Trash and Recycled Material Identification using Convolutional Neural Networks (CNN)

Rumana Sultana
School of Eng. and Technology
Western Carolina University
Cullowhee, NC
rsultana1@catamount.wcu.edu

Dr. Robert Adams
Associate Professor
School of Eng. and Technology
Western Carolina University
Cullowhee, NC
radams@email.wcu.edu

Dr. Yanjun Yan
Associate Professor
School of Eng. and Technology
Western Carolina University
Cullowhee, NC
yyan@email.wcu.edu

Dr. Paul Yanik
Associate Professor
School of Eng. and Technology
Western Carolina University
Cullowhee, NC
pyanik@wcu.edu

Dr. Martin Tanaka
Associate Professor
School of Eng. and Technology
Western Carolina University
Cullowhee, NC
mtanaka@wcu.edu

***Abstract*—**The aim of this research is to improve municipal trash collection using image processing algorithms and deep learning technologies for detecting trash in public spaces. This research will help to improve trash management systems and help to create a smart city. Two Convolutional Neural Networks (CNN), both based on the AlexNet network architecture, were developed to search for trash objects in an image and separate recyclable items from the landfill trash objects, respectively. The two-stage CNN system was first trained and tested on the benchmark TrashNet indoor image dataset and achieved great performance to prove the concept. Then the system was trained and tested on outdoor images taken by the authors in the intended usage environment. Using the outdoor image dataset, the first CNN achieved a preliminary 93.6% accuracy to identify trash and non-trash items on an image database of assorted trash items. A second CNN was then trained to distinguish trash that will go to a landfill from the recyclable items with an accuracy ranging from 89.7% to 93.4% and overall 92%. A future goal is to integrate this image processing based trash identification system in a smart trash can robot with a camera to take real-time photos that can detect and collect the trash all around it.

Keywords*—*CNN, AlexNet, Image Classification, Deep Learning, Object detection.

I. INTRODUCTION

A city is best loved by people who live in it when it is healthy and hygienic. But in the era with a growing population, more and more people are moving into the city area, creating more trash than before and it is very difficult to maintain the cleanliness of the city. If we look at the south Asian countries, we can easily understand how challenging it is. Though first world countries have a well-established trash management system as they have enough funds to invest and maintain such a trash management system, most of the developing countries cannot do this properly and yet they are the majority of the world population. That’s why trash management has been a crucial issue worldwide. Overflowing of trash bins is a common scenario in most of the developing countries. Also, there is a tendency among people of these countries to dump the trash not inside the trash can but outside the can. The surrounding area of the trash can becomes a breeding place for germs. This is very unhygienic and awkward. Passing by a roadside trash bin in that situation is obviously not a good experience for people, uninviting for newcomers, and especially unhealthy for kids and senior citizens. Uncollected trash and litter along highways or other areas in developed countries pose a serious problem for the residents in terms of hygiene, neighborhood appeal, and environment protection. The World Health Organization [1] has indicated that 842,000 deaths per year globally are attributable to “unsafe water supply, sanitation and hygiene”. Of this total 361,000 are children under age five, mostly in low-income countries. Automatic trash collection systems, in addition to improving public health, will also reduce the cost of collecting trash, which is a big amount in both developed and developing countries. For example, CBSNewYork [2] published that New York city pays $300 million per year for collecting trash.

Recent progress in deep learning research has contributed greatly to unparalleled improvements in computer vision. Convolutional neural networks (CNN) are one of the most powerful deep-learning algorithms which has many applications in image classification, segmentation, and detection [3-6]. Therefore, in this paper CNN is proposed to perform trash detection and recognition. Chu et al. [7], proposed a multilayer hybrid deep-learning system (MHS) that can sort trash disposed of by individuals in an urban public area. The system can automatically sort trash items as recyclable or otherwise. They used the AlexNet CNN [3] to extract key image features and optical sensors to detect other numerical feature information. This system used multilayer perceptrons (MLP) to classify the trash object by consolidating information collected from diverse channels. The proposed MHS achieved a mean accuracy higher than 90%, but the system can classify only 22 fixed items of trash in public areas. Other trash items on the road or in a park would not be counted in their system.

