Classificação de imagens utilizando Redes Neurais Convolucionais

Murillo Freitas Bouzon

***Resumo*—Redes Neurais Artificias é um conceito que surgiu já há algum tempo, porém é uma área que atualmente está em constante crescimento graças a popularização do *Deep Learning*. Um dos modelos de *Deep Learning* muito utilizado é o modelo de Redes Neurais Convolucionais, onde aplica-se convoluções ao invés de matrizes de pesos entre as camadas escondidas. Esse modelo tem sido utilizado recentemente em diversas aplicações na literatura, principalmente para a classificação de imagens. Sendo assim, neste trabalho foi utilizado Redes Neurais Convolucionais para a classificação de imagens, implementando o modelo LeNet para classificação da base MNIST e utilizando modelos pré-treinados para a classificação de imagens de base ImageNet. Os resultados mostraram que o modelo implementado da LeNet obteve um valor de precisão de aproximadamente 0.97, dependendo de quantas épocas forem utilizadas para o treinamento.**

***Index Terms*—Redes Neurais Convolucionais, Machine Learning, Classificação de Imagens, LeNet**

I. INTRODUÇÃO

O uso de Redes Neurais Artificiais se tornou comum para resolver problemas de classificação e regressão, utilizando uma base de dados para o treinamento da rede para ela aprender como classificar os dados de forma corretamente.

Para esses tipos de problemas, a descrição da solução é bem definida, tornando-fácil a modelagem da resolução desse problema para uma Rede Neural. Porém, o verdadeiro desafio aparece para problemas que são fáceis de resolver pelos humano mas difíceis de serem descritos formalmente, como reconhecimento e fala ou imagens.

Desta forma, o *Deep Learning* surgiu como solução para esses problemas, permitindo o aprendizado e o entendimento da solução de forma hierárquica, onde cada camada da rede representa uma parte da solução total. Assim, a Rede Neural aprende ao juntar conhecimento de cada camada, permitindo que o computador aprenda conceitos complexos juntando conceitos mais simples aprendidos em camadas anteriores.

Devido ao fato das redes tradicionais, como MLP, não convergirem quando possuem muitas camadas, surgiram novos modelos de redes neurais artificiais, sendo um dos modelos mais famosos as Redes Neurais Convolucionais, baseado na organização dos neurônios no córtex visual animal, sendo utilizado para diversas aplicações como reconhecimento facial e reconhecimento de objetos.

As Redes Neurais Convolucionais se tornaram bastante populares para resolver diversos problemas não-intuitivos. Um exemplo é o trabalho de [1], onde foi utilizado modelos tradicionais de CNNs (*Convolutional Neural Network*) para classificação de trilha sonora, utilizando modelos de *Deep Learning* pré-treinados. Foi concluído que os modelos utilizados obtiveram bons resultados na tarefa de classificação de áudio, apontando que grandes bases de dados colaboram com isso.

Outro trabalho que também utilizou CNNs foi o [2], porém aqui ela foi utilizada para classificação de objetos das bases NORB e CIFAR10, e reconhecimento de números escritos a mão da base de dados MNIST. Foram obtidos como resultado uma taxa de erro de $2.53\%, 19.51\% e 0.35\%$.

Dada a importância e a popularização das Redes Neurais Convolucionais no cenário atual, este trabalho tem como objetivo a implementação do modelo de Rede Neural Convolucional LeNet para classificação dos dígitos da base de dados MNIST e utilizar modelos de redes pré-treinados para classificação de imagens da base Image-Net.

II. CONCEITOS FUNDAMENTAIS

Nesta seção serão apresentadas as teorias relacionadas ao método desenvolvido neste trabalho.

A. Redes Neurais Convolucionais

As Redes Neurais Convolucionais funcionam da mesma forma que as Redes Neurais Artificiais, porém a multiplicação pela matriz de pesos é substituída por uma convolução em pelo menos uma das camadas, o que permite que a rede processe imagens e vídeos de grandes tamanhos.

