06 August 2023
AIGC-5502: Image Processing & Computer Vision
Humber College Faculty of Applied Sciences & Technology

Automatic Waste Classification Using Computer Vision

N01601584: Terry Lay

N01613230: Ruoyu Zhu

Abstract

This research paper presents an in-depth review of the state-of-the-art computer vision techniques applied to classify various types of waste. This includes a discussion of related works, methodologies, experimental results, and potential future directions to enhance waste management strategies. A successful model will demonstrate the possibility of leveraging artificial intelligence and computer vision to develop a fully automated waste segregation system. The proposed solution uses transfer learning to fine-tune VGGNet-16 and achieved a testing accuracy of 95.31% over 12 categories of waste.

Introduction

Waste management is essential for protecting the environment and promoting sustainability. It refers to various processes and actions needed to properly dispose of waste. Unfortunately, ineffective recycling and composting are some of the biggest challenges faced in this industry. People often fail to realize that not all waste can be disposed of in the same manner, and it can be difficult to determine what materials need to be separated before recycling. Fig. 1 below shows some examples of different types of waste.

Figure 1: battery, biological, cardboard, and plastic waste

According to the Government of Canada, *Plastic waste and pollution reduction* (2021), Canadians dispose of over 3 million tonnes of plastic waste annually and only 9% of it is properly recycled. The rest is thrown into landfills, waste-to-energy facilities, or the environment. To avert potential health and environmental hazards, it is important to segregate waste appropriately. This improves the quality of recycling and minimizes the amount of material that ends up in landfills.

Computer vision is a subfield of artificial intelligence. It enables machines to interpret and understand visual data by replicating a human's ability to perceive, interpret and make sense of videos and images. Advancements in deep learning, particularly Convolutional Neural Networks (CNNs), have helped achieve remarkable progress in tackling problems related to image classification and object recognition.

There are many existing models that have been trained on large datasets to help computers learn general patterns and features present in data. These can be used in transfer learning setups to boost the performance on specific tasks by reducing training times and improving generalization. With the help of various techniques such as image preprocessing and data augmentation, this paper will shed light on the use of transfer learning to fine-tune VGGNet-16, ResNet-50, and MobileNetV2 as potential solutions to overcome the waste segregation problem.

Related Work

Automatic waste classification is widely researched in the field of computer vision. Many of the past contributions propose similar approaches in their solutions. Nafiz et al.(2023) developed a fully automated waste segregation system called ConvoWaste which utilized a pre-trained Inception-Resnet V2 model with several extra layers. This model was pre-trained on 1.4 million photos and over 1,000 classes. They were able to achieve an accuracy of 98% for six categories of waste based on 14,400 images of plastic, metal, glass, medical waste, organic, and e-waste. This model was designed to classify general types of waste but could be extended to include more groups of recyclable materials.

Agarwal et al.(2020) trained a model to identify seven categories of resin codes from images of plastic waste. The Resin Identification Code (RIC) is a system introduced in 1988 in the US as an indicator of the type of material a plastic product is made of. The dataset consisted of 3200 training, 400 testing, and 400 validation images. This paper did not use transfer learning in the proposed solution. Instead, researchers extracted embeddings of the images using Siamese and Triplet loss networks. They then fit a K-Nearest Neighbour algorithm and assigned predicted categories based on the top K closest matches to the embedded test image. Their results showed a high accuracy of 99.7%, but the model suffers from high computation costs and may not be feasible in real world applications.

Wu et al.(2022) analyzed the effect of transfer learning on the performance of VGGNet-16 and ResNet-50 for the classification of organic versus residual waste. With 22,010 images collected, they fine-tuned the models and achieved accuracies of 96.6% with ResNet-50 and 95.6% with VGGNet-16. Their results were comparable with the previous studies mentioned, but it would be interesting to examine how these pre-trained models would perform beyond a binary classification problem. In this paper, the dataset is extended to include a greater variety of waste categories. The performance of VGGNet-16 and ResNet-50 will also be studied in addition to MobileNetV2 over 12 different classes.

Methodology

This section presents a comprehensive overview of the employed methodologies for waste classification. The dataset is the Household Garbage Classification dataset from Kaggle. It comprises of 12 classes: paper, cardboard, biological, metal, plastic, green-glass, brown-glass, white-glass, clothes, shoes, batteries, and trash. The number of images is broken down in table 1.

