



Novo
Ensino
Suplementar

Classificação de Imagens no CIFAR-10 com CNN e Grad-CAM: Uma Abordagem Interativa em Streamlit

Mariana Lins
UFAL – NES Turma 2025.4
marianalins901@gmail.com

September 28, 2025

Abstract

Este trabalho apresenta o desenvolvimento de redes neurais para classificação de imagens no dataset CIFAR-10 utilizando apenas um canal de cor. Foram implementados e comparados um modelo MLP e uma CNN, com técnicas de regularização (Dropout e L2). Avaliou-se o desempenho em métricas de acurácia e matriz de confusão, além da aplicação de Grad-CAM para interpretabilidade. Também foi criada uma interface interativa em Streamlit, permitindo upload de imagens externas e visualização de predições. Os resultados indicam que a CNN supera o MLP em desempenho, validando a hipótese de que convoluções extraem melhor os padrões espaciais.

1 Introdução

As redes neurais convolucionais (CNNs) são amplamente aplicadas em tarefas de visão computacional, destacando-se pela capacidade de capturar padrões espaciais. O CIFAR-10 é um dataset clássico com 60.000 imagens de 32x32 pixels em 10 classes balanceadas. Neste trabalho, analisamos o impacto de usar apenas um canal de cor e comparamos o desempenho de um MLP e de uma CNN.

2 Pergunta e Hipótese

Pergunta: a CNN supera o MLP na classificação do CIFAR-10 com apenas um canal de entrada? **Hipótese:** sim, pois a CNN pode explorar relações espaciais que o MLP ignora.

3 Descrição do Dataset

O CIFAR-10 contém 50.000 imagens de treino e 10.000 de teste, divididas em 10 classes. Normalizamos os pixels para o intervalo $[0,1]$ e usamos apenas o canal vermelho.



Figure 1: Exemplos de imagens do CIFAR-10 em escala de cinza (canal R).

4 Metodologia

4.1 Modelos

O MLP possui camadas densas (256, 128) com ReLU, Dropout e saída softmax. A CNN contém dois blocos Conv+MaxPool seguidos de uma camada densa (128 neurônios) e saída softmax.

4.2 Treinamento

Ambos os modelos foram treinados em TensorFlow/Keras com otimizador Adam, loss *categorical_crossentropy*, *EarlyStopping* e *ModelCheckpoint*. O treino foi repetido com três seeds (42, 123, 999).