O primeiro modelo de CNN foi proposto em [3], chamado de LeNet-1. A arquitetura dessa rede é composta de camadas esparsas, convolucionais e de *max-pooling*. As camadas mais baixas são compostas por camadas convolucionais e camadas de *max-pooling* alternadas, enquanto as camadas mais altas são redes MLP tradicionais. A Figura 1 apresenta um exemplo da LeNet com uma imagem de entrada $32x32$, utilizada na classificação da base de dados MNIST.

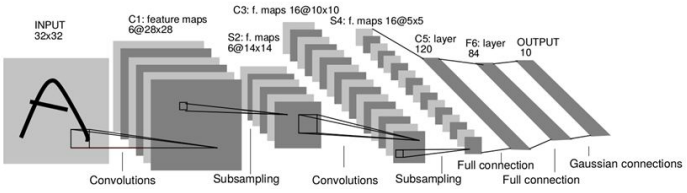


Figura 1: Ilustração da Rede Neural Convolucional LeNet.

III. METODOLOGIA

Para a metodologia deste trabalho, foi implementada a rede LeNet-1 para a classificação de imagens da base MNIST. Para isso, foi utilizada a biblioteca Keras [4], uma biblioteca para fácil implementaçãode Redes Neurais.

Desta forma, criou-se uma rede seguindo a arquitetura mostrada na Figura 1, onde a primeira camada é uma camada

convolucional, com um *kernel* $5x5$ e o número de filtros $= 6$. A segunda camada é uma camada de *max-pooling* com tamanho de *pool* $= 2$. A terceira camada é uma outra camada convolucional com *kernel* $5x5$ e 16 filtros seguido de uma camada de *max-pooling* de tamanho 2. As últimas 3 camadas são MLP tradicionais com 120, 84 e 10 neurônios respectivamente.

Além disso, utilizou-se modelos pré-treinados para classificar imagens da ImageNet. Os modelos utilizados foram VGG16 [5], VGG19 [6] e ResNet50 [7].

A implementação foi feita em *Python*, com o auxílio da biblioteca numpy.

IV. BASE DE DADOS

Para validar se a implementação foi feita corretamente, foram realizados experimentos em duas bases de dados, sendo descritas a seguir:

A. MNIST Dataset

A base de dados MNIST é uma base de dígitos manuscritos possuindo 60000 amostras para treinamento e 10000 observações para teste, sendo apresentada em [8]. Esta base se tornou bastante popular na comunidade científica, sendo utilizada em teste para técnicas de aprendizado e métodos de reconhecimento de padrões sem esforços adicionais para o pré-processamento da base. A Figura 2 mostra um exemplo da observação dessa base.

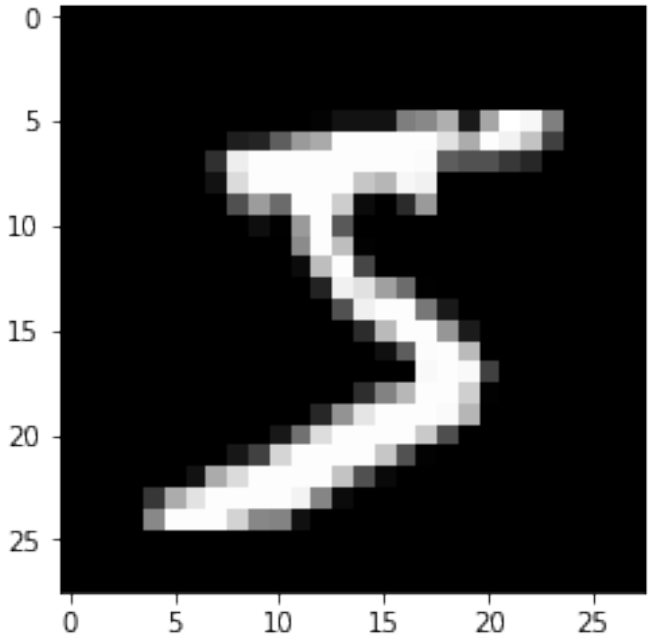


Figura 2: Exemplo do número 5 manuscrito da base MNIST.

