

SCC0910 - Tópicos Avançados em Ciências de Computação I

Trabalho 2

Alunos:

Mateus de Souza Santos - 11366913

Pedro Henrique Dias Junqueira de Souza - 11294312

Wellington Matos Amaral - 11315054

Resumo

O presente trabalho teve como objetivo construir um classificador binário de imagens para detecção de fogo em imagens, tendo como foco imagens de florestas, para isso utilizando Redes Neurais Convolucionais (CNNs). Para isso foi inicialmente implementada uma rede convolucional rasa de 12 camadas sequenciais (4 convoluções, 2 *max pooling* e 1 camada densa como *output*) e o otimizador SGD (*stochastic gradient descent*) com loss *binary cross entropy*, treinada com o *dataset* “Forest Fire Images” para o desenvolvimento do modelo. A acurácia alcançada pelo modelo foi de 88% no conjunto de testes demonstrando com margem para aprimoramento, que foi refinada com o uso de Fine Tuning na rede Resnet50, obtendo acurácia de 100% no conjunto de dados de teste utilizado (o mesmo do primeiro modelo). Esse modelo pode ter o potencial para ser utilizado no monitoramento de incêndios em áreas florestais.

Introdução

Este trabalho tem por objetivo a construção de um classificador de imagens fazendo o uso de Redes Neurais Convolucionais (CNNs) que seja capaz de identificar a presença ou ausência de fogo em imagens, com foco especificamente em imagens de florestas.

Se alcançada uma acurácia significativa, o modelo poderia ser utilizado na construção de sistemas de monitoramento de incêndios florestais baseados em câmeras, que podem ser implementados com um custo relativamente baixo e grande facilidade, sem a necessidade de hardware específico, fazendo o uso apenas de câmeras que possam enviar fotos a um servidor onde o modelo classifica as imagens.

O código-fonte do primeiro modelo construído pode ser acessado através [deste link](#), e, o do segundo, [neste link](#).

Resultados Esperados

Conforme descrito na seção de introdução, o objetivo do projeto é desenvolver um classificador binário para identificar incêndios florestais. Para tal, foi escolhido o dataset [Forest Fire Images](#), disponível publicamente na plataforma Kaggle. Este conjunto de dados é composto por 5000 imagens, sendo 4000 imagens para treinamento, 950 para validação e 50 para testes. Alguns exemplos estão ilustrados a seguir:



Imagem 1: Exemplos de entrada

Tanto os dados de treinamento quanto os dados de validação e teste possuem uma boa distribuição das classes do problema (presença ou ausência de fogo), com exatamente 50% do dataset sendo ocupado por cada classe.

Este dataset é bastante simples e não oferece grande dificuldade para classificação.

Arquitetura

Arquitetura do Primeiro Modelo

No primeiro experimento, foi implementada uma rede convolucional rasa, definida conforme a imagem 2, a seguir:

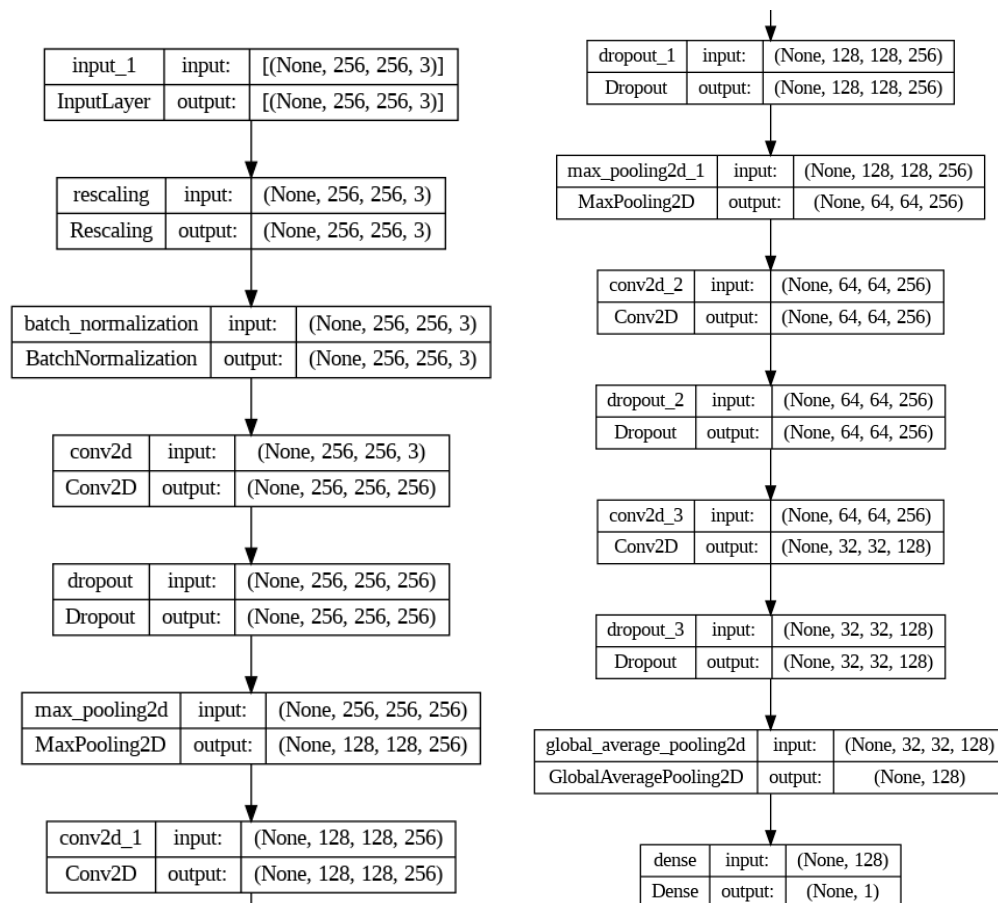


Imagem 2: Arquitetura do primeiro modelo

As imagens do dataset possuem resolução de 1920x1080, desta forma foi necessário adicionar à rede a formatação das entradas antes do processamento. Inicialmente, as imagens foram redimensionadas para atingirem a resolução de 256x256, desta forma, ainda que haja perda de informações, o tamanho da entrada será menor e assim exige menos processamento. Após o redimensionamento a entrada é normalizada. E então é iniciado o processamento.

Após alguns experimentos, o grupo optou por implantar a seguinte arquitetura: 12 camadas sequenciais, sendo elas: 4 convoluções, 2 max pooling e uma camada densa como output.

O otimizador escolhido foi o SGD (*stochastic gradient descent*), contudo poderia ser outro, como por exemplo Adam. Visto que esse é um problema de classificação binária, foi utilizada a loss é *binary cross entropy*, foram aplicados *dropout* para evitar *gradiente vanishing* ou *exploding*, nas convoluções, optou-se por utilizar relu como função de ativação, pois é uma das mais utilizadas atualmente.

Arquitetura do Segundo Modelo

No modelo mais recente, foi implementado um fine tuning da arquitetura Resnet50, descrita na imagem a seguir:

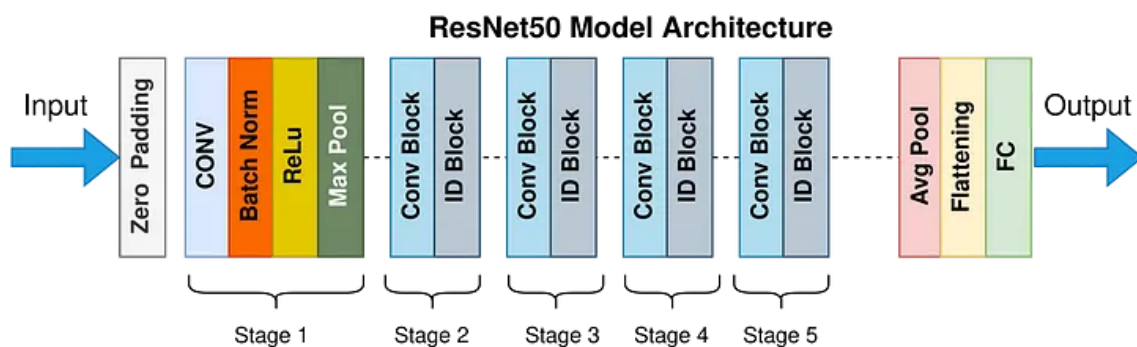


Imagem 3: Arquitetura da rede Resnet50. Disponível em <https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758>

A Resnet50 é uma rede bastante profunda, contando com 50 camadas de convolução. O seu principal elemento é o uso de conexões residuais ou *skip connections*, que viabilizam a passagem de informações de uma camada para as subsequentes sem perda de informação, o que é uma abordagem poderosa para solucionar o problema da perda de desempenho causada pela adição de muitas camadas a uma rede.