Bai et al. [8] presented a garbage pickup robot which can detect trash accurately and autonomously on the grass. They used a deep neural network, ResNet [9], for trash recognition and a navigation strategy to guide the robot to move around. With the trash recognition and automatic navigation functions, the robot can clean trash on the ground in parks or schools automatically. Their trash recognition accuracy reached above 95%. But the robot can detect trash only on grass. So, the trash on road or parking areas could not be identified by the robot.

The above two research efforts achieved very high accuracy in using CNN architectures. Based on these works, we are proposing a system which can identify trash items from any public spaces such as along a road, in a parking lot, in a recreation area or park, a community space, etc. Our ultimate goal is to build a trash collecting robot that self-navigates in a park or public space, looking for objects on the ground. This paper is the first step towards that goal, to identify the objects from images. Figure 1 displays a flow chart of the decision-making process. People passing-by may throw objects into the trash robot, in which case the object will be classified as trash or recyclable and stored into separate inner bins. In addition to having a trash receptacle, the robot will be equipped with a camera to capture images and decide whether to take an object or not. It will pick up any object that it perceives to be either trash or recyclable. After grabbing an object, the trash collecting robot would then bring the object inside itself to examine more closely. With a clearer image of the object, the robot would then classify the object into one of two categories: trash or recyclable. The robot will classify recyclable items as metal, plastic, glass, or fiber. Fiber includes any paper or cardboard item. In this research effort, we trained the AlexNet CNN [3] to classify images firstly as either "take" or "non-take" and secondly as either trash or recyclable. Testing of these CNN's with real outdoor images in public spaces produced quite accurate results.

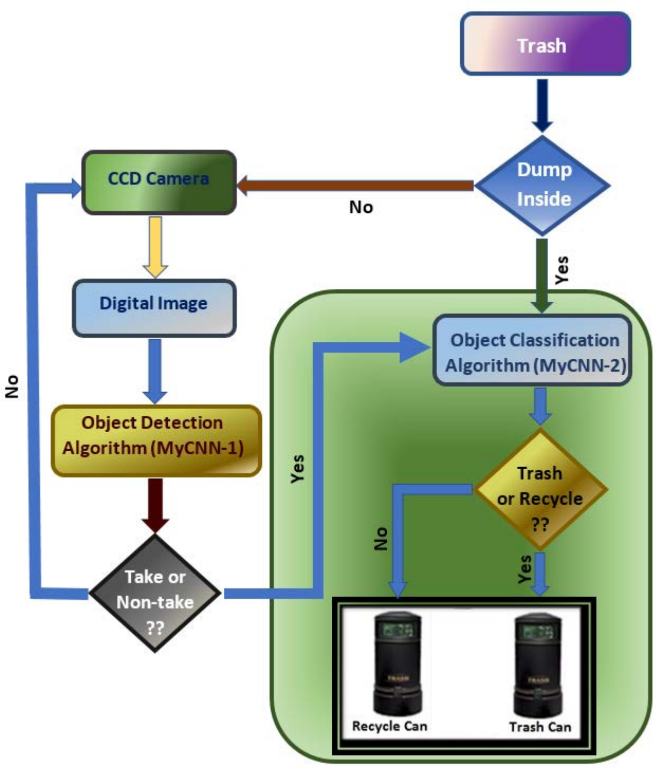


Figure 1: Flow Chart of the Decision-Making Process

II. Approach (Theory)

A **Convolutional Neural Network (ConvNet/CNN)** is a Deep Learning algorithm. It takes an image as input, then assigns learnable weights and biases to various features/objects in the image and finally differentiates one from the other. Convolutional neural networks (CNN) have been used by other researchers to analyze digital images for object recognition or classification [3-6]. Instead of primitive methods such as filtering by a hand-engineered process, after proper training, CNNs can filter/categorize images into different classes based on the input-output pairs. The architecture of a CNN is similar with the internal connecting system of neurons in the human brain. It is a combination of convolutional layers, pooling layers, fully connected layers, and normalization layers [3-6]. In the convolutional layers, each kernel generates a feature map by convolving the input image with moving kernels with certain window size and stride size. The ReLU (rectifier linear unit) is applied on the output to avoid gradient vanishing, and then pooling is applied to reduce noise and feature dimensions. After multiple convolutional layers and pooling, the features are then flattened to be fed into the fully connected layers, where each layer consists of sets of nodes (artificial neurons) in columns, and the output of every node (activation neuron) of a layer is mapped to the input of all nodes in the next layer.