Table 1: Images per category in the dataset

Category	Number of Images
Brown glass	607
Green glass	629
Trash	697
Metal	769
White glass	775
Plastic	865
Cardboard	891
Batteries	945
Biological	985
Paper	1050
Shoes	1977
Clothes	5325
Total	**15515**

To prepare the data, splitfolders was first used to split the data into 80% training and 20% validation sets. The training set consisted of 12,417 images and the testing set consisted of 3,108 images. The image arrays were then converted to tensors and resized to be 224x224 in dimension. Normalization was applied using a mean of [0.485, 0.456, 0.406] and standard deviation of [0.229, 0.224, 0.225] for all three models.

Overfitting is a common problem in deep learning where a model performs well on training data but fails to generalize on unseen data. Various techniques were employed to help prevent the problem of overfitting. The first technique was to transform the training images when loading the data. This included randomly flipping them horizontally with a probability of 50%, vertically with a probability of 20%, and rotating them -20° to 20°. The motivation behind this was to encourage generalization by including diversity in the data through random noise. Fig. 2 below shows a batch of images after transformations.

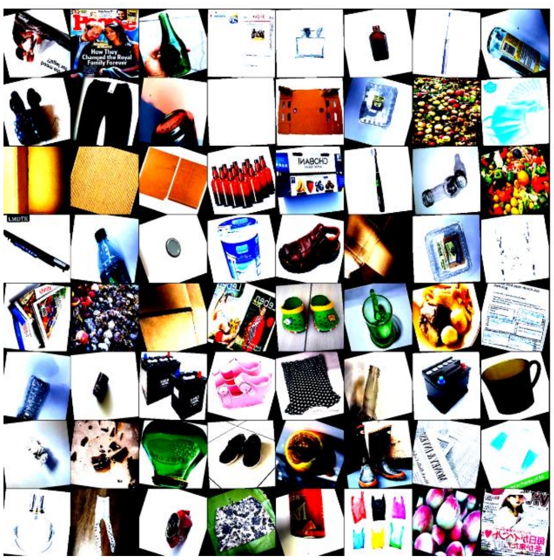


Figure 2: Batch of transformed images

Working with imbalanced data is also a concern when it comes to overfitting. Underrepresented categories in the training set leads to poor generalization of the model as it tends to show bias towards the majority classes. This dataset had some class imbalances with clothing and shoes taking up more than a third of the dataset. To minimize the effect of this, ImbalancedDatasetSampler was used when loading the data into batches.

According to the documentation in *Torchsampler*, it is an easy-to-use PyTorch sampler that can:

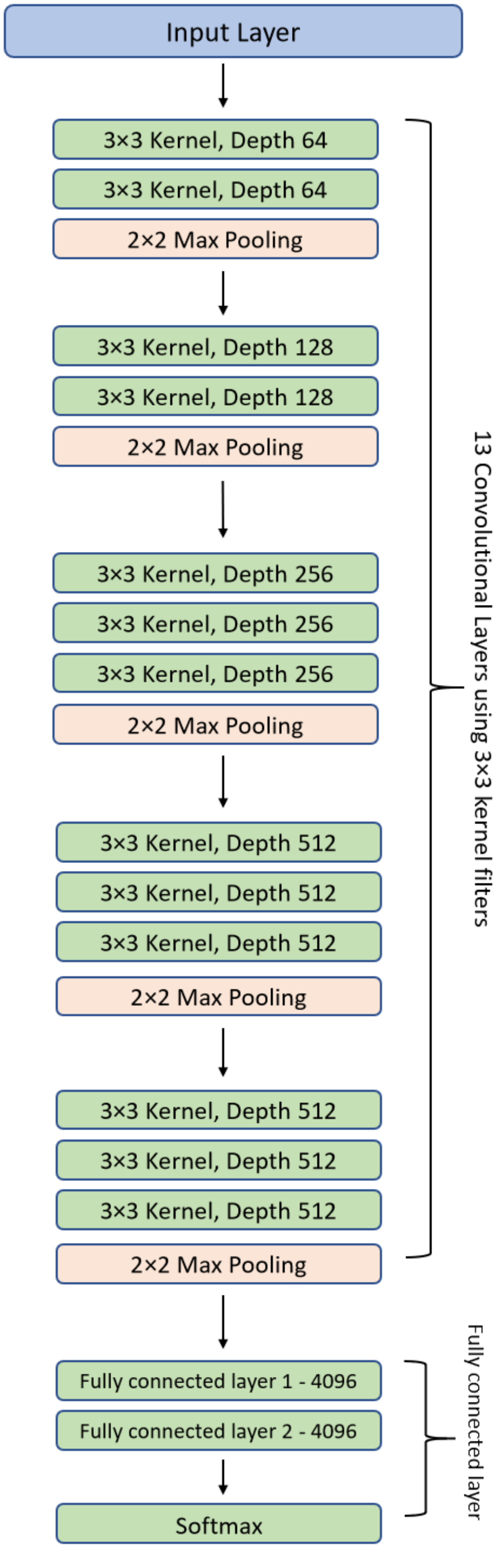
- rebalance class distributions when sampling from the imbalanced dataset
- automatically estimate sampling weights
- avoid creating a new balanced dataset
- mitigate overfitting when used in conjunction with data augmentation techniques

