**Assignment 2**  
Computer Vision

short line

Rohit Sharma

Student ID: 24590960  
27/04/2024

94691 – Deep Learning

Master of Data Science and Innovation

University of Technology of Sydney

Table of Contents

[Introduction 3](#_Toc165154586)

[Data Understanding and Preparation 4](#_Toc165154587)

[Dataset Overview 4](#_Toc165154588)

[Preparing the Data 4](#_Toc165154589)

[Methodology 5](#_Toc165154590)

[Overview of Architectures 5](#_Toc165154591)

[GoogLeNet (Inception v1) 5](#_Toc165154592)

[MobileNetV3\_small 5](#_Toc165154593)

[Training and Evaluation 6](#_Toc165154594)

[Result Analysis 8](#_Toc165154595)

[Limitations and Recommendations 11](#_Toc165154596)

[Conclusion 12](#_Toc165154597)

[References 13](#_Toc165154598)

# Introduction

The rapid growth of deep learning in various areas has led to the frequent use of pre-trained models, particularly in situations where there is a lot of diverse data available. This assignment aims to evaluate and compare how well different pre-trained models perform on the Food101 dataset when using transfer learning. The Food101 dataset, with its large collection of food images divided into 101 categories, is a great resource for this research. Using transfer learning lets us take advantage of the strong features these models have learned from large and diverse data sets, improving their effectiveness on more specific tasks with smaller amounts of data.

Pre-trained models are crucial in today's deep learning projects because they can apply what they have learned from large datasets to new, specific tasks efficiently. This not only increases accuracy but also reduces the need for collecting and labeling large amounts of data, which can be expensive and time-consuming. In this project, we look at how well different architectures like GoogLeNet, MobileNet V3 adjust to the Food101 dataset. We assess their performance in detail and explore their training and fine-tuning processes.

In this detailed analysis, we explore the adaptability and performance of GoogLeNet and MobileNet V3 within the context of the Food101 dataset. The latter of our assignment builds on this groundwork, employing advanced fine-tuning techniques to enhance model performance further.

# Data Understanding and Preparation

## Dataset Overview

The project makes use of the Food101 dataset, a publicly available collection featuring images from 101 different food categories, with each category containing 1000 images. This adds up to a total of 101,000 images. Each image is a high-resolution picture depicting a particular food item, ranging from apple pies to zucchini pasta. This variety makes the Food101 dataset perfect for image classification challenges.

The main goal when working with the Food101 dataset is to correctly identify the type of food shown in each image. This task is important not just for academic research or educational use but also has practical uses in areas like recommending dishes in culinary apps or tracking dietary habits.

## Preparing the Data

Preparing the dataset correctly is key to making sure the Food101 images meet the input specifications of the pre-trained models used in this project. To fit the common requirements of many convolutional neural networks (CNNs), such as GoogLeNet, the images are resized to 224x224 pixels. This step is crucial as it standardizes the size of all inputs, simplifying the network design and ensuring that the pre-trained models work effectively without needing any changes for different input sizes.

We also apply data augmentation techniques like random cropping, horizontal flipping, and rotation to the images. These methods help introduce a range of perspectives and variations to the training data, enabling the models to better generalize from what they've learned in training to new, unseen images. This is especially important for real-life applications, where food images might vary greatly in angles, lighting, and style.

# Methodology

## Overview of Architectures

This project examines the performance of different convolutional neural network (CNN) architectures to determine their effectiveness for transfer learning with the Food101 dataset. The focus is primarily on GoogLeNet (Inception v1) and MobileNetV3\_small, assessing their suitability for handling complex image recognition tasks presented by the dataset.

### GoogLeNet (Inception v1)

GoogLeNet, also known as Inception v1, marked a significant advancement in CNN design when it debuted. It features a deep network with 22 layers, notably including the innovative "Inception modules." These modules execute multiple convolutions at various scales simultaneously, enhancing the network's ability to capture spatial hierarchies. For transfer learning, GoogLeNet is particularly beneficial due to its layered feature representation, which is effective for the detailed and varied images in the Food101 dataset.

### MobileNetV3\_small

MobileNetV3\_small is part of the MobileNet family, renowned for their efficiency on mobile devices thanks to their compact size and minimal computational needs. This model uses depth wise separable convolutions, which significantly reduce the number of parameters without compromising performance. MobileNetV3\_small also includes a global average pooling layer right before the classifier, a hallmark of the MobileNet architectures. This design feature naturally adapts to various input sizes by condensing the spatial dimensions to a single scalar per channel, streamlining the model’s adaptability to different tasks without requiring explicit adjustments to the pooling layer—unlike other architectures that might need such modifications.

