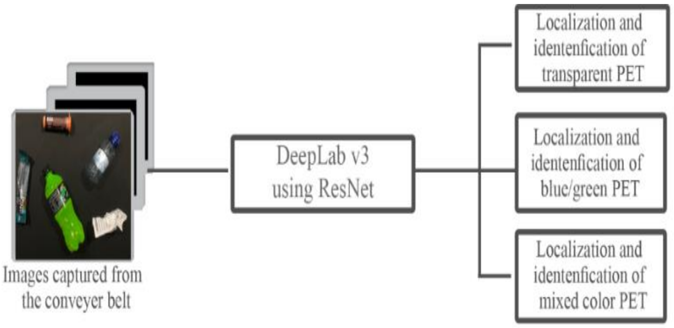
Images captured from the conveyer belt
DeepLab v3 using ResNet
Localization and identenfication of transparent PET
Localization and identenfication of blue/green PET
Localization and identenfication of mixed color PET

Fig. 1. Proposed System Architecture.

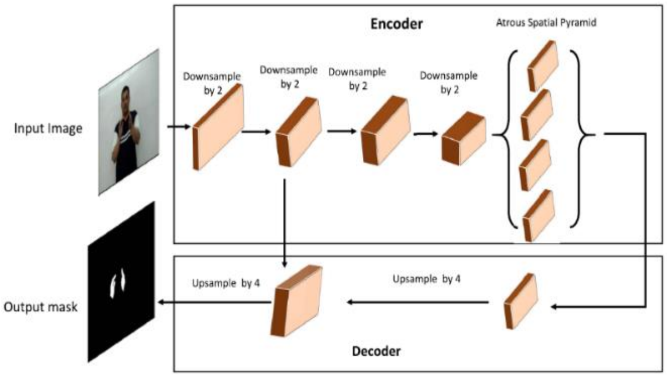

Fig. 2. Simplified Structure of DeepLabv3+ [56].

The latter performs four parallel operations. These are one convolution with kernel 1X1 and three atrous convolutions with kernel 3X3 and rates equal to 6, 12 and 18, respectively. This results in extracting 4 features at different scales. The feature map learned at the end of the encoder is a stack of the obtained 4 features. To alleviate the localization problem, fully connected Conditional Random Field (CRF) [23] is employed. It is a statistical approach that models the relation between pixels by estimating the cost of assigning a pair of labels to a pair of pixels (pairwise cost). Its main function is to clear out invalid predictions by coupling neighbor pixels and privileging same label assignment for nearby pixels.

This leads to refining the segmentation result. Furthermore, DeepLabv3+ adopts the encoder-decoder architecture. It aims to refine the edges obtained by the segmentation. More specifically, the encoder is responsible for extracting the features and the decoder allows retrieving the spatial resolution. The encoder consists of two modules which are ResNet with atrous convolution component and the ASPP. The decoder merges and up-samples the learned features and the result of the encoder after up-sampling. We train DeepLabv3+ using labeled captured images. Images captured from the conveyer belt are fed to DeepLabv3+ [8]. The corresponding mask images are provided at the output. Mask images indicate the label of each pixel of the input image. The label could be transparent, blue and green, mixed color, or others.

V. Experiments

In order to evaluate the performance of the proposed approach, a dataset is collected. It includes images of size 720X960X3 pixels captured from 420X594 mm scene using a camera. The scene contains PET and non-PET materials on a black background representing the conveyer belt. The waste materials can be overlapped or not. Three types of PET are considered. These are transparent, blue and green, and mixed color PET. The dataset is labeled manually accordingly. The performance of the proposed approach is assessed using five performance measures. These are the standard performance measures used for semantic segmentation that take into consideration both the categorization and localization performances [57]. Namely, we will use the Global Accuracy [58], Mean Accuracy [58], Mean Intersection over Union (Mean IoU) [10], Weighted Intersection over Union (Weighted IoU) [58], and Mean BFscore [59]. In order to assess the performance of the proposed system, we intend to conduct three experiments.

A. Experiment 1

In this experiment, we try to empirically figure out the best hyperparameter configuration for both Resnet-50 and Resnet-18 [47] when they are used as backbone models for the DeepLabv3+ in the context of PET sorting. In this regard, we train two DeepLabv3+ models. Precisely, Resnet-50 and ResNet-18 are trained using 60% of the data, validated on 20%, and tested on the remaining 20%. For each considered model, various configurations were tested. In particular, the optimizer, the learning rate, and the L2 regularization parameter were tuned. This results in 6 configurations for each model. Table I and Table III report the details of each considered configuration with respect to Resnet-50 and Resnet-18 [47], respectively. Moreover, for this experiment the size of batch is set to 2 and the number of epochs is set to 30. In order to determine the best model, the testing performance results of each configuration are reported. These are Global Accuracy, Mean Accuracy, Mean IoU, Weighted IoU, and Mean BFscore.

