Garbage Classification: Traditional Machine Learning vs Deep Learning

Adwait R Sambare
40162664
Concordia University
adwait395@gmail.com

Kshitij Yerande
40194579
Concordia University
kshitij.yerande@mail.concordia.ca

Siddhartha Jha
40201472
Concordia University
siddharthajha07@gmail.com

Abstract

Image classification problems have found numerous uses around the world in day-to-day life. Using deep convolution neural networks for the purpose of image classification has proven to even provide better precision and accuracy than what humans are capable of. Classifying garbage into multiple classes such as cardboard, glass, metal and so on is a typical multi- class classification problem. Transfer learning using deep pre- trained models that have been trained already on very large data sets which have many classes are particularly useful in an image classification problem like this. In this paper, traditional machine learning approach using HOG features is presented against deep learning approach using transfer learning(ResNet50) and a shallow convolution neural network for the classification of images of different classes of garbage. Both the approaches are evaluated, their evaluation metrics are presented and results are interpreted to identify the most suited model for this image classification problem.

Index Terms*-Deep Learning, Transfer Learning, neural net- works, image classification, Convolutional neural networks, ResNet-50, Histogram of Gradients, SVM*

1. Introduction

Garbage classification is crucial for minimizing harm to the environment. Although traditional methods of classification are in place, humans often fail to strictly adhere to them. Therefore, it is necessary to check and verify the garbage, resulting in an additional step performed by humans. With advancements in machine learning and artificial intelligence (AI), convolution neural networks (CNN) and hand crafted features using linear classifier have been created that can learn to classify garbage images. To train a CNN to identify a large number of classes, large data sets are required which can be a limitation is many cases. Transfer learning using pre-trained models allows for very deep models to be adapted to new data sets, proving to be much faster than training a network from scratch. This paper compares the traditional machine learning approach using a histogram of gradients with SVM classifier with deep learning approaches using a shallow neural network and transfer learning approach using ResNet50.

2. Related Work

Convolutional neural networks (CNNs) were used in 2010 to classify the 1.3 million high-resolution images in the LSVRC-2010 ImageNet training set into 1000 different classes, marking a turning point in image classification problems [3]. CNNs have been used for image-related computational tasks since the late 1980s but regained popularity due to computational power advancements [7]. CNNs have become the de facto for classification tasks. Organic waste was observed to be classified with better precision than other waste types using deep CNN architectures. Different transfer learning approaches have been compared for waste classification and segregation using deep neural networks such as AlexNet, GoogleNet, VGG, ResNet, and DenseNet [5]. Transfer learning with weight parameter tuning can achieve the best results on waste classification datasets. Implementation of these transfer learning methods for garbage classification has proven useful and optimized memory consumption through reducing the number of layers can improve predictor performance [9].

3. Solution Approaches

In paper we will discuss three solution appraches:

- **SVM with HOG features:** SVM classifier trained on histogram of gradients of images.
- **Transfer Learning**: Model trained on pre-trained ResNet-50 with final layer substitution and dropout.
- **Custom-CNN**: CNN architecture with three convolution blocks and intermediate pooling layers.

3.1. SVM with HOG features

A class of objects such as a cardboard, plastic etc. vary so much in color and shape. Structural cues like shape give a more robust representation. Gradients of specific directions captures some notion of shape. To allow for some variability in shape, we'll use features known as Histogram of Oriented Gradients (HOG) to classify waste items [4].

The concept behind HOG is that we arrange the pixels into small cells rather than using each pixel's specific gradient direction. We compute all of the gradient directions for each cell and divide them into various orientation bins.

3.1.1 Feature Extraction

To extract HOG features we can specify the **number of orientations**, **pixels_per_cell** , and **cells_per_block**. In our model we have used 9 orientation with 6 pixels per cell and 3 cells per block. The number of orientations is the number of orientation bins that the gradients of the pixels of each cell will be split up in the histogram. The pixels_per_cells is the number of pixels of each row and column per cell over each gradient the histogram is computed. The cells_per_block specifies the local area over which the histogram counts in a given cell will be normalized.

3.1.2 SVM Training

Support vector machine is a descriptive traditional machine learning classifier which linearly as well as non-linearly classifies various classification problems. It outperforms all its alternatives because support vectors are considered for separating different types of data using hyper plane [8]. The model is trained for 1000 epochs with $gamma = 0.001$ and $kernel = rbf$. A kernel is a function that takes the original non-linear problem and transforms it into a linear one within the higher-dimensional space. RBF(Radial Basis Function) is the default kernel used within the sklearn SVM classification algorithm.