A estrutura da rede é dividida em blocos, com cada bloco contendo várias camadas de convolução, normalização e ativação. Além desses blocos, as últimas camadas da rede são para *pooling* e classificação.

O fine tuning feito utiliza inicialmente a Resnet50 congelada, mantendo seus pesos pré treinados, e, após isso, são adicionadas outras camadas, sendo uma de *pooling*, reduzindo a dimensionalidade, algumas camadas densas, uma camada de *dropout* para evitar *overfitting* e uma camada densa de saída

A imagem a seguir ilustra a arquitetura final da rede:

Model: "FeatureExtractionModel"

Layer (type)	Output Shape	Param #
inputLayer (InputLayer)	[(None, 256, 256, 3)]	0
resnet50 (Functional)	(None, None, None, 2048)	23587712
poolingLayer (GlobalAverage Pooling2D)	(None, 2048)	0
dense_15 (Dense)	(None, 1024)	2098176
dense_16 (Dense)	(None, 512)	524800
dropout_5 (Dropout)	(None, 512)	0
dense_17 (Dense)	(None, 256)	131328
outputLayer (Dense)	(None, 1)	257
activationLayer (Activation)	(None, 1)	0
Total params: 26,342,273		
Trainable params: 2,754,561		
Non-trainable params: 23,587,712		

Imagem 4: Camadas da rede treinada

Resultados

Resultados do primeiro modelo

A arquitetura implementada alcançou uma acurácia de 88% no conjunto de testes, desta forma, ainda que para um classificador binário não seja um resultado ideal, o grupo estava bastante otimista em relação ao refinamento da rede, pois, uma arquitetura bastante simples alcançou resultados bastante satisfatórios.

A seguir temos os gráficos de perda e acurácia durante o treinamento:

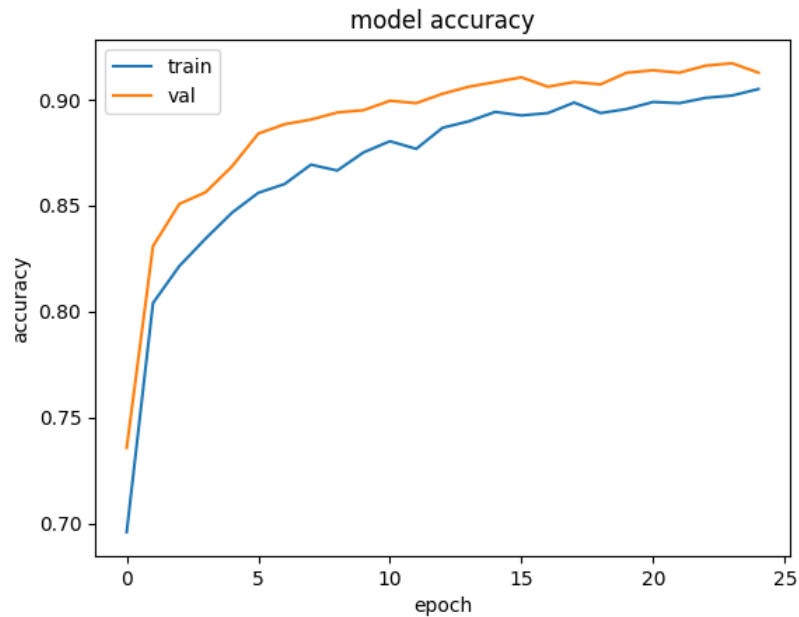


Imagem 3: Resultados dos treinamentos com primeiro modelo

Resultado do segundo modelo

A nova arquitetura implementada obteve resultados melhores ao atingir uma acurácia de 100%, superando o baseline inicial. O resultado é satisfatório e mostra que o modelo é promissor para a utilização em detecção de incêndios florestais, ainda que seja necessário mais estudos para confirmar se há robusta qualidade no modelo.