TABLE I. THE CONSIDERED CONFIGURATIONS FOR RESNET-50

	Optimizer	Learning rate	L2 regularization
Configuration 1	SGDM	1e-3	0.005
Configuration 2	ADAM	1e-3	0.05
Configuration 3	SGDM	Initially:1e-3 Learn Rate Drop Period: 5 Learn Rate Drop Factor: 0.2	0.001
Configuration 4	SGDM	Initially:1e-3 Learn Rate Drop Period: 6 Learn Rate Drop Factor: 0.5	0.001
Configuration 5	SGDM	Initially:1e-3 Learn Rate Drop Period: 4 Learn Rate Drop Factor: 0.05	0.001
Configuration 6	SGDM	Initially:5e-2 Learn Rate Drop Period: 5 Learn Rate Drop Factor: 0.2	0.001

TABLE II. WEIGHTED IOU AND MEAN BFSCORE WHEN USING RESNET-50

	Weighted IoU	Mean BFscore
Configuration 1	0.9476	0.8744
Configuration 2	0.7539	0.7084
Configuration 3	0.9687	0.8933
Configuration 4	0.9233	0.8420
Configuration 5	0.9268	0.8146
Configuration 6	0.6333	0.6591

TABLE III. THE CONSIDERED CONFIGURATIONS FOR RESNET-18

	Optimizer	Learning rate	L2 regularization
Configuration 1	SGDM	1e-3	0.005
Configuration 2	SGDM	1e-4	0.001
Configuration 3	SGDM	Initially:5e-3 Learn Rate Drop Period: 5 Learn Rate Drop Factor: 0.2	0.001
Configuration 4	SGDM	Initially:1e-2 Learn Rate Drop Period: 6 Learn Rate Drop Factor: 0.03	0.001
Configuration 5	SGDM	5e-3	0.01
Configuration 6	SGDM	Initially:2e-2 Learn Rate Drop Period: 1 Learn Rate Drop Factor: 0.3	0.1

TABLE IV. WEIGHTED IOU AND MEAN BFSCORE WHEN USING RESNET-18

	Weighted IoU	Mean BFscore
Configuration 1	0.9252	0.8337
Configuration 2	0.9165	0.7698
Configuration 3	0.7863	0.6905
Configuration 4	0.8002	0.7077
Configuration 5	0.8122	0.7328
Configuration 6	0.7978	0.7012

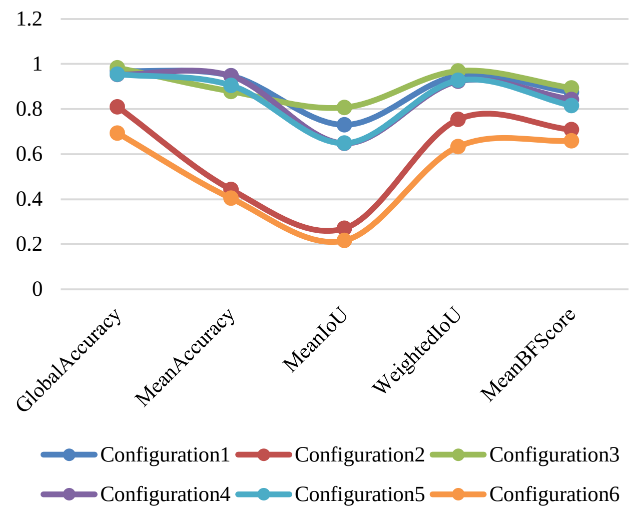

Fig. 3. Proposed System Performance when using Resnet-50 as Backbone.

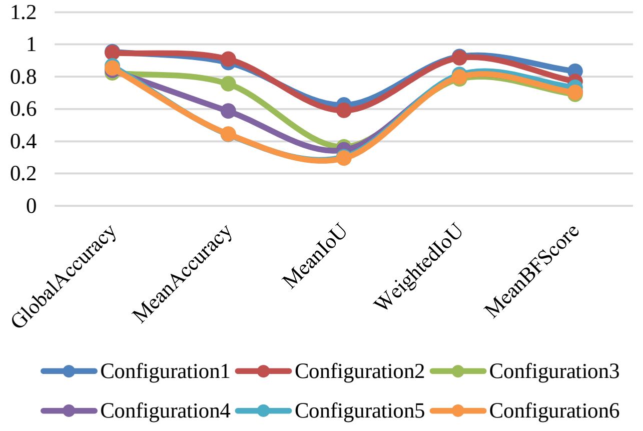

Fig. 4. Proposed System Performance when using ResNet-18 as Backbone.

Fig. 3 displays the performance measures of the system when using Resnet-50 as backbone for the DeepLabv3+ semantic segmentation approach. Similarly, Fig. 4 shows these performances when using Resnet-18. Table II and Table IV report Weighted IoU and Mean BFscore for Resnet-50 and Resnet-18, respectively.

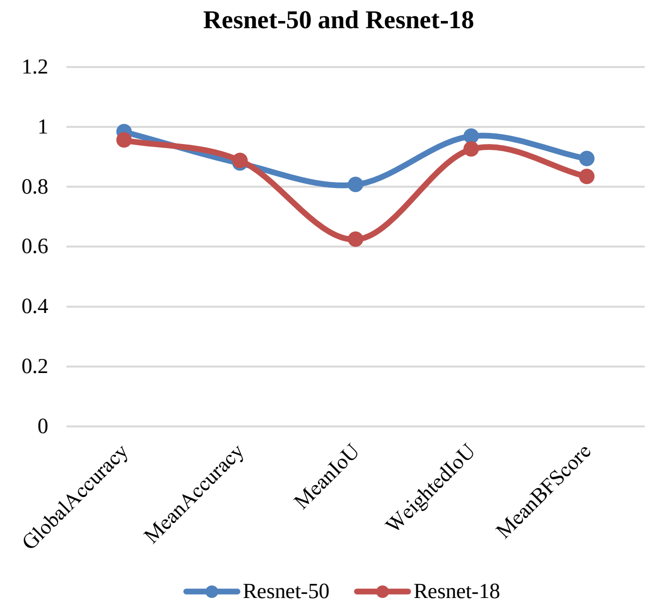

Fig. 5. Resnet-50 and Resnet-18 Models.