4.3 Interface Interativa

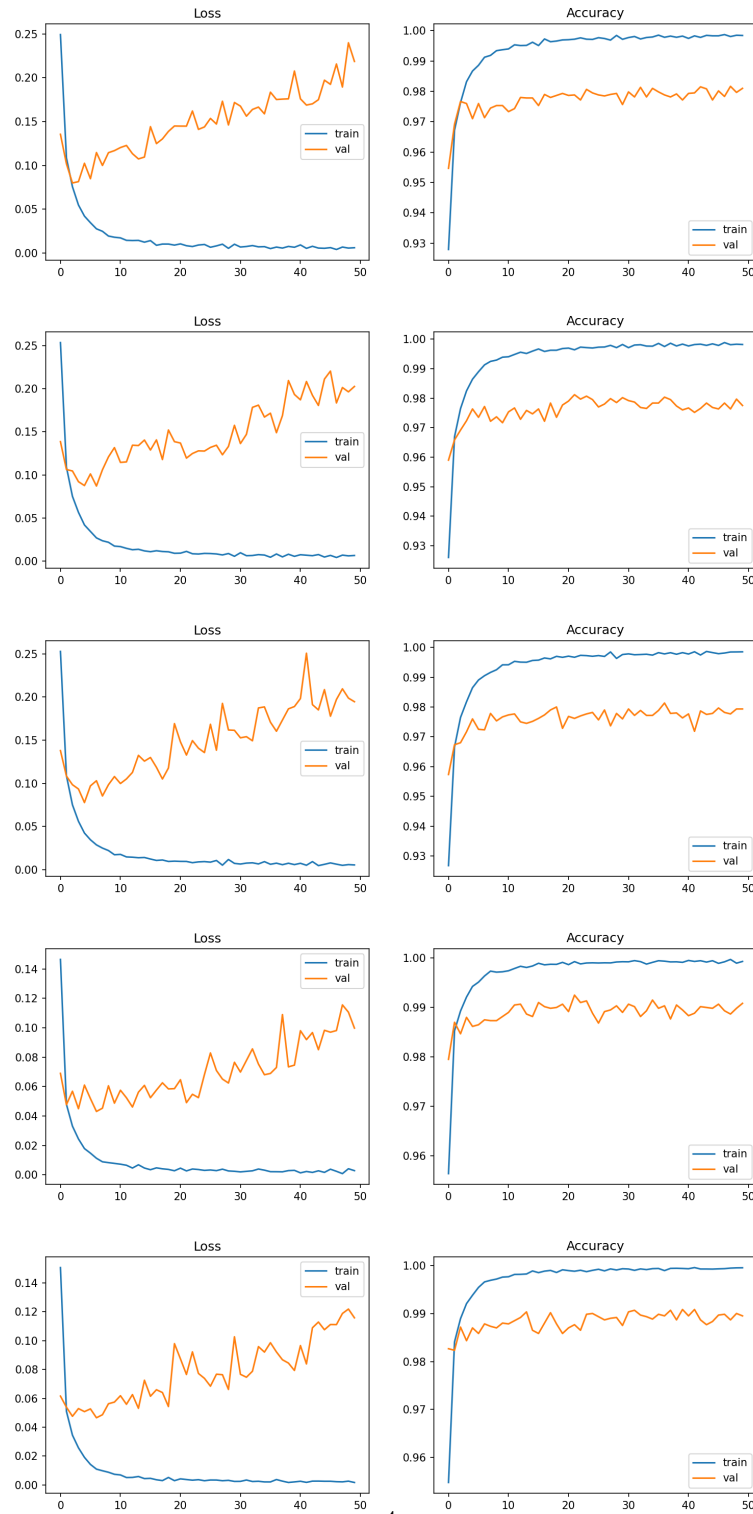
Foi desenvolvida uma aplicação em Streamlit que permite upload de imagens, exibe probabilidades por classe e mostra Grad-CAM.

O MLP possui aproximadamente 150 mil parâmetros e a CNN cerca de 545 mil, ambos bem abaixo do limite de 1 milhão definido no enunciado.

5 Resultados

5.1 Curvas de Treinamento

A Figura 2 apresenta as curvas de treino e validação (acurácia e perda) para os modelos MLP e CNN. Nota-se que a CNN converge mais rapidamente e mantém melhor desempenho na validação.



4

Figure 2: Curvas de treino e validação para MLP (esquerda) e CNN (direita).

5.2 Visualização Grad-CAM

A Figura 3 mostra exemplos de mapas de ativação gerados pelo Grad-CAM para a CNN. Nota-se que a rede concentra sua atenção nas regiões mais discriminativas da imagem.