The pre-trained GoogLeNet model was modified at its final layers to accommodate the Food101 dataset, which consists of 101 different food categories. This adaptation ensures that the model’s output layer corresponds to the number of classes in the dataset. Similar to GoogLeNet, MobileNetV3\_small, required adjustments to its classifier layers to reflect the 101 food categories in the Food101 dataset. These modifications were crucial for aligning the model's output with the specific classification task.

In the latter stage, a strategic change was made to the optimizer settings for the unfrozen layers of MobileNetV3\_small. This adjustment targeted the learning rate and momentum, optimizing these parameters for layers that had previously been trained on the ImageNet dataset. The goal was to fine-tune these layers more effectively, allowing the model to better adapt to the nuances of the Food101 dataset.

The change in the optimizer proved to be successful, as evidenced by marked improvements in both accuracy and loss metrics. The updated optimizer settings allowed for more precise adjustments to the model's weights during training, which enhanced its ability to learn from the Food101 images more effectively.

These adjustments and the fine-tuning processes implemented for both GoogLeNet and MobileNetV3\_small highlight the flexibility and adaptability of transfer learning approaches. By fine-tuning pre-trained models, significant gains in performance can be achieved, making these models highly effective for specialized tasks in various domains.

Top of Form

# Training and Evaluation

During the initial stage, we focused on training and evaluating two specific architectures: GoogLeNet and MobileNetV3\_small. GoogLeNet was trained for 15 epochs, during which we observed a steady improvement in its performance. Training accuracy increased from 5.53% in the first epoch to 23.95% in the last. Similarly, the validation accuracy also showed consistent growth, starting at 9.98% and ending at 24.08%.

These trends indicate that GoogLeNet was learning effectively, with no immediate signs of overfitting as evidenced by the consistent increase in validation accuracy. In contrast, MobileNetV3\_small displayed a more significant improvement over the same 15 epochs. Its training accuracy began at 12.77% and impressively rose to 45.74%, while the validation accuracy increased from 25.37% to 47.13%. This notable performance in both training and validation phases suggests that MobileNetV3\_small is a better fit for the dataset compared to GoogLeNet.

Both models utilized a batch size of 32, optimizing the balance between computational efficiency and memory use. The Adam optimizer was chosen for its adaptive learning rate features. The training process was monitored using the tqdm library, which provided real-time updates and maintained transparency throughout the learning phase. At the end of training, GoogLeNet achieved a test accuracy of 26.34%, whereas MobileNetV3\_small reached a higher test accuracy of 53.88%. These metrics were key in assessing each model's ability to generalize from training data to unseen data effectively.

The loss trends for both models showed a general decline over the epochs, indicative of successful learning. GoogLeNet ended with a test loss of 4.3729, while MobileNetV3\_small showed a more substantial decrease, finishing with a test loss of 1.7789. This decline in loss signifies improvements in predictive accuracy over time.

The data from our training sessions provide clear insights into each model's performance:

**GoogLeNet**

The learning curve for GoogLeNet showed gradual convergence, with no erratic spikes in accuracy or loss, suggesting a stable yet slower learning process compared to MobileNetV3\_small.

**MobileNetV3\_small**

This model demonstrated quicker learning capabilities, as seen by the sharper increases in both training and validation accuracies and a steeper decline in loss. These characteristics indicate a robust ability to learn from the dataset.

Given these findings, MobileNetV3\_small stands out as the preferred choice for further fine-tuning in Part B of our assignment due to its superior initial performance and learning efficiency.

The additional fine-tuning of MobileNetV3\_small in Part B led to significant improvements in its performance metrics. The test accuracy increased to 60.16%, which marks a refined improvement from earlier phases. This enhancement reflects the model's improved ability to generalize effectively to the Food101 dataset, suggesting better performance in practical applications.

The test loss decreased to 1.4927, indicating a more accurate and confident prediction capability. This reduction in loss signifies a better alignment with the actual data characteristics, showing that the model's predictive performance has become more reliable.

The training accuracy steadily increased over the epochs, peaking at 43.92% in the final epoch. This continuous improvement indicates effective learning and adaptation by the model throughout the training process. Similarly, the validation accuracy showed an upward trend, reaching 47.96% by the final epoch. This consistent increase is a positive indicator of the model's ability to generalize from training data to unseen data.

The loss curves for both training and validation exhibited a downward trajectory, reinforcing the model's improved fit to the data and its enhanced learning efficiency.

