GUBJSE: ISSN: 2409-0476

A Comparative Analysis of Convolutional Neural Networks for Trash Classification

Md. Atikuzzaman, Md. Parvez Hossain, Md. Zahidul Islam and Syed Ahsanul Kabir

***Abstract*—**In the era of the twenty-first century, automation is the biggest fi eld of research. Although several works have been done successfully, there is still so much to do. Increasing trash could be the biggest threat to a better human life in the future. To automate the task, machines need to understand the type of trash to be recycled or separate similar trash for recycling. For this process, the perfect classi cation of trash plays a vital role in making a better life and a cleaner world. There are so many popular convolutional neural network (CNN) models for image classi cation. In this work, we examine and analyze the outcomes of several Residual Network(ResNet) and Visual Geometry Group (VGG) CNN models on a trash dataset. Our main investigation is the accuracy evaluation for different VGG and ResNet models where the training dataset, test dataset, number of epochs, and batch size are the same for all the models. Finally, we compare VGG and ResNet models with each other. We have got the peak accuracy for ResNet152 among all ResNet models and the peak accuracy on VGG16 among all VGG models. And we have got the maximum accuracy on ResNet152 among all the ResNet and VGG models which is about 94%.

***Index Terms*—**CNN, Trash Detection, ResNet, VGG, Trash Classi cation

I. Introduction

AMOUNT of solid waste all over the world is increasing day by day. Each year, according to the World Bank report, nearly four billion tons of waste are increasing worldwide. Most of them are from urban areas. It is estimated that by 2025 the amount of waste will increase by 70 percent [1]. In [1], It has also shown that the amount of waste will increase rapidly within the next 25 years. As the number of industries in the urban area grows, the production of waste increases as well. Disposing of this massive amount of waste is becoming a big issue. Solid waste includes plastic, metal, paper, glass, wood, etc. There are several ways to dispose of this waste, such as Landfill, burning, etc. Land-filling waste is becoming a problem for the people living around landfill site. On the other hand, burning waste will be the reason for air pollution, which will be a threat to human health. By considering these issues, we believe that recycle of waste will be a way to solve these issues and protect the environment and people's health. The current way of recycling trash/waste needs a Man forced to separate trash/waste by hand-picking, which is a threat to human health who are involved with the process of separating waste because of existing injurious stuff. We've built an automated system to solve this issue, which can sort the waste/garbage/waste. This method takes less time and is more reliable than the manual process. After sorting the trash, with the help of the recycling process, it can convert waste into fuel and energy, which will help to increase the growth of the economy [2].

On a trash dataset, we trained various convolutional neural network models (ResNet [3] and VGG [4]). ResNet34, ResNet50, ResNet101, and ResNet152 are residual neural networks that have been trained. The VGG11, VGG16, and VGG19 models have also been trained. Following that, we compared all of the convolutional neural network models.

The rest of the paper is organized as follows. Section II discusses some related works. Section III discusses the dataset. Section IV discusses the Methodology followed by V Result Analysis, discussion, and conclusion.

II. Literature Review

There are many works done on classification using CNN [5], [6], [7].

Anh H.Vo et al. [8] developed a convolutional neural network model namely A novel framework for trash classification using deep transfer learning similar to ResNet and they get a very high accuracy of 94% for the DNN-TC model. They train their model using both Trashnet and VN-trash datasets.

Adedeji et al. proposed a trash classification system [9] using dataset from [10]. In their work, they had got an accuracy rate of 87 percent. Their proposed classification system the model was developed using the ResNet-50 CNN model.

Krizhevsky et al. have proposed an CNN architecture named AlexNet [11] won ImageNet Large-Scale Visual Recognition

DOI: https://doi.org/10.3329/gubjse.v8i1.62327

This paper was received on 3 June 2022, revised on 21 June 2022, and accepted on 27 August 2022.