3.2. Pre-trained ResNet-50

In this paper we have used ResNet-50 pre-trained model trained on ImageNet data set of 1.2 million images with 1000 classes [2]. The ImageNet data set contains some classes which are similar to our classes in the dataset and we would like to utilize the knowledge gained from ImageNet to train our model. The ResNets network architecture uses the Residual Learning method in its components. This method helps to keep the values of the parameters from falling into a state of saturation (vanishing gradient) by adding the value of the shortcut [6].

The 50-layer ResNet architecture includes the following elements [1]:

- A 7×7 kernel convolution alongside 64 other kernels with a 2-sized stride.
- A max pooling layer with a 2-sized stride.

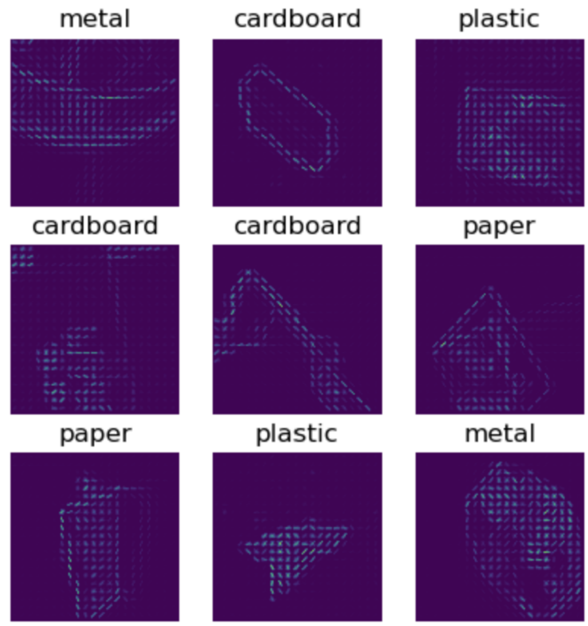

Figure 1. HOG images

- 9 more layers—3×3,64 kernel convolution, another with 1×1,64 kernels, and a third with 1×1,256 kernels. These 3 layers are repeated 3 times.
- 12 more layers with 1×1,128 kernels, 3×3,128 kernels, and 1×1,512 kernels, iterated 4 times.
- 18 more layers with 1×1,256 cores, and 2 cores 3×3,256 and 1×1,1024, iterated 6 times.
- 9 more layers with 1×1,512 cores, 3×3,512 cores, and 1×1,2048 cores iterated 3 times.
- Average pooling, followed by a fully connected layer with 1000 nodes, using the softmax activation function.

3.2.1 Training Steps

- Images are normalized and resized to 128x128 dimensions and are loaded into Pytorch Dataloader with specified batch size.
- ResNet-50 model pre-trained weights are loaded from Pytorch.
- **Final Layer:** The final linear layer of ResNet-50 pretrained model is defined as $2048 \longrightarrow 512$ and $512 \longrightarrow 6$ with activation function as ReLu and dropout 0.2.
- While training the final layer on our data set the pre-trained weights are not updated as back-propogation is disabled for those layers.

- The data is trained for 40 epochs with learning rate of 0.001.
- Trained model is saved on disk.

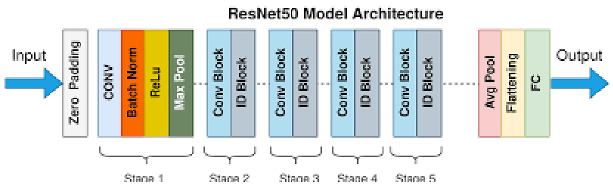

Figure 2. ResNet-50 Architecture

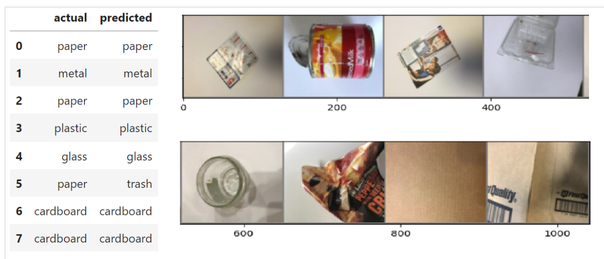

Figure 3. ResNet-50 Predictions

3.3. Vanilla Custom-CNN Model

The Custom-CNN model uses a 3 convolution layer architecture with intermediary pooling layers followed by 3 linear layers. The architecture of the model is as follows which takes a 128x128 RGB images as input :

