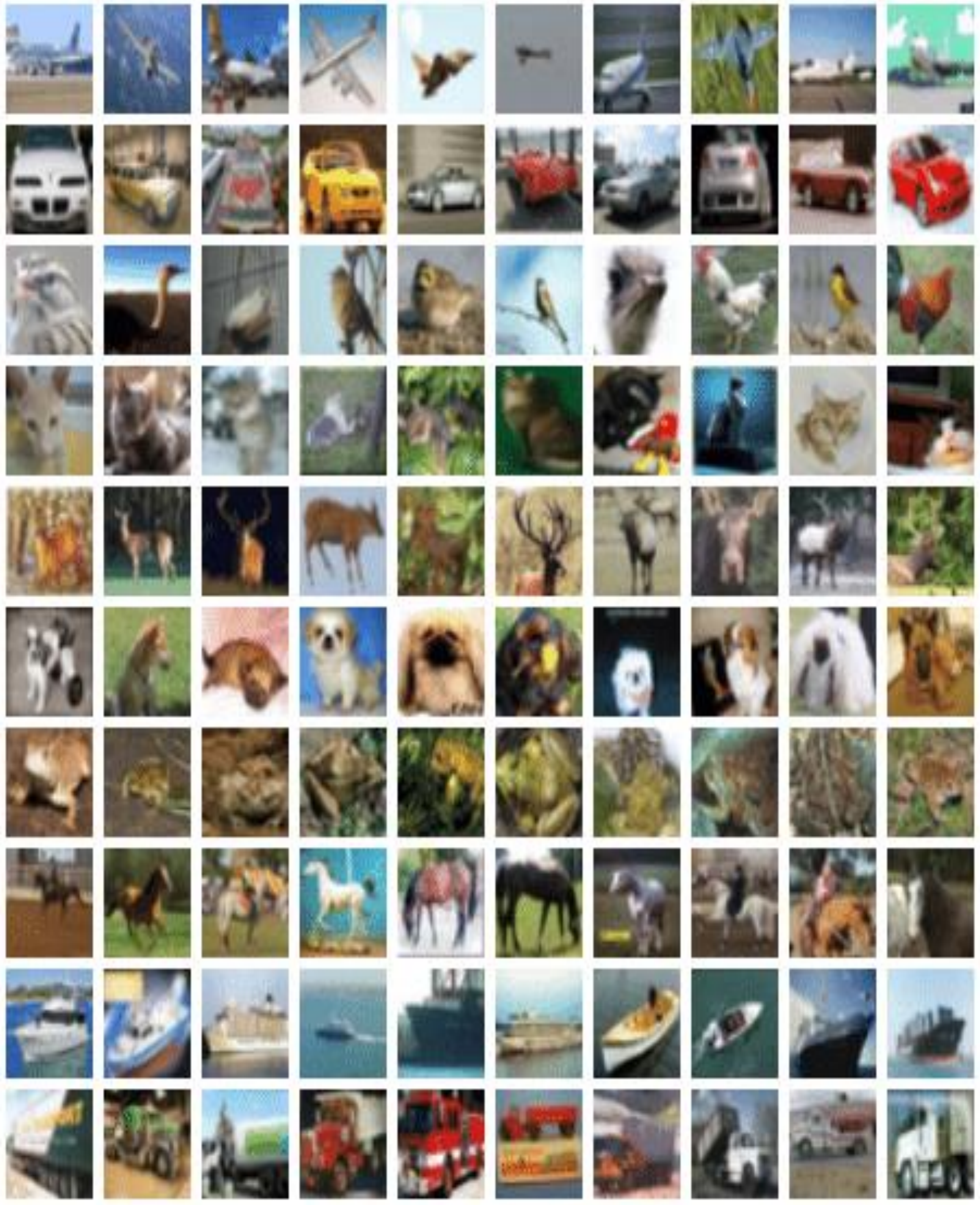
# Analysis and Comparison of Transfer Learning and CNN Approaches on Image Classification

[Google Colab Link]

Name: Friday Odeh Ovbiroro

Student No: 22034665



# INTRODUCTION

This project aims at exploring the captivating realm of transfer learning as it pertains to image classification tasks. Transfer learning, a formidable technique in deep learning, involves harnessing the knowledge ingrained in a pre-trained model, particularly emphasizing its weights, for a novel and related task. The crux of my investigation revolves around unraveling the principles and significance of transfer learning, with a specific emphasis on its integration within the Keras framework.

The CIFAR-10 dataset meticulously crafted by the Canadian Institute for Advanced Research is used in this project. This dataset acts as a guiding beacon, allowing me to evaluate the efficacy of transfer learning vis-à-vis the conventional methodology of building a model from scratch. CIFAR-10 comprises 60,000 32x32 color images, segregated into a 50,000-image training set and a 10,000-image test set, spanning 10 diverse classes. These classes depict a varied array of objects and scenes, posing a formidable challenge for image classification tasks.

The presentation will pivot around several key focal points:

- An overview of transfer learning and its pivotal role in the landscape of machine learning.
- The selection of the pre-trained model employed in the project and a thorough exploration of its original purpose.
- Delving into the fine-tuning steps applied to tailor the pre-trained model to the specific task at hand.
- A comparison of results between transfer learning and the process of crafting a model from the ground up.
- limitations and exploration of potential areas for enhancement.

# Revolutionizing Machine Learning: The Crucial Role of Transfer Learning

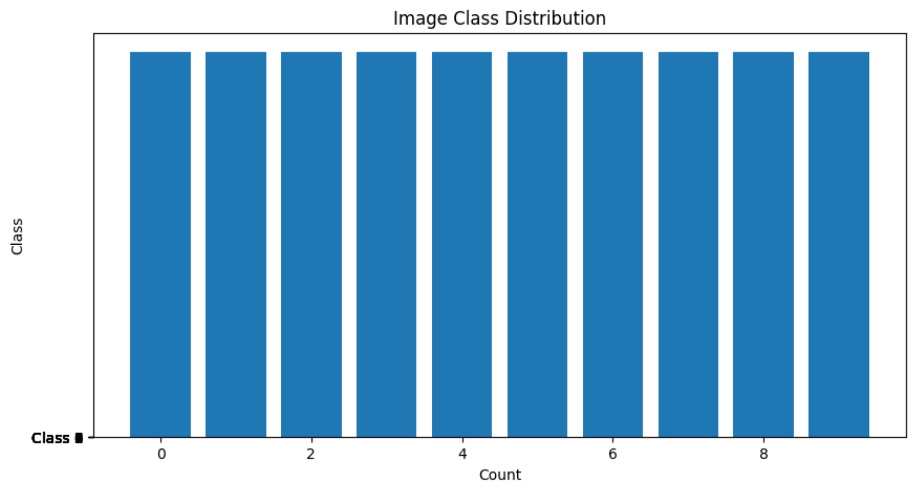
Transfer learning, a foundational pillar in machine learning, is reshaping our approach to novel tasks by seamlessly applying knowledge acquired from one domain to another. The versatility of this approach spans diverse domains, including image and speech recognition to natural language processing. This adaptability empowers models to confront new challenges efficiently, conserving computational resources and time.

The application of transfer learning comes with multifaceted advantages. It significantly enhances model performance, especially in scenarios where labeled data is scarce. By building upon existing knowledge, models achieve superior generalization, adapt swiftly, and attain heightened accuracy. Moreover, transfer learning proves to be a valuable ally in overcoming challenges posed by data scarcity, providing a viable path to improved performance even in resource-constrained environments.

Within the realm of transfer learning, two prevalent adaptations come to the forefront: fine-tuning and feature extraction. Fine-tuning allows for the adjustment of specific layers within a pre-trained model to align with the nuances of the new task. Conversely, feature extraction involves harnessing the insights acquired by the early layers of the pre-trained model as a fixed feature extractor for the new task. These modifications underscore the inherent flexibility in transfer learning, offering bespoke adjustments to optimize model performance based on the distinct requirements of the target task.

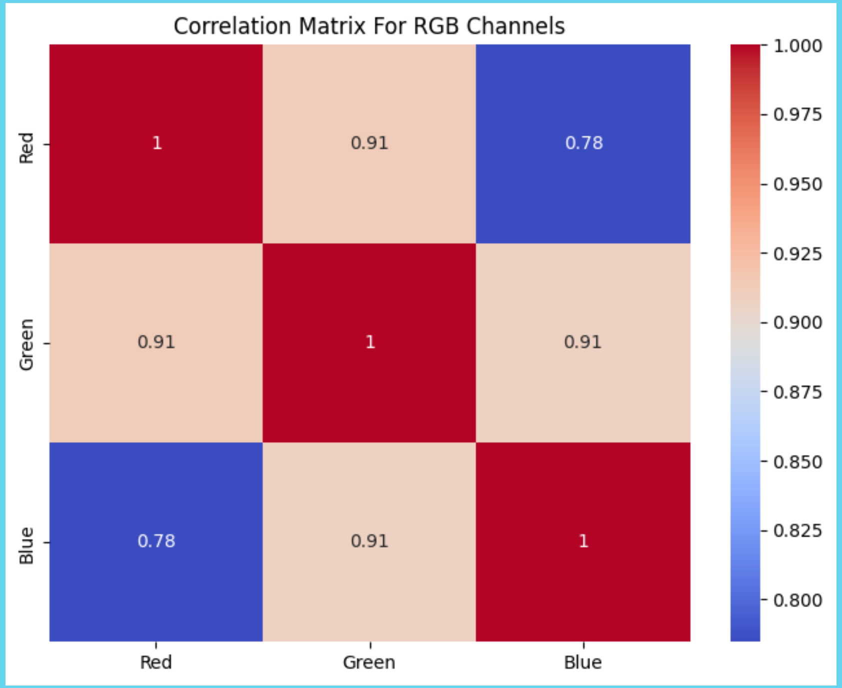
# Exploratory Data Analysis (EDA)

Before building the models, exploratory data analysis (EDA) was carried out. This essential step helped to deeply understand the dataset, which is key in fine-tuning the preprocessing required for this project.




**Analysis of Class Distribution:**

Investigating the class distribution in the CIFAR-10 dataset showed an even distribution across different classes, indicating a balanced dataset. This balance is important for training models to prevent bias towards overrepresented classes. It ensures equal representation of each category, allowing for a fair evaluation of the model's performance across different classes.




**Analysis of Color Channel Correlations in CIFAR-10 Dataset:**

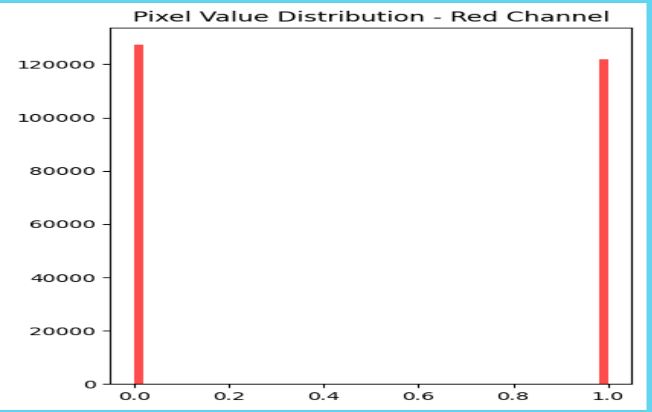
Analyzing the CIFAR-10 dataset’s color channel interactions, we found strong correlations among the red, green, and blue channels. Notably, the red and green channels shared a high correlation coefficient of 0.91, suggesting a significant linear relationship and information overlap. A similar high correlation was observed between the green and blue channels. However, the correlation between the red and blue channels was somewhat lower at 0.78, indicating a strong but less redundant relationship compared to the other channel pairs