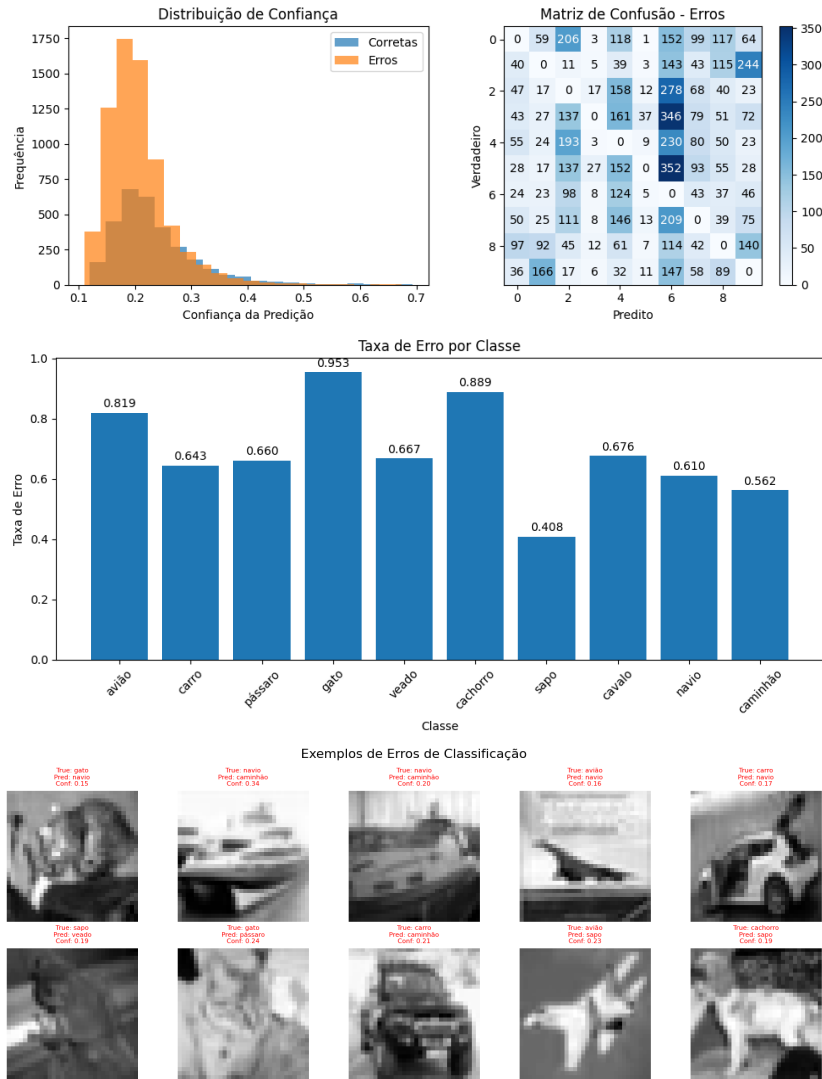


Figure 3: Exemplos de visualizações Grad-CAM para três classes do CIFAR-10.

A matriz de confusão mostra que a CNN obteve desempenho superior ao MLP:

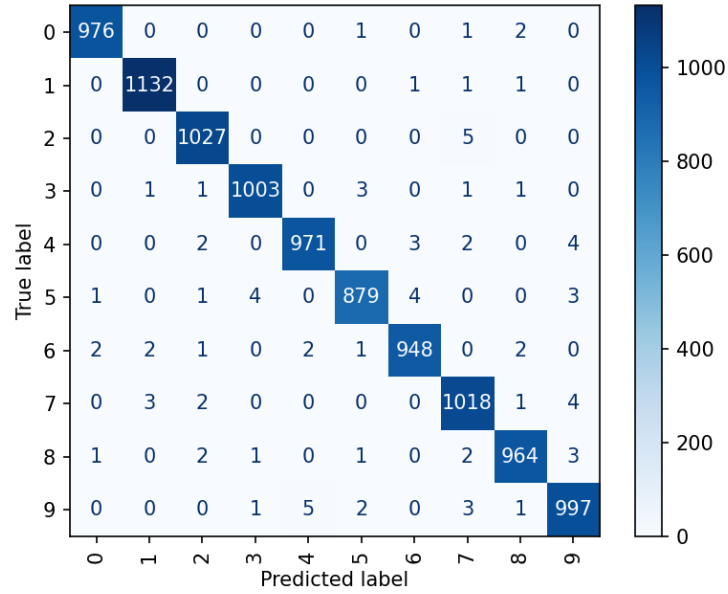


Figure 4: Matriz de confusão para a CNN.

Também foram analisados erros mais frequentes, como confusões entre classes visuais semelhantes:

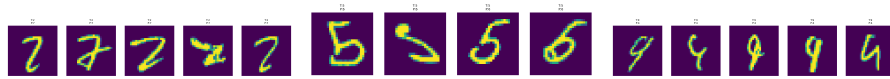


Figure 5: Exemplos de erros da CNN: (esquerda) 2 vs 7, (meio) 5 vs 6, (direita) 9 vs 4.

5.3 Tabela Comparativa

A Tabela 1 resume os resultados de acurácia média obtidos pelos modelos.

model	seed	acc	f1
MLP	42	0.981	0.9808789966722863
CNN	42	0.9904	0.9902577433529576
MLP	123	0.9774	0.9771796332755404
CNN	123	0.991	0.9909445043486886
MLP	999	0.9806	0.9804345606853804
CNN	999	0.9915	0.9914106674641365

Table 1: Comparação de desempenho entre MLP e CNN.

6 Conclusão

Os experimentos confirmaram que a CNN supera o MLP em desempenho, mesmo utilizando apenas um canal do CIFAR-10. Além disso, a interface em Streamlit e o Grad-CAM contribuíram para maior interpretabilidade. Como próximos passos, sugere-se testar outros canais (G, B), redes mais profundas e técnicas de data augmentation.