

## UM ESTUDO COMPARATIVO ENTRE DIFERENTES TAMANHOS DE RESNET EM TERMOS DE EFICIÊNCIA DE CLASSIFICAÇÃO NO DATASET CIFAR-10

RAFAEL DUARTE DOS SANTOS<sup>1</sup>; RENAN MARCILI LEMONS<sup>2</sup>; RUHAN ÁVILA DA CONCEIÇÃO<sup>3</sup>

<sup>1</sup>Universidade Federal de Pelotas – rdsantos@inf.ufpel.edu.br

<sup>2</sup>Universidade Federal de Pelotas – rmlemons@inf.ufpel.edu.br

<sup>3</sup>Instituto Federal Sul-rio-grandense – ruhanconceicao@ifsul.edu.br

### 1. INTRODUÇÃO

A inteligência artificial (IA) e as redes neurais artificiais se tornaram cada vez mais populares com o passar dos dias. Primeiro, a construção de modelos cada vez mais sofisticados foi possível graças aos avanços tecnológicos que aumentaram o acesso ao processamento e à memória. Grandes quantidades de dados também estão prontamente disponíveis, principalmente pela internet e pelas mídias sociais, o que deu a esses modelos o combustível de que precisam para serem treinados.

O objetivo da inteligência artificial é criar máquinas capazes de realizar operações que exigem inteligência humana, como aprendizado, reconhecimento de padrões e tomada de decisão. Dentro da IA, o aprendizado de máquina é um subconjunto que permite que os sistemas aprendam com dados, aumentando seu desempenho ao longo do tempo. O aprendizado profundo vai um passo além e usa redes neurais artificiais multicamadas, ou redes neurais profundas, para modelar dados complexos e extrair padrões (SICHMAN, 2021).

As redes neurais artificiais, ou RNAs, são amplamente empregadas em uma ampla gama de setores, incluindo bancos, saúde, automação industrial e, mais recentemente, assistentes virtuais e carros autônomos. A classificação de imagens, na qual o objetivo é identificar com precisão o que uma imagem representa, como a diferença entre um cachorro e um gato, é um exemplo típico de uma aplicação de redes neurais artificiais (RNAs). Ao longo do tempo, várias redes neurais especializadas, como ResNet, VGGNet, InceptionNet e EfficientNet, foram desenvolvidas para essa tarefa (HE et al., 2015).

Essas redes são distintas umas das outras devido às suas características individuais (GOODFELLOW et al., 2016). Por exemplo, as conexões residuais foram apresentadas pela primeira vez pela ResNet. Devido à sua profundidade, redes como a ResNet têm melhor desempenho quando comparadas a outras arquiteturas para tarefas complicadas de categorização de imagens.

### 2. METODOLOGIA

Python é amplamente utilizado para projetos de ciência de dados e inteligência artificial devido à sua facilidade de uso e adaptabilidade, permitindo a rápida implementação de modelos de aprendizado de máquina e algoritmos. Dentro desse ambiente, a biblioteca PyTorch (PYTORCH, 2024), desenvolvida pela Facebook AI Research, destaca-se pela criação de redes neurais, oferecendo suporte nativo para GPU e CPU, o que acelera o treinamento de modelos em grandes conjuntos de dados. Sua principal característica é a construção de grafos computacionais dinâmicos, facilitando a criação e depuração

de modelos complexos, como CNNs (Redes Neurais Convolucionais) e RNNs (Redes Neurais Recorrentes).

O Google Colab (GOOGLE, 2024) é uma plataforma acessível que permite a execução de código Python diretamente na nuvem, facilitando o desenvolvimento sem a necessidade de configurar um ambiente local. Ele oferece a possibilidade de usar GPUs e TPUs para acelerar o treinamento de modelos de aprendizado profundo, essenciais para grandes volumes de dados e redes complexas. Além disso, a integração com o Google Drive facilita o gerenciamento de projetos, enquanto suas opções de hardware (CPU, GPU, TPU) permitem que os desenvolvedores escolham a configuração ideal para otimizar o desempenho e a produtividade.

Neste trabalho, utilizou-se um conjunto de dados (*dataset*) popular para aplicações de visão computacional. O CIFAR-10 tem 60.000 imagens coloridas de 32x32 pixels organizadas em 10 grupos diferentes, incluindo carros, pássaros, aviões e gatos. Para este estudo, os modelos ResNet são treinados e avaliados em tarefas de classificação de imagens usando o CIFAR-10. Ele atua como um *benchmark* para avaliar o desempenho do modelo durante o treinamento, permitindo a comparação de redes neurais com profundidades variadas em termos de precisão e eficiência (KRIZHEVSKY, 2012).

#### **Hiperparâmetros utilizados:**

- *Batch size* de 256: indica quantas amostras são manipuladas antes das configurações de rede serem atualizadas. Um tamanho de lote maior pode agilizar o treinamento e permitir um uso mais eficiente dos recursos do computador.
- *Learning rate* de 0,0005: O ritmo no qual o modelo modifica seus pesos em resposta ao erro observado é controlado por esse parâmetro. Esse tipo de valor moderado promove uma convergência mais estável.
- Número de Épocas de 50: É o número de vezes que o modelo é executado em todo o conjunto de dados de treinamento. Isso permite modificações iterativas de desempenho.

Entre os modelos empregados, está o ResNet 18 que é uma arquitetura com 18 camadas, mais leve e mais eficiente no reconhecimento de padrões, tornando-o perfeito para aplicativos que exigem menos recursos de processamento. Um modelo intermediário com 50 camadas que atinge uma mistura justa entre precisão e complexidade é chamado ResNet 50. O ResNet 152 conta com 152 camadas, este modelo está entre os mais profundos da família ResNet e é usado para extrair padrões intrincados de grandes conjuntos de dados. Esses modelos permitiram comparações entre vários níveis de profundidade e complexidade de rede avaliando o desempenho em tarefas de categorização de imagens.

### **3. RESULTADOS E DISCUSSÃO**

Conforme pode ser observado na Tabela 1, houve diferenças notáveis na precisão e duração do treinamento dos resultados obtidos com os vários modelos ResNet. O modelo ResNet 18 de 18 camadas levou 41 minutos para treinar e atingiu uma precisão de 80,84%. O modelo ResNet 50 de 50 camadas levou 112 minutos para treinar e atingiu uma precisão de 83,55%. O ResNet 152, o modelo mais profundo, com a melhor precisão de 85,94%, mas precisou de 262 minutos para treinar. Redes neurais mais profundas são capazes de registrar padrões

mais complexos, portanto, adicionar camadas a uma rede neural geralmente resulta em maior precisão. Mas benefícios relativos nem sempre são alcançados adicionando mais camadas. Por exemplo, comparado ao ResNet 18, o ResNet 50 tem 178% mais camadas, mas sua precisão aumentou em apenas 3%. No entanto, após adicionar 744% mais camadas, o ResNet 152 só melhorou a precisão em 6%. Isso demonstra que, à medida que uma rede se aprofunda, as vantagens de adicionar mais camadas diminuem nesta tarefa em específico.

