

TRANSFERÊNCIA DE HABILIDADE PARA DIFERENTES REPRESENTAÇÕES EM REDES NEURAIS DE CONVOLUÇÃO

PEDRO LEMOS BALLESTER¹; RICARDO MATSUMURA ARAUJO²

¹Universidade Federal de Pelotas – plballester@inf.ufpel.edu.br

²Universidade Federal de Pelotas – ricardo@inf.ufpel.edu.br

1. INTRODUÇÃO

Com o aumento da capacidade computacional nos últimos anos e com o desenvolvimento de máquinas e sistemas altamente paralelizáveis, a possibilidade de aplicação de sistemas com aprendizado supervisionado aumentou. Esses sistemas são caracterizados pela habilidade de aprender um comportamento através do mapeamento de entradas por funções para posteriores classificações e previsões. Este aprendizado se torna possível através da criação de grandes conjuntos de dados compostos por entradas para o modelo e suas respectivas respostas esperadas, modificando os coeficientes das funções conforme as entradas são colocadas no sistema.

Desde 2012, Aprendizado Profundo tem sido amplamente utilizado, superando o estado da arte em áreas como reconhecimento de voz, jogos, reconhecimento de objetos e restauração de imagens. Aprendizado Profundo consiste na utilização de grandes conjuntos de dados, como *ImageNet* (RUSSAKOVSKY, DENG et al., 2015), e na criação de sistemas com diversas camadas hierárquicas de conhecimento para a solução de problemas através de treinamento.

Em especial Redes Neurais de Convolução, um dos modelos de Aprendizado Profundo, é eficiente em reconhecimento de objetos (KRIZHEVSKY et al., 2012) através de seu aprendizado supervisionado. Além disso, em (YU et al., 2015), demonstrou-se que Redes Neurais de Convolução são boas também para classificação de objetos representados de forma abstrata, nesse caso desenhos feitos a mão com traços muito simples que representavam objetos mais complexos.

Porém, para corretamente endereçar cada uma das tarefas citadas anteriormente, faz-se necessário um estudo para a escolha da melhor arquitetura da rede, composta por diversas camadas, cada uma responsável por extrair informações dos dados de entrada, organizadas de diversas maneiras diferentes. Além disso, uma metodologia de criação do conjunto de dados deve ser empregada, garantindo que este seja representativo o suficiente do universo de possibilidades da aplicação desejada a ser utilizado. Além disso, diversos outros fatores interferem no treinamento apropriado de uma rede neural com muitas camadas, como a taxa de aprendizagem e a escolha do algoritmo otimizador – responsável por encontrar os valores dos coeficientes da rede, impossibilitando a rápida aplicação da rede em diversos problemas.

A transferência de habilidade surge como uma solução para o problema do treinamento demorado e complexo. Utilizando redes que foram treinadas com conjuntos de dados amplos, é possível reaplicar o conhecimento agregado nos neurônios para outras situações, com pouco ou nenhum (YOSINSKI et al., 2014) treinamento adicional. Esse processo se intensifica em redes neurais de convolução, visto que o conhecimento agregado em múltiplas camadas hierárquicas de convolução é vasto se o treinamento inicial for suficientemente amplo.

Em diversas situações a transferência é realizada através do congelamento das camadas de convolução da rede, fazendo somente o treinamento de camadas envolvidas na classificação. Porém essa característica no treinamento inutiliza diversos neurônios que são responsáveis por tarefas muito específicas.

Nesse trabalho, propõe-se uma metodologia de transformação do conhecimento agregado em todas as camadas da rede neural para os mesmos objetos conhecidos, porém com uma representação diferente. Utiliza-se como estudo de caso desenhos abstratos feitos a mão de objetos conhecidos pela rede, que como apresentado em (BALLESTER; ARAUJO, 2016) a rede é incapaz de reconhecer sem o devido treinamento.

O objetivo é verificar se utilizando uma rede treinada com fotos e realizando poucas etapas de treinamento, torna-se possível a classificação de desenhos cuja rede conhecia como fotos.

Este trabalho pode ser aplicado no caso do conjunto de dados utilizado ser insuficiente para o treinamento da rede, porém existe um conjunto similar, menos abstrato da mesma aplicação, como pinturas e modelos 3D com baixo detalhamento.

2. METODOLOGIA

Para a realização do trabalho, foi necessária a utilização de dois conjuntos de dados principais. O primeiro, *ImageNet*, possui milhares de imagens de mil classes diferentes (e.g. gato persa, texugo, trator), permitindo a rede armazenar uma grande quantidade de conhecimento através de seus neurônios, este que será utilizado para realizar as classificações. O segundo, *Berlin-TU* (EITZ et al, 2012), conjunto que possui 250 classes de objetos que foram desenhados a mão, cada uma com 80 imagens diferentes.

Como o foco do trabalho é na transferência de habilidade da rede entre objetos conhecidos mas em diferentes representações, foi necessário encontrar a relação entre os dois conjuntos de dados. Para isso, utilizou-se a *WordNet* (MILLER, 1995), encontrando classes que possuíam somente uma referência semântica entre os dois conjuntos, prevenindo ambiguidade. Exemplificando, mesmo *Berlin-TU* contendo a classe cachorro, esta não foi utilizada, visto que o conjunto *ImageNet* possui diversas classes, cada uma representando uma raça de cachorro. Foram encontradas 57 relações (e. g. zebra, escorpião, ampulhetas).

Utilizando uma rede no modelo *AlexNet* (KRIZHEVSKY, 2012) treinada com o *ImageNet*, foi possível carregar seus pesos e realizar um treinamento posterior. Esse treinamento é realizado somente com desenhos, assim sendo possível analisar a capacidade da rede em rapidamente transferir seu conhecimento de fotos para desenhos.

Com as 57 classes encontradas, um subconjunto do conjunto *Berlin-TU* contendo somente as classes de interesse foi criado. Originalmente, cada classe continha 80 imagens, essas foram separadas em dois grupos. De cada classe, 60 das 80 imagens foram colocadas em um grupo, utilizado durante o treinamento da rede neural. As 20 imagens restantes foram colocadas em outro grupo, utilizado para avaliar a qualidade do treinamento, chamado conjunto de teste.

Na etapa de treinamento foi realizada somente uma época - uma apresentação de cada imagem do conjunto de dados – havendo um aumento substancial na qualidade da classificação. O algoritmo otimizador utilizado foi o *Gradient Descent Optimizer*, sendo importante destacar a taxa de aprendizado, fixada em 10^{-4} , visto que pequenas modificações em seu expoente causavam grande diminuição na qualidade da classificação. Além disso, utilizando os mesmos parâmetros, foi trei-

nada uma segunda rede idêntica com somente 29 das 57 classes para descobrir se o conhecimento das 29 classes influenciaria positivamente a classificação das 28 restantes.

Uma rede com os pesos gerados randomicamente também foi treinada com os desenhos para servir de comparação entre o treinamento envolvendo transferência de habilidade e o sem transferência.

3. RESULTADOS E DISCUSSÃO

Para o acompanhamento do rápido crescimento da classificação, foram realizadas modificações na maneira com que os resultados são analisados. Inicialmente, ao invés da aleatorização das imagens de treino, elas foram ordenadas de forma que nenhuma imagem de uma classe seja vista novamente até que todas as outras classes tenham sido mostradas. Além disso, o *batch* – número de imagens sendo apresentadas simultaneamente antes de uma etapa de treinamento – foi colocado como 1.

Com a apresentação de apenas 15 imagens de cada classe o resultado da classificação já passava dos 60% de acerto do teste utilizando transferência de habilidade. Com uma época, valor considerado pequeno em treinamentos de redes neurais profundas, a quantidade de acertos alcançava 74.1% do conjunto de testes. Antes do treinamento posterior a rede respondia corretamente a classe 3.5% das vezes.

Também foi realizado o treinamento utilizando a rede inicializada randomicamente durante uma época, mesma quantidade do treinamento com transferência. Após diversas inicializações e mudanças de parâmetros, não foram obtidos conjuntos de parâmetros que causassem alguma mudança na classificação a ponto de se tornar melhor que aleatório.

Considerando a rede treinada com 29 classes. A melhoria na classificação das 29 classes foi equivalente a da rede que foi treinada com todas as classes, porém das 28 restantes o resultado foi inferior. Entretanto, houve uma melhