# Image Classification and Analysis on the CIFAR-10 Dataset

Dhruba Nandi

STATS 507 Project Proposal

## Overview

The CIFAR-10 dataset is a foundational benchmark in computer vision research, containing 60,000 32x32 RGB images evenly distributed across 10 classes. It is extensively used to evaluate the performance of novel architectures and algorithms. This project aims to evaluate two state-of-the-art models, Microsoft's ResNet-152 and Google's EfficientNet-B2, to analyze their effectiveness in image classification tasks on CIFAR-10.

The motivation for this project is to compare the design philosophies of these two architectures—ResNet-152's deep residual learning and EfficientNet-B2's compound scaling—to gain insights into their relative strengths, weaknesses, and applicability to small-scale image classification tasks. This analysis will contribute to a better understanding of trade-offs between accuracy, computational efficiency, and parameter optimization.

This project will use the CIFAR-10 dataset from Hugging Face, and pretrained models for ResNet-152 and EfficientNet-B2, fine-tuning them for performance evaluation. It will explore the effects of data augmentation, model depth, and computational efficiency, providing actionable insights for real-world applications in image classification.

## Prior Work

The CIFAR-10 dataset has been a pivotal benchmark in computer vision, with many state-of-the-art models developed and evaluated using it:

- Krizhevsky and Hinton [2009] introduced CIFAR-10 and used simple CNNs to establish baseline performance metrics.
- He et al. [2016] proposed ResNet, which uses residual connections to train very deep networks. Variants like ResNet-20 and ResNet-110 were benchmarked on CIFAR-10, demonstrating significant improvements in accuracy.
- Zagoruyko and Komodakis [2016] developed Wide ResNet-28-10, which expanded the width of residual layers instead of increasing depth, achieving better results with reduced training time.

- Sabour et al. [2017] introduced Capsule Networks (CapsNet), replacing pooling layers with dynamic routing mechanisms, preserving spatial hierarchies, and achieving comparable results on CIFAR-10.
- Huang et al. [2017] proposed DenseNet-40, which connects each layer to every subsequent layer in a dense manner. DenseNet demonstrated excellent performance on CIFAR-10 while requiring fewer parameters than conventional CNNs.
- Dosovitskiy et al. [2021] introduced the Vision Transformer (ViT), including ViT-H/14, which achieved state-of-the-art accuracy of 99.5% on CIFAR-10 by leveraging self-attention mechanisms and large-scale pretraining.

Potential methods for achieving the project goals include data augmentation techniques like random cropping, flipping, and color jittering to improve model generalization, and fine-tuning pretrained ResNet-152 and EfficientNet-B2 models to adapt them for CIFAR-10 classification. Performance will be evaluated using metrics like accuracy, precision, recall, and alongside computational efficiency analysis.

# Preliminary Results

The CIFAR-10 dataset consists of 50,000 training images and 10,000 testing images across 10 classes, with balanced distribution. Each image has a resolution of 32x32 pixels, which introduces challenges such as capturing fine-grained patterns and reducing overfitting.

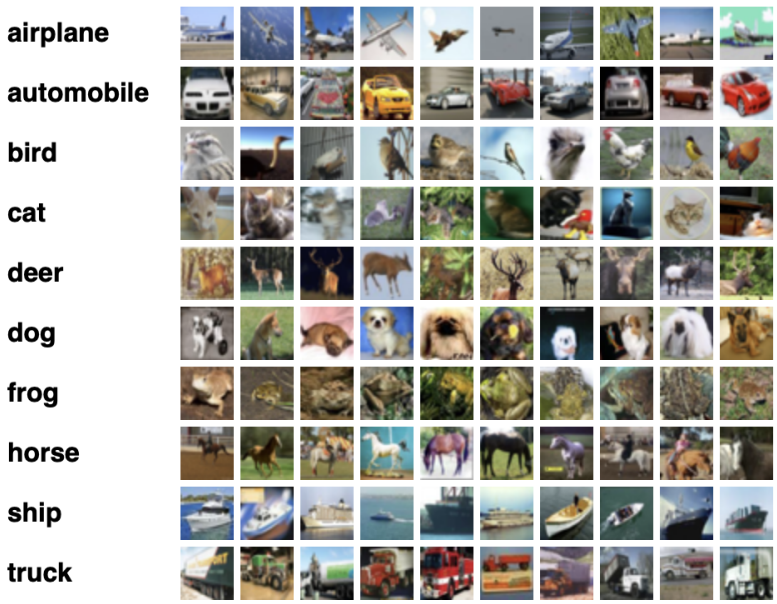


Figure 1: Sample images from the CIFAR-10 dataset.

Basic CNN experiments achieved 80% accuracy, revealing limitations in feature extraction and generalization for smaller images. Overfitting emerged as a bottleneck in deeper

models, necessitating effective augmentation techniques. Initial exploration involved augmentations like random cropping and flipping, which showed promise in improving training stability.

Preliminary experiments with lightweight ResNet variants suggested a trade-off between model depth and computational efficiency. Tools such as Hugging Face's `datasets` library were used for dataset handling, while PyTorch was utilized for initial model training and evaluation.

# Project Deliverables

The project will result in a modular Python repository, hosted on GitHub, with well-documented code for dataset preprocessing, model fine-tuning, and evaluation. A 2-page IEEE-style report will summarize the methodology, results, and insights, including visualizations of training curves, confusion matrices, and comparative performance metrics.

Sub-goals include developing a robust preprocessing pipeline with data augmentation, fine-tuning ResNet-152 and EfficientNet-B2 models for optimal accuracy, and analyzing their computational trade-offs in terms of memory usage and inference time. Training and validation results will be visualized to compare both models and draw conclusions about their applicability to real-world scenarios.

# Timeline

- **Week 1-2:** Literature review, dataset exploration, and implementation of preprocessing pipelines with augmentation. Train baseline models for initial insights.
- **Week 3-4:** Fine-tune ResNet-152 and EfficientNet-B2, optimize hyperparameters, and validate performance.
- **Week 5:** Analyze model results, prepare visualizations, and finalize the GitHub repository and IEEE-style report.

This project will provide insights into the performance and efficiency of ResNet-152 and EfficientNet-B2 on CIFAR-10, demonstrating their strengths and limitations while highlighting practical considerations for deploying these architectures.

# References

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

Alex Krizhevsky and Geoffrey Hinton. Learning multiple layers of features from tiny images. In *Technical Report*, 2009.

Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. Dynamic routing between capsules. In *Advances in neural information processing systems*, pages 3856–3866, 2017.

Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2016.