In mathematical form, the fully connected function can be expressed by the following equations of a forward pass and backward pass propagation rules [9]. The components of input vector **x** are the outputs of layer 1 that can be expressed as $a_j(1) = x_j$, where j=1, 2, 3….,n_1 and $n_1 = n$ is the dimensionality of **x**. The computation performed by neuron i in layer l is given by

$$z_i(l) = \sum_{j=1}^{n_{l-1}} W_{ij}(l) a_j(l-1) + b_j(l) \tag{1}$$

where i=1,2, 3,…., n_l and l=2,……, L and $z_i(l)$ represent the net input to neuron i in the layer l, which is formed using all outputs from layer l-1. $b_j(l)$ is the bias value associated with the i^{th} neuron in the l^{th} layer. The activation value of neuron i in layer l is given by $a_i(l) = h\big(z_i(l)\big)$ for i =1,2,3,….., n_l where h is an activation function. The value of network output node i is $a_j(L) = h\ \big(z_i(L)\big)$ for i=1,2,3,….., n_L. These are all the operations required to map the input of a fully connected feedforward network to its output. The relationship between the net input and the output of any neuron in any layer (except the first layer) is the same which could be denoted by $\delta_j(l)$ for any node j in any hidden layer. $\delta_j(l)$ can be expressed as

$$\delta_j(l) = \frac{\partial E}{\partial z_j(l)} \tag{2}$$

where E is the error of the j^{th} neuron. As we will be proceeding backward in the network, we need the relationship between $\delta_j(l)$ and $\delta_j(l+1)$ so that we can start with $\delta_j(L)$ and find $\delta_j(L-1)$, $\delta_j(L-2)$, and finally reach at layer 2. The equation can be expressed by,

$$\delta_j(l) = h'\left(z_j(l)\right) \sum_i W_{ij}(l+1)\, \delta_i(l+1) \qquad (3)$$

Through some algebra it can be shown that the rate of change of error with respect to network weights is:

$$\frac{\partial E}{\partial w_{ij}(l)} = a_j(l-1)\delta_i(l) \qquad (4)$$

Similarly, the rate of change of error with respect to biases is:

$$\frac{\partial E}{\partial b_i(l)} = \delta_i(l) \qquad (5)$$

These rates of change are used in the backpropagation model to update the weights and biases:

$$w_{ij}(l) = w_{ij}(l) - \alpha a_j(l-1)\delta_i(l) \qquad (6)$$

$$b_i(l) = b_i(l) - \alpha \delta_i(l) \qquad (7)$$

The accuracy of CNN can be enhanced by fine tuning dimensional parameters and local architecture structure [8]. Various CNN architectures of different variations have emerged in recent years [3]. AlexNet [3] is used in this work due to its in-field processing capabilities and low computational cost. AlexNet [3] was introduced in the 2012 ImageNet Challenge (ILSVRC) by significantly reducing the image classification top-5 error from 26% to 15.3%. It is well recognized for its highly capable architecture.

In this study, we used the MATLAB® version of AlexNet that consists of 25 layers including 5 Convolutional Layers and 3 Fully Connected Layers as shown in Table 1. Multiple Convolutional Kernels are used to extract required features in an image. Many kernels of the same size are used in a single convolutional layer. The output of the last Fully Connected Layer is fed into the 1000-way softmax function corresponding to 1000 class labels. Cross-Channel Normalization layer is associated with the fourth and eighth layers. Max-pooling layers are placed after the Cross-Channel Normalization layers and the sixteenth layer. The ReLU nonlinearity is used after each convolutional layer. The neurons in the fully connected layers are connected to all neurons in the previous layer, with 4096 neurons each [3]. The number of the neurons in the last Fully Connected Layer (layer 23) is set to the number of categories (02) discussed in Section III.