Md. Atikuzzaman is with the Department of Computer Science and Engineering, Green University of Bangladesh, Dhaka, Bangladesh. E-mail: atik@cse.green.edu.bd

Md. Parvez Hossain is with the Department of Computer Science and Engineering, Green University of Bangladesh, Dhaka, Bangladesh. E-mail: parvez@cse.green.edu.bd

Md. Zahidul Islam is with the Department of Information and Communication Technology, Islamic University, Kustia, Bangladesh. E-mail: zahidimage@gmail.com

Syed Ahsanul Kabir is with the Department of Computer Science and Engineering, Green University of Bangladesh, Dhaka, Bangladesh. E-mail: kabir@cse.green.edu.bd

Challenge(ILSVRC),2012. Their proposed CNN architecture is relatively simple. It is quite popular because of its well-performing nature. After winning the ImageNet challenge, became the state of the art in the domain of image classification.

Rahaman et al. have presented a rule-based Bangla sign language classification system [12].

Donovan, J. presented the ”Auto-Trash” project [13] in the TechCrunch disrupt hackathon, 2016. In their work, auto sorting trashcan can differentiate between recycling and compost waste by a camera and Raspberry Pi. They developed their project by using the TensorFlow framework owned by Google.

Mittal, Gaurav, et al. developed a smartphone app named ”Spotgarbage” which can detect trash/garbage by using deep learning [14]. In their project, allow people to identify and report trash/waste in their Surroundings.

III. Dataset and Data Collection

For the analysis, we have used a dataset of trash images which was made by Mindy Yang and Gary Thung [10]. This is a small trash image dataset having 1989 trash images. It consists of 4 different classes of trash images. These are glass, plastic, paper, metal, cardboard, and paper. We have resized all the trash images into 512x384. Some trash images same have shown in figure 1.

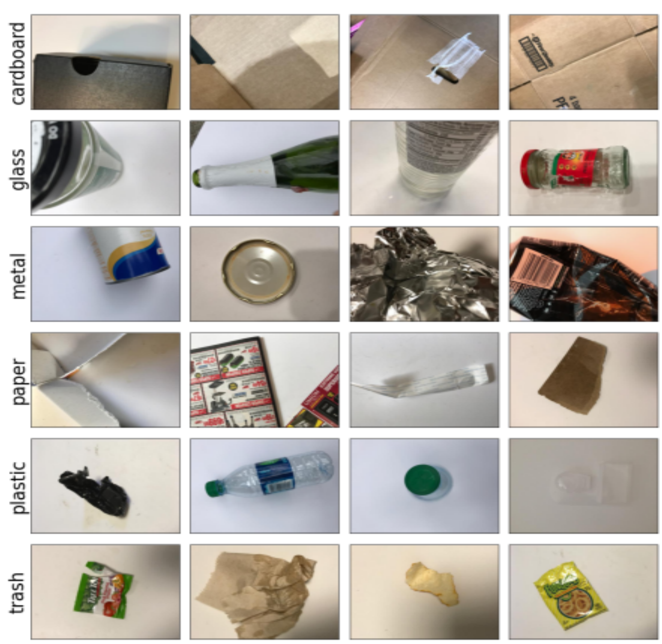

Fig. 1: Sample images of different classes from Trash dataset

Data augmentation methods have been applied to trash images for the small size in the pre-processing stage. The reason for selecting this process is the different orientations of the trash images. These methods include random shearing of the images, random scaling of the images, random rotation of the images, random translation of the images, and random brightness control of the images. These methods were applied to maximize the size of the dataset.

IV. Methodology

Our experiment has been done based on several ResNet CNN models. These are: ResNet34, ResNet50, ResNet101 and ResNet152 [4]. Also use VGG11, VGG16 and VGG19 [3] CNN models. We carry out research and investigate the accuracy rate among these where the training dataset, test dataset, number of epochs, and batch size are the same for all the models. ResNe34 and VGG16 architecture shown in figure 2.

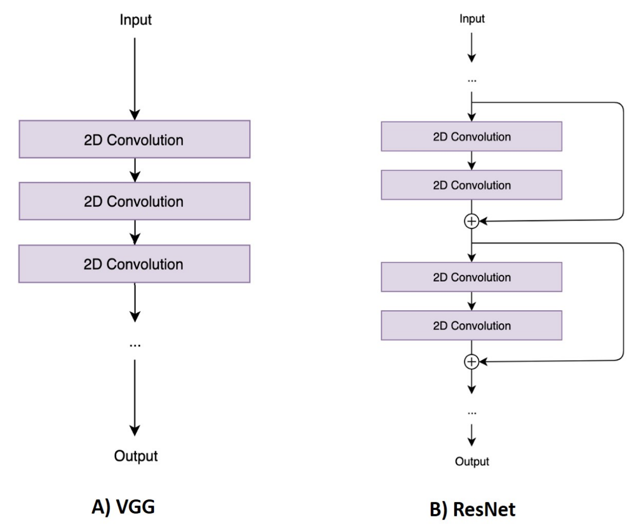

Fig. 2: Layer Architecture of VGG and ResNet

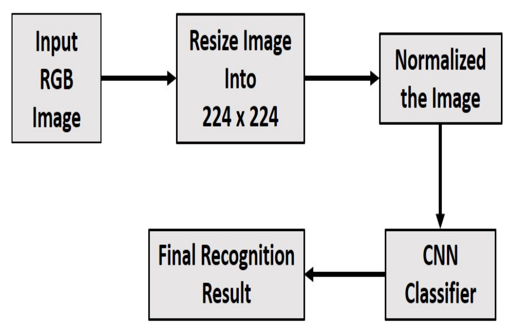

Fig. 3: System Architecture

A. System Architecture

The whole classification system is done in five steps. These are Input image, Image resizes, Normalization, CNN classification, and Final recognition which is shown in figure 3.

1) Input Image: The first step of this system is the input image. It takes an RGB image as its input.

2) Image resize: The second step is image resizing. After taking the RGB image of 512 x384 as input it resizes the RGB image into size 224 x 224.

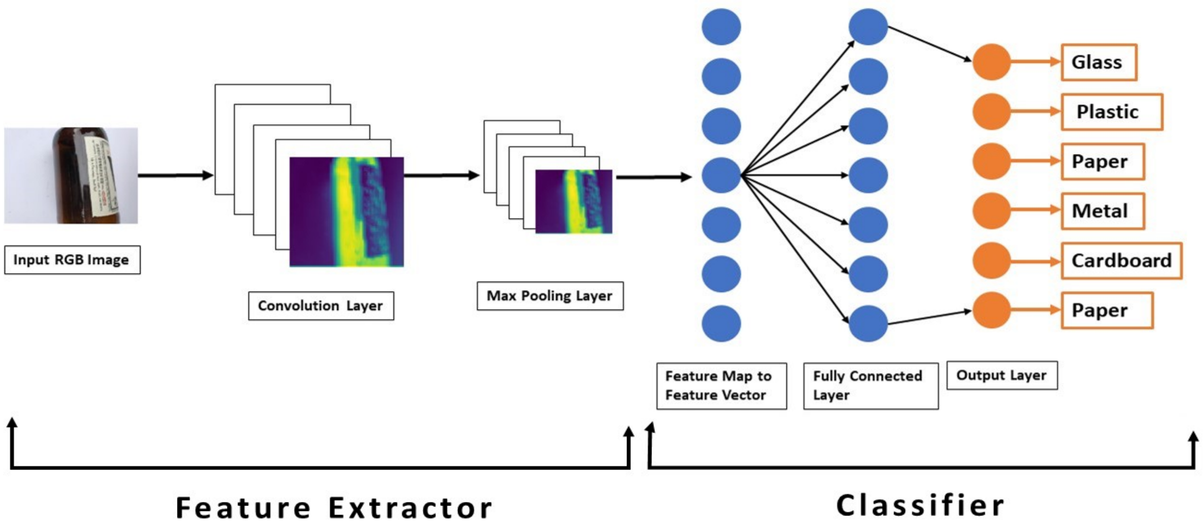

Fig. 4: General CNN classifier with different layers

3) Normalization: For all VGG and ResNet models, we used a mini-batch size 16 and the activation function ReLU.

