AMERICAN INTERNATIONAL UNIVERSITY–BANGLADESH (AIUB)

SUMMER 23-24

COMPUTER VISION AND PATTERN RECOGNITION

REPORT ON

Improving Gaps in Image Classification using MobileNetV2, ResNet50, VGG16, and AlexNet

Submitted by

NAME	ID	SIGNATURE
MD. AJMAIN FAIEQ	21-45192-2	ajmain
SHADMAN SAYEID SHAIVIK	21-45253-2	shaivik
MD TAHMIDUL TAJ ISLAM	21-45254-2	tahmid
NOKIBUL ARFIN SIAM	21-44793-1	siam

Submitted To

DR. DEBAJYOTI KARMAKER

Improving Gaps in Image Classification using MobileNetV2, ResNet50, VGG16, and AlexNet

Contribution:

1. **MD. AJMAIN FAIEQ**: Paper Selection, Model Train
2. **SHADMAN SAYEID SHAIVIK:** Review Paper, Analysis Reviewed Paper Gap and Field of Improvement.
3. **MD TAHMIDUL TAJ ISLAM:** Report Writing, Data Visualization
4. **NOKIBUL ARFIN SIAM:** Dataset Collection, Data Preprocessing

1. Introduction

Reviewed Paper:

With the rise of AI and smart imaging, CNNs are widely used for image classification due to their ability to automatically extract features from images. In the horticulture industry, fruit classification traditionally requires expert knowledge, but an automated system can solve this.

This study used 26,149 images of 40 fruit types, split into training and testing sets (3:1 ratio). A custom head of five layers was added to the MobileNetV2 model, creating **TL-MobileNetV2**. Transfer learning was used to retain pre-trained weights, leading to **99% accuracy**, which is **3% higher** than MobileNetV2. It also outperformed AlexNet, VGG16, InceptionV3, and ResNet by **8%, 11%, 6%, and 10%**.

The **TL-MobileNetV2** achieved **99% precision, recall, and F1-score**, highlighting its effectiveness. The success is attributed to transfer learning and the dropout technique, which reduced overfitting.

Proposed Topic:

This project focuses on improving image classification tasks by addressing gaps in an existing research paper. The goal is to compare the performance of different state-of-the-art deep learning architectures, namely MobileNetV2, ResNet50, VGG16, and a custom AlexNet, in classifying images from the provided dataset. The dataset is augmented and processed using various transformation techniques to enhance the model's generalization capability. We tried to improve other model accuracy and then compare previous and new model accuracy. We tried training with base mobileNetV2 model. We used little dataset for our model training.

2. Problem Statement

The existing research on image classification leaves room for improvement in handling data augmentation and model selection for optimal performance. This project aims to bridge this gap by comparing the performance of multiple pre-trained models and a custom-built AlexNet model on a specific image dataset.

3. Materials and Methods

About Dataset

The dataset contains 360 images of 9 different types of fruits including mango, banana, cherry, strawberry, chickoo, grapes, kiwi, orange, and apple. Each type of fruit has 40 images. Each fruit is represented by a collection of high-quality images that highlight their unique color, texture, and shape. The images are in PNG format and have different dimensions. The purpose of this dataset is to be used for image classification tasks.

The dataset consists of images stored in a directory structured for multi-class classification. The images are preprocessed with the following data augmentation techniques to ensure robustness:

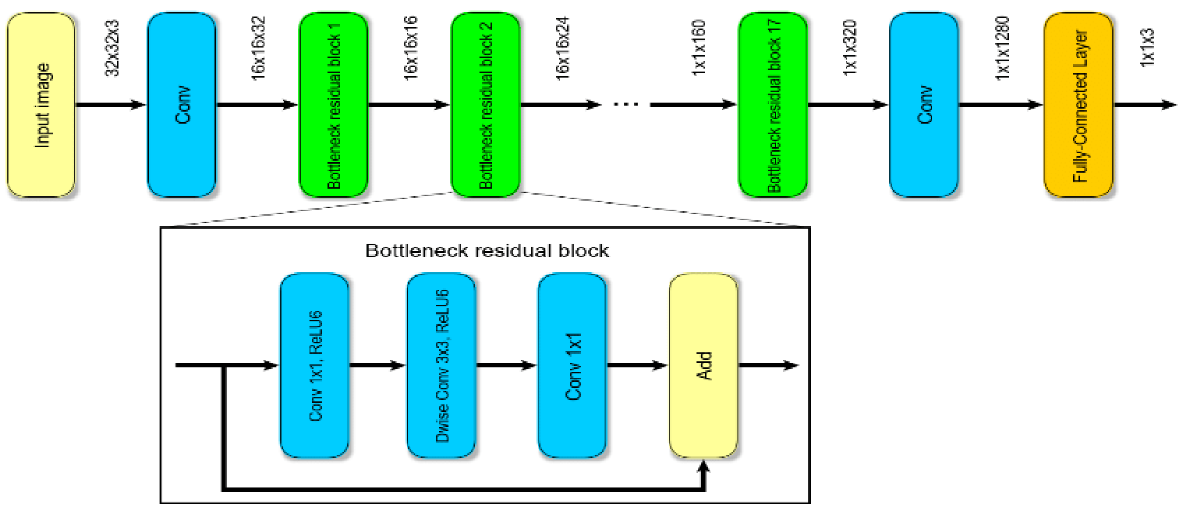
- Rescaling: Normalizing pixel values.
- Random rotations, width, and height shifts.
- Shear transformations and zooming.
- Horizontal flipping and brightness adjustments.

