ModelT Transfer Learning Feature Extraction Without Data Augmentation Results

Uma imagem com texto, captura de ecrã, Website, Página web

Descrição gerada automaticamente

**Powered By:**

**Gonçalo Ferreira, nº 2222051**

**José Delgado, nº 2222049**

# Introdução

Neste modelo com a técnica de transfer learning usamos a vgg19, a principal diferença entre a vgg16 e a vgg19 reside no número de camadas convolutional. A VGG16 possui 13 camadas convolutional e 3 camadas fully connected, totalizando 16 camadas treináveis. Em contrapartida, a vgg19 tem 16 camadas convolutional e 3 camadas fully connected, somando 19 camadas treináveis. Essa diferença de três camadas adicionais na vgg19 permite que a rede capture características mais complexas dos dados, aumentando o número total de parâmetros da rede e, consequentemente, a sua capacidade de representação. No entanto, isso também implica em maior complexidade computacional, maior tempo de treino e um potencial aumento no risco de overfitting se não for adequadamente gerenciada.

No início do projeto usamos três optimizadores distintos o RMSprop, o SGD (Stochastic Gradient Descent) e o Adam. Após alguns testes não gostamos dos resultados do SGD e acabamos por usar mais o RMSprop e depois o Adam, sendo por fim o Adam o optimizador que mais usamos. O Adam é um algoritmo de otimização popular usado em deep lerning, ele pode ser visto como a combinação do RMSprop e do SGD, ele combina os melhores recursos do momento (para acelerar a convergência) e da adaptação do lerning rate para cada parâmetro com base no histórico de gradientes (para controlar a atualização de maneira adaptativa).

Em relação aos dados, dividimos o conjunto completo em três partes distintas, treino, teste e validação. Essa abordagem permitiu-nos treinar o modelo sequencialmente com cada uma das partes do conjunto de treino, avaliando o desempenho em um conjunto de validação separado após cada fase de treino. Posteriormente, cada um desses conjuntos foi dividido em quatro partes igualmente proporcionais, garantindo que cada parte representasse um quarto do tamanho original do conjunto, realizamos esta operação com o objetivo de os treinos serem menos demorados, conseguindo obter feedback mais rápido pelo modelo.

Não colocamos todos os testes que fizemos porque por vezes era uma alteração pequena em valores, como por exemplo o lerning rate, que variávamos várias vezes entre 1e-6 e 1e-2 e o Dropout, que variávamos dependendo do overfitting, etc... Para chegar aos resultados obtidos tivemos de fazer vários testes e não conseguimos registar todos, ao longo do relatório relatamos os mais importantes.

# Resultados

Resultado 1

A screenshot of a computer program

Description automatically generated

A graph of training and validation loss

Description automatically generated

Este foi o primeiro treino que nós fizemos com os modelos T, os resultados são muito fracos o validation accuracy está quase nos 70% o training accuracy está nos 0.95%, temos, portanto, overfitting.

Este resultado já era esperado, pois a arquitetura do modelo é muito complexa, o que faz com que ele aprenda os padrões do conjunto de treino muito bem, mas falhe em generalizar para novos dados.

Este treino contém outro problema que é o tamanho definido das imagens, no início definimos as imagens de tamanho 32x32 o que faz com que o output do feature extraction da vgg19, os seus feature maps se tornem demasiado pequenos ou mesmo inválidos

Resultado 2

A screenshot of a computer program

Description automatically generated

A graph of training and validation loss

Description automatically generated

O melhor resultado que obtivemos com imagens de 32x32 foi o apresentado acima, reduzimos a complexidade da arquitetura para evitar o overfiting, e mudamos o optimizador para o Adam.

Também usamos calbacks, colocamos Early Stopping é uma técnica utilizada durante o treino de um modelo para evitar o overfitting, ela monitoriza uma métrica especificada (neste caso, 'val\_loss' que é a perda no conjunto de validação) e interrompe o treino se a métrica parar de melhorar por um número especificado de épocas (definido pelo parâmetro patience), ou seja para que o modelo não continue a treinar desnecessariamente. O ReduceLROnPlateau é uma técnica que ajusta o lerning rate durante o treino do modelo. Esta técnica também acompanha uma métrica especificada (neste caso, 'val\_loss').

Quando a métrica não mostra melhoria por um número determinado de épocas (definido pelo parâmetro patience), o lerning rate é reduzida por um fator especificado (definido pelo parâmetro factor). Isso permite que o modelo continue a aprender de maneira mais refinada, ajudando a alcançar uma convergência mais estável e precisa. Por fim temos a técnica de droupout, é uma técnica de regularização em redes neurais que combate o overfitting, durante o treino, o dropout desativa aleatoriamente um conjunto de unidades neurais em uma camada, forçando a rede a aprender características mais robustas e generalizáveis. Essencialmente, isso impede que unidades individuais se tornem muito dependentes umas das outras, promovendo uma representação mais diversificada dos dados.

A figura abaixo ilustra o funcionamento da técnica de dropout.

A diagram of a network

Description automatically generated

Resultado Final

A white rectangular object with a black border

Description automatically generated

A screenshot of a computer code

Description automatically generated

A imagem acima mostra a arquitetura obtida no final, nela podemos ver que o input já é de (4, 4, 512), o que significa que definimos o tamanho das imagens para 150x150, assim as features calculadas pela vgg19 já não irão ter tamanho inválidos.

Nos últimos treinos aumentamos o valor da variável weight\_decay, pois o modelo estava com overffiting. O parâmetro weight\_decay, também conhecido como regularização L2, adiciona uma penalização à função loss do modelo com base na magnitude dos pesos, incentivando o modelo a manter pesos pequenos. Ao penalizar pesos grandes, ajuda a evitar que o modelo se ajuste excessivamente aos dados de treino, promovendo melhor generalização para dados novos, o inverso ou seja baixar o valor do parâmetro significa menos regularização e mais overfitting, mas também mais flexibilidade e menos underfitting.

A figura abaixo ilustra a resposta do modelo a diferentes níveis de regularização, mostrando os efeitos de underfitting, regularização apropriada e overfitting.

A diagram of a weight decay

Description automatically generated

As figuras abaixo ilustram a melhoria após mudar o parâmetro de weight\_decay, a primeira figura mostra a pré mudança e a segunda figura mostra a pós mudança. Talvez o modelo iria continuar com overfiting se a rede continuasse a treinar, mas as escalas de diferenças mudaram, ou seja, o overfiting no segundo exemplo só começou depois dos 90%.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A figura abaixo ilustra o resultado final com os datasets de teste e de validação, conseguimos observar que o validation accuracy varia entre 88% e 89%

A screenshot of a computer

Description automatically generated