Table 1: MATLAB AlexNet Layer Configuration

Layer	Type
1	Data (227x227x3 Size Images)
2	96 kernels of size 11x11x3 Convolutions
3	ReLU
4	Cross Channel Normalization
5	3x3 Max Pooling
6	256 kernels of size 5x5x48 Convolutions
7	ReLU
8	Cross Channel Normalization
9	3x3 Max Pooling
10	384 kernels of size 3x3x256 Convolutions
11	ReLU
12	384 kernels of size 3x3x192 Convolutions
13	ReLU
14	256 kernels of size 3x3x192 Convolutions
15	ReLU
16	3x3 Max Pooling
17	4096 Fully Connected Layer
18	ReLU
19	50% Dropout
20	4096 Fully Connected Layer
21	ReLU
22	50% Dropout
23	1000 Fully Connected Layer
24	Softmax
25	Classification Output

III. METHOD AND RESULTS

We developed a set of procedures for training a Convolutional Neural Network (CNN) to classify objects as either trash, recyclable or other. We used the AlexNet CNN architecture and trained it with take or not take images on the public spaces for a smart robot trash can to decide to grab an object or not. More specifically, "take" means the identified item is a trash item to be grabbed, and "non-take" means the identified item is not a trash item that should not be grabbed.

We trained AlexNet to perform a set of four tests as follows:

1. Trained AlexNet with TrashNet [11] images and tested the resulting CNN with a subset of TrashNet images in 5 categories (metal, plastic, glass, paper, cardboard).
2. Tested the same CNN using an indoor camera in real time focusing of trash objects.
3. Trained AlexNet with outdoor images to classify as either "take" or "non-take". Tested the resulting CNN with a subset of outdoor images.
4. Trained AlexNet with outdoor images to classify as either landfill trash or recyclable. Tested the resulting CNN with a subset of outdoor images.

Tests 1 and 2 were preliminary tests to confirm accuracy of the AlexNet CNNs. We downloaded a publicly available CNN called Deep Learning Toolbox Model for AlexNet Network [12] for use in developing an algorithm in MATLAB. We modified the AlexNet architecture by changing the number of neurons in the last Fully Connected Layer to suit our requirements. We also downloaded a publicly available database, named TrashNet [11], of trash images taken in an indoor environment and separated over 2000 images into 5 categories (metal, plastic, glass, paper, cardboard) to train the CNN. The preliminary training results on the TrashNet indoor images confirmed the applicability of CNN in this application with good accuracy.

Tests 3 and 4 are practical tests that could be implemented on the final trash robot design. The outdoor "take" and "non-take" images were all taken by us from the surroundings of human living areas on a college campus. Every image used to train the CNN is a real scenario of trash in our area. For the second task, we trained another AlexNet CNN using the "take" item images to further classify the items into landfill trash or recyclable.

The detailed procedure concerning each of these tests are described below.

Test 1 – Five Categories in a Controlled Indoor Setting

As a primary test, deep learning methods for implementing Convolutional Neural Networks (CNN) in MATLAB were used to train AlexNet in the 5 categories (metal, plastic, glass, paper, cardboard) of images. We tested the accuracy of our trained version of AlexNet using a subset of the TrashNet images. Results are shown in Table 2. Accuracy of detection exceeded 80 % for all 5 categories. It should be noted that the images used in this test were taken in a controlled indoor environment with a consistent lighting background. That helps to explain why the results shown above are very accurate. Several examples of the images used in the test are shown in Figure 2.

Table 2: Results of CNN classification using TrashNet indoor images

Category	Total count of images	Count of correctly detected images	Accuracy (%)
Metal	41	39	91.68
Plastic	48	38	81.25
Paper	59	53	89.83
Cardboard	40	37	92.5
Glass	50	46	92
Overall	238	213	89.50

Category: Metal

(a) Example of training images used as metal

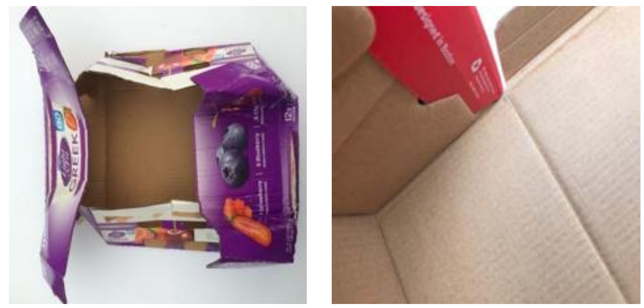
Category: Plastic

(b) Example of training images used as plastic

Category: Paper

(c) Example of training images used as paper

Category: Cardboard



(d) Example of training images used as cardboard

Category: Glass

(e) Example of training images used as glass

Figure 2. Examples of training images used to detect five categories of recyclable materials.