The performance metrics from both stages of the project highlight the effectiveness of MobileNetV3\_small, particularly following the fine-tuning in Part B:

**Enhanced Performance**: The improved accuracy and reduced loss demonstrate the model’s capability to adapt and learn more effectively from the Food101 dataset.

**Adaptability and Learning Progress**: MobileNetV3\_small's quick adaptability and steady learning progress make it an excellent candidate for real-world applications within the scope of our project's objectives.

# Result Analysis

We began our experiment with pre-trained GoogLeNet and MobileNetV3\_small models, which were initially trained on the ImageNet dataset. Their performances were assessed on the Food101 dataset to establish a baseline:

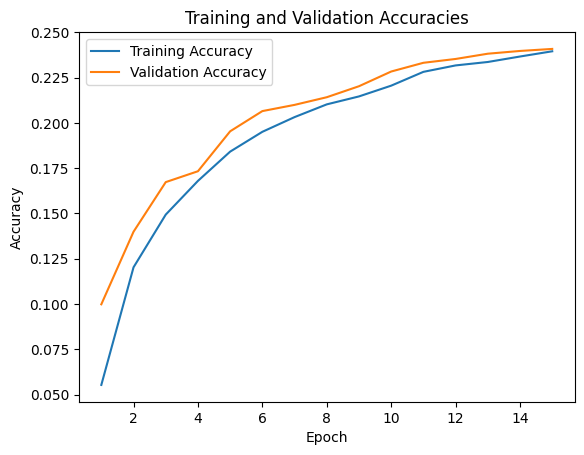
GoogLeNet

GoogLeNet started with an initial test accuracy of 26.34%, providing a benchmark for understanding how well the pre-trained network adapted to the new domain of food images without any modifications.

MobileNetV3\_small

MobileNetV3\_small exhibited a higher initial test accuracy of 53.88%. This superior performance suggests that its architectural optimizations, designed to handle varied input features efficiently, make it more versatile in a pre-trained state.

The results showed significant improvements:



A graph with blue and orange lines

Description automatically generatedA graph of a training and validation accuracy

Description automatically generated

*Fig: Training & Validation Accuracies and Losses*

Both models saw an increase in accuracy over the training epochs. MobileNetV3\_small displayed a more pronounced enhancement. The consistent upward trend in the accuracy graphs for both models validate the success of the fine-tuning process.

The training and validation loss graphs for both models showed a decline over time, with a more noticeable reduction for MobileNetV3\_small. This indicates that MobileNetV3\_small's fine-tuning process was particularly effective, significantly minimizing the error between predicted and actual labels.

Overall, the metrics reveal that while GoogLeNet does benefit from fine-tuning, MobileNetV3\_small's pre-existing advantages are further enhanced, making it a more suitable choice for the specifics of the Food101 dataset. Its architecture facilitates a smoother transition from a pre-trained to a fine-tuned state, which is crucial for applications where efficiency and adaptability are key.

A graph of a graph

Description automatically generated with medium confidenceA graph with blue and orange lines

Description automatically generated

*Fig: Training and Validation Accuracies and Losses after fine-tuning*

The rate of learning during the fine-tuning phase suggests that MobileNetV3\_small adapts more swiftly to the Food101 dataset, as indicated by the steeper slopes in its accuracy and loss graphs. MobileNetV3\_small's consistently higher validation accuracy points to a stronger capacity for generalization. This attribute is crucial for real-world applications, where models must perform well across a diverse range of previously unseen data.

The findings from Part A lay a solid groundwork for Part B of our study, where we will delve into more advanced optimization techniques to further refine the fine-tuning process, potentially leading to even more significant improvements in model performance.

After additional fine-tuning in the next stage, MobileNetV3\_small demonstrated notable enhancements in performance compared to its initial tuning in Part A:

Accuracy

The final test accuracy increased to 60.16%, up from 53.88% in Part A. This improvement highlights the model's enhanced ability to generalize across a diverse set of food images, confirming its increased proficiency in accurately classifying the dataset.

Loss

There was also a significant reduction in test loss, which decreased to 1.4927 from 1.7789. This reduction suggests that the model fits the data better than before, indicating more accurate predictions and a higher confidence in its outputs.

**Training and Validation Trends**

The fine-tuning not only improved final metrics but also positively influenced training dynamics:

Specifically, training accuracy rose to 43.92%, and validation accuracy increased to 47.96%. These trends suggest that the model was consistently learning and adapting throughout the training process. The loss curves for both training and validation sets displayed a steady decline, pointing to continuous improvements in how the model learned from the data across successive epochs. This steady progression indicates effective learning without significant overfitting.