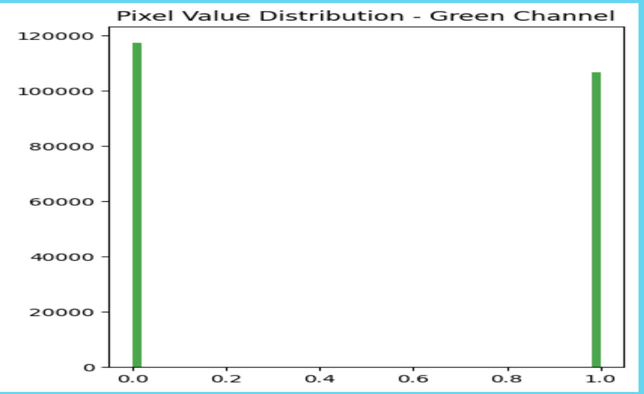
# Exploratory Data Analysis Cont'd

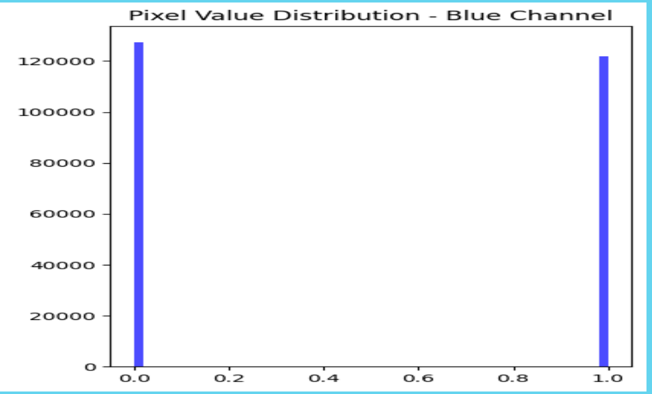
**Histogram of Pixel Intensities in Different Color Channels:**

Examining the histograms of pixel intensities in the CIFAR-10 dataset's red, green, and blue channels reveals distinct peaks at both the lower and upper extremes, indicating a bimodal distribution. This suggests that many pixels are either at the high end (likely representing main subjects in the images with rich saturation) or at the low end (possibly backgrounds with little saturation). The lack of a significant number of mid-range intensity pixels implies that the images in this dataset tend to have stark contrasts, showing little gradual transition between colors. This characteristic can pose challenges in image processing tasks or during the training of machine learning models, as it leads to less nuanced color distinctions between different categories.

The clear skew towards the extremes of the intensity spectrum highlights the importance of careful normalization in the preprocessing phase of this project.

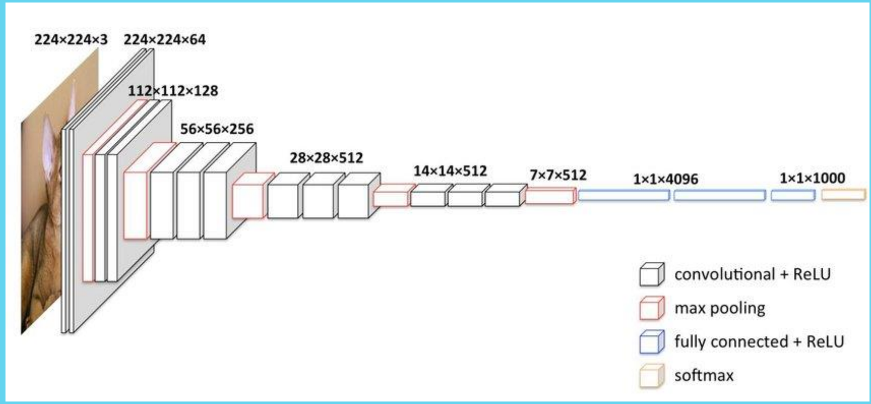


# Strategic Choice: Opting for VGG16 in Image Recognition

In this project, the selected pretrained model is VGG16 (Visual Geometry Group 16), a convolutional neural network (CNN) architecture celebrated for its outstanding performance in the ImageNet classification challenge. Boasting 16 layers, each equipped with 3x3 filters, VGG16 seamlessly amalgamates simplicity and efficacy, rendering it an ideal candidate for diverse transfer learning applications.




**Purpose and Versatility:**

VGG16's roots trace back to its remarkable feats in the ImageNet competition, where it underwent rigorous training on over 1.2 million labeled images spanning diverse categories. Proficient in object detection and classification, VGG16 demonstrates its prowess by accurately categorizing 1000 images with an impressive accuracy of 92.7%. It stands as a cornerstone in image classification algorithms, particularly esteemed for its seamless integration into transfer learning frameworks.

Within this project's scope, I leverage VGG16 as the pre-trained model, delving into its reservoir of knowledge from the ImageNet challenge. Through the strategic utilization of pre-existing knowledge and weights, the VGG16 model adeptly learns and adapts to the unique intricacies and complexity of the new image recognition task.