4) CNN classification: Although the VGG and ResNet are similar in many ways there is a difference in layer architecture. In VGG there is no shortcut connection from one layer to another but in ResNet, there is a shortcut connection. All the models we used in this analysis then went through 6 fully connected layers and a softmax layer. A General CNN classifier with different layers is shown in figure 4. A details architecture of ResNet-50 is shown in Table I which is used to train the model and other models are almost similar to this model architecture.

TABLE I: Details ResNet-50 architecture used in our experiments.

Layer	Output size		Kernel	Repeat
Stage 1 (Freeze)	112 x 112	Convolution	7 x 7, 64, stride 2 3 x 3, MP, stride 2	1
Stage 2 (Freeze)	56 x 56	Convolution block	1 x 1, 64 3 x 3, 64 1 x 1, 256	3
Stage 3 (Freeze)	28 x 28	Convolution block	1 x 1, 128 3 x 3, 128 1 x 1, 512	4
Stage 4	14 x 14	Convolution block	1 x 1, 256 3 x 3, 256 1 x 1, 1024	6
Stage 5	7 x 7	Convolution block	1 x 1, 512 3 x 3, 512 1 x 1, 2048	3
	1 x 1	Global average pooling		
Average pooling	7 x 7	Average pooling Flatten Dense (512) + ReLU Dropout (0.5) Dense(#classes)		
Output Layer blackheight		Softmax		

5) Final recognition: Final recognition for all the models is determined by the Classification Output (cross-entropy)layer.

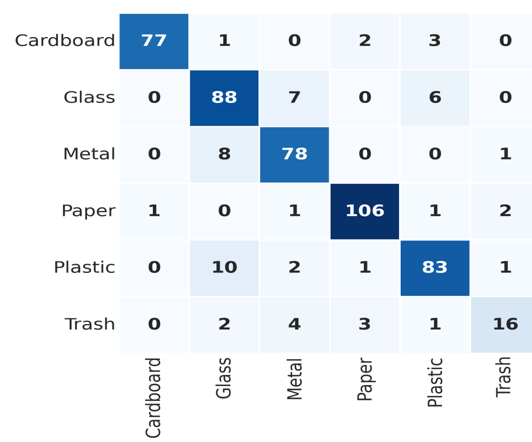

Fig. 5: Confusion Matrix for ResNet34

V. Result Analysis

A. *Result Analysis for ResNet*

We have used six classes of waste for this work. These are cardboard, glass, metal, paper, plastic, and trash. We have shown the confusion matrix for resnet34 model in figure 5, resnet50, resnet101 and resnet152 models respectively in figure 6, figure 7 and figure 8. For all ResNet CNN models, we have calculated the recall, precision, and overall accuracy

