

# **Trabalho 2**

## **Tópicos Avançados em Computação**

**Guilherme Machado Rios - 11222839**

**São Carlos / Junho**

## 1. Resumo

Este relatório apresenta uma comparação detalhada do desempenho das redes neurais convolucionais ResNet e MobileNet no conjunto de dados CIFAR10. Para realizar a comparação, ambas as redes foram treinadas utilizando as mesmas condições de treinamento, incluindo hiperparâmetros e número de épocas de treinamento. Os resultados foram avaliados utilizando diversas métricas de desempenho, tais como acurácia de classificação, tempo de treinamento e consumo de memória.

## 2. Introdução

Este trabalho tem como objetivo realizar uma análise comparativa entre duas proeminentes arquiteturas de redes neurais convolucionais, a MobileNet e a ResNet. A MobileNet, projetada com foco em eficiência em dispositivos móveis, e a ResNet, desenvolvida para superar desafios de treinamento em redes profundas, são amplamente reconhecidas e utilizadas em tarefas de visão computacional. Neste estudo, serão exploradas as características, as vantagens e as diferenças dessas arquiteturas, além de uma análise comparativa de seus desempenhos em termos de acurácia.

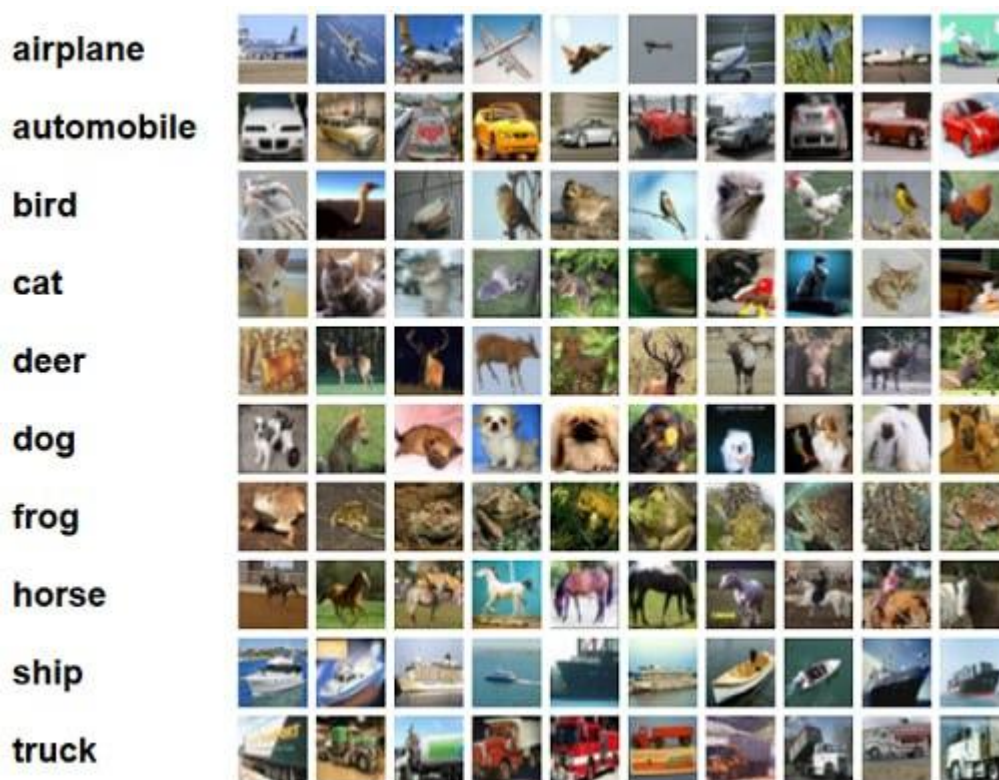
O código fonte do trabalho pode ser encontrado neste [repositório](#).

## 3. O dataset

CIFAR10 é um conjunto de dados que consiste em 60.000 imagens coloridas de 32x32 pixels em 10 classes, com 6.000 imagens por classe. O conjunto de dados é dividido em 50.000 imagens de treinamento e 10.000 imagens de teste. As 10 classes são:

- avião
- automóveis
- pássaro
- gato
- cervo
- cão
- sapo
- cavalo
- navio
- caminhão

Aqui estão alguns exemplos de imagens no conjunto de dados CIFAR10:



O conjunto de dados é comumente usado para tarefas de classificação de imagem e é um conjunto de dados de referência para testar o desempenho de vários algoritmos de aprendizado de máquina. É um conjunto de dados desafiador devido ao tamanho pequeno das imagens e à semelhança entre algumas das classes.