# Fine-Tuning Expertise: Tailoring VGG16 for CIFAR-10 Precision

## Modifying the Top 5 Layers for Better Task-Specific Feature Detection in the CIFAR-10 Dataset

In the meticulous refinement phase of this analysis, a finetuning strategy was adapted to the pre-trained VGG16 model for the intricacies of the CIFAR-10 dataset. To fully unlock the potential of transfer learning, a conscious decision was made to selectively freeze the initial convolutional layers of the VGG16 architecture. This strategic move aimed to safeguard the wealth of pre-trained knowledge woven into these layers, originally cultivated from the expansive ImageNet dataset, capturing intricate hierarchical features.

Additionally , to make some parts of the VGG16 model more flexible to learn complexities of the Cifar10 dataset, changes were made not only the dense layers but also the top 5 layers that process the images. As a result, these layers were able to learn and adapt the complexities from new data. These unfrozen layers, encompassing both the dense layers and the top 5 layers responsible for high-level feature extraction, played a pivotal role in capturing more task-specific features. This precision fine-tuning step was thoughtfully chosen to adapt the model's learned representations to the unique characteristics and nuances inherent in the CIFAR-10 dataset.

The acumen of this fine-tuning strategy aimed to strike a delicate balance—leveraging the generalization power of pre-trained convolutional features while adapting the model's higher-level representations for optimal performance in classifying CIFAR-10 images. The decision to unfreeze the top 8 convolutional layers acted as a catalyst, enabling the model to dynamically adapt and fine-tune these higher-level features. The result: a bespoke model, meticulously tuned to achieve superior results in our targeted image classification task.

# Data Preparation for Model Training and Optimization

**One-Hot Encoding for Class Labels:**

The conversion of class labels into one-hot encoded vectors played a pivotal role in facilitating multi-class classification. This adept process transforms numerical class labels into a binary matrix representation, laying the foundation for effective model training.

**Normalization for Standardized Input:**

An integral measure centered around normalizing the image data to ensure pixel values lie within the 0 to 1 range. This normalization, achieved by dividing pixel values by the maximum value of 255, optimizes the model's learning process by establishing a standardized input.

**Strategic Allocation of Validation Set:**

Strategically earmarking 20% of the training data to compose a validation set proved instrumental. This distinct dataset serves the critical role of monitoring the model's performance on previously unseen data, offering valuable insights for fine-tuning and acting as a safeguard against overfitting.

**Optimization with SGD Algorithm:**

For the optimization stage, the Stochastic Gradient Descent (SGD) algorithm was chosen for its simplicity and efficiency. SGD is particularly effective for large-scale datasets, offering steady and gradual improvements to the model's performance. Its straightforward approach to iteratively updating model parameters makes it a reliable choice for a wide range of machine learning problems..

# Model Summary :Transfer learning vs. building a model from scratch

## MODEL FROM SCRATCH

Model: "sequential"

| Layer (type) | Output Shape | Param # |
|---|---|---|
| conv2d (Conv2D) | (None, 32, 32, 64) | 1792 |
| max_pooling2d (MaxPooling2D) | (None, 16, 16, 64) | 0 |
| conv2d_1 (Conv2D) | (None, 14, 14, 32) | 18464 |
| max_pooling2d_1 (MaxPooling2D) | (None, 7, 7, 32) | 0 |
| dropout (Dropout) | (None, 7, 7, 32) | 0 |
| flatten (Flatten) | (None, 1568) | 0 |
| dense (Dense) | (None, 256) | 401664 |
| dropout_1 (Dropout) | (None, 256) | 0 |
| dense_1 (Dense) | (None, 10) | 2570 |

Total params: 424490 (1.62 MB)
Trainable params: 424490 (1.62 MB)
Non-trainable params: 0 (0.00 Byte)

The convolutional neural network (CNN) built from the ground up boasts a sequential design, seamlessly integrating two sets of convolutional and max-pooling layers to extract robust features. Subsequent to these layers, a flatten layer takes the lead, paving the way for a dense layer accompanied by dropout for regularization. This intricate design culminates in an output layer primed for predictions. The strategic placement of dropout layers serves as a precautionary measure to curb overfitting. Notably, the model encompasses a grand total of 424,490 trainable parameters, underscoring its adaptability and formidable learning capacity.