As shown in Fig. 3 and Fig. 4, configuration 3 allowed obtaining the best performance for Resnet-50. Actually, since the data is unbalanced Weighted IoU and Mean BFscore reflect better the performance of the system. Thus, configuration 3 outperformed the other configurations with a Weighted IoU of 0.9687, and Mean BFscore of 0.8933. This is confirmed by the results reported in Table II. More specifically, configuration 3 uses stochastic gradient descent (SGDM) as optimizer, a learning rate initially set to 0.001, and increasing by a factor of 0.2 every 5 epochs, and an L2 regularization of 0.001. According to [60], SGDM is expected to give better results. Moreover, the considered learning rate gave better result by avoiding missing the optimal weights while training the network. Furthermore, the L2 regularization of 0.001 avoided both over-fitting and under- fitting situations. In fact, in case of a large value, the model doesn't fit well, while in case of a small value, the training time is too long. Concerning Resnet-18, the best performances in terms of Weighted IoU and Mean BFscore were obtained when adopting configuration 1 which consists of a constant learning rate of 0.001, and a L2 regularization of 0.005. In fact, by avoiding missing optimal values for the model, and not over-fitting it, these two hyperparameters yielded a Weighted IoU equal to 0.9252, and a Mean BFscore equal to 0.8337. This result is confirmed by Table IV where configuration 1 yielded better results.

B. *Experiment 2*

In this experiment, we try to empirically determine which Deep Learning model, Resnet-50 or Resnet-18 [47] is more effective as backbone model for the DeepLabv3+ when used for PET sorting. In this regard, we take into consideration the best obtained results for both models according to experiment 1. Namely, we consider the results lead by configuration 3 for Resnet-50 and the one lead by configuration 1 for Resnet-18. Fig. 5 displays the performance measures of Resnet-50 and Resnet-18 on the testing sets, respectively. As mentioned previously, since the data is unbalanced, Weighted IoU and Mean BFscore are more suitable to assess the performance of the system. Thus, in Table V, we report Resnet-50 and Resnet-18 performances in terms of Weighted IoU and Mean BFscore. To further investigate the obtained results, Fig. 6 shows the comparison between Resnet-50 and Resnet-18 in terms of Weighted IoU with respect to each considered class. Similarly, Fig. 7 displays the comparison between Resnet-50 and Resnet-18 in terms of Mean BFscore with respect to each considered class. Finally, Fig. 8 displays sample semantic segmentation results obtained using Resnet-18 and Resnet-50. Taking into account, the best obtained results for both Resnet-50 (configuration 3), and Resnet-18 (configuration 1), we compare the two models when used as backbone for deepLabv3+ semantic segmentation approach.

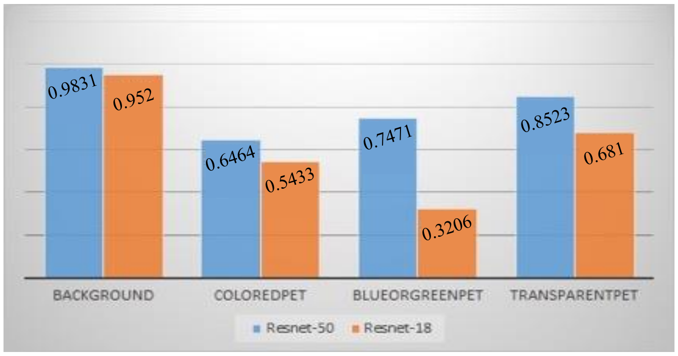

Fig. 6. Comparison between Resnet-50 and Resnet-18 in Terms of Weighted IoU with respect to each Considered Class.

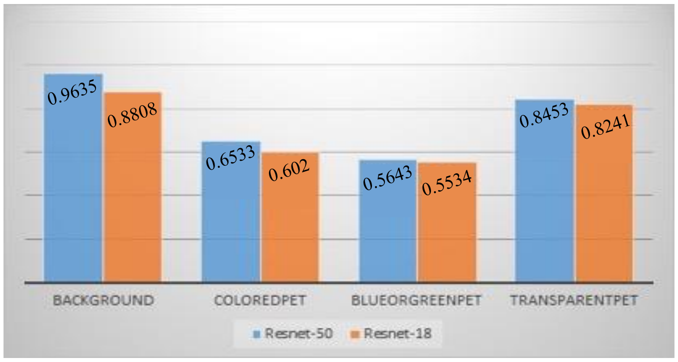

Fig. 7. Comparison between Resnet-50 and Resnet-18 in Terms of Mean BFscore with respect to each Considered Class.

TABLE V. TESTING WEIGHTED IOU AND MEAN BFSCORE COMPARISON BETWEEN RESNET-50 AND RESNET-18

	Resnet-50	**Resnet-18**
Weighted IoU	0.9687	0.9252
Mean BFscore	0.8933	0.8337

As shown, in Table V, Resnet-50 outperforms Resnet-18 with Weighted IoU equal to 0.97 and Mean BFscore equal to 0.89 for the testing results. The reason that Resnet-50 performs better than Resnet-18 can be explained by the fact that a deeper network learns more abstract features which yields better segmentation results.