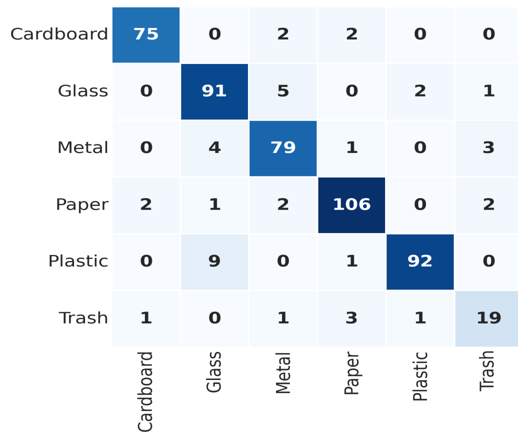

Fig. 6: Confusion Matrix for ResNet50

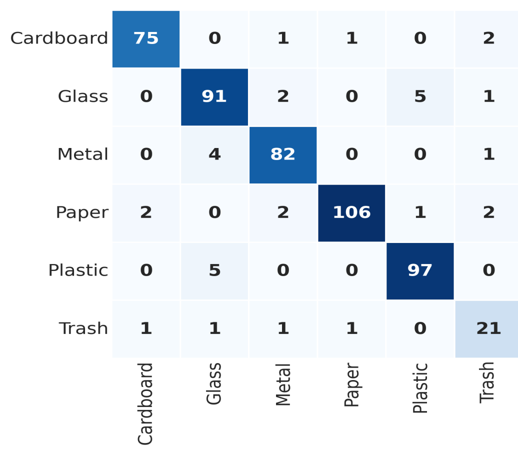

Fig. 7: Confusion Matrix for ResNet101

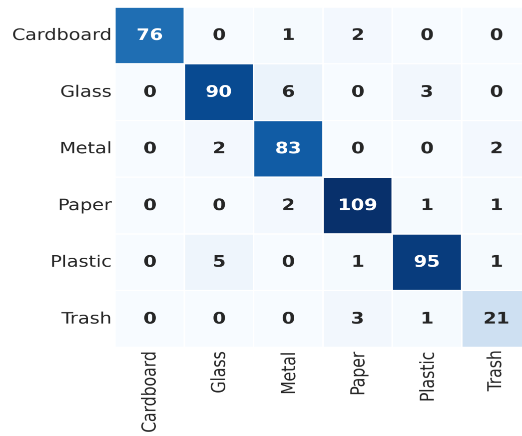

Fig. 8: Confusion Matrix for ResNet152

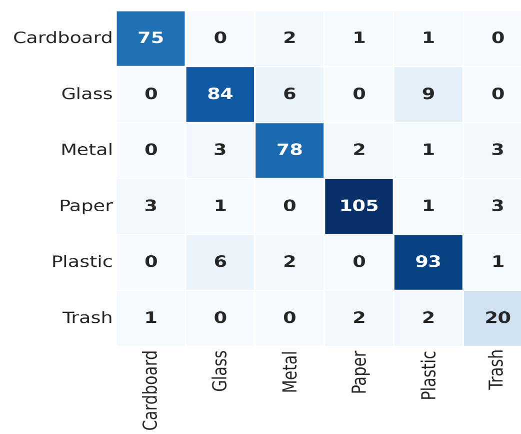

Fig. 9: Confusion Matrix for VGG11

for 20 epochs. We have shown these values for ResNet34 in Table II and for resnet50, resnet101 and resnet152 models respectively in Table III, Table IV and Table V. We have got overall accuracy of 88.713% in resnet34. Overall accuracy for all other ResNet models are shown in Table IX.

B. Result Analysis for VGG

We have shown the confusion matrix of VGG11 model in the figure 9 and for VGG16 and VGG19 respectively shown in figure 10 and figure 11. For all VGG CNN models, we have calculated the recall, precision and overall accuracy for 20 epoch. We have shown these values for VGG11 in Table VI, VGG16 in Table VII and VGG19 Table VIII. We have got overall accuracy of 90.099% in VGG11. Overall accuracy for all other VGG models are shown in Table IX.

C. Comparison

We have used six classes for this task. We have used several ResNet models and three VGG models. We have shown the comparison among all the ResNet and VGG models we have used in figure 12 and table IX.

TABLE II: Recall and Precision for ResNet34

Class	Recall	Precision
Cardboard	98.718%	92.771%
Glass	80.734%	87.129%
Metal	84.783%	89.655%
Paper	94.643%	95.495%
Plastic	88.298%	85.567%
Trash	80%	51.538%

TABLE III: Recall and Precision for ResNet50

Class	Recall	Precision
Cardboard	98.718%	92.771%
Glass	80.734%	87.129%
Metal	84.783%	89.655%
Paper	94.643%	95.495%
Plastic	88.298%	85.567%
Trash	80%	51.538%