A seguir tem os gráficos de acurácia e *loss* do modelo:

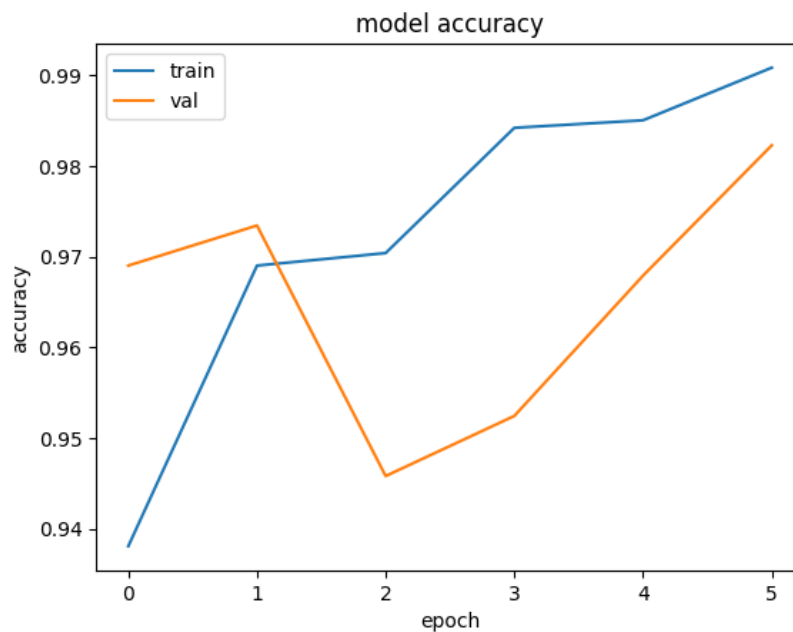


Imagem 4: Resultado dos treinamentos com segundo modelo

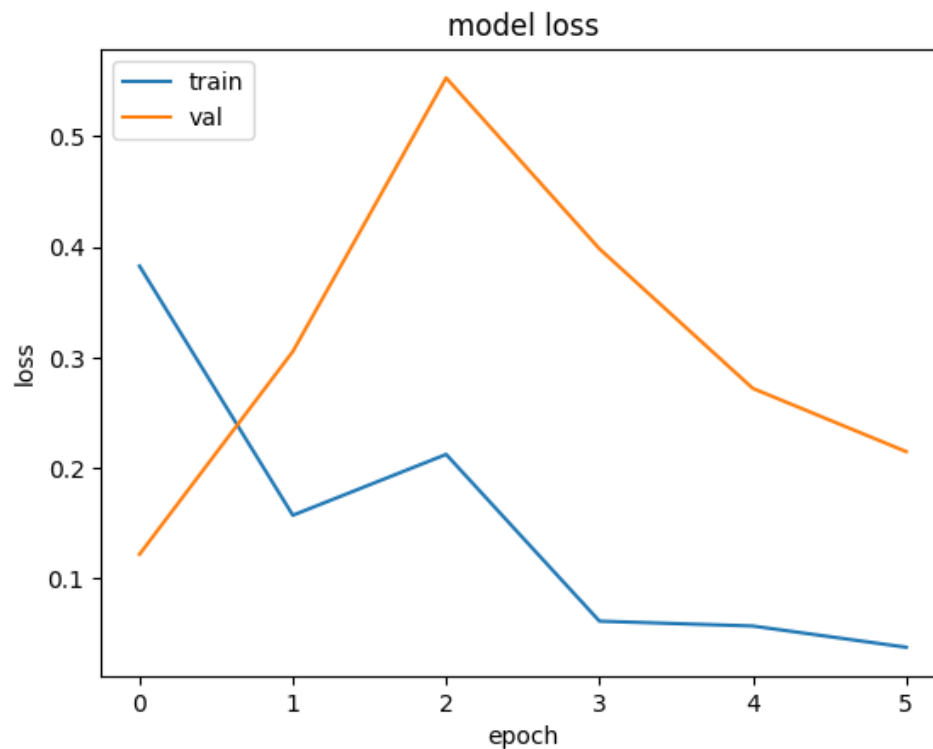


Imagem 5: Visualização da *loss* durante o treinamento da rede

Referências

Geetha, S., Abhishek, C.S. & Akshayanat, C.S. Machine Vision Based Fire Detection Techniques: A Survey. *Fire Technol* 57, 591–623 (2021).

<https://doi.org/10.1007/s10694-020-01064-z>

B. Uğur Töreyn, Yiğithan Dedeoğlu, Uğur Gündükbay, A. Enis Çetin, Computer vision based method for real-time fire and flame detection

Mohnish Sai Prasad. <https://www.kaggle.com>, 2023. Forest Fire Images. Disponível em:

<https://www.kaggle.com/datasets/mohnishsaiprasad/forest-fire-images>. Acesso em: 01 maio 2023.