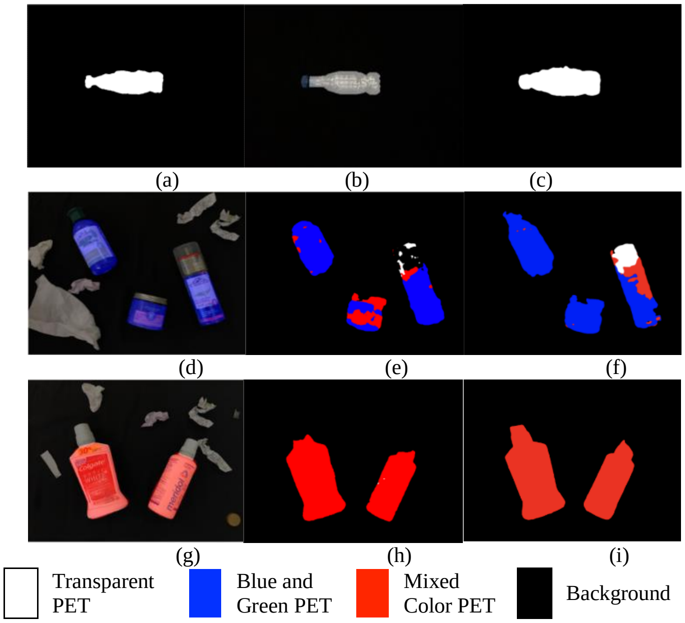

Fig. 8. Sample Semantic Segmentation Results Obtained using Resnet-18 and Resnet- 50. (a), (d), and (g) are the Original Images, (b),(e),and (h) The Segmentation Results Obtained when using Resnet-18, and (c), (f) and (i) The Segmentation Results Obtained when using Resnet-50.

As shown from Fig. 6 and Fig. 7, the performance is not the same for all considered categories. In fact, the background and Transparent PET classes have the higher performances. This is explained by the fact that background consists of black homogeneous color corresponding to the conveyer belt. Obviously, this is an easy segmentation problem. Concerning Transparent PET category, the obtained result can be attributed to the fact that this category is represented by a larger number of pixels in the dataset. In fact, if the model is trained with larger training set, the classification results are expected to be better. Alternatively, Blue or Green PET, and Colored PET are less represented in the training data. Furthermore, these two categories have large intra-class variance. In fact, in addition to the container shape variance, they are characterized by the color variance, whereas the background and the transparent PET categories have the same color per category. To better illustrate the obtained results, we can see from Fig. 8 (e) and Fig. 8 (f) that some blue or Green PET pixels are segmented as Mixed color PET or transparent PET. Similarly, from Fig. 8 (h) and Fig. 8 (i), we observe that some parts of the background are segmented as mixed color PET.

C. *Experiment 3*

In this experiment, we intend to compare DeepLabv3+ semantic segmentation model to the state-of-the-art approaches on the waste sorting dataset. Namely, Fully Connected Network (FCN) [61], Unet [62], and Segnet [63] semantic segmentation approaches are considered. In this experiment, we consider Resnet-50 as backbone for DeepLabv3+ since it achieved better performance than Resnet-18. For the-state-of-the-art approaches, several hyperparameter configurations are considered. Moreover, in this experiment, the batch size is set to 2 for Unet [62], and Segnet [63]. It is the largest possible value due to the memory size constraint. Alternatively, since FCN [61] uses a smaller size of the images (lower resolution), it is possible to increase the batch size 3. After considering the above mentioned configuration, the best obtained performance with respect to each approach is considered for the purpose of comparison with DeeepLabv3+. Fig. 9 displays the performance comparison between DeepLabv3+ [47], FCN [61], Unet [62], and Segnet [63]. As depicted in Fig. 9, DeepLabv3+ outperforms the other deep learning segmentation approaches with weighted IoU equal to 0.9687 and a Mean BFscore of 0.8933. The second best is FCN, whereas Unet and Segnet perform poorly in terms of Mean BFscore.

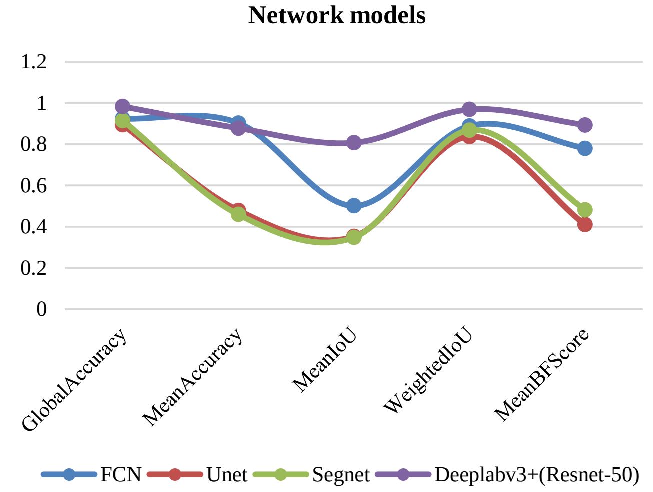

Fig. 9. Performance Comparison of DeepLabv3+ [47], FCN [61], Unet [62], and Segnet [63].