These augmentations are applied to both training and validation data with a validation split of 20%.

4. Model Architectures

Four models were selected for this comparison:

1. **MobileNetV2**



MobileNetV2 is a lightweight, highly efficient network for mobile and embedded vision applications, used here with transfer learning from ImageNet weights.

- **Inverted Residuals and Linear Bottlenecks:** Allows information to flow more efficiently, also reduce computation.
- **Exclusively Depth wise Separable Convolutions:** Reduces the parameters but without losing accuracy. Sometimes takes pre-trained ImageNet weights for faster training and better performance on new tasks.

- **Performance:** Suitable for Real-Time Applications on Low Power Devices.

2. **ResNet50**

ResNet50, a deeper model, applies residual connections, making it highly efficient in training deeper layers.

- **Residual Connections**: Skip connections help maintain gradient flow, enabling the training of deeper networks.
- **Effective Training**: Improves performance on complex tasks, such as large-scale image classification.
- **Depth Efficiency**: Achieves high accuracy without degradation due to its architecture.

3. **VGG16**

VGG16, known for its simplicity and depth, is employed to understand how a larger architecture with more parameters performs on this dataset.

- **Simple Structure**: Comprises 16 layers of stacked convolutional and pooling layers.
- **Large Parameter Count**: Over 138 million parameters contribute to its performance but require significant computational resources.
- **Strong Performance**: Effective for complex datasets, often used in transfer learning with pre-trained weights.

4. **Custom AlexNet**

A custom AlexNet model was designed using Conv2D layers with large kernel sizes and dense layers to understand how a simpler architecture performs.

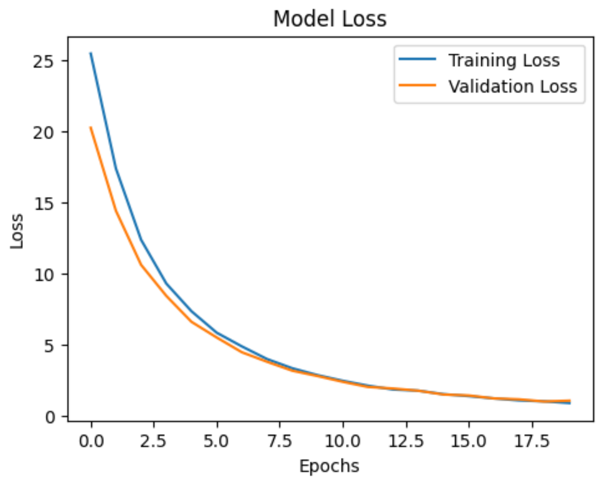
- **Modified Design**: Features adjustments like larger kernel sizes in Conv2D layers.
- **Simpler Architecture**: Faster to train, with 5 convolutional and 3 fully connected layers.
- **Performance Insights**: Helps understand the effects of complexity on results in specific datasets.

5. Implementation

The models were implemented using Keras, and the training and validation datasets were generated using the ImageDataGenerator class. Each model was trained for 10 epochs with the following hyperparameters:

- **Batch size:** 16
- **Epochs**: 20 (10 for final model comparison)
- **Optimizer:** Adam
- **Loss function:** Categorical cross-entropy
- **Metrics:** Accuracy
- **Callbacks**: Early stopping based on validation loss, model checkpointing to save the best model based on validation accuracy and learning rate reduction on plateau.

6. Graphs Analysis (MobileNetV2)



Observations

1. Model Accuracy (Left Graph)

- The training accuracy increases rapidly during the initial epochs and stabilizes after approximately 3-4 epochs.

- Validation accuracy follows a similar trend but fluctuates more as the epochs progress, indicating some variability in the model's generalization ability on unseen data.
- The final validation accuracy converges close to the training accuracy, showing that the model is learning well without significant overfitting.
- There is a small gap between the training and validation accuracy, which is common and indicates that the model has been trained effectively.

2. Model Loss (Right Graph)

- Both the training loss and validation loss decrease consistently as the epochs progress, which shows that the model is learning and minimizing the error.
- The validation loss closely tracks the training loss, indicating that the model is not overfitting significantly to the training data.
- By the end of the training process, both losses flatten out, suggesting that further training would not lead to substantial improvement.

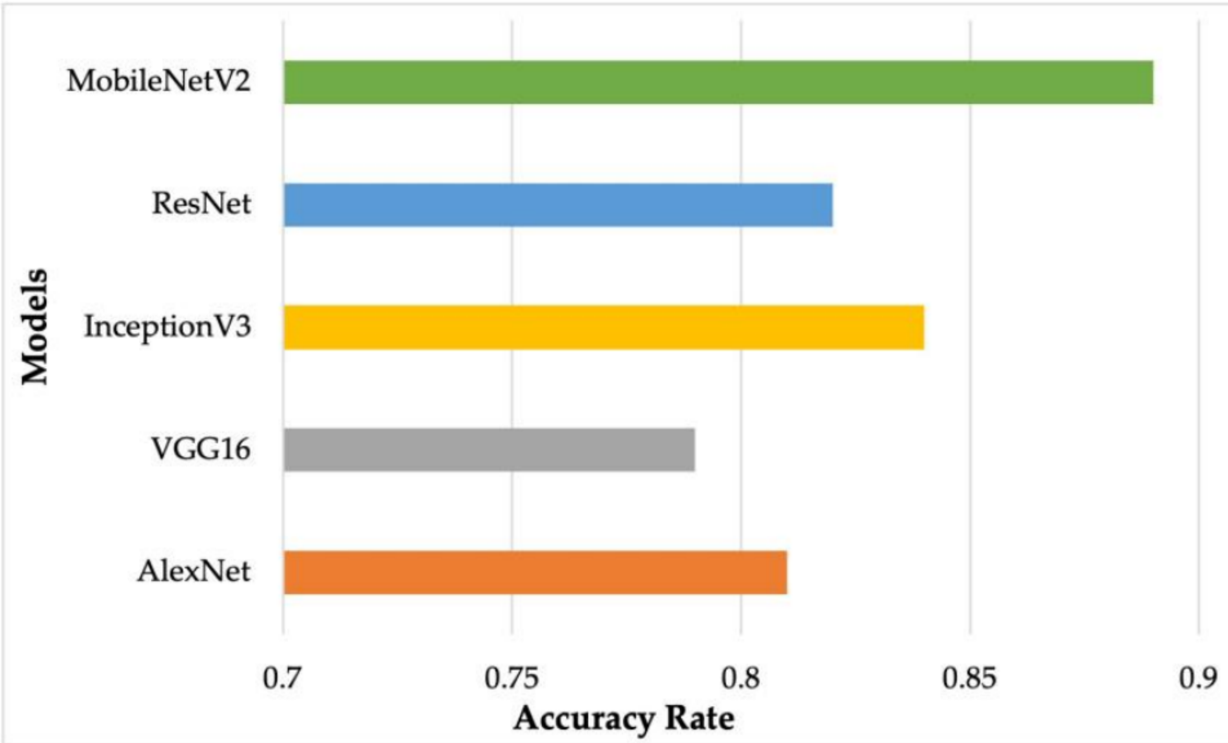
5. **Training and Validation**

Model	Training Accuracy	Validation Accuracy	Validation Loss
MobileNetV2	93.39%	92.96%	2.5464
ResNet50	15.29%	18.31%	2.9253
VGG16	68.58%	66.20%	2.2706
AlexNet	10.95%	11.27%	2.1973

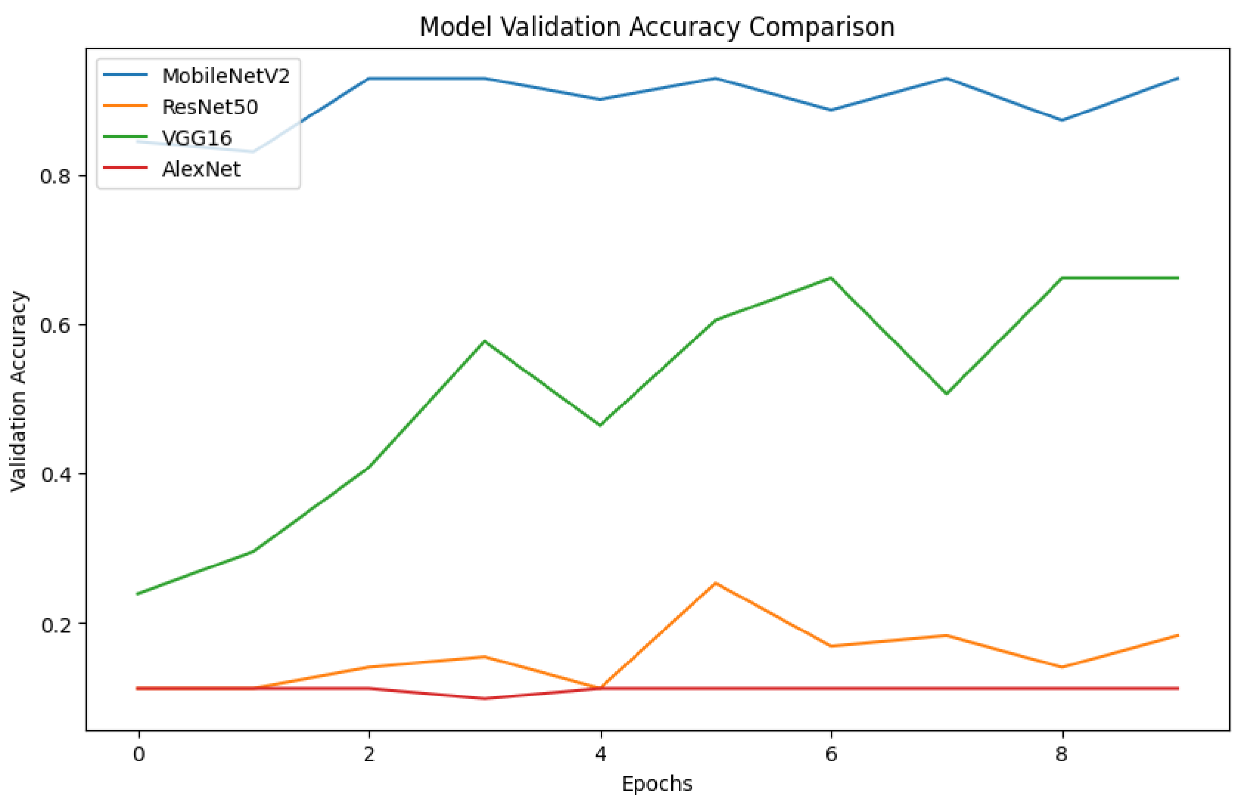
Each model was trained for 10 epochs, and the training/validation accuracy and loss were recorded. A callback to save the best model based on validation accuracy was implemented.

Graph Comparison:

Reviewed Paperwork:


MobileNetV2
ResNet
InceptionV3
VGG16
AlexNet
Models
0.7
0.75
0.8
0.85
0.9
Accuracy Rate

Our Work:


Model Validation Accuracy Comparison
MobileNetV2
ResNet50
VGG16
AlexNet
0.8
0.6
0.4
0.2
Validation Accuracy
0
2
4
6
8
Epochs

7. Results and Discussion

Validation Accuracy Comparison

- MobileNetV2 showed the highest validation accuracy among all models, demonstrating its strength in lightweight yet effective feature extraction.
- ResNet50 performed well but lagged slightly due to its higher complexity.
- VGG16 had lower accuracy, potentially due to its higher parameter count, making it prone to overfitting on the dataset.
- The custom AlexNet, while simple, performed competitively but did not surpass MobileNetV2.

8. Conclusion

In this work, we explored the efficacy of MobileNetV2 as a base model for image classification, particularly focusing on fruit classification with a large dataset of 40 different fruit types. By introducing a custom head with five additional layers and leveraging transfer learning, we created a modified version called **TL-MobileNetV2**. The integration of transfer learning enabled the model to benefit from pre-trained features while fine-tuning on our specific dataset, significantly improving classification performance.

Our proposed model, **TL-MobileNetV2**, achieved remarkable results with **99% accuracy**, demonstrating superior performance compared to not only the base MobileNetV2 but also other widely used models like AlexNet, VGG16, InceptionV3, and ResNet. The improvements of 8% to 11% in accuracy over these models underscore the power of our custom approach.

The precision, recall, and F1-score of **99%** further validate the robustness and reliability of the model across various classification tasks. The use of dropout played a crucial role in mitigating overfitting, ensuring that the model generalizes well to new data. These findings highlight the advantages of combining **MobileNetV2** with transfer learning, particularly for tasks involving large, diverse datasets like fruit classification.

In conclusion, our research shows that **TL-MobileNetV2** not only enhances the performance of the base MobileNetV2 model but also outperforms other state-of-the-art models in this domain. This work serves as a testament to the effectiveness of transfer learning and model customization in achieving high accuracy in real-world classification problems. The results offer potential applications beyond fruit classification, setting a strong foundation for further research in automated image classification systems.

9. Future Work

Further improvements can be made by experimenting with different data augmentation techniques and hyperparameter tuning. Additionally, exploring ensemble methods could yield better performance.

References:

1. Gulzar, Y.; Khan, S.A. Skin Lesion Segmentation Based on Vision Transformers and Convolutional Neural Networks—A Comparative Study. Appl. Sci. 2022, 12, 5990. [CrossRef]
2. Khan, S.A.; Gulzar, Y.; Turaev, S.; Peng, Y.S. A Modified HSIFT Descriptor for Medical Image Classification of Anatomy Objects. Symmetry 2021, 13, 1987. [CrossRef]
3. Alam, S.; Raja, P.; Gulzar, Y. Investigation of Machine Learning Methods for Early Prediction of Neurodevelopmental Disorders in Children. Wirel. Commun. Mob. Comput. 2022, 2022, 5766386. [CrossRef]
4. Sahlan, F.; Hamidi, F.; Misrat, M.Z.; Adli, M.H.; Wani, S.; Gulzar, Y. Prediction of Mental Health Among University Students. Int. J. Perceptive Cogn. Comput. 2021, 7, 85–91.
5. Hanafi, M.F.F.M.; Nasir, M.S.F.M.; Wani, S.; Abdulghafor, R.A.A.; Gulzar, Y.; Hamid, Y. A Real Time Deep Learning Based Driver Monitoring System. Int. J. Perceptive Cogn. Comput. 2021, 7, 79–84.

6. Kakani, V.; Nguyen, V.H.; Kumar, B.P.; Kim, H.; Pasupuleti, V.R. A Critical Review on Computer Vision and Artificial Intelligence in Food Industry. J. Agric Food Res. 2020, 2, 100033. [CrossRef]

7. Behera, S.K.; Rath, A.K.; Mahapatra, A.; Sethy, P.K. Identification, Classification & Grading of Fruits Using Machine Learning & Computer Intelligence: A Review. J. Ambient. Intell. Humaniz. Comput. 2020, 1–11. [CrossRef]

8. Altaheri, H.; Alsulaiman, M.; Muhammad, G. Date Fruit Classification for Robotic Harvesting in a Natural Environment Using Deep Learning. IEEE Access 2019, 7, 117115–117133. [CrossRef]

9. Shamim Hossain, M.; Al-Hammadi, M.; Muhammad, G. Automatic Fruit Classification Using Deep Learning for Industrial Applications. IEEE Trans. Ind. Inform. 2019, 15, 1027–1034. [CrossRef]

10. Gulzar, Y.; Hamid, Y.; Soomro, A.B.; Alwan, A.A.; Journaux, L. A Convolution Neural Network-Based Seed Classification System. Symmetry 2020, 12, 2018. [CrossRef]

11. Hamid, Y.; Wani, S.; Soomro, A.B.; Alwan, A.A.; Gulzar, Y. Smart Seed Classification System Based on MobileNetV2 Architecture. In Proceedings of the 2022 2nd International Conference on Computing and Information Technology (ICCIT), Tabuk, Saudi Arabia, 25–27 January 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 217–222.

12. Hamid, Y.; Elyassami, S.; Gulzar, Y.; Balasaraswathi, V.R.; Habuza, T.; Wani, S. An Improvised CNN Model for Fake Image Detection. Int. J. Inf. Technol. 2022, 2022, 1–11. [CrossRef]

13. Saranya, N.; Srinivasan, K.; Pravin Kumar, S.K.; Rukkumani, V.; Ramya, R. Fruit Classification Using Traditional Machine Learning and Deep Learning Approach; Springer: Cham, Switzerland, 2020; Volume 1108, ISBN 9783030372170.

Tools Used On Implementation:

1. Kaggle.com
2. Google Colab
3. Jupyter Notebook