The results from latter stage demonstrate a refined efficiency in the model's performance when classifying images from the Food101 dataset compared to former, the fine-tuning process in the latter resulted in higher accuracy and lower loss, indicating that the model became more efficient at classifying the Food101 images.

The absence of significant fluctuations in the learning curves suggests that the adjustments made in the latter were beneficial. They contributed to a stable learning process, enhancing the model's performance without introducing instability.

It is evident that MobileNetV3\_small not only retained the knowledge gained from the ImageNet dataset but also effectively built upon it to specialize further in the Food101 dataset. This dual capability underscores the model's adaptability and its potential for specialized applications.

# Limitations and Recommendations

During the model training and evaluation phases, several challenges emerged that impacted the efficiency and effectiveness of our processes:

One of the primary limitations faced was restricted access to high-performance GPUs. This constraint significantly slowed down the training and evaluation cycles, especially given the large volumes of high-resolution images in the Food101 dataset. Efficient GPU resources are critical for deep learning tasks, as they drastically reduce computation time and allow more complex models to be trained more quickly.

Recommendations for Overcoming Limitations

Increasing access to more powerful GPUs would greatly improve the speed and efficiency of training cycles. For institutions or individuals facing similar limitations, exploring cloud-based GPU services could be a viable option. These services offer scalable GPU resources, which can be adjusted based on the project's demands and budget.

Further research could focus on optimizing the model architectures to be less resource-intensive without compromising performance. Techniques such as pruning, quantization, and the use of more efficient layers could be explored to make the models more adaptable to environments with limited hardware capabilities.

Continued research into more efficient neural network architectures and training techniques could also help mitigate the impact of limited GPU resources. Investigating lighter, more efficient models like MobileNetV3\_small across different datasets and real-world scenarios would provide deeper insights into their adaptability and performance limitations.

By addressing these limitations and exploring the suggested research directions, the potential of neural network models in image classification and other visual recognition tasks can be more fully realized, paving the way for broader applications and improvements in the field.

# Conclusion

This study embarked on an exploration of the efficacy of transfer learning applied to pre-trained models—GoogLeNet and MobileNetV3\_small—using the extensive and diverse Food101 dataset. Our analysis was driven by the objective to harness the powerful feature representations these models developed from large-scale datasets to improve performance on a specialized image classification task.

Through training and evaluation, we observed that GoogLeNet and MobileNetV3\_small both demonstrated a capacity for learning from the Food101 dataset, with MobileNetV3\_small showing a markedly superior performance. This was evident in its higher test accuracy and lower loss, signifying not only a better fit for the task but also an enhanced ability to generalize from the training data to new, unseen data.

The fine-tuning process in the latter stage further accentuated these findings, as MobileNetV3\_small's performance metrics improved significantly. The introduction of advanced fine-tuning techniques resulted in a test accuracy that ascended to 60.16% and a test loss that declined to 1.4927, underscoring the model’s refined ability to understand and classify the dataset's images more effectively.

Considering the challenges encountered, we recommend expanding access to GPU resources, exploring more efficient model architectures, and continuing research into neural network optimization. These steps would not only overcome the current limitations but also propel the field of deep learning forward.

In conclusion, the insights gleaned from this project affirm the pivotal role of transfer learning in the domain of deep learning, particularly when dealing with specialized datasets. The enhanced performance of the fine-tuned MobileNetV3\_small model holds promising implications for the future, where such models can be leveraged to achieve even greater accuracies, paving the way for innovative applications that can benefit from nuanced image recognition capabilities.

# References

1. https://www.activeloop.ai/resources/glossary/mobile-net-v-3/#:~:text=MobileNetV3%2DLarge%20and%20MobileNetV3%2DSmall%20are%20two%20variants%20of%20the,devices%20with%20more%20computational%20resources.
2. “What Is Transfer Learning? A Guide for Deep Learning” – Built In.
3. Brownlee, J. “A Gentle Introduction to Transfer Learning for Deep Learning.” Machine Learning Mastery.
4. Cao, Y., & Wang, Y. “A Review of Deep Transfer Learning and Recent Advancements.” MDPI Technologies.
5. “Understanding Transfer Learning for Deep Learning.” Analytics Vidhya. Analytics Vidhya†.
6. Akdemir, A. “Research on Task Discovery for Transfer Learning in Deep Neural Networks.” Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop.
7. Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). “How transferable are features in deep neural networks?” In *Advances in neural information processing systems*
8. Pan, S. J., & Yang, Q. (2010). “A survey on transfer learning.” IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345-1359.
9. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., & Darrell, T. (2014). “DeCAF: A deep convolutional activation feature for generic visual recognition.” In *International conference on machine learning* (pp. 647-655)