In order to better analyze the obtained results, a comparison of the four considered semantic segmentation approaches with respect to each category in terms of IoU and Mean BFscore is displayed in Fig. 10 and Fig. 11, respectively. As it can be seen from Fig. 10, DeepLabv3+ gives the best IoU performance with respect to all categories. This means that it is able to localize the PET container with respect to all categories better than the other approaches. Moreover, we can observe from Fig. 11 that DeepLabv3+ has the highest Mean BFscore with respect to most categories. However, the semantic segmentation performance is not the same with respect to all categories. This is the case for all considered segmentation approaches. For a better illustration of the results, we show sample segmentation results of DeepLabv3+ [47], FCN [61], Unet [62], and Segnet [63]. As shown in Fig. 12, Deeplabv3+ outperforms the other semantic segmentation approaches for the sample image representing the transparent PET. In fact, it localizes and identifies better the boundaries of the containers. This can be accredited to fully connected Conditional Random Field (CRF) module. Similar result can be observed from Fig. 13. In fact, although DeepLabv3+ miss - segmented some Blue and Green PET pixels as Mixed Color PET (Fig. 13(b)), it performs better than the other segmentation approaches. FCN is the second best (Fig. 13(c)). However, Unet (Fig. 13(d)) and Segnet (Fig. 13(e)) are not able to segment the Blue and Green PET image shown in Fig. 13 (a). Actually, these two approaches classified the corresponding pixels as transparent PET. It means they were not able to capture the visual characteristics of the Blue or Green category.

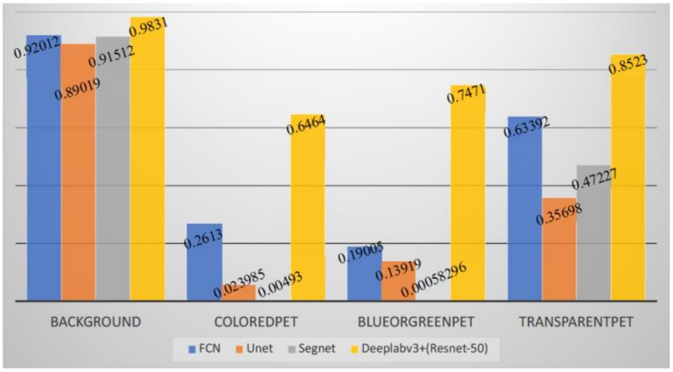

Fig. 10. Performance Comparison of the Four Considered Semantic Segmentation Approaches with respect to each Category in Terms of IoU.

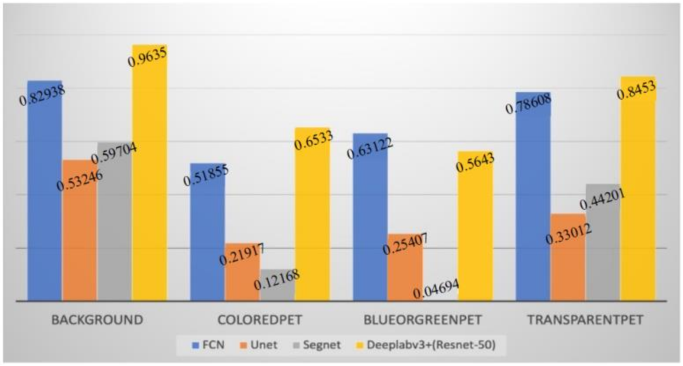

Fig. 11. Performance Comparison of the Four Considered Semantic Segmentation Approaches with respect to each Category in Terms of Mean BFscore.

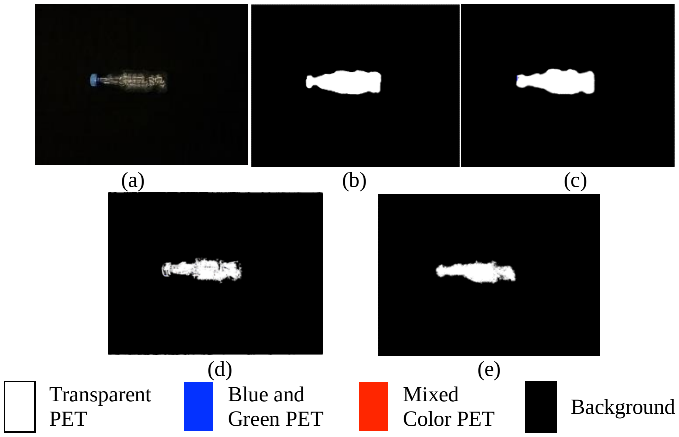

Fig. 12. Transparent PET Sample Image (a) Segmentation Results of (b) DeepLabv3+, (c) FCN (d) Unet, and (e) Segnet.

As showcased in Fig. 14, DeepLabv3+ yields better segmentation results than the other approaches. For this case too, FCN is the second best, and Unet and Segnet perform poorly. Similar analysis can be conducted on Fig. 15. Although the background (the black conveyer belt and other non-PET materials) is correctly segmented by DeepLabv3+, FCN miss-segmented some pixels as transparent PET, and Unet and Segnet miss-segmented a large number of pixels as transparent PET. Actually, the confusion between the conveyer belt and the transparent PET can be explained by the fact that the transparency of this material makes them appear as black. Nevertheless, DeepLabv3+ is able to learn the appropriate visual feature that in engendered good segmentation result.