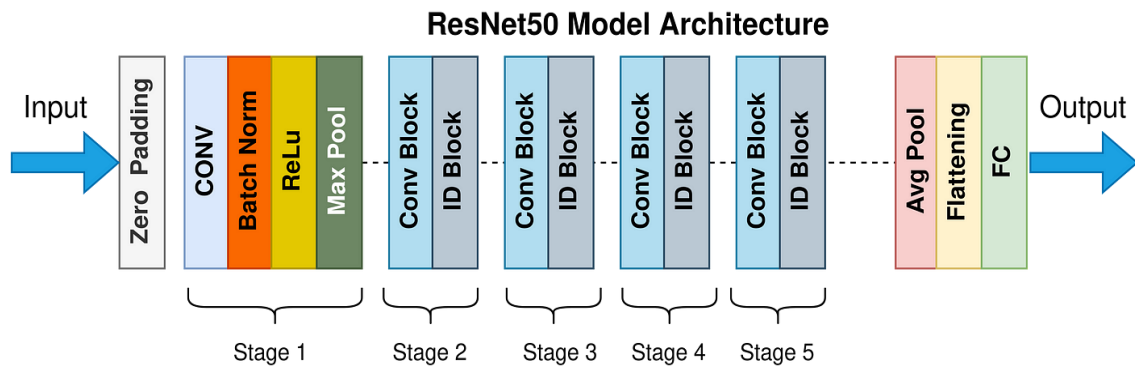
## 4. O modelos:

### 4.1. ResNet

A Rede Neural Residual (ResNet) é um modelo arquitetural de destaque proposto em 2015. Essa abordagem utiliza blocos residuais para aprender a diferença entre a entrada e a saída esperada, proporcionando uma solução eficaz para o problema do *vanishing gradient* com o aumento do número de camadas. Os blocos residuais são compostos por conexões de atalho que preservam as informações da entrada, permitindo sua transmissão para camadas subsequentes.

A arquitetura da ResNet consiste em camadas residuais empilhadas em combinação com camadas de pooling e convolucionais. Essa abordagem arquitetural revolucionária tem tido um impacto significativo no campo da visão computacional, viabilizando o treinamento eficiente de redes neurais profundas. A presença das conexões de atalho nos blocos residuais permite a preservação das informações cruciais e a mitigação dos problemas relacionados à propagação do gradiente durante o treinamento, resultando em um desempenho aprimorado e maior capacidade de aprendizado.

A abordagem de aprendizado residual superou desafios anteriores relacionados ao treinamento de redes profundas, abrindo novas perspectivas para avanços em diversas áreas, como reconhecimento de objetos, segmentação de imagens e análise de imagens médicas. Com base em seu impacto, é evidente que a ResNet representa um marco fundamental no campo da visão computacional, impulsionando o progresso e ampliando os limites do que é possível alcançar por meio do uso de redes neurais profundas.

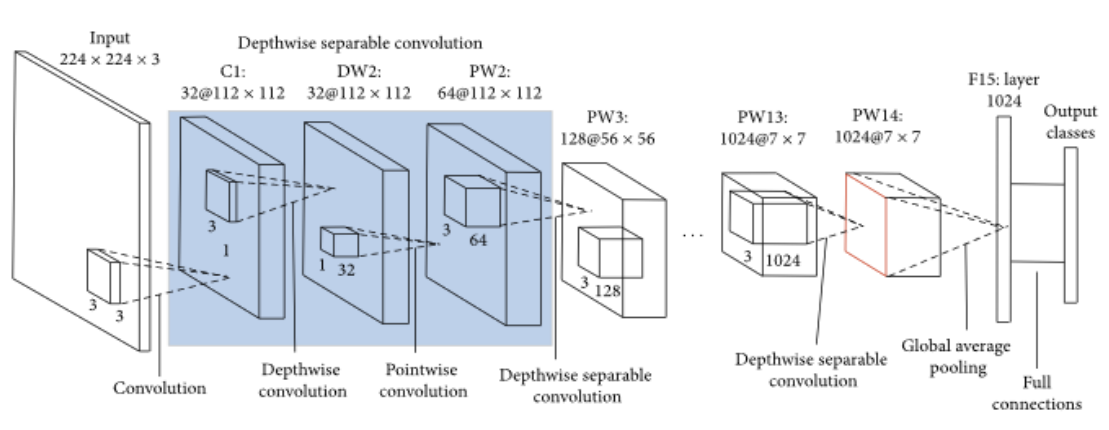


## 4.2. MobileNet

A MobileNet é uma arquitetura de rede neural convolucional desenvolvida especialmente para dispositivos móveis com recursos computacionais limitados. Seu objetivo é oferecer eficiência e bom desempenho em tarefas de visão computacional. Para alcançar isso, a MobileNet utiliza técnicas avançadas, como convoluções separáveis em profundidade e convoluções em camadas estreitas.