## VGG16 TRANSFER LEARNING

Model: "sequential_2"

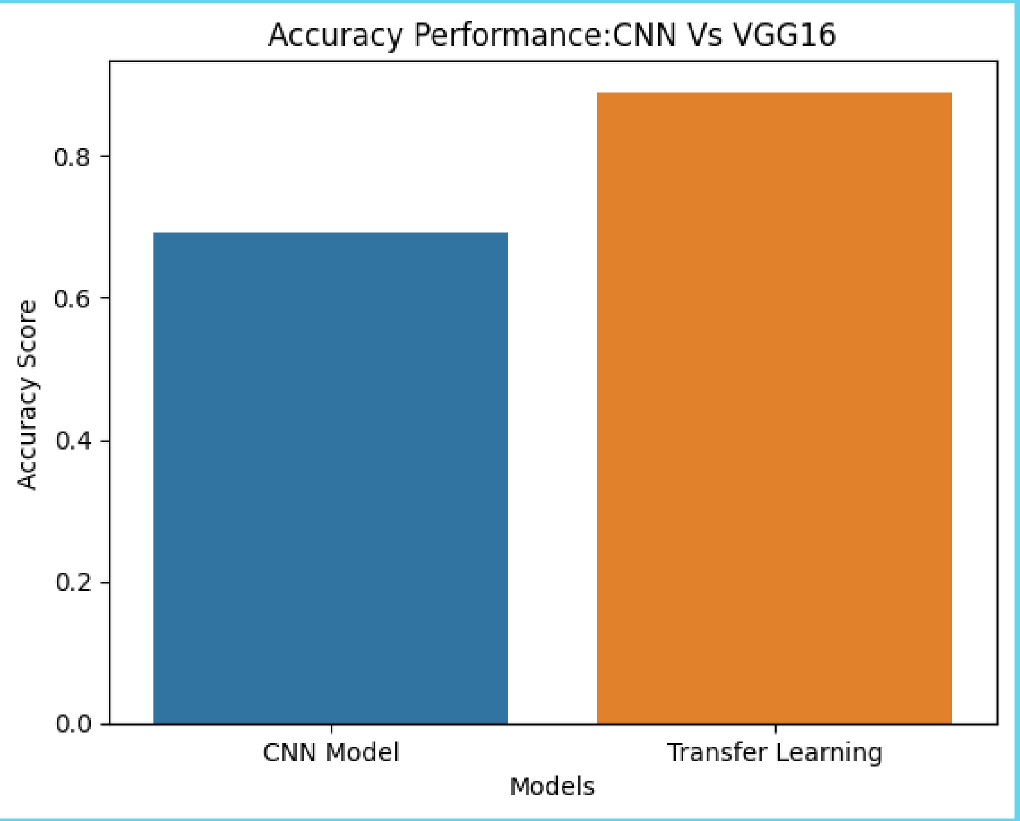
| Layer (type) | Output Shape | Param # |
|---|---|---|
| vgg16 (Functional) | (None, 2, 2, 512) | 14714688 |
| flatten_2 (Flatten) | (None, 2048) | 0 |
| dense_4 (Dense) | (None, 256) | 524544 |
| dense_5 (Dense) | (None, 10) | 2570 |

Total params: 15241802 (58.14 MB)
Trainable params: 5246730 (20.01 MB)
Non-trainable params: 9995072 (38.13 MB)

The transfer learning model embraced in this endeavor unfolds with a sequential architecture, strategically incorporating the renowned VGG16 architecture as a potent feature extractor. Sequential layers seamlessly follow, including flattening and dense layers meticulously designed for optimal classification performance Remarkably, the model boasts a substantial complexity, wielding a total of 5,246,730 trainable parameters and 9,995,072 non trainable parameters

# Performance Insights :VGG16 Transfer Learning vs. CNN Model

The thorough evaluation of transfer learning juxtaposed with building a model from the ground up on the CIFAR-10 dataset reveals illuminating insights, with a profound focus on two pivotal metrics: accuracy and the confusion matrix.