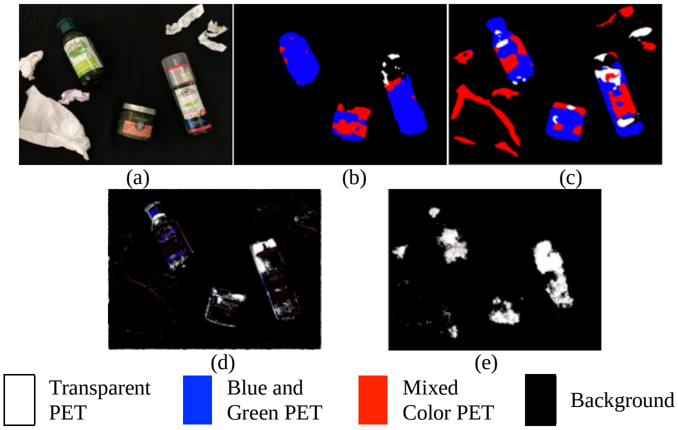

Fig. 13. Blue or Green PET Sample Image (a) Segmentation Results of (b) DeepLabv3+, (c)FCN (d) Unet, and (e) Segnet.

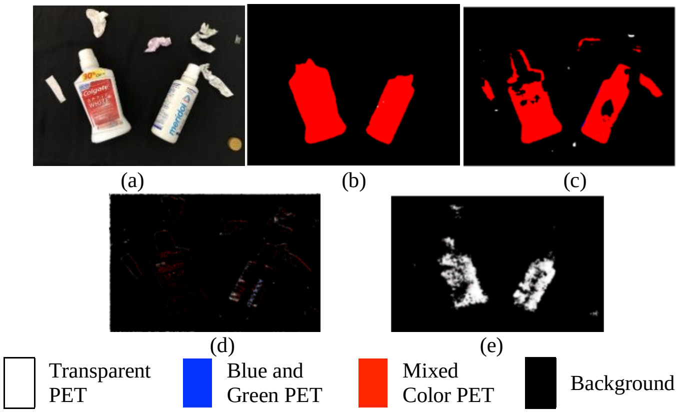

Fig. 14. Colored PET Sample Image (a) Segmentation Results of (b) DeepLab v3+, (c) FCN (d) Unet, and (e) Segnet.

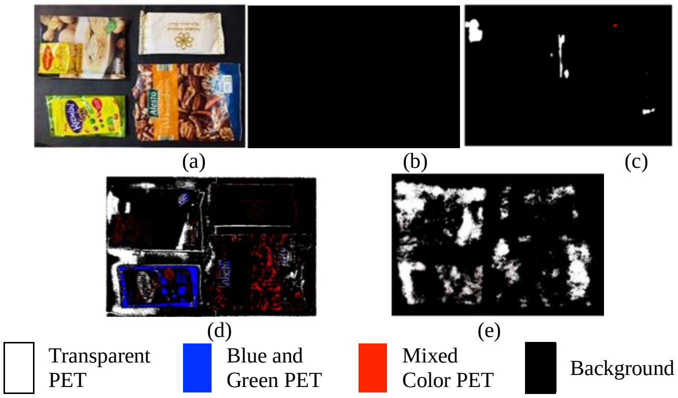

Fig. 15. Non-PET Sample Image (a) Segmentation Results of (b) DeepLab v3+, (c) FCN (d) Unet, and (e) Segnet.

VI. Conclusion and Future Work

Plastic containers are one of the most common types of waste. In order to be recycled, they need to be sorted according to their type since the quality of recycled plastic depends on the quality of waste separation. Of particular interest is Polyethylene Terephthalate (PET). In fact, recycling centers sort plastic waste into PET and non-PET, and further sort PET into transparent PET, blue and green PET, and mixed color PET. For this purpose, mechanical systems have been used. They need to recognize and localize PET materials in order to move them to the appropriate waste bin. In this context, we proposed to design a computer vision system to locate and recognize PET waste materials in a captured waste image using a deep learning network architecture called DeepLabv3+. The conducted experiments showed that increasing the number of layers of Resent from 18 to 50 yields better semantic segmentation results. Furthermore, DeepLabv3+ outperformed the other considered approaches on the PET sorting dataset. As future works, we suggest to use Resnet with even larger number of layers, and to investigate ways to decrease the frame processing time.

References

[1] "Trends in Solid Waste Management," 2020. https://datatopics.worldbank.org/what-a-waste/trends_in_solid_waste_management.html (accessed Oct. 10, 2020).

[2] J. N. Hahladakis and E. Iacovidou, "Closing the loop on plastic packaging materials: What is quality and how does it affect their circularity?," Sci. Total Environ., vol. 630, pp. 1394–1400, 2018, [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0048969718307307.

[3] "Polymer waste management: pet recycling." http://polymerwastemanagement.blogspot.com/2007/11/pet-recycling.html (accessed Oct. 10, 2020).

[4] L. Bartolome, M. Imran, B. G. Cho, W. A. Al-Masry, and D. H. Kim, "Recent Developments in the Chemical Recycling of PET," in Material Recycling - Trends and Perspectives, InTech, 2012, pp. 65–84.