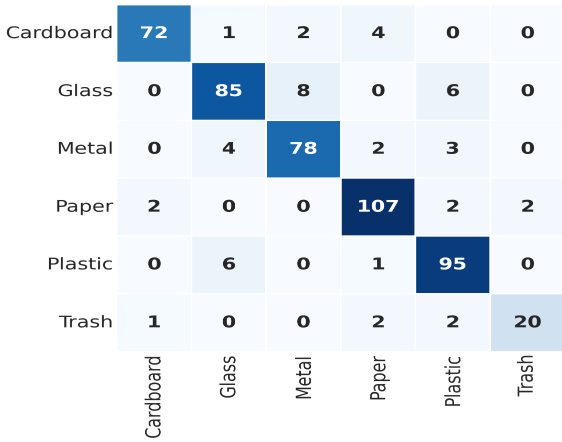

Fig. 10: Confusion Matrix for VGG16

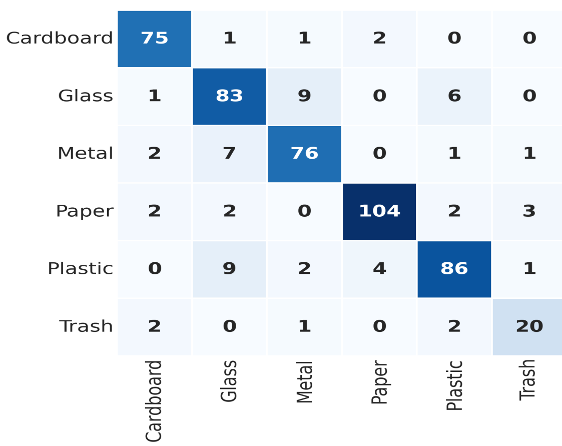

Fig. 11: Confusion Matrix for VGG19

1) ResNet: We have used four different ResNet models in our work. We have got different accuracy in different ResNet models. We have shown a comparison among all the ResNet models we have used in our work in figure 13.

2) VGG: We have used four different VGG models in our work. We have got different accuracy in different VGG models. We have shown a comparison among all the VGG models we have used in our work in figure 14.

TABLE IV: Recall and Precision for ResNet101

Class	Recall	Precision
Cardboard	98.718%	92.771%
Glass	80.734%	87.129%
Metal	84.783%	89.655%
Paper	94.643%	95.495%
Plastic	88.298%	85.567%
Trash	80%	51.538%

TABLE V: Recall and Precision for ResNet152

Class	Recall	Precision
Cardboard	98.718%	92.771%
Glass	80.734%	87.129%
Metal	84.783%	89.655%
Paper	94.643%	95.495%
Plastic	88.298%	85.567%
Trash	80%	51.538%

TABLE VI: Recall and Precision for VGG11

Class	Recall	Precision
Cardboard	94.937%	94.937%
Glass	89.362%	84.848%
Metal	88.636%	89.655%
Paper	95.455%	92.92%
Plastic	86.916%	91.176%
Trash	74.074%	80%

VI. Discussion

In our work, we have examined several CNN models. For ResNet model we have used ResNet34, ResNet50, ResNet101 and ResNet152. Among all the ResNet models we have got the highest accuracy on the ResNet152 model which was 93.861%. For VGG models we have used VGG11, VGG16, and VGG19. Among all the VGG models we have got the highest accuracy on the VGG16 model which was 90.495%.

As we all know, if the layer number is increased accuracy is also increased. But we have noticed that VGG19 gave less accuracy than VGG16 which is an exceptional case.

A. Limitations

This system work for only individual image of trash. It can't detect and classify multiple trash from a single video or image.

B. Future Work

As the dataset we used was comparatively small, in the future we will add more trash images to increase the size of

TABLE VII: Recall and Precision for VGG16

Class	Recall	Precision
Cardboard	96%	91.139%
Glass	88.542%	85.859%
Metal	88.636%	89.655%
Paper	92.241%	96.396%
Plastic	87.963%	93.137%
Trash	100%	80%

TABLE VIII: Recall and Precision for VGG19

Class	Recall	Precision
Cardboard	91.463%	94.937%
Glass	81.373%	83.838%
Metal	85.939%	87.356%
Paper	94.545%	92.035%
Plastic	88.66%	84.314%
Trash	80%	80%

TABLE IX: Accuracy comparison among several train models

SL.	Models	Train Accuracy	Test Accuracy
1	ResNet34	94.63%	88.71%
2	ResNet50	95.02%	91.49%
3	ResNet101	96.83%	93.47%
4	ResNet152	97.12%	93.86%
5	VGG11	95.67%	90.09%
6	VGG16	96.46%	90.49%
7	VGG19	94.89%	87.92%