- **Convolution Layer 1:** kernel size 11 with depth 16 and zero padding followed by batch normalization and ReLu activation function.
- **Pooling Layer 1:** size 3 and stride 2
- **Convolution Layer 2:** kernel size 7 with depth 32 and zero padding followed by batch normalization and ReLu activation function.
- **Pooling Layer 2:** size 3 and stride 2
- **Convolution Layer 3:** kernel size 3 with depth 32 and zero padding followed by batch normalization and ReLu activation function.
- **Linear Layer 1:** Linear layer with 0.5 dropout and ReLu activation connected to 256 neurons.
- **Linear Layer 2:** Linear layer with 0.5 dropout and ReLu activation connected to 128 neurons.
- **Final Linear Layer 3:** Linear layer connected to 6 neurons.

3.3.1 Training Steps

- Images are normalized and resized to 128x128 dimensions and are loaded into Pytorch Dataloader with specified batch size.
- Model is trained on the training images.
- The data is trained for 50 epochs with learning rate of 0.001.
- Trained model is saved on disk.

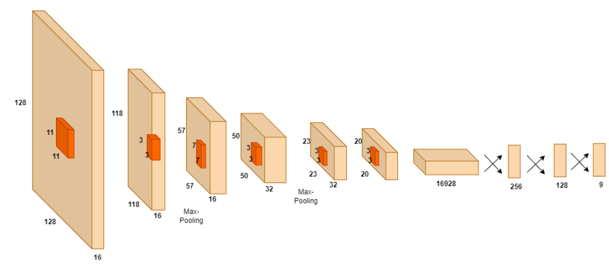

Figure 4. Custom-CNN architecture

4. Experimental Results

On training the three models on our data set, each model was tested on test set comprising of 794 images. The performance of each model was observed by examining four metrics accuracy,precision ,recall and f-score. Below is the table presenting the results of testing:

Model/Metric	Accuracy	F-score	Recall	Precision
ResNet-50 (Pre-trained)	80.2	0.793	0.788	0.802
Vanilla Custom-CNN	69.00	0.686	0.686	0.69
SVM - HOG features	56.4	0.535	0.526	0.618

Table 1. Experimental Results

4.1. SVM with HOG features

It is observed that the SVM model trained using histogram of gradients has accuracy of 56.4%. The confusion matrix is presented in Figure 5. The model is trained relatively faster than the convolution models. However the model is not able to classify most of the objects correctly. From the confusion matrix we can see that the model is predicting metal and plastic as glass. As histogram of gradients give notion of shape, the training data contains plastic bottles and glass bottles therefore the model is getting confused between glass and plastic models, however most of the objects have similar shape but different texture therefore the model is classifying most of the data inaccurately.

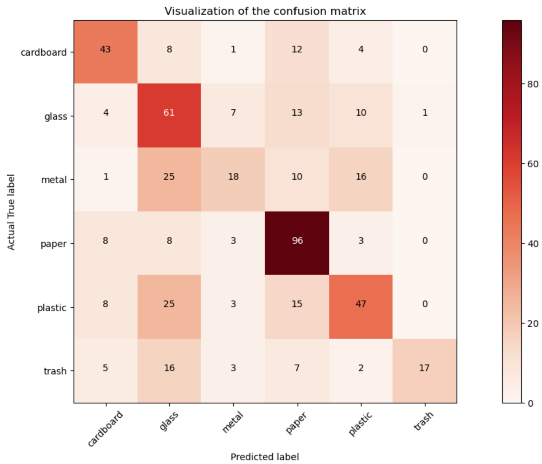

Figure 5. Confusion Matrix - HOG features with SVM

4.2. Pre-trained ResNet-50

Our Pre-trained ResNet-50 model performs the best and gives accuracy of 80.2%. This is because the ResNet-50 is trained on ImageNet data set comprising of 1.2 million images containing 1000 classes. Some of these classes are same as our data set and other are similar in nature. Therefore ResNet-50 is able to learn the features of our data set and able to give high accuracy. From the confusion matrix in figure 6 we can see that the model is able to predit most of the classes correctly. There are few incorrect predictions for plastic, metal and glass however the overall results are satisfactory.