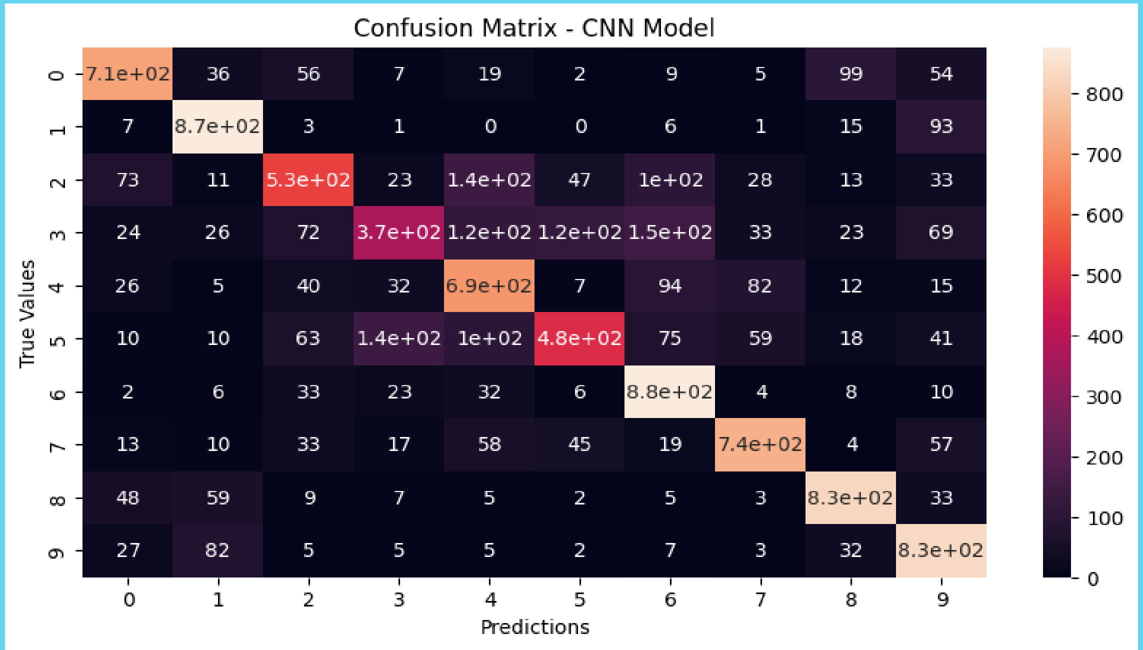

**<u>Accuracy Unveiled:</u>**

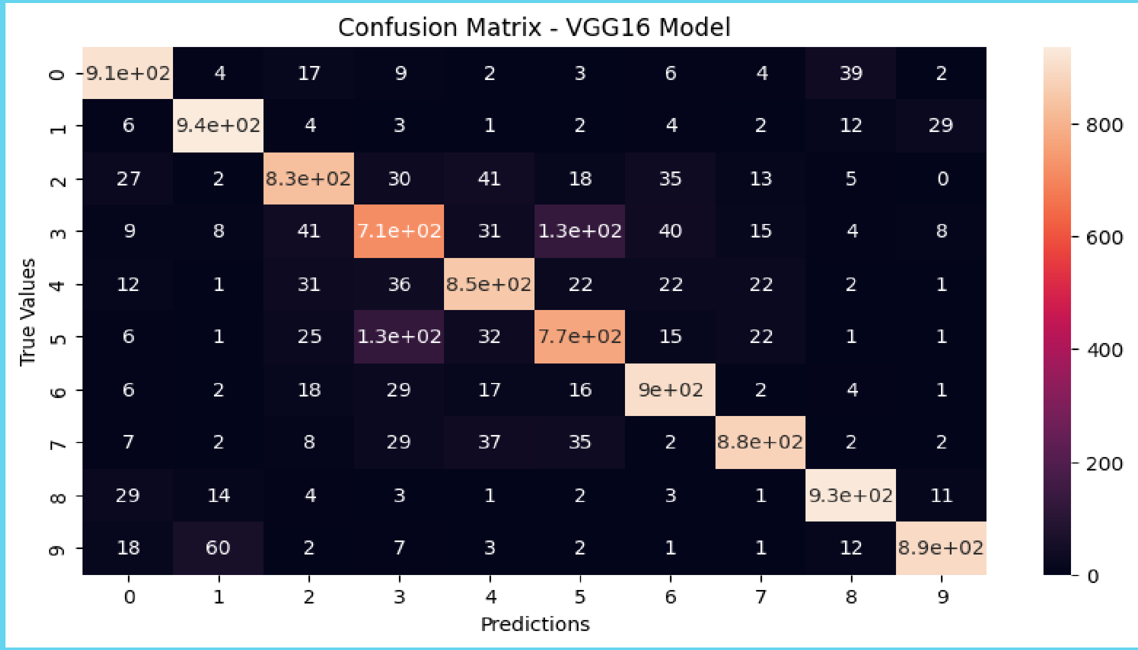
The CNN Model (depicted in blue) showcases praiseworthy performance, boasting an accuracy of 69%. This achievement stands out, especially considering the balanced nature of the CIFAR-10 dataset and the meticulous construction of the model from scratch.