Test 2 – Five Categories using a Real-Time Camera

Then, we proceeded to a camera focused on a white display board background in an indoor office capturing images of 13 different objects. Each object was rotated to obtain images of each object at 10 different viewing angles, for a total of 260 captured images. A set of 10 images used to test one object is shown in Figure 3. Each image was tested using our trained version of AlexNet (from Test 1), which classified each image into one of 5 categories (metal, plastic, glass, paper, cardboard).

A summary of the test results is shown in Table 3. The trained AlexNet was able to identify 7 out of 13 objects correctly with an accuracy of 90 % or higher and was able to identify 3 out of 13 objects correctly with an accuracy between 70 and 80 %. The CNN had some difficulty with three objects: brown paper, clear glass, and an orange plastic box. The brown paper was confused for cardboard and plastic and it was classified as 70% cardboard and 30% plastic. One of the possible reasons for this is that the CNN is unable to differentiate the color between cardboard and paper as the color is very close for both cases. In case of clear glass, the CNN was showing 60% glass and 40% plastic. The reason could be same reflection of light on glass and plastic. The same problem was showing for a clear plastic cup.

The MATLAB program used to implement the real time detection presents a figure of the real time image with the detected object listed in the figure title. This is shown in figure 4 in which the captured images of the 13 objects that were used to test the CNN. These figures show some sample results. For example, in Figure 4 object 1, a plastic bag is correctly identified in the plastic category. In Figure 4 we show a variety of objects. Some are correctly identified; some are not.

Figure 3: One full test procedure for an object

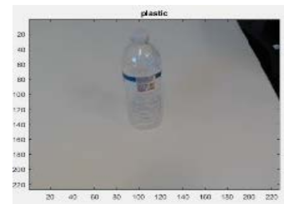
Table 3: Results of AlexNet CNN classification using indoor camera, trained on TrashNet images

Object number	Object	Actual Category	Detected Category
1	Green Plastic bag	Plastic	100% Plastic
2	Plastic bottle	Plastic	100% Plastic
3	Plastic box	Plastic	100% Plastic
4	White glass mug	Glass	100% Glass
5	Red Plastic cup	Plastic	80% Plastic, 20% metal
6	Brown paper	Paper	70% cardboard, 30% plastic
7	White paper with writing	Paper	70% Paper, 30% cardboard
8	Clear glass	glass	60% glass, 40% plastic
9	green soda can	metal	90% metal, 10% plastic
10	clear plastic cup	plastic	70% plastic, 30% glass
11	3D printed green object	plastic	100% plastic
12	Box	Cardboard	100% Cardboard
13	Orange plastic box	Plastic	20% plastic, 80% paper or cardboard