The sampler was passed into DataLoader. In each epoch, the loader sampled the entire dataset and weighed the samples inversely to the probability of the class appearing.

Once the data was loaded, VGGNet-16, ResNet-50, and MobileNetV2 were initialized for training. These models were pre-trained on the ImageNet dataset containing over 14 million images across 1000 object classes. To fine-tune them, the last layer for each model was removed and retrained to fit the waste dataset. According to Simonyan and Zisserman (2014), the VGGNet-16 model improves on AlexNet and replaces large filters with sequences of smaller 3x3 filters.

The architecture of VGGNet-16 is a 16-layer deep neural network with a total of 138 million parameters. There are a total of five configurations with each configuration's depth increasing with more added layers. Adjusting the final layer of this model yielded 134,260,544 non-trainable parameters and 49,164 trainable parameters. Figure 2 breaks down the layers in the VGGNet-16 architecture.

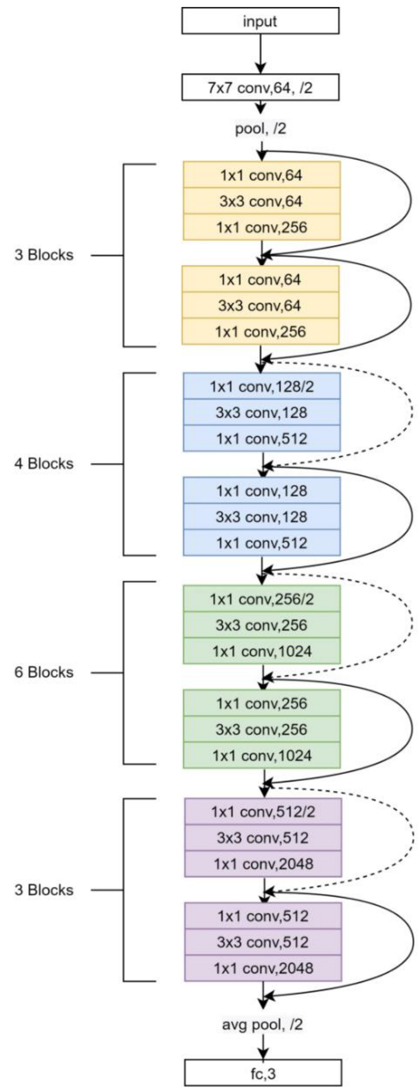
Figure 2: VGGNet-16 architecture



Note: From "Transfer learning using vgg-16 with deep convolution al neural network for classifying images" by Tammina, S., 2019, *In ternational Journal of Scientific and Research Publications (IJSRP),* 9(10), 143-150.

Introduced by He et al. (2016),ResNet-50 as a 50-layer CNN with 48 convolutional layers, one MaxPool layer and one average pool layer for a total of 24.6 million parameters. Retraining the final layer resulted in 23,508,032 non-trainable parameters and 1,115,788 trainable parameters. Figure 3 breaks down the layers in the ResNet-50 architecture.

