

Transfer Learning

Rodrigo Rusa
208592
rodrigorusa@gmail.com

Thales Eduardo Nazatto
074388
tenazatto@gmail.com

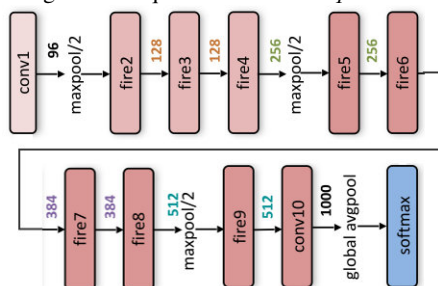
I. INTRODUÇÃO

Transfer Learning é uma técnica de Aprendizado de Máquina no qual aplica-se o aprendizado obtido na solução de determinado problema para a resolução de outros problemas diferentes. Como realizar o aprendizado de uma rede por completo pode consumir muito tempo de processamento, recursos e dados, a utilização dessa técnica agiliza a descoberta de informações. Em casos de aprendizado com imagens essa técnica funciona muito bem, pois as primeiras camadas dessas redes são formadas por filtros que extraem informações das imagens como contornos, arestas e cor, que são informações necessárias em qualquer problema com imagens.

A transferência do aprendizado pode ser realizado de várias maneiras. Os *frameworks* atuais de aprendizado de máquina, como o *Keras*, permitem configurar quais camadas serão transferidas o conhecimento já adquirido e quais serão ajustadas com a nova base de dados em questão. As primeiras camadas das redes convolucionais, conhecidas como *CNNs*, apresentam conhecimento mais genérico, como filtros, e as últimas camadas apresentam conhecimento mais específico do problema. Dessa forma, se o problema tem uma base de dados similar a base de dados do conhecimento que será transferido, apenas as últimas camadas necessitam de ajustes. Caso a base de dados seja diferente, o ajuste será necessário em mais camadas da rede.

Com o objetivo de aprender aplicar essa técnica de transferência de aprendizado para a resolução de problemas a partir de *CNNs* foi utilizado nesse trabalho a rede *SqueezeNet* [1] treinada com a base de dados *ImageNet* [2] para o treinamento de uma rede capaz de prever as classes da base *CIFAR-10* [3]. *SqueezeNet* é uma rede *CNN* pequena que atingiu a mesma acurácia que a *AlexNet* [4] sobre a base *ImageNet* utilizando 50x menos parâmetros. A Figura 1 ilustra sua arquitetura.

Figure 1: Arquitetura da rede *SqueezeNet*



ImageNet é uma grande base de dados com mais de 14 milhões de imagens de mais de 30 mil classes. É uma base muito utilizada para transferência de aprendizado, pois possui parâmetros obtidos de muitos dados com boa acurácia. Em contrapartida, a base de dados *CIFAR-10* é uma base de dados menor comparada com a *ImageNet*, possuindo 60.000 imagens de 10 classes, sendo 6.000 imagens de cada classe. Suas classes são: **airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck**.

II. METODOLOGIA

A fim de explorar a técnica de transferência de aprendizado foram conduzidos experimentos de forma que o número de camadas a serem treinadas (ajustadas) a cada experimento aumenta, avaliando-se a acurácia no conjunto de validação. Para tais experimentos foi utilizado a ferramenta *Jupyter Notebook* com o *framework Keras* para treinamento e validação das redes.

Os dados da base de dados *CIFAR-10* foram normalizados pela divisão de cada pixel pelo seu valor máximo 255 por se tratarem de imagens, ficando no intervalo entre 0 e 1. Além disso, os rótulos das imagens foram categorizados utilizando a técnica *One Hot Encoding* que transforma os rótulos em vetores de *bits*.

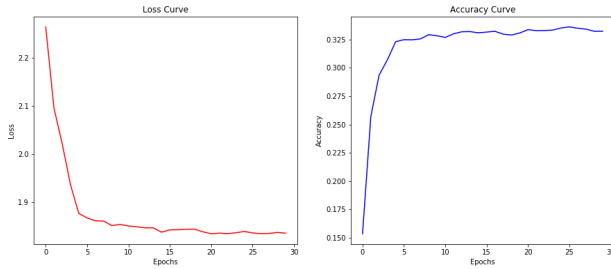
Para todos os experimentos conduzidos foi utilizado 80% dos dados para treinamento e 20% para validação. O conjunto de teste foi utilizado apenas ao final no melhor modelo encontrado. A seção a seguir apresenta os experimentos realizados e os resultados obtidos.

III. EXPERIMENTOS E DISCUSSÃO

No primeiro experimento realizado foram removidas as camadas finais da rede *SqueezeNet*, responsáveis pela classificação das classes da base *ImageNet* e foi adicionada novas camadas para a classificação das 10 classes da base de dados *CIFAR-10*. As camadas anteriores foram congeladas com os pesos obtidos da base *ImageNet*. A rede foi treinada utilizando otimizador *Adam* com *learning rate* de 0.001 por 30 épocas. A Figura 2 ilustra as curvas do erro (*loss*) e da acurácia para o conjunto de treinamento.