Object 1

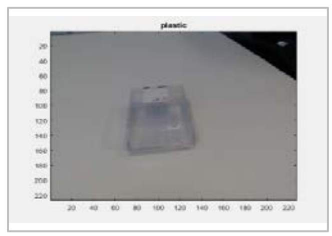
Correct detection of a plastic bag

Object 2



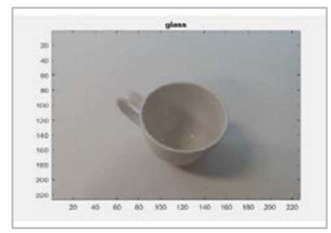
Correct detection of a plastic bottle

Object 3



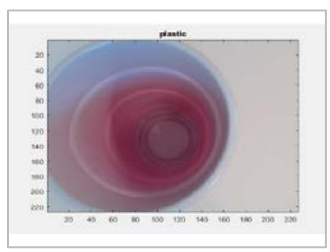
Correct detection of a plastic box

Object 4

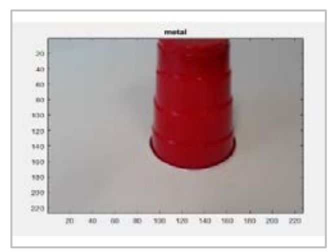


Correct detection of a white mug

Object 5

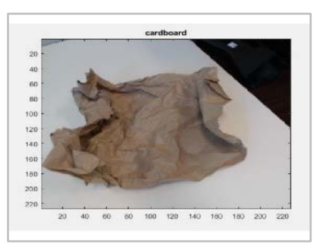


Correct detection of a plastic cup

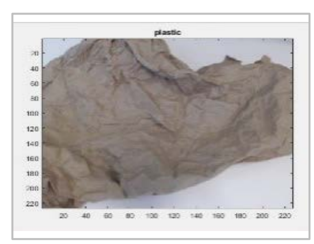


Incorrect detection of a plastic cup as metal

Object 6

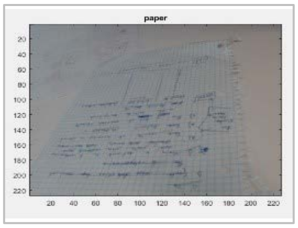


Incorrect detection of brown paper as cardboard

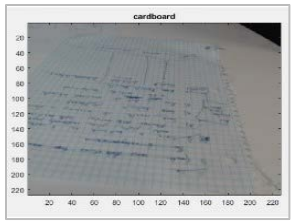


Incorrect detection of brown paper as plastic

Object 7

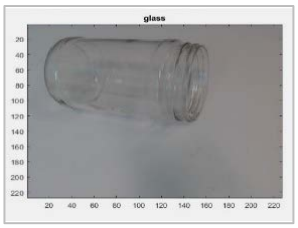


Correct detection of writing paper

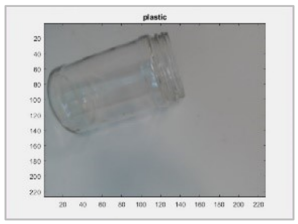


Incorrect detection of writing paper as cardboard

Object 8

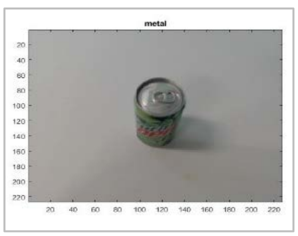


Correct detection of a clear glass jar

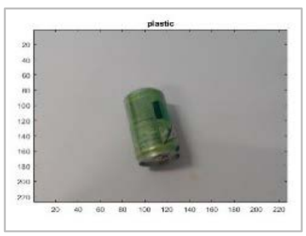


Incorrect detection of a clear glass jar as plastic

Object 9

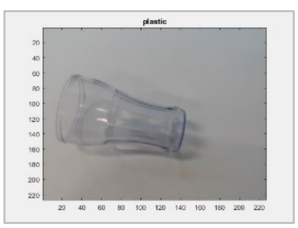


Correct detection of a green soda can

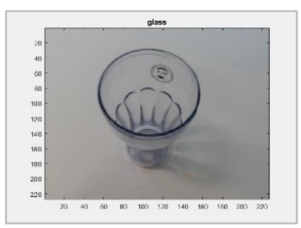


Incorrect detection of a green soda can as plastic

Object 10

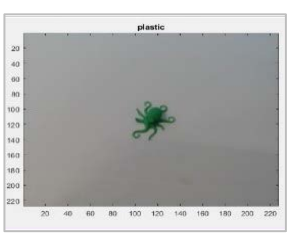


Correct detection of a clear plastic cup



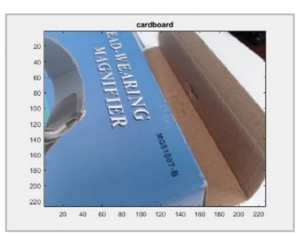
Incorrect detection of a clear plastic cup as glass

Object 11



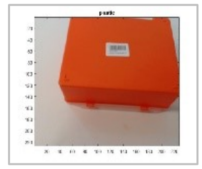
Correct detection of 3d printed object

Object 12

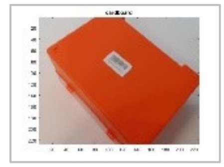


Correct detection of a box

Object 13



Correct detection of an orange box



Incorrect detection of an orange box as cardboard

Incorrect detection of an orange box as paper

Figure 4: Example of images of 13 objects used for testing

Test 3 – Classifying Outdoor Images as either Take or Non-take

To develop training for outdoor images, we took 1054 digital pictures of outdoor scenes, and then put them into two categories of “take” or “not take”. For the first outside test we captured the photos of items mainly on the grass, on a sidewalk, on the road, or in a flower bed. We then trained AlexNet on these 1054 images. Images in the “take” category included trash and recyclable items. Images in the “non-take” category included pictures of grass, birds, trees, sidewalk, etc. We tested the CNN on 316 similar types of images from these two categories (210 “take” images and 106 “non-take” images). The overall classification accuracy was 93.6 %. 97.6 % of the “take” items were correctly identified and 85.8% of the “non-take” items were correctly identified. Table 4 shows the results of this test. Figures 5 and 6 show a few examples of training and test images used for classification of outdoor objects. When we test with a given image, the CNN algorithm makes a decision of “take” or “non-take” and displays the image with a title indicating the decision, as shown in Figure 7.

Table 4: Results of CNN classification using outside images with 2 categories

Category	Total count of images	Count of correctly detected images	Accuracy (%)
“take”	210	205	97.6
“non-take”	106	91	85.9
Overall	316	296	93.6

“take” images

(a) Take image on grass

“non-take” images

(b) Non-take image on road

(c) Take image on grass

(d) Non-take image on road

Figure 5: Example of training images for outdoor object classification

"take" images

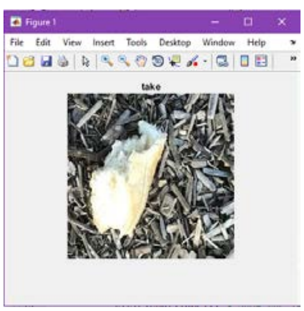
(a) Take image on grass

"non-take" images

(b) Non-take image on road

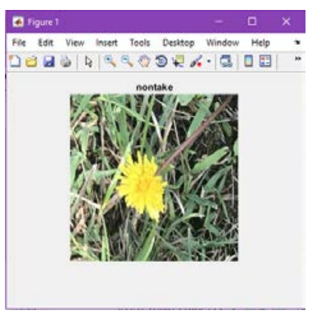
Figure 6: Example of test images for outdoor object classification

"take" images



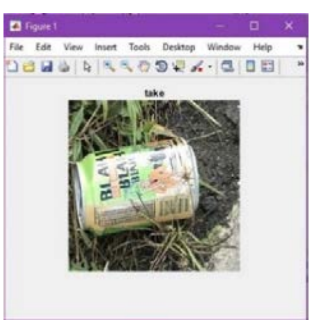
(a) Correct detection of take object in a park

"non-take" images



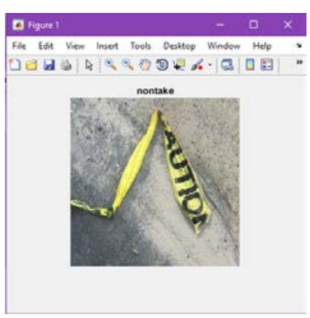
(b) Correct detection of non-take object on grass

"take" images



(c) Correct detection of take object beside road

"non-take" images



(d) Correct detection of non-take object on road

Figure 7: Sample output of test images for outdoor object detection

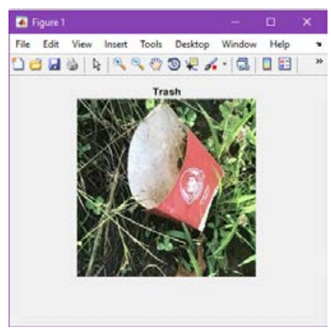
Test 4 – Classifying Outdoor Images as either Trash or Recyclable

Then we split the "take" image database into two categories "trash" and "recyclable" and trained another AlexNet CNN with 700 outdoor images from the "take" category. We then tested the CNN on 175 similar types of images from these two categories (107 "trash" images and 68 "recycle" images). The results (shown in Table 5) were as accurate as before. The overall classification accuracy was 92%. 89.7% of the "Recycle" items were correctly identified and 93.5% of the "Trash" items were correctly identified. Figure 8 shows several example of output images with decision of "trash" or "recycle" indicated in the image title.

Table 5: Results of CNN classification using outside images with 2 categories

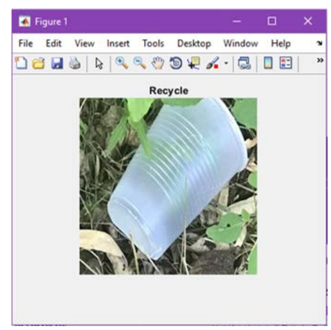
Category	Total count of images	Count of correctly detected images	Accuracy (%)
"recycle"	68	61	89.7
"trash"	107	100	93.5
Overall	175	161	92

"trash" images



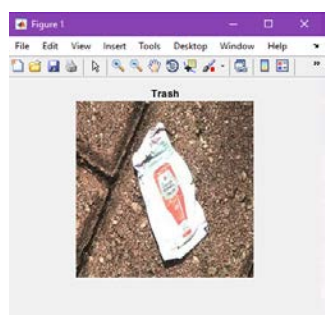
(a) Correct detection of trash object on grass

"recycle" images



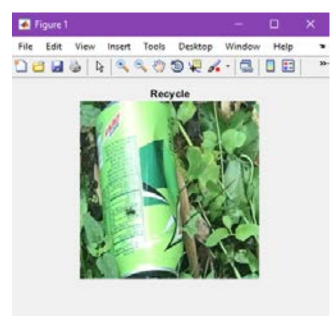
(b) Correct detection of recycle object on grass

"trash" images



(c) Correct detection of trash object on sidewalk

"recycle" images



(d) Correct detection of recycle object on grass

Figure 8: Sample output of test images for outdoor object classification

IV. CONCLUSION

In this study we developed an CNN-based algorithm for detecting trash and non-trash, as well as further differentiating landfill and recyclable items in the trash category, for the purpose of developing an automatic trash collection system. Results were positive with an accuracy of detection ranging from 89.7% to 93.5%. Integrating this image processing-based classification into smart trash cans will be more suitable for cleaning garbage on public spaces than the existing cleaning mechanisms used by road sweeper trucks or vacuum cleaning. Experimental results proved that the proposed algorithm can recognize garbage and recycled material accurately. This algorithm can serve as a powerful tool for designing a trash can robot for cleaning the garbage on a big lawn in a park or school. Future work will consist of using our two-stage trained CNN in an algorithm that can work with a microcontroller and a camera to move a trash can robot around a public space and identify an object on the ground, then pick and sort the trash as landfill or recyclable.

REFERENCES

[1] World Health Organization Website, https://www.who.int/water_sanitation_health/diseases-risks/diseases/en/

[2] CBSNewYork, https://newyork.cbslocal.com/2018/01/29/nyc-garbage-pickup-charges/

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in Proceedings of Neural Information Processing System Conference, Lake Tahoe, CA, USA, December 2012.

[4] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, andY. Lecun, “OverFeat: integrated recognition, localization and detection using convolutional networks,” in Proceedings of International Conference on Learning Representations, Banff,Canada, April 2014.

[5] M. D. Zeiler and R. Fergus, “Visualizing and understanding convolutional neural networks,” in Proceedings of 13th European Conference on Computer Vision (ECCV), Zurich, Switzerland, September 2014.

[6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of Conference on Computer Vision and Pattern Recognition, Las Vegas Valley, NV, USA, June 2016.

[7] Chu, Y., Huang, C., Xie, X., Tan, B., Kamal, S., & Xiong, X. (2018). “Multilayer Hybrid Deep-Learning Method for Waste Classification and Recycling,” Computational Intelligence and Neuroscience, 2018.

[8] J. Bai, S. Lian, Z. Liu, K. Wang and D. Liu, "Deep Learning Based Robot for Automatically Picking Up Garbage on the Grass," in IEEE Transactions on Consumer Electronics, vol. 64, no. 3, pp. 382-389, Aug. 2018.

[9] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning forimage recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, 2016, pp. 770–778.

[10] Gonzales, R.C. and Woods, R.E., “Digital Image Processing,” 4th edition, Pearson, 2018.

[11] Trashnet, Dataset of images of trash, https://github.com/garythung/trashnet.

[12] AlexNet Tool Box, Mathworks, https://www.mathworks.com/matlabcentral/fileexchange/59133-deep-learning-toolbox-model-for-alexnet-network.