Figure 3: ResNet-50 architecture

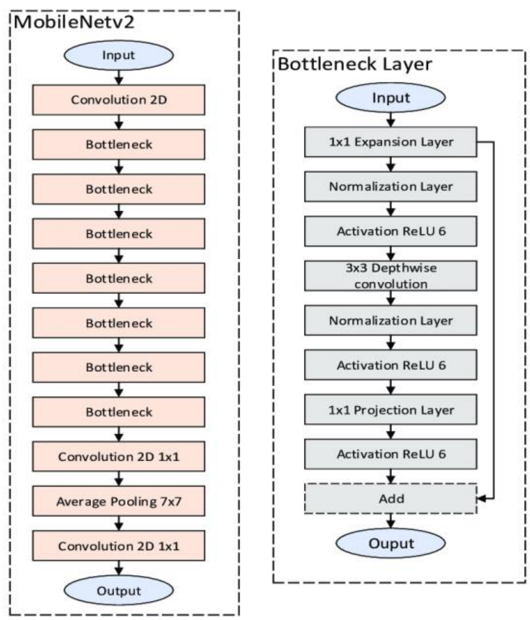


Note: From "A classification of Arab ethnicity based on face image using deep learning approach" by Al-Humaidan, N. A., & Prince, M, 2021, *IEEE Access*, 9.

MobileNetV2 is described as a simple and not computationally intensive CNN which works especially well for mobile applications.

This makes it useful in real-world settings since it is efficient and lightweight. According to Sandler et al. (2018), the model contains initial fully convolution layer with 32 filters, followed by 19 residual bottleneck layers. The total parameters in MobileNetV2 are 3.5 million. Adjusting the final layer resulted in 2,223,872 non-trainable parameters and 1,281,000 trainable parameters.

Figure 5: MobileNet architecture



Note: From "Design space exploration of a sparse mobilenetv2 using high-level synthesis and sparse matrix techniques on FPGAs" by Tragoudaras, A., 2022, *Sensors*, 22(12), 4318.

Experiments and Results

In this section, the experimental setup will be presented including details of the hyperparameters and training procedures. Subsequently, an analysis of the classification results of the three CNN models – ResNet-50, VGGNet-16, and MobileNet will be reported.

Each model was trained over five epochs with a batch size of 64. The optimizer for the model was Adam with a learning rate of 0.001 and cross entropy loss function.

A learning curve represents the relationship between a model's performance and the number of epochs. This study measured both accuracy and loss over training iterations. Fig. 6 and 7 below show the model accuracy and loss during training for the VGGNet-16 model.

Figure 6: VGGNet-16 training loss

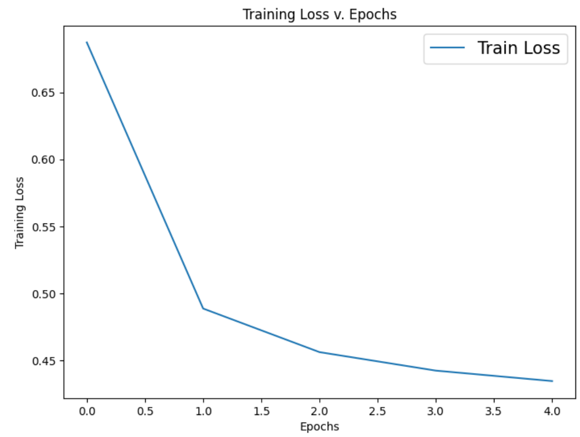
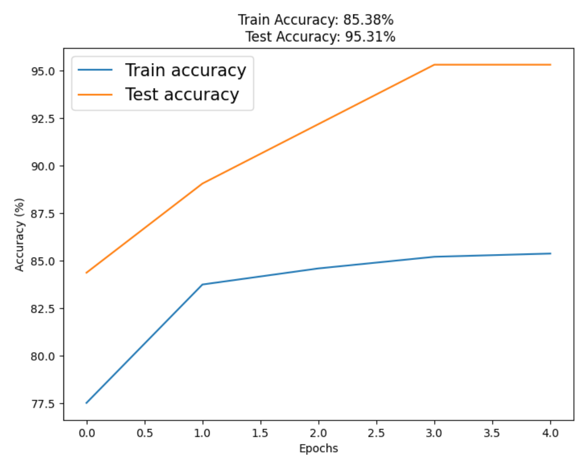


Figure 7: VGGNet-16 training and testing accuracy



As the VGGNet-16 model iteratively processed the training data, its performance steadily improved without significant fluctuations or erratic behaviour. This can be attributed to the benefits of transfer learning, which allowed the model to leverage knowledge gained from pre-training on a large-scale dataset before fine-tuning on the waste classification dataset. The smoothness of the learning curve indicates strong generalizability of VGGNet-16 which was able to learn meaningful features from the data and achieved a testing accuracy of 95.31% in the final epoch.

The second model tested was ResNet-50, which achieved a testing accuracy of 91.86% in the last epoch.

Figure 8: ResNet-50 training loss

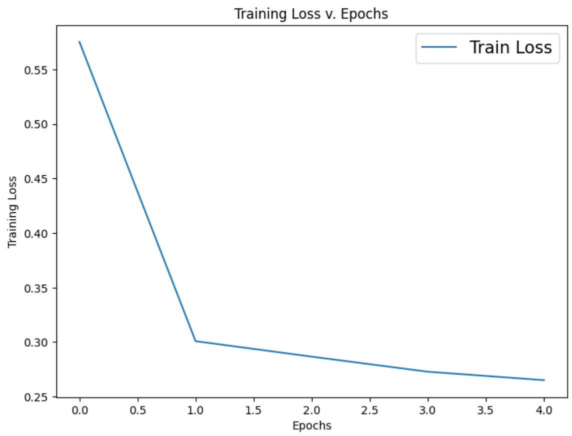
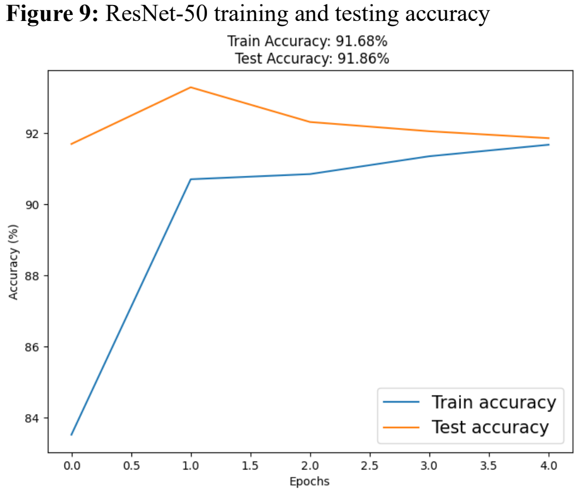


Figure 9: ResNet-50 training and testing accuracy



The loss curve shows a steady decrease which is a sign that the model was able to learn from the training set, but the fluctuation in the testing accuracy indicates some signs of possible overfitting.

The final model trained used MobileNetV2 and achieved a testing accuracy of 91.37% in the fifth epoch. MobileNetV2's performance was comparable with ResNet-50. The testing accuracy appears to fluctuate around 91% without much steady improvement. VGGNet-16 had the best performance overall in terms of learning and testing accuracy. The learning curves for the MobileNetV2 model can be found in Fig. 10 and 11

Figure 10: MobileNet training loss

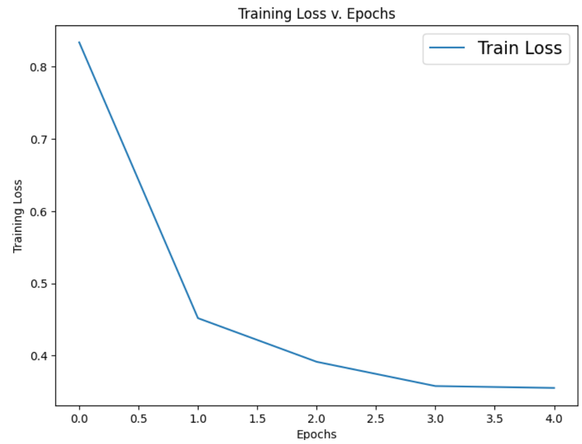
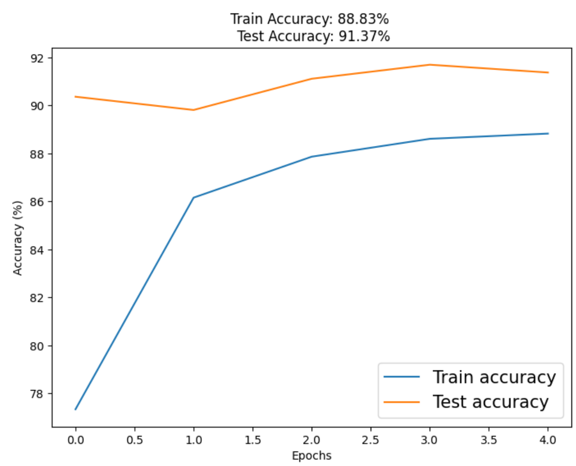


Figure 11: MobileNet training and testing accuracy



Discussion on Results

The three models showed promising results in detecting categories of waste. MobileNetV2 and ResNet-50 performed similarly and achieved accuracies between 91%-92%. These results were quite surprising considering that MobileNetV2 has significantly less parameters than ResNet-50. VGGNet-16 outperformed both models with a high accuracy of 95.31%. Increasing the number of epochs would provide a clearer understanding of each model's behaviour but based on this study, VGGNet-16 is the best model to choose for this use case.

The selected model from this study has comparable results with previous findings of other researchers.

Table 2: Comparison with previous work

Model	Categories	Accuracy
ConvoWaste	6	98%
Resin Code Identifier	5	99.7%
Organic Waste Classifier	2	96.6%
Household Waste Classifier	**12**	**95.31%**

Although ConvoWaste achieved an accuracy of 98%, this model only categorized six types of waste compared to 12 in the household waste classifier. The Resin Code Identifier had a remarkable accuracy of 99.7%, but could only label five types of plastic and used a model that was computationally expensive. The work done by researchers who built the organic waste classifier also used ResNet-50 and VGGNet-16. They achieved an accuracy of 96.6% but their model was only able to identify organic from residual waste. Their results showed ResNet-50 outperformed VGGNet-16 by 1%, but VGGNet-16 performed much better than both ResNet-50 and MobileNet for the household waste dataset with 12 categories.

In summary, results showed that transfer learning is extremely powerful in reducing training times while still providing optimal performance. VGGNet-16 seemed to perform best when there were many categories, but both MobileNetV2 and ResNet-50 showed promising results, nonetheless. This can be enhanced if trained over more epochs. Additionally, adding several layers could also prove to improve the performance of the models. A more comprehensive comparison could be done by tuning hyperparameter values such as weight decay and batch size.

Conclusion

By showcasing the efficacy of transfer learning in waste classification and offering a comparative analysis of ResNet-50, VGGNet-16, and MobileNetV2, this study can guide the selection and deployment of CNN models in real-world waste sorting applications, fostering a more sustainable approach to waste management and resource recovery. This can be accomplished by building a relationship between software and hardware to create a system that places categorized materials into their appropriate waste bins.

Future work on this study would be to expand the number of categories in the model to include even more types of recycling, e-waste and organic materials to broaden the impact of leveraging computer vision for waste management strategies.

References

[1] Canada, E. a. C. C. (2023, April 14). *Plastic waste and pollution reduction*. Canada.ca.https://www.canada.ca/en/environment-climate-change/services/managing-reducing-waste/reduce-plastic-waste.html

[2] Nafiz, M. S., Das, S. S., Morol, M. K., Al Juabir, A., & Nandi, D. (2023, January). Convowaste: An automatic waste segregation machine using deep learning. In *2023 3rd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)* (pp. 181-186). IEEE.

[3] Agarwal, S., Gudi, R., & Saxena, P. (2020). Application of computer vision techniques for segregation of plasticwaste based on resin identification code. *arXiv preprint arXiv:2011.07747*.

[4] Wu, F., & Lin, H. (2022). Effect of transfer learning on the performance of VGGNet-16 and ResNet-50 for the classification of organic and residual waste. *Frontiers in Environmental Science*, *10*, 2129.

[5] *torchsampler*. (2022, May 23).PyPI. https://pypi.org/project/torchsampler/

[6] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for Large-Scale image recognition. *Computer Vision and Pattern Recognition*. http://export.arxiv.org/pdf/1409.1556

[7] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *Computer Vision and Pattern Recognition.* https://doi.org/10.1109/cvpr.2016.90

[8] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., &

Chen, L. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. *Computer Vision and Pattern Recognition*. https://doi.org/10.1109/cvpr.2018.00474

[9] Tammina, S. (2019). Transfer learning using vgg-16 with deep convolutional neural network for classifying images. *International Journal of Scientific and Research Publications (IJSRP)*, 9(10), 143-150.

[10] Al-Humaidan, N. A., & Prince, M. (2021). A classification of Arab ethnicity based on face image using deep learning approach. *IEEE Access*, 9, 50755-50766.

[11] Tragoudaras, A., Stoikos, P., Fanaras, K., Tziouvaras, A., Floros, G., Dimitriou, G., ... & Stamoulis, G. (2022). Design space exploration of a sparse mobilenetv2 using high-level synthesis and sparse matrix techniques on FPGAs. *Sensors*, 22(12), 4318.