[5] C. H. Park, H. S. Jeon, H. S. Yu, O. H. Han, and J. K. Park, "Application of electrostatic separation to the recycling of plastic wastes: Separation of PVC, PEL and ABS," Environ. Sci. Technol., vol. 42, no. 1, pp. 249–255, 2008.

[6] A. Picon, O. Ghita, P. F. Whelan, and P. M. Iriondo, "Fuzzy spectral and spatial feature integration for classification of nonferrous materials in hyperspectral data," IEEE Trans. Ind. Informatics, vol. 5, no. 4, pp. 483–494, 2009.

[7] Y. Tachwali, Y. Al-Assaf, and A. R. Al-Ali, "Automatic multistage classification system for plastic bottles recycling," Resour. Conserv. Recycl., vol. 52, no. 2, pp. 266–285, 2007, [Online]. Available: https://dspace.aus.edu/xmlui/bitstream/handle/11073/106/35.232-2005.07.pdf?sequence=1&isAllowed=y.

[8] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 4, pp. 834–848, 2017.

[9] Z. Li, W. Yang, S. Peng, and F. Liu, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," arXiv:2004.02806, 2020, [Online]. Available: http://arxiv.org/abs/2004.02806.

[10] I. Ulku and E. Akagunduz, "A Survey on Deep Learning-based Architectures for Semantic Segmentation on 2D images," arXiv:1912.10230, 2019.

[11] A. Kanezaki, "Unsupervised image segmentation by backpropagation," in IEEE international conference on acoustics, speech and signal processing (ICASSP), Tokyo, 2018.

[12] V. Lempitsky, A. Vedaldi, and A. Zisserman, "A pylon model for semantic segmentation," in Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems, Granada, 2011.

[13] "A 2019 Guide to Semantic Segmentation | by Derrick Mwiti | Heartbeat." https://heartbeat.fritz.ai/a-2019-guide-to-semantic-segmentation-ca8242f5a7fc (accessed Sep. 29, 2020).

[14] M. Siam, S. Elkerdawy, M. Jagersand, and S. Yogamani, "Deep semantic segmentation for automated driving: Taxonomy, roadmap and challenges," in IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, Yokohama, 2018.

[15] A. Garcia-Garcia, S. Orts-Escolano, S. O. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A Review on Deep Learning Techniques Applied to Semantic Segmentation," arXiv:1704.06857, 2017.

[16] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in European conference on computer vision, Zurich, 2014.

[17] Y. Zhang, Z. Qiu, T. Yao, D. Liu, and T. Mei, "Fully Convolutional Adaptation Networks for Semantic Segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake, 2018, [Online]. Available: http://arxiv.org/abs/1804.08286.

[18] L. C. Chen, Y. Yang, J. Wang, W. Xu, and A. L. Yuille, "Attention to Scale: Scale-Aware Semantic Image Segmentation," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016.

[19] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik, "Hypercolumns for object segmentation and fine-grained localization," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Boston, 2015.

[20] P. Arbeláez, J. Pont-Tuset, J. Barron, F. Marques, and J. Malik, "Multiscale combinatorial grouping," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, 2014.

[21] J. R. R. Uijlings, K. E. A. Van De Sande, T. Gevers, and A. W. M. Smeulders, "Selective search for object recognition," Int. J. Comput. Vis., vol. 104, no. 2, pp. 154–171, 2013.

[22] R. Girshick, J. Donahue, T. Darrell, J. Malik, U. C. Berkeley, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, 2014, [Online]. Available: http://arxiv.

[23] P. K.˙henbu˙hl and V. Koltun, "Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials," in Advances in Neural Information Processing Systems, Granada, 2011.

[24] C. Farabet, C. Couprie, L. Najman, and Y. Lecun, "Learning hierarchical features for scene labeling," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1915–1929, 2013.

[25] J. Dai, K. He, and J. Sun, "Convolutional feature masking for joint object and stuff segmentation," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Boston, 2015.

[26] D. Eigen and R. Fergus, "Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-Scale Convolutional Architecture," in IEEE International Conference on Computer Vision (ICCV), Santiago, 2015.

[27] M. Cogswell, X. Lin, S. Purushwalkam, and D. Batra, "Combining the Best of Graphical Models and ConvNets for Semantic Segmentation," arXiv:1412.4313, vol. 2, 2014, [Online]. Available: http://arxiv.org/abs/1412.4313.

[28] M. A. Zulkifley, M. M. Mustafa, A. Hussain, A. Mustapha, and S. Ramli, "Robust identification of polyethylene terephthalate (PET) plastics through bayesian decision," PLoS One, vol. 9, no. 12, pp. 1–21, 2014.

[29] S. Ramli, M. M. Mustafa, A. Hussain, and D. A. Wahab, "Histogram of intensity feature extraction for automatic plastic bottle recycling system using machine vision," Am. J. Environ. Sci., vol. 4, no. 6, pp. 583–588, 2008.

[30] J. Bobulski and J. Piatkowski, "PET waste classification method and plastic waste database WaDaBa," in Advances in Intelligent Systems and Computing, Ukraine, 2018.

[31] K. Özkan, S. Ergin, S. Işik, and I. Işikli, “A new classification scheme of plastic wastes based upon recycling labels,” Waste Manag., vol. 35, pp. 29–35, 2015.