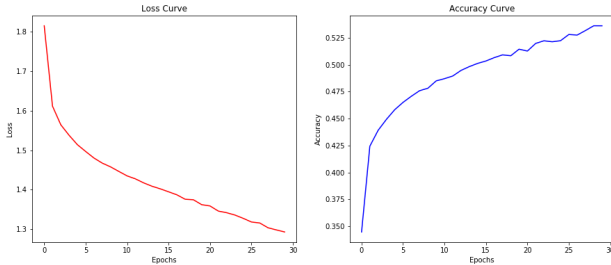
Avaliando o modelo treinado no conjunto de validação foi obtido uma acurácia de 38,32%. Com o resultado, podemos concluir que congelando a maioria das camadas da rede *SqueezeNet* não permitiu a rede adaptar seus pesos para as características da base *CIFAR-10*, resultando em baixa acurácia.

Figure 2: Curvas de erro e acurácia ajustando apenas as últimas camadas da rede *SqueezeNet*



Desse modo, no segundo experimento realizado aumentou-se o número de camadas a serem ajustadas: Além das camadas descongeladas no primeiro experimento foram também descongeladas as camadas dos dois últimos módulos *fire* (*fire* 8 e *fire* 9). Para o treinamento da nova rede com os mesmos parâmetros utilizados no primeiro experimento obteve-se as curvas de erro e acurácia ilustradas na Figura 3.

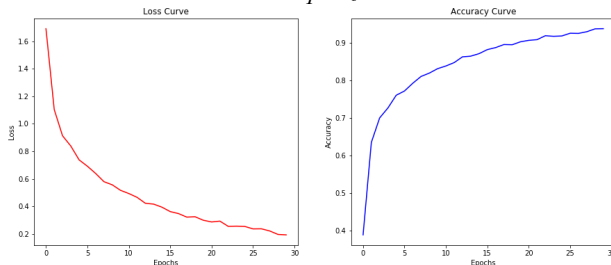
Figure 3: Curvas de erro e acurácia ajustando mais camadas da rede *SqueezeNet*



Nela foi alcançado uma acurácia de 50,15% no conjunto de validação. Comparada com a rede do experimento anterior, concluímos que uma rede com mais camadas descongeladas apresentou melhor acurácia. Porém, a acurácia ainda apresenta-se baixa.

Por fim, com o objetivo de melhorar a acurácia obtida foi treinada uma nova rede realizando um ajuste fino em todas as camadas da *SqueezeNet*, ou seja, descongelando todas as suas camadas. Utilizando os mesmos parâmetros definidos anteriormente, as seguintes curvas de erro e acurácia foram obtidas:

Figure 4: Curvas de erro e acurácia ajustando todas as camadas da rede *SqueezeNet*

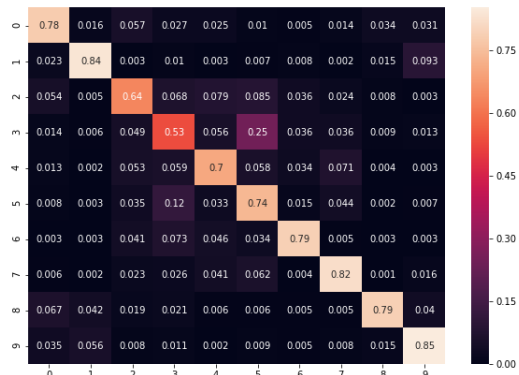


Para essa configuração da rede foi obtido um melhor resultado com 74,29% de acurácia. Podemos concluir que

treinar toda a rede para a base de dados *CIFAR-10* apresentou melhor resultado do que treinar apenas algumas camadas. Por se tratar de uma base de dados com um número menor de classes comparado com a base *ImageNet*, os pesos treinados para prever imagens da base *ImageNet* não conseguiram ser generalizados para a base *CIFAR-10*.

Encontrado o melhor modelo de rede para a base de dados em questão, com o ajuste fino de todas as camadas da rede *SqueezeNet*, foi testado o conjunto de teste no qual foi atingido uma acurácia de 74,77%. A matriz de confusão para esse conjunto de teste é ilustrada na Figura 5.

Figure 5: Matriz de confusão para o conjunto de teste



IV. CONCLUSÕES

Nesse trabalho foi explorada a técnica *Transfer Learning*, com o objetivo de transferir o aprendizado obtido da rede *SqueezeNet* a partir da base *ImageNet* para prever as classes da base *CIFAR-10*. Foram experimentadas três transferências de aprendizado diferentes, partindo de um ajuste apenas das últimas camadas da *SqueezeNet* até o ajuste completo onde foram treinadas todas as camadas da rede. A partir dos resultados obtidos concluiu-se que o treinamento de todas as camadas apresentou melhor acurácia, pois a rede com os pesos treinados para a base *ImageNet* não conseguiu generalizar para uma base menor, como a *CIFAR-10*.

Notou-se também que essa técnica de transferência de aprendizado é muito útil em casos de uma base de dados pequena e similar a base usada no treinamento da rede, pois não é necessário mais dados para alcançar bons resultados. Em imagens essa técnica é muito empregada, pois as primeiras camadas dessas redes, conhecidas como *CNNs*, possuem pesos que são filtros de extração de informações úteis em problemas desse domínio. Além disso, essa técnica economiza recursos como tempo de processamento e dados sem perda de acurácia.

REFERENCES

- [1] SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5Mb model size (<https://arxiv.org/pdf/1602.07360.pdf>)
- [2] *ImageNet* dataset www.image-net.org
- [3] *CIFAR-10* dataset <https://www.cs.toronto.edu/~kriz/cifar.html>
- [4] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In NIPS, 2012.