B. ImageNet

A base de dados ImageNet é uma base de imagens de objetos variados apresentada em [9], possuindo um total de $14,197,122$ imagens de 28 categorias. Para este trabalho, foram utilizadas 3 classes dessa base, sendo elas a classe *laranja*, *tigre* e *cerveja*. A Figura 3 apresenta exemplos das classes selecionadas dessa base.

(a) Exemplo da classe laranja.

(b) Exemplo da classe tigre.

(c) Exemplo da classe cerveja.

Figura 3: Exemplos de imagens da ImageNet.

V. EXPERIMENTOS E RESULTADOS

Para avaliar o método implementado, realizou-se dois experimentos. No primeiro experimento foi testado a rede LeNet implementada para a classificação dos dígitos da base de dados

MNIST, apresentada na Seção IV-A. Para isso, foi variado o número de épocas de treinamento entre $5, 10e15$, calculando as métricas *precision, recall e Fscore* para cada época. Os resultados são apresentado na Tabela I, mostrando que com 15 épocas foi obtido o valor de precisão mais alta. Os valores de *Precision, Recall e Fscore* se mantiveram os mesmo devido ao número de *Falsos Positivos* ser igual ao de *Falsos Negativos*.

Épocas	Precision	Recall	Fscore
5	0.9736	0.9736	0.9736
10	0.9744	0.9744	0.9744
15	0.9756	0.9756	0.9756

Tabela I: Resultados do primeiro experimento.

Para o segundo experimento, utilizou-se as redes pré-treinadas VGG16, VGG19 e ResNet50 para classificar imagens da base ImageNet, apresentada na Seção IV-B. Foram selecionadas 3 classes, sendo elas *laranja*; *tigre* e *cerveja*. Utilizou-se 10 imagens ao todo, sendo 5 da ImageNet e as outras 5 de domínio público. A Figura 4 apresenta os resultados desse experimento. Os resultados mostram que o modelo VGG19 foi o pior dos três modelos testados, obtendo uma acurácia média de 93.3% de todas as classes.

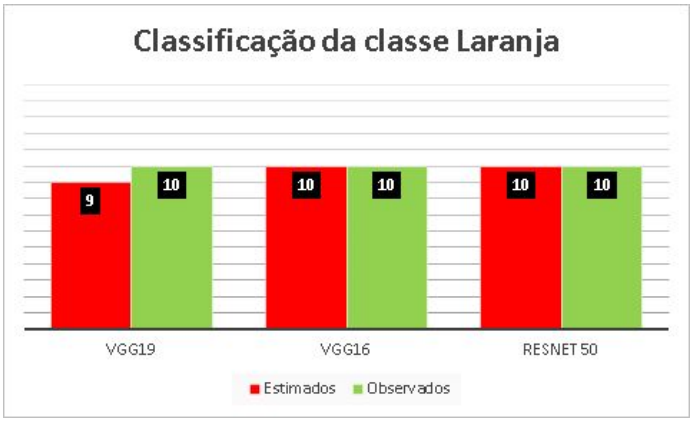
VI. CONCLUSÃO

Neste trabalho foi utilizado Redes Neurais Convolucionais para classificação de imagens, utilizando a biblioteca Keras para implementar o modelo LeNet e para utilizar os modelos pré-treinados VGG16, VGG19 e ResNet50.