[32] E. Scavino, D. A. Wahab, A. Hussain, H. Basri, and M. M. Mustafa, “Application of automated image analysis to the identification and extraction of recyclable plastic bottles,” J. Zhejiang Univ., vol. 10, no. 6, pp. 794–799, 2009.

[33] MathWorks, Image Processing Toolbox Use Guide, 2nd ed. Natick: The Math Works Inc, 1997.

[34] K. Fukunaga, Introduction to statistical pattern recognition, 2nd ed. San Diego: Academic Press, 1990.

[35] J. Canny, “A Computational Approach to Edge Detection,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 8, no. 6, pp. 679–698, 1986.

[36] N. Otsu, “A threshold selection method from gray-level histograms,” Automatica, vol. 11, no. 3, pp. 285–296, 1975.

[37] K. Pearson, “On lines and planes of closest fit to systems of points in space,” London, Edinburgh, Dublin Philos. Mag. J. Sci., vol. 2, no. 11, pp. 559–572, 1901.

[38] H. Hotelling, “Analysis of a complex of statistical variables into principal components,” J. Educ. Psychol., vol. 24, no. 6, pp. 417–441, 1933.

[39] B. Schölkopf, A. Smola, and K.-R. Müller, “Kernel principal component analysis,” in 7th International Conference in Artificial Neural Networks (ICANN’97), Lausanne, 1997.

[40] R. Fisher, “The Use Of Multiple Measurements In Taxonomic Problems,” Ann. Eugen., vol. 7, no. 2, pp. 179–188, 1936.

[41] M. Lee, H. Shen, J. Z. Huang, and J. S. Marron, “Biclustering via Sparse Singular Value Decomposition,” Biometrics, vol. 66, no. 4, pp. 1087–1095, 2010.

[42] M. Belkin and P. Niyogi, “Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering,” in Neural Information Processing Systems Foundation (NIPS) 2001, Vancouver, 2001.

[43] V. Vapnik, The Nature of Statistical Learning Theory, 2nd ed. New York: Springer, 2000.

[44] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 2nd ed. New Jersey: Prentice-Hall, Upper Saddle River, 2002.

[45] H. White, “Artificial Neural Networks: Approximation and Learning Theory,” 1992.

[46] O. Adedeji and Z. Wang, “Intelligent Waste Classification System Using Deep Learning Convolutional Neural Network,” Procedia Manuf., vol. 35, pp. 607–612, 2019, [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2351978919307231.

[47] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016.

[48] O. Russakovsky et al., “Imagenet large scale visual recognition challenge,” Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, 2015.

[49] M. W. Rahman, R. Islam, A. Hasan, N. I. Bithi, M. M. Hasan, and M. M. Rahman, “Intelligent waste management system using deep learning with IoT,” J. King Saud Univ. - Comput. Inf. Sci., 2020, doi: https://doi.org/10.1016/j.jksuci.2020.08.016.

[50] M. Yang and G. Thung, “Classification of Trash for Recyclability Status,” San Jose, 2016. [Online]. Available: http://cs229.stanford.edu/proj2016/report/ThungYang-ClassificationOfTrashForRecyclabilityStatus-report.pdf.

[51] B. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in Advances in Neural Information Processing Systems 25 (NIPS 2012), Lake Tahoe, 2012.

[52] D. G. Lowe, “Distinctive image-features from scale-invariant keypoints,” Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, 2004.

[53] W. L. Mao, W. C. Chen, C. T. Wang, and Y. H. Lin, “Recycling waste classification using optimized convolutional neural network,” Resour. Conserv. Recycl., vol. 164, no. 105132, 2021, [Online]. Available: https://doi.org/10.1016/j.resconrec.2020.105132.

[54] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, 2017.

[55] H. W. Mwangi and M. Mokoena, “Using Deep Learning to Detect Polyethylene Terephthalate (PET) Bottle Status for Recycling,” Glob. J. Comput. Sci. Technol., vol. 19, no. 4, pp. 27–31, 2019.

[56] S. Aly and W. Aly, “DeepArSLR: A Novel Signer-Independent Deep Learning Framework for Isolated Arabic Sign Language Gestures Recognition,” IEEE Access, vol. 8, pp. 83199–83212, 2020, doi: 10.1109/ACCESS.2020.2990699.

[57] “Evaluate semantic segmentation,” 2017. https://www.mathworks.com/help/vision/ref/evaluatesemanticsegmentation.html (accessed Nov. 07, 2020).

[58] T. Ghosh, L. Li, and J. Chakareski, “Effective Deep Learning for Semantic Segmentation Based Bleeding Zone Detection in Capsule Endoscopy Images,” Proc. - Int. Conf. Image Process. ICIP, no. September 2019, pp. 3034–3038, 2018.

[59] E. Fernandez-Moral, R. Martins, D. Wolf, and P. Rives, “A New Metric for Evaluating Semantic Segmentation: Leveraging Global and Contour Accuracy,” IEEE Intell. Veh. Symp. Proc., vol. 2018-June, pp. 1051–1056, 2018.

[60] L. Luo, Y. Xiong, Y. Liu, and X. Sun, “Adaptive gradient methods with dynamic bound of learning rate,” arXiv, no. 2018, pp. 1–19, 2019.

[61] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, 2015.

[62] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9351, pp. 234–241, 2015.

[63] V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 12, pp. 2481–2495, 2017.