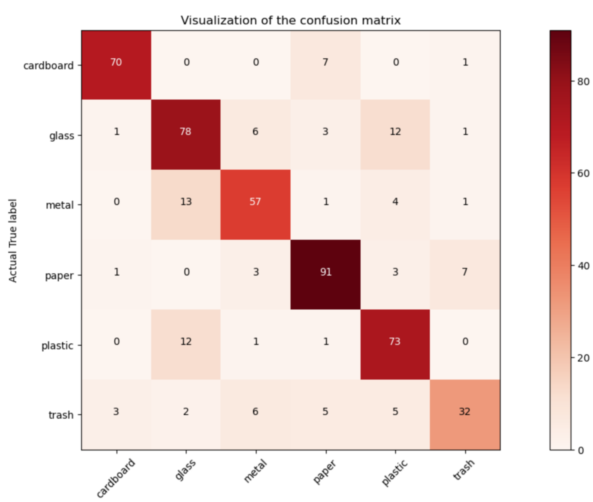

Figure 6. Confusion Matrix - ResNet-50

4.3. Vanilla Custom-CNN Model

Our Custom-CNN model achieves an accuracy 69%. From the confusion matrix in figure 7 we can see that the model incorrectly predicts metal and plastic and makes some mistakes in predicting all the classes. Thus the overall accuracy is less. The reason could be that the model architecture is not very deep and for plastic and glass the images contains bottles and both are white and transparent in nature therefore the model has difficulty in classifying plastic correctly. Similar reason could apply for metals. Improving the model architecture by introducing more convolution layers could help improve the performance of model.

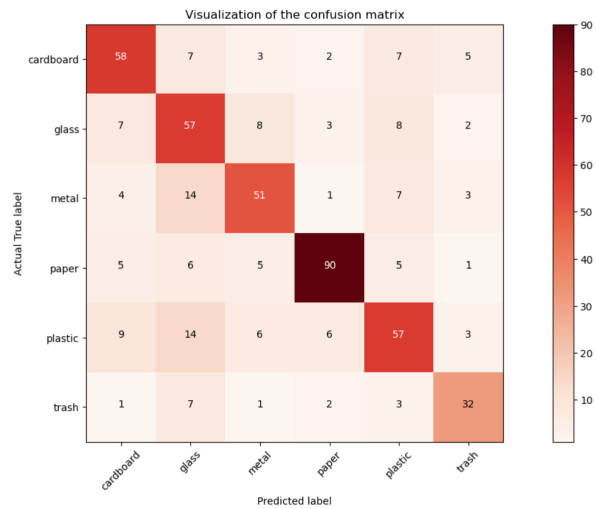

Figure 7. Confusion Matrix - Vanilla Custom-CNN

5. Conclusion

Transfer learning approach to image classification has proven to be very effective as compared to other approaches resulting in high accuracy and precision for the problem statement used in this paper. This is most likely due to its very deep architecture and huge training data of natural images in ImageNet dataset. The predictor performance of ResNet-50 is relatively good with yielding accuracy of almost 80%. The vanilla custom CNN model while showing reasonable test performance can be improved by using deep CNN architecture and more training data. The performance of the model could be improved by tuning the hyper parameters as there are no signs of over fitting. The feature-based model using HOG features with SVM classifier is a good approach in case of limited computing capability however the model has limitations as it tries to classify based on HOG features which consider the structure and shape of the object in images and in real world many objects in images from different categories can have similar shape and structure.

Github Link: https://github.com/Kshitij13579/Waste-Item-Classification

References

[1] Datagen. Resnet-50: The basics and a quick tutorial, 2018. 2

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. *CoRR*, abs/1603.05027, 2016. 2

[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. 1

[4] Mithi. Vehicle detection with hog and linear svm, 2017. 2

[5] Saifali Noorani and Milind Fernandes. Evaluation of convolutional neural networks for waste identification. In *2017 International Conference on Computing Methodologies and Communication (ICCMC)*, pages 204–207, 2017. 1

[6] Luyl-Da Quach, Nghi Pham Quoc, Nhien Huynh Thi, Duc Chung Tran, and Mohd Fadzil Hassan. Using surf to improve resnet-50 model for poultry disease recognition algorithm. In *2020 International Conference on Computational Intelligence (ICCI)*, pages 317–321, 2020. 2

[7] Wang Z. Rawat W. Deep convolutional neural networks for image classification: A comprehensive review. neural comput., 2017. 1

[8] Md. Golam Sarowar, Md Abdur Razzak, and Md Abdullah Al Fuad. Hog feature descriptor based pca with svm for efficient accurate classification of objects in image. In *2019 IEEE 9th International Conference on Advanced Computing (IACC)*, pages 171–175, 2019. 2

[9] Kaushal Yagnik, Mohit Garg, and Narayanan Krishnan. Spotgarbage: smartphone app to detect garbage using deep learning. pages 940–945, 09 2016. 1