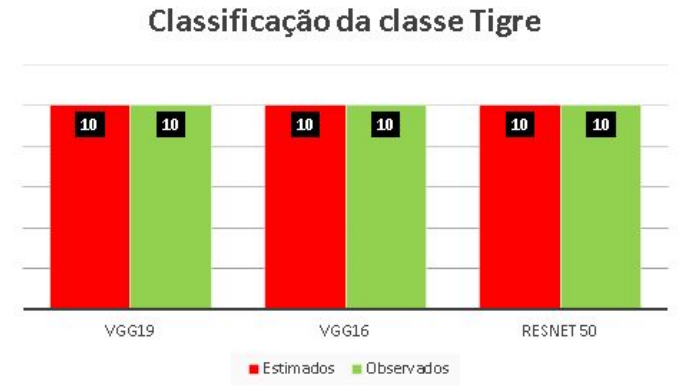
Para a avaliação, foram feitos dois experimento. O primeiro experimento mediu a precisão da LeNet ao classificar os dígitos manuscritos da base MNIST. Os resultados mostraram que foi possível obter uma precisão de até 0.9756, utilizando 15 épocas para o treinamento.

Para o segundo experimento, utilizou-se os modelos pré-treinados para classificar as classes de imagens *laranja*, *tigre* e *cerveja*. Os resultados dos modelos VGG16 e ResNet50 foram de uma acurácia de aproximadamente 97% e para o modelo VGG19 de 93.3%, obtendo uma boa taxa de acerto para a classificação de imagens.

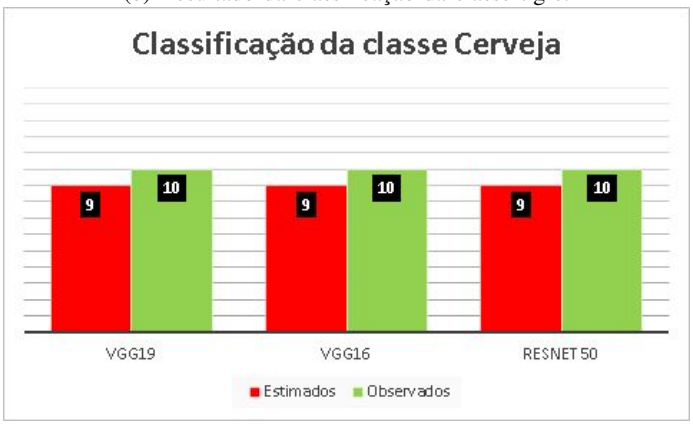
As Redes Neurais Convolucionais mostraram ser uma técnica eficaz na classificação de imagens multi-classes, obtendo altas taxas de acertos, mostrando o porque de ser uma técnica tão popular na literatura, sendo de fácil implementação e resultados no estado-da-arte.



(a) Resultado da classificação da classe laranja.



(b) Resultado da classificação da classe tigre.



(c) Resultado da classificação da classe cerveja.

Figura 4: Resultados do segundo experimento.

REFERÊNCIAS

[1] Shawn Hershey, Sourish Chaudhuri, Daniel P. W. Ellis, Jort F. Gemmeke, Aren Jansen, R. Channing Moore, Manoj Plakal, Devin Platt, Rif A. Saurous, Bryan Seybold, Malcolm Slaney, Ron J. Weiss, and Kevin W. Wilson. Cnn architectures for large-scale audio classification. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 131–135, 2016.

[2] Dan C. Ciresan, Ueli Meier, Jonathan Masci, Luca Maria Gambardella, and Jürgen Schmidhuber. Flexible, high performance convolutional neural networks for image classification. In *IJCAI*, 2011.

[3] Yann LeCun, Bernhard E. Boser, John S. Denker, Donnie Henderson, Richard E. Howard, Wayne E. Hubbard, and Lawrence D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1:541–551, 1989.

[4] François Chollet et al. Keras. https://keras.io, 2015.

[5] Xiangyu Zhang, Jianhua Zou, Kaiming He, and Jian Sun. Accelerating very deep convolutional networks for classification and detection. *CoRR*, abs/1505.06798, 2015.

[6] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015.

[8] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. 1998.

[9] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.