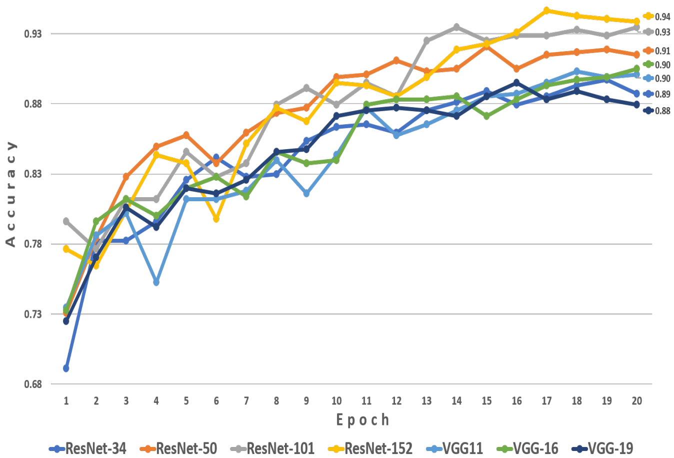

Fig. 12: Comparison among all ResNet and VGG models

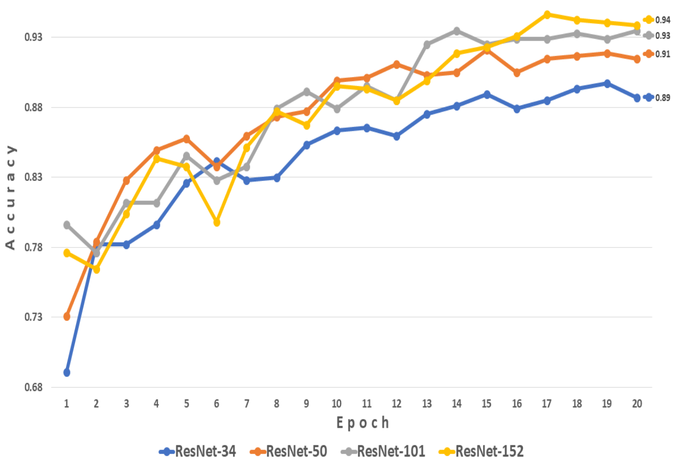

Fig. 13: Comparison among all ResNet models

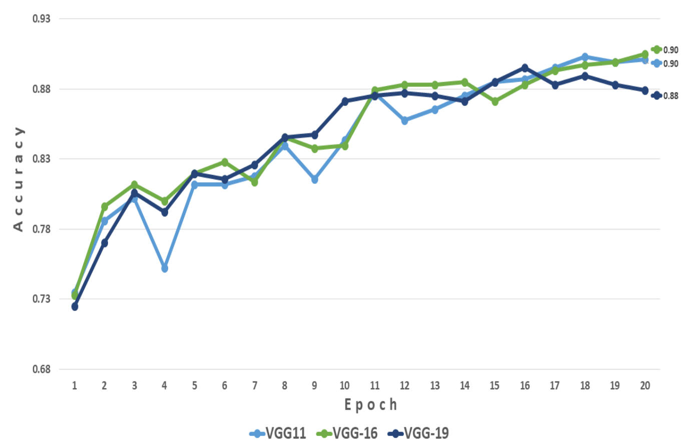

Fig. 14: Comparison among all VGG models

the dataset. We will also try to classify and detect multiple trash from an individual video or image.

VII. Conclusion

At the human level, CNNs have reached the result of image classification. The number of loaded layers will raise feature levels as more layered CNN networks extract low, medium, and high-level features and classifiers in an end-to-end multi-layer manner. Look at the ResNet results to see how important the loaded layer is. When the deeper network began to converge, a degradation issue emerged: as the number of network layers increased, performance got saturated (which was unexpected) and then gradually degraded. Overfitting or adding extra layers to a deep network that results in reduced precision does not cause this deterioration. Deeper VGG models for this dataset are not optimizable, according to a study of training accuracy.

Acknowledgment

This work was supported in part by the Center for Research, Innovation, and Transformation (CRIT) of Green University of Bangladesh (GUB).

References

[1] D. Hoornweg and P. Bhada-Tata, "What a waste: a global review of solid waste management. washington (dc): World bank; 2012."

[2] N. Johansson and H. Corvellec, "Waste policies gone soft: An analysis of european and swedish waste prevention plans," *Waste management*, vol. 77, pp. 322–332, 2018.

[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[5] M. Atikuzzaman, M. Asaduzzaman, and M. Z. Islam, "Vehicle number plate detection and categorization using cnns," in *2019 International Conference on Sustainable Technologies for Industry 4.0 (STI)*. IEEE, 2019, pp. 1–5.

[6] M. Atikuzzaman, T. R. Rahman, E. Wazed, M. P. Hossain, and M. Z. Islam, "Human activity recognition system from different poses with cnn," in *2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI)*. IEEE, 2020, pp. 1–5.

[7] M. Pervej, S. Das, M. P. Hossain, M. Atikuzzaman, M. Mahin, and M. A. Rahaman, "Real-time computer vision-based bangla vehicle license plate recognition using contour analysis and prediction algorithm," *International Journal of Image and Graphics*, p. 2150042, 2021.

[8] A. H. Vo, M. T. Vo, T. Le *et al.*, "A novel framework for trash classification using deep transfer learning," *IEEE Access*, vol. 7, pp. 178 631–178 639, 2019.

[9] O. Adedeji and Z. Wang, "Intelligent waste classification system using deep learning convolutional neural network," *Procedia Manufacturing*, vol. 35, pp. 607–612, 2019.

[10] M. Yang and G. Thung, "Classification of trash for recyclability status," *CS229 Project Report*, vol. 2016, 2016.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[12] M. A. Rahaman, M. P. Hossain, M. M. Rana, M. A. Rahman, and T. Akter, "A rule based system for bangla voice and text to bangla sign language interpretation," in *2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI)*. IEEE, 2020, pp. 1–6.

[13] J. Donovan, "Auto-trash sorts garbage automatically at the techcrunch disrupt hackathon," 2016.

[14] G. Mittal, K. B. Yagnik, M. Garg, and N. C. Krishnan, "Spotgarbage: smartphone app to detect garbage using deep learning," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2016, pp. 940–945.

Md. Atikuzzaman was born in Madaripur, Bangladesh, in 1995. He received the B.Sc. in Computer Science and Engineering degree from Green University of Bangladesh (GUB) in 2019. At present he is working as a lecturer of department of Computer Science and Engineering in Green University of Bangladesh (GUB) from September 2019. His research interest includes image Processing, Computer Vision, Data Science and Machine Learning.

Md. Parvez Hossain was born in Cumilla, Bangladesh, in 1994. He received the B.Sc. in Computer Science and Engineering degree from Green University of Bangladesh (GUB) in 2019. At present he is working as a lecturer of department of Computer Science and Engineering in Green University of Bangladesh (GUB) from September 2019. His research interest includes image Processing, BlockChain, Computer Vision, Bioinformatics and Machine Learning.

Dr. Zahidul Islam was born in Bangladesh. He received the Ph.D. in Computer Engineering from Chonnam National University, Gwangju, South Korea. He received his B.Sc. and M.Sc. in Applied Physics Electronic Engineering from University of Rajshahi. At present he is working as a Professor at Information and Communication Technology, Islamic University. His research interest includes Computer vision, 3D object, human and motion tracking, Articulated body tracking, Genetic algorithm etc.

Sayed Ahsanul Kabir was born in Bangladesh. He received the M.Sc. in Strategic Management IT from Central College, UK and M.Sc. in Computer Science from University of Dhaka. He received the B.Sc. in Applied Physics Electronics from University of Dhaka. At present he is working as an Associate Professor, Department of Computer Science and Engineering in Green University of Bangladesh. His research interest includes image Processing, Computer Vision, and Machine Learning.