Image Classification Using MobileNetV2: A Research Overview

Author: Gursimran Sujlana

Abstract

This research focuses on the application of the MobileNetV2 architecture for image classification tasks. MobileNetV2, known for its efficiency in mobile and embedded vision applications, is leveraged to classify a diverse set of food images into predefined categories. The model is fine-tuned on a custom dataset and evaluated for its accuracy. Additionally, the study incorporates the use of an external dataset for further analysis of food items and their calorie content.

1. Introduction

Image classification is a fundamental task in computer vision with applications spanning various fields such as medical imaging, autonomous driving, and retail. Deep learning models, particularly Convolutional Neural Networks (CNNs), have significantly improved the performance of image classification tasks. MobileNetV2, a lightweight deep learning model, offers a balance between accuracy and computational efficiency, making it suitable for deployment on devices with limited resources.

2. Literature Review

The advent of deep learning has revolutionized image classification, with models like AlexNet, VGG, and ResNet setting benchmarks. MobileNetV2, introduced by Google, improves upon its predecessor by using depthwise separable convolutions and linear bottlenecks, making it both fast and accurate. Previous studies have demonstrated the effectiveness of MobileNetV2 in various applications, including real-time object detection and medical image analysis.

3. Methodology

3.1 Data Preparation

The dataset used for this study consists of images organized into training and testing directories. Each image is labeled according to its class. ImageDataGenerator from TensorFlow's Keras API is used for data augmentation and preprocessing.

- **Training Data**: 19,098 images belonging to 34 classes, augmented using random transformations to enhance the model's generalization capabilities.
- **Validation Data**: 4,775 images belonging to 34 classes, preprocessed similarly to the training data but without augmentation.

3.2 Model Architecture

The MobileNetV2 architecture, pretrained on the ImageNet dataset, serves as the base model. The top classification layer is removed, and a custom head is added to adapt the model to our specific classification task.

- **Base Model**: MobileNetV2 without the top layer, set to non-trainable to retain pretrained weights.
- **Custom Head**: Consists of a Global Average Pooling layer, followed by a Dense layer with 1024 units and ReLU activation, a Dropout layer for regularization, and a final Dense layer with softmax activation for classification into 34 categories.

3.3 Model Compilation and Training

The model is compiled using the Adam optimizer and categorical cross-entropy loss. It is trained for 10 epochs on the training dataset, with accuracy as the evaluation metric.

3.4 Model Evaluation

The trained model is evaluated on the test dataset to assess its accuracy and generalization performance. The evaluation metrics include test loss and test accuracy. The model achieved an accuracy of 85% on the test dataset.

4. Results

The trained model achieved an accuracy of 85% on the test dataset. The food items and their respective calorie counts were sorted and analyzed. The following data was obtained:

4.1 Food Item Calories

```mathematica
Food Item          Calories
Apple Pie          295
Baked Potato       160
Burger             295-800
Butter Naan        260
Chai               30
Chapati            104
Cheesecake         257
Chicken Curry      200-300
Chole Bhature      450
Crispy Chicken     290-350
Dal Makhani        245
Dhokla             45
Donut              250-500
Fried Rice         240
Fries              365
Hot Dog            150-300
Ice Cream          200
Idli               39
Jalebi             150
Kaathi Rolls       150-350
Kadai Paneer       290
Kulfi              200
Masala Dosa        133
Momos              35
Omelette           154
Paani Puri         10
Pakode             75
Pav Bhaji          600
Pizza              285
Samosa             262
Sandwich           200-500
Sushi              200-400
Taco               170-250
Taquito            250
```

Copy code: mathematica, Copy code

4.2 Class Indices

arduino
Copy code

```
{
 'Baked Potato': 0,
 'Crispy Chicken': 1,
 'Donut': 2,
 'Fries': 3,
 'Hot Dog': 4,
 'Sandwich': 5,
 'Taco': 6,
 'Taquito': 7,
 'apple_pie': 8,
 'burger': 9,
 'butter_naan': 10,
 'chai': 11,
 'chapati': 12,
 'cheesecake': 13,
 'chicken_curry': 14,
 'chole_bhature': 15,
 'dal_makhani': 16,
 'dhokla': 17,
 'fried_rice': 18,
 'ice_cream': 19,
 'idli': 20,
 'jalebi': 21,
 'kaathi_rolls': 22,
 'kadai_paneer': 23,
 'kulfi': 24,
 'masala_dosa': 25,
 'momos': 26,
 'omelette': 27,
 'paani_puri': 28,
 'pakode': 29,
 'pav_bhaji': 30,
 'pizza': 31,
 'samosa': 32,
 'sushi': 33
}
```

5. Discussion

The results indicate that MobileNetV2, with transfer learning, is effective for image classification tasks. The model benefits from pretrained weights, reducing the need for extensive training on large datasets. However, the performance can be further improved by fine-tuning more layers of the base model and increasing the training epochs.

6. Conclusion

This study demonstrates the applicability of MobileNetV2 for image classification. The model achieves competitive accuracy with minimal computational resources, making it suitable for deployment in real-world applications. Future work includes exploring more advanced data augmentation techniques and integrating additional features such as attention mechanisms to enhance model performance.

7. References

- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4510-4520).
- Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1251-1258).
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In Advances in Neural Information Processing Systems (pp. 1097-1105).