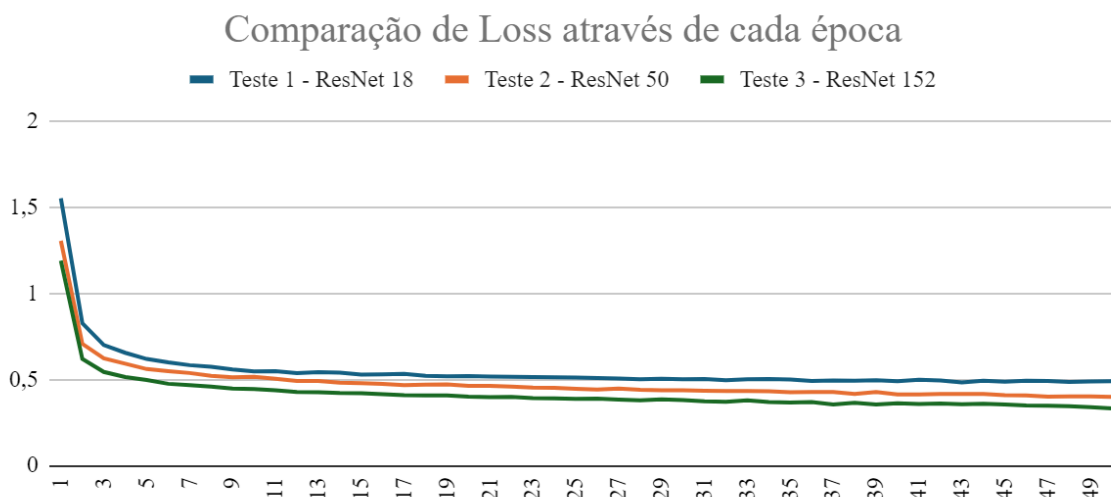
Tabela 1 - Comparativo entre os modelos

Teste	Acurácia	Camadas	Tempo de Treinamento (min)	Acréscimo de Camadas	Acréscimo de Acurácia
ResNet18	80,84%	18	41	-	-
ResNet50	83,55%	50	112	178%	3%
ResNet152	85,94%	152	262	744%	6%

Ao comparar o erro (*Loss*) dos vários modelos, todos eles mostraram uma redução considerável ao longo do treinamento. Como pode ser observado na Figura 1, sendo o maior *loss* exibido primeiro pelo modelo ResNet 18, indicando um desempenho inicial abaixo da média. No entanto, houve um declínio perceptível durante o treinamento, mostrando que, embora a rede fosse mais simples, ela ainda conseguia aprender efetivamente. O ResNet 50 demonstrou uma queda consideravelmente mais perceptível em sua Perda inicial em comparação ao ResNet 18, indicando maior eficácia na modificação de seus parâmetros para minimizar erros.

Por outro lado, dos três, o modelo ResNet 152, pré-treinado, apresentou a menor perda inicial, o que indica uma forte capacidade de aprendizado imediato, mesmo com o treinamento realizado apenas na última camada. Ele demonstrou o menor valor de perda final na conclusão do treinamento, demonstrando que sua profundidade permite registrar com mais precisão padrões mais complexos. Ao comparar os modelos, o ResNet 50 e o ResNet 152 são particularmente bons na redução de perda; no entanto, o ResNet 152 tem um custo de treinamento maior e exibe o melhor desempenho geral.

Figura 1 - Comparativo entre os modelos



#### 4. CONCLUSÕES

Com base nos testes conduzidos, selecionar um modelo de rede neural requer um compromisso entre o grau de precisão necessário e o poder de processamento disponível. O modelo mais leve, ResNet 18, tem o menor tempo de treinamento, mas sua precisão e redução de perdas não são tão boas quanto as dos modelos mais profundos. No entanto, com um tempo de treinamento mais apropriado, o ResNet 50 provou ser uma opção intermediária eficaz, atingindo boa precisão e uma grande redução na perda. Embora o ResNet 152 mostre a melhor precisão e a menor perda final, ele necessita de um tempo de treinamento substancialmente maior.

Em conclusão, pode-se dizer que o modelo ResNet 50 é a melhor escolha para aplicações que exigem uma compensação entre precisão e economia de tempo. No entanto, o ResNet 152 é a melhor opção quando a máxima precisão é necessária e os recursos de processamento não são uma barreira.

#### 5. REFERÊNCIAS BIBLIOGRÁFICAS

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. Deep Learning. Cambridge: MIT Press, 2016. Disponível em: <http://www.deeplearningbook.org>. Acessado em: 29 set. 2024.

GOOGLE. Conheça o Colab. Acessado em: 03 out. 2024. Disponível em: <https://colab.research.google.com/>.

HE, Kaiming; ZHANG, Xiangyu; REN, Shaoqing; SUN, Jian. Deep Residual Learning for Image Recognition. *arXiv*, 2015. Disponível em: <https://arxiv.org/abs/1512.03385>. Acessado em: 29 set. 2024.

KRIZHEVSKY, Alex. Learning Multiple Layers of Features from Tiny Images. *University of Toronto*, Toronto, 2012.

PYTORCH. Acessado em: 03 out 2024. Disponível em: <https://pytorch.org/>.

SICHMAN, Jaime Simão. Inteligência Artificial e sociedade: avanços e riscos. *Estudos Avançados*, São Paulo, v.35, n.101, p.37–50, jan. 2021. Disponível em: <https://doi.org/10.1590/s0103-4014.2021.35101.004>. Acessado em: 01 out. 2024.