Enhanced Garbage Image Recognition System

Wei Chen, Yan Xuan, Yiqing Wang

[*] College of Engineering, Northeastern University

Abstract- Due to the rapid development of cities, the amount of garbage is also increasing. If it is not sorted and recycled, it will have adverse effects on people's health and living environment. When it comes to waste recycling, the key to effective waste management is ensuring that waste is properly sorted and goes through the different recycling and resource recovery processes. Therefore, in this project, we use YOLO(You Only Look Once) and CNN(Convolutional neural network) algorithms to classify junk images. The goal is to realize an intelligent garbage recognition system based on deep learning to help the garbage disposal industry to carry out efficient and automatic garbage disposal and classification. The training samples are used to train the neural network, and then the test samples are input to test the accuracy of the neural network, which verifies the accuracy of the convolutional neural network in identifying junk pictures. And use the same data to train and test the YOLO algorithm, and compare the accuracy of the recognition results.

Index Terms- CNN algorithm, YOLO algorithm, Trash Detection, Image Recognition

I. INTRODUCTION

With the development of science and technology, people in modern society pay more and more attention to the classification, recycling and treatment of garbage. In terms of garbage disposal, how to classify unsorted messy garbage has always been a problem that plagues environmentalists. In this paper, we propose to use deep learning methods to solve this problem. We use deep learning algorithms to identify garbage images, so as to achieve the purpose of classification and processing. The algorithm chosen to be used is the YOLO(You Only Look Once) algorithm. In order to make its effect matchable, we also chose the CNN(Convolutional neural network) algorithm.
Data Source: Garbage Classification https://www.kaggle.com/datasets/asdasdasasdas/garbage-classification

A. CNN(Convolutional neural network)

Convolutional Neural Networks also known as CNNs or ConvNets, are a type of feed-forward artificial neural network whose connectivity structure is inspired by the organization of the animal visual cortex. Small clusters of cells in the visual cortex are sensitive to certain areas of the field. Individual neuronal cells in the brain respond or fire only when certain orientations of edges are present. Some neurons activate when shown vertical edges, while others fire when shown horizontal or diagonal edges. in deep learning to evaluate visual information. These networks can handle a wide range of tasks involving images, sounds, texts, videos, and other media. Convolutional neural network is a feedforward neural network consisting of several convolutional layers and pooling layers. Convolutional neural networks are widely used in image processing. Its network structure is shown in Figure 1.

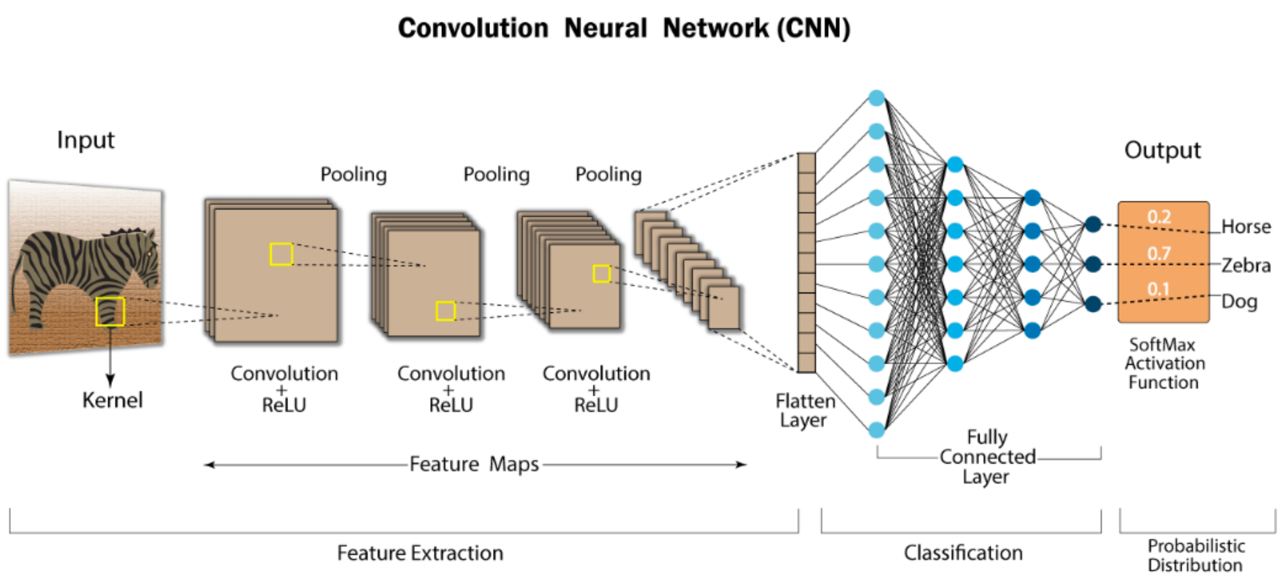

Figure 1. Convolutional neural network(CNN)

Convolutional Neural Networks (CNNs) have an input layer, an output layer, numerous hidden layers, and millions of parameters, allowing them to learn complicated objects and patterns. It uses convolution and pooling processes to subsample the given input before

applying an activation function, where all of them are hidden layers that are partially connected, with the completely connected layer at the end resulting in the output layer. The output shape is similar to the size of the input image.
Convolution is the process of combining two functions to produce the output of the other function. The input image is convoluted with the application of filters in CNNs, resulting in a Feature map. Filters are weights and biases that are randomly generated vectors in the network. Instead of having individual weights and biases for each neuron, CNN uses the same weights and biases for all neurons. Many filters can be created, each of which catches a different aspect from the input. Kernels are another name for filters.

B. YOLO(You Only Look Once)

On the whole, the Yolo algorithm uses a separate CNN model to achieve end-to-end target detection. The whole system is shown in Figure 2: first, the input image is resized to 448x448, then sent to the CNN network, and finally the network prediction result is processed to obtain the detected target. Compared with the R-CNN algorithm, it is a unified framework, which is faster, and the training process of Yolo is also end-to-end.

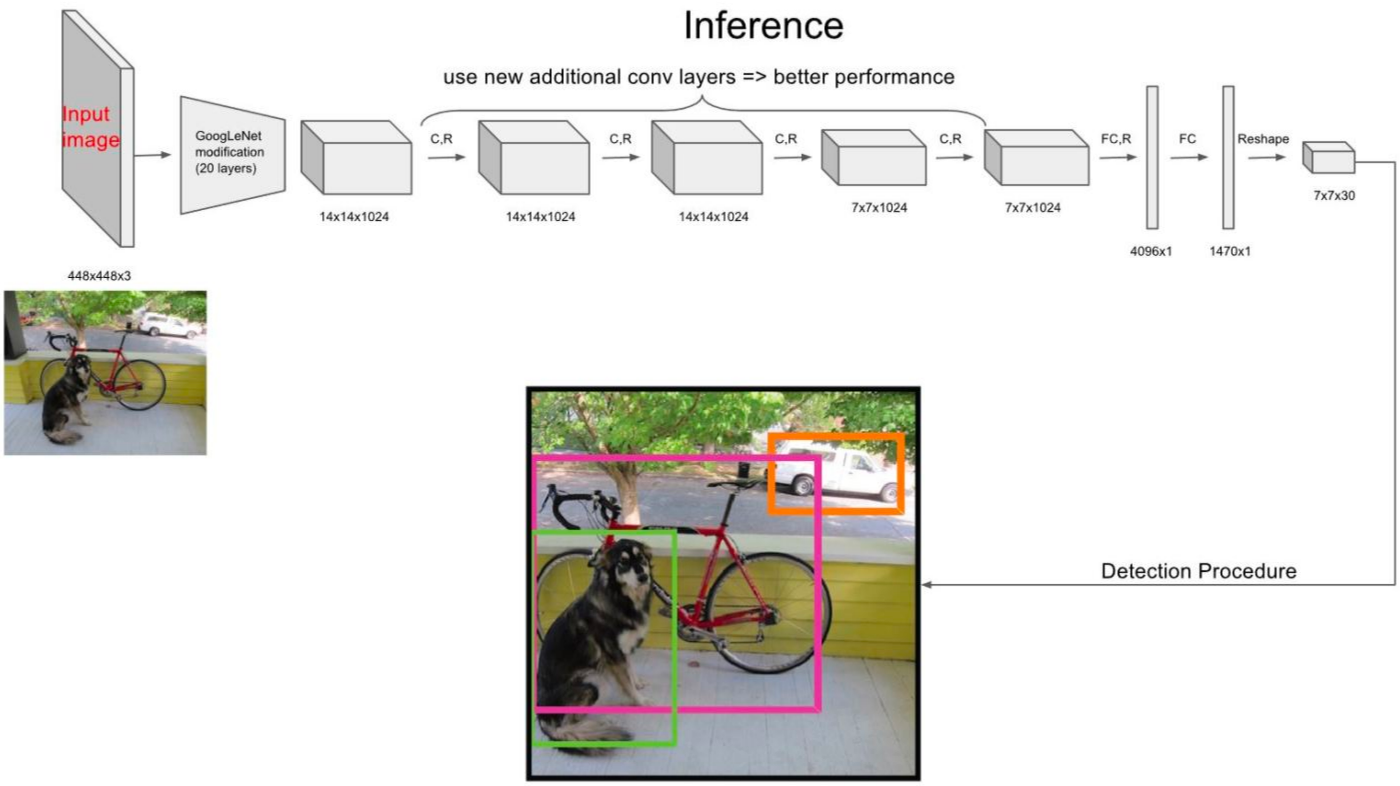

Figure 2. YOLO(You Only Look Once)

In terms of network design, Yolo uses a convolutional network to extract features, and then uses a fully connected layer to obtain predicted values. The network structure refers to the GooLeNet model, which includes 24 convolutional layers and 2 fully connected layers, as shown in Figure 8. For the convolutional layer, 1x1 convolution is mainly used for channel reduction, followed by 3x3 convolution. For the convolutional layer and the fully connected layer, the Leaky ReLU activation function is used: Max(x,0.1x). But the last layer uses a linear activation function.
The advantage of the Yolo algorithm is that it uses a CNN network to achieve detection. It is a single-pipeline strategy. Its training and prediction are end-to-end, so the Yolo algorithm is relatively simple and fast. And because Yolo is convolving the entire picture, it has a larger field of view when detecting targets, and it is not easy to misjudge the background. In addition, Yolo has strong generalization ability, and the model has high robustness when doing migration.

II. EXPERIMENTAL VERIFICATION AND ANALYSIS

A. CNN(Convolutional neural network)

After training on the Convolutional Neural Network, The performance of the model is shown below.

```
1/1 [==============================] - 0s 15ms/step
Maximum Probability:  0.9678255
Classified: plastic
```

Figure 3. Visualization of model results Part1

```
CARDBOARD : 0.29 %
GLASS : 0.27 %
METAL : 0.0 %
PAPER : 2.61 %
PLASTIC : 96.78 %
TRASH : 0.04 %
```

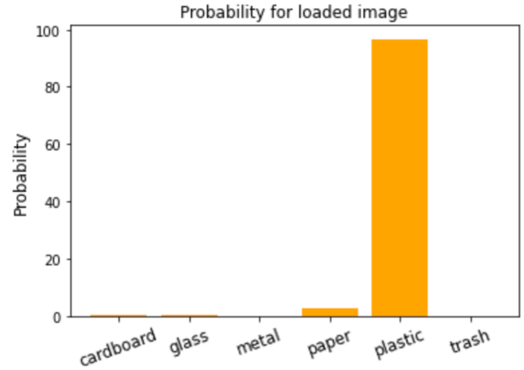

Figure 4. Visualization of model results Part2

As shown in Figure 3, the convolutional neural network algorithm model accurately identified the pictures we randomly selected from the test set and classified them as plastic waste. In Figure 4, we can see a more specific display, the possibility of the model judging for each type of garbage when judging the classification of garbage: CARDBOARD: 0.29%, GLASS: 0.27%, METAL: 0.0%, PAPER: 2.61 %, PLASTIC : 96.78%, TRASH : 0.04%, it can be seen that in the process of recognizing the currently loaded picture, the convolutional neural network algorithm model not only recognizes successfully, but also has a high recognition accuracy. Figure 5 shows more prediction results.

Figure 5. Randomly sample some predictions from the test images

After getting a large number of successful prediction results, we use Accuracy Graph and Loss Graph to measure the performance of the model. As shown in the visualization results in Figures 7 and 8 below, we can see that, through testing and verification, during the training process, the accuracy of the CNN model is gradually improving, and the training loss is gradually decreasing. The accuracy score of our CNN trained model is 76.26 percent.

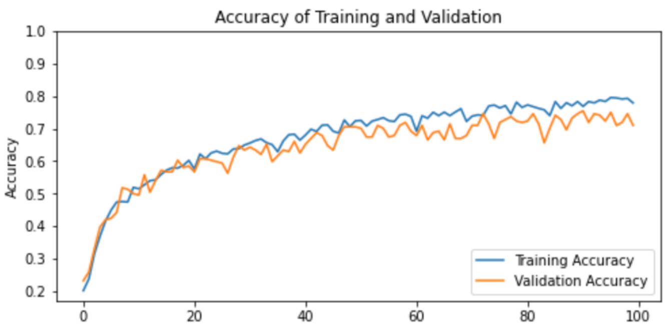

Figure 7. Accuracy of Training and Validation

Figure 8. Loss of Training and Validation

B. YOLO(You Only Look Once)

First we create a custom yaml model which contains the path of training data and testing data and all the names of categories. We train this custom model using original weights for 100 epochs. The performance of results are shown below as figure 9.

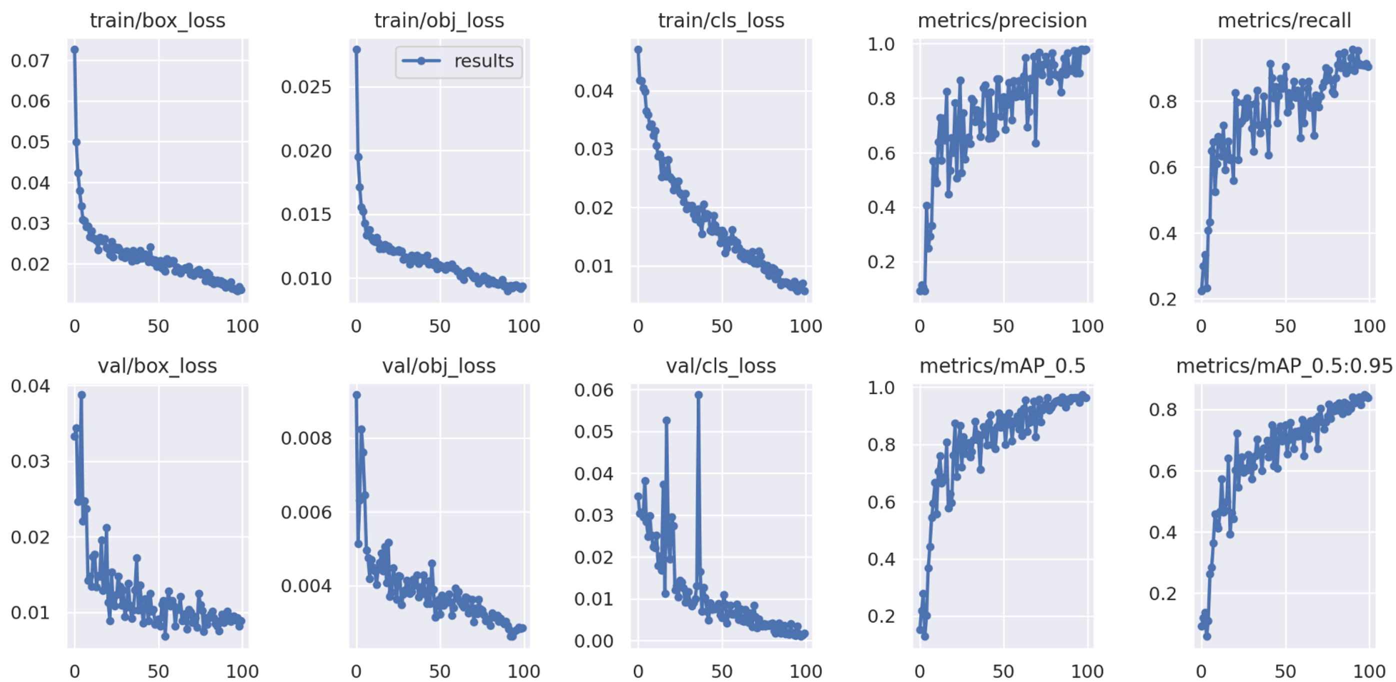

Figure 9 The performance of results

We know that there are three kinds of Loss functions of YOLOv5, which are Classification loss, Confidence loss and Localization loss. From figure 9, we can see that train/box_loss which is the mean value of the CIoU loss function, is very low, meaning that the boundary box is accurate. Train/obj_loss is low which means the target detection is accurate. Train/cls_loss is also low which means we get more accurate classification. We can also find that as the training time increases, the precision and recall also increase.

The AP value is an important evaluation index to measure the performance of the target detection model classifier. The larger the AP value, the better the performance of the classifier, and the smaller the performance of the classifier, the worse the performance of the classifier; the size of the AP value is equal to the area enclosed by the P-R curve and the coordinate axis area.

mAP represents the average value of AP values of all label categories, and the larger the mAP value, the better the model performance. mAP combines the trade-off between precision (Precision) and recall (Recall), and considers false positives (FP) and false negatives (FN), so this property makes mAP a reasonable evaluation index for the performance of most detection models.We can see that mAP_0.5 and mAP_0.5:0.95 are both close to 1 which means our model has a good performance.

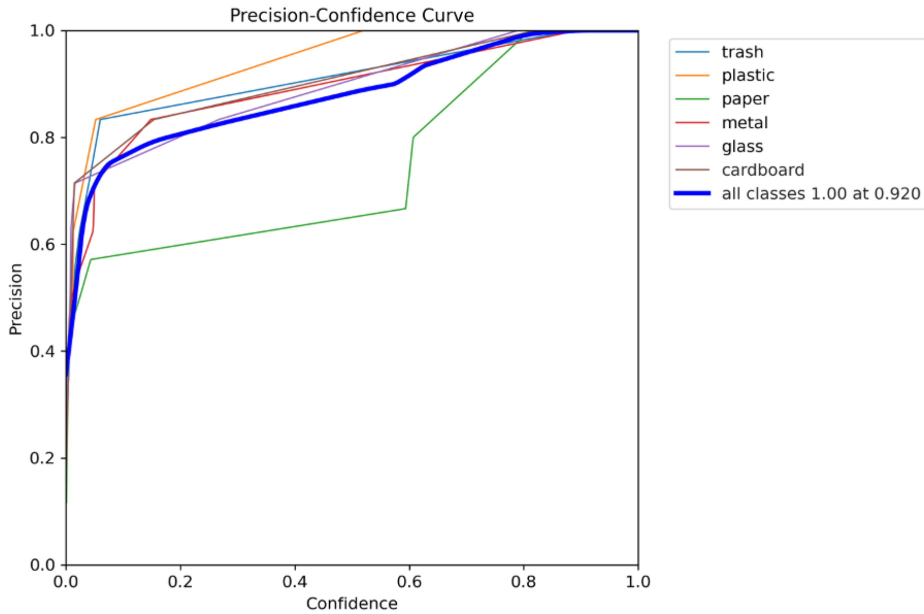

Figure 10 Precision and Confidence's curve

Figure 10 is P_curve showing the relationship between Precision and Confidence. Confidence is representing the confidence that the model believes that the detection box is a certain type of activity. We can see that the value of all classes is 0.92. and figure 11 is F1_curve, F1 score is also known as Balanced Score. it is defined as the harmonic mean of the correct rate and recall rate. In the case of β=1, the value range of F1-Score is 0 to 1, our F1 score of all classes is nearly 0.94 which is good.

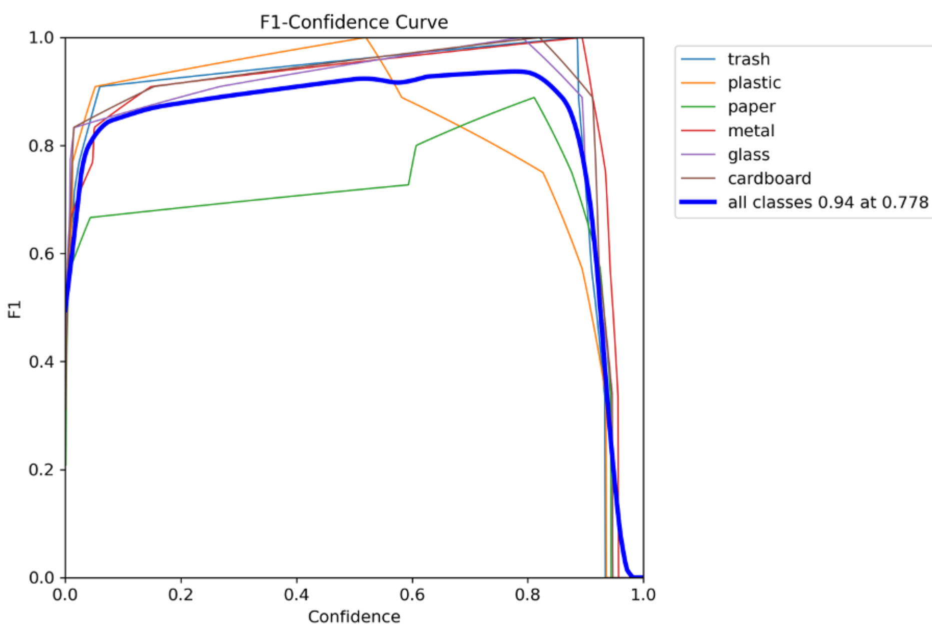

Figure 11 F1 score

From the training result, we know that our total mAP is higher than 0.97. trash: 0.995, plastic: 0.995, paper: 0.866, metal: 0.995, glass: 0.995, cardboard: 0.995. Figure 12 is the prediction image of our testing data, we can see that the boundary of detection is really accurate and nearly all the categories has high accuracy which proves our model has a good performance, we can also see that model has difficulties classify paper and plastic, which is also proven from the image of confusion matrix. Considering the similarity of the two materials, maybe we can use a more accurate and discriminative dataset to get better performance.

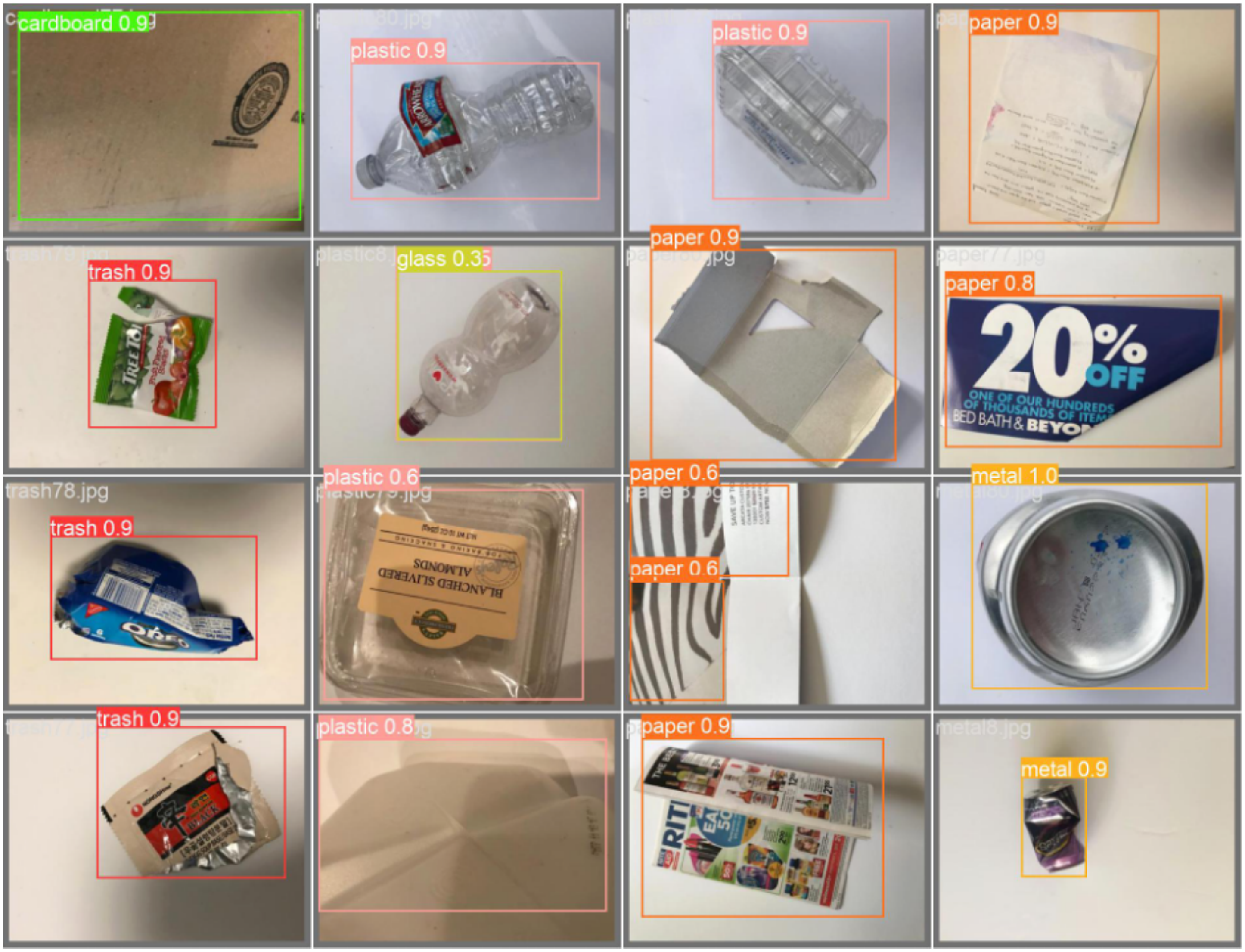

Figure 12 Training result shows

III. CONCLUSION

In this final project, we present, split, and test a dataset of more than 2000 images across six categories. We used CNN and YOLOv5 structure for model training. In the simulation prediction of the CNN model, we visualize the prediction effect of the model according to the prediction accuracy of the training model, and see the specific process of model training and the final performance of the model according to the model Accuracy and Loss of Training and Validation chart. We can see that the accuracy of the model is finally about 0.76. After comparing with other documents, it is found that the accuracy of the CNN network based on this data set is indeed not that high, and the accuracy of 0.76 is already a relatively excellent score.

In the YOLOv5 model, we are based on the three Loss functions of YOLOv5, namely Classification loss, Confidence loss and Localization loss; so we first used these three indicators as evaluation criteria. At the same time, the AP value is an important evaluation index to measure the performance of the target detection model classifier, so we also use this parameter as a standard. Based on the analysis of these four metrics, we obtain a model that is trained to have accurate bounding box and object detection, as well as accurate classification. We can also find that as the training time increases, the precision and recall also increase; and both mAP_0.5 and mAP_0.5:0.95 are close to 1, which means our YOLOv5 model has good performance.

According to our data segmentation and model training, the final accuracy of the CNN-based model did not meet expectations. After searching for information, it is found that the dataset performs well in ResNet, but there are also problems such as insufficient CNN layer settings, insufficient number of epochs, and insufficient training set data. This brings us to the importance of choosing a model and a richer dataset.

REFERENCE

[1] International Journal of Scientific and Research Publications, IJSRP Research Paper Format. https://www.ijsrp.org/ijsrp-paper-format.php
[2] Xiaoxiao Jiang. Target Detection: YOLO Principle and Implementation. https://zhuanlan.zhihu.com/p/32525231
[3] tf.keras.preprocessing.image.ImageDataGenerator, https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/image/ImageDataGenerator
[4] Classifying Recycling Material, https://www.kaggle.com/code/olmaneuh/classifying-recycling-material
[5] Basics of CNN in Deep Learning, https://www.analyticsvidhya.com/blog/2022/03/basics-of-cnn-in-deep-learning/
[6] Trash Management System, https://www.kaggle.com/code/maninair/trash-management-system
[7] Model conversion overview, https://www.tensorflow.org/lite/models/convert