In stark contrast, the VGG16 Transfer Learning Model takes the spotlight with a significantly higher accuracy of 86%. This substantial disparity underscores the prowess of transfer learning, capitalizing on pre-trained models for a remarkable performance boost.

The impressive accuracy of the transfer learning model finds its roots in the utilization of extensive and diverse knowledge acquired through prior training on vast datasets. This advantage empowers the model to discern and generalize features more effectively than its CNN counterpart, which commenced feature learning from a blank canvas.

# Confusion Matrix :Transfer learning vs. CNN Model




The scrutiny of confusion matrices unravels a comprehensive comparison of classification prowess between our the CNN Model built from scratch and the formidable VGG16 Transfer Learning Model. The CNN model shows evenly distributed misclassifications across classes, revealing certain stumbling blocks in accurate predictions. This illuminates areas necessary for improvement in feature extraction and class discrimination.

In contrast, the VGG16 model's confusion matrix displays a more pronounced diagonal accurate pattern, a testament to its superior classification accuracy—especially in classes 3,4,5 and 6. This spotlight on specific classes suggests the model utilizes the benefits of advanced feature detectors cultivated during its extensive pre-training stage, resulting in a better accuracy of precision in class predictions.

In the final evaluation, the VGG16 model stands out as the top performer. It does really well because of the advanced features it learned during its initial training. However, both models show how they could be improved for even better accuracy on specific types of data. This sets the scene for further work, where the goal of getting even better results continues.

# Conclusion

This project conducts a comparison analysis, pitting a CNN model constructed from the ground up against VGG16 transfer learning model, VGG16, using CIFAR-10 dataset. The VGG16 model asserted its dominance with superior classification accuracy, as evidenced not only in numerical accuracy metrics but also in the pronounced concentration of true positives within the confusion matrix. This triumph can be largely attributed to VGG16's nuanced feature extraction capabilities, finely tuned through pre-training on an expansive image dataset.

**Constraints and Limitations**

Despite promising outcomes, our study recognizes several limitations that warrant acknowledgment:

- Dataset Balance: The analysis unfolded on a balanced dataset, potentially deviating from real-world scenarios where data imbalances are prevalent.
- Computational Constraints: Limited computational resources imposed constraints, hindering a comprehensive exploration of hyperparameter tuning and the investigation of more intricate model architectures.

**Strategies for Enhanced Performance**

- Data Augmentation Symphony: Infuse diversity into the training dataset by employing a myriad of augmentation techniques for a more robust training set.
- Harmonic Hyperparameter Tuning: Conduct an exhaustive exploration, akin to fine-tuning instruments, to identify optimal model parameters.
- Fine-Tuning Overture: Retune additional layers of the VGG16 model, orchestrating a specific adaptation for CIFAR-10 nuances.
- Class-Specific Refinement: Direct focused attention towards underperforming classes, orchestrating tailored adjustments for improved balanced accuracy.

# References

1. Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., ... & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics, and transfer learning. IEEE Transactions on Medical Imaging, 35(5), 1285-1298.

2. Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1717-1724.

3. Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2018). Places: A 10 million Image Database for Scene Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(6), 1452-1464.

4. Wang, J., Yang, Y., Mao, J., Huang, Z., & Huang, C. (2016). Cnn-rnn: A unified framework for multi-label image classification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2285-2294.

5. Zhang, Y., & Yang, Q. (2018). A Survey on Multi-Task Learning. arXiv:1707.08114.

6. Kornblith, S., Shlens, J., & Le, Q. V. (2019). Do better ImageNet models transfer better?. arXiv:1805.08974.

7. Caruana, R. (1997). Multitask learning. Machine Learning, 28(1), 41-75.