A técnica de convolução separável em profundidade divide a operação de convolução em duas etapas separadas, reduzindo a quantidade de parâmetros e operações computacionais necessárias. Isso permite que a MobileNet seja executada de forma eficiente em dispositivos móveis. Além disso, as convoluções em camadas estreitas utilizam filtros menores, diminuindo ainda mais o custo computacional, enquanto capturam características importantes das imagens.

Essas técnicas são aplicadas em blocos repetidos ao longo da arquitetura da MobileNet, resultando em um equilíbrio adequado entre eficiência e desempenho. Essa arquitetura é uma solução valiosa para a execução de tarefas de visão computacional em dispositivos móveis com recursos limitados.



## **5. Metodologia**

### **5.1. Dados**

Os dados foram obtidos a partir da própria biblioteca do PyTorch. Para serem usados nos modelos, são transformados em tensores do PyTorch e normalizados de acordo com o dataset.

### **5.2. Características dos Modelos**

#### **5.2.1. ResNet**

A arquitetura da ResNet é projetada para superar o desafio de treinar redes profundas com sucesso. A ResNet utiliza blocos residuais, que são a base fundamental dessa arquitetura e possibilitam a construção de redes extremamente profundas.

Cada bloco residual consiste em uma sequência de camadas convolucionais, onde a entrada original é adicionada ao resultado dessas camadas. Essa conexão direta, também conhecida como "skip connection", permite que as informações originais fluam inalteradas pelas camadas convolucionais. Essa estratégia é implementada por meio de uma operação de adição ou concatenação, combinando a saída do bloco com a entrada original.

Os blocos residuais podem ter diferentes configurações, mas geralmente são compostos por duas ou três camadas convolucionais. Os blocos básicos são compostos por duas convoluções com filtros de tamanho 3x3 e o mesmo número de canais de entrada e saída. Já os blocos bottlenecks, presentes nas versões mais profundas da ResNet, adicionam uma camada adicional com filtros de 1x1 para reduzir a dimensão dos canais intermediários, economizando recursos computacionais.

A ResNet também utiliza camadas de ativação, normalização e pooling, além das camadas convolucionais nos blocos residuais. Camadas de ativação, como ReLU (Rectified Linear Unit), são aplicadas para introduzir não-linearidade às saídas das camadas convolucionais. Camadas de normalização, como Batch Normalization, são adicionadas para estabilizar o treinamento da rede. Já as camadas de pooling, como max pooling, são usadas para reduzir a dimensionalidade espacial das características.

#### **5.2.2. MobileNet**

A arquitetura da MobileNet é composta por uma série de camadas convolucionais organizadas em blocos repetidos, projetados especificamente para atender às restrições de recursos computacionais em dispositivos móveis.

Cada bloco da MobileNet é composto por duas etapas principais: uma convolução em profundidade (depthwise convolution) e uma convolução em ponto (pointwise convolution). A convolução em profundidade é responsável por extrair características em um conjunto reduzido de canais, reduzindo a quantidade de parâmetros e o custo computacional. Em seguida, a convolução em ponto é aplicada para combinar essas características em um espaço dimensional mais amplo.

A arquitetura da MobileNet também permite ajustar a largura (width) dos blocos, controlando o número de canais nas camadas convolucionais. Esse ajuste influencia diretamente o número total de parâmetros da rede e o equilíbrio entre eficiência e desempenho.

### **5.3. Treinamento**

Ambos os modelos foram construídos em PyTorch utilizando-se Adam como otimizador e uma taxa de aprendizagem de 0.001 por 20 épocas.

## **6. Resultados**

Ao comparar os resultados obtidos pelas arquiteturas MobileNet e ResNet nessa, observa-se que ambas alcançam desempenhos muito bons. A MobileNet, projetada para dispositivos móveis com restrições de recursos computacionais, oferece uma solução eficiente, obtendo uma acurácia de aproximadamente 84.18%. Por outro lado, a ResNet, focada em superar o desafio de treinar redes profundas, atinge uma acurácia próxima mas ligeiramente inferior, chegando a cerca de 83.68%. Essa diferença mostra que a MobileNet pode ser uma escolha muito eficiente, ainda mais quando trazemos número de parâmetros para a discussão. Vale ressaltar que é esperado que a ResNet atinja resultados melhores ao treinarmos o modelo por mais épocas. A escolha entre essas arquiteturas dependerá das necessidades específicas do aplicativo e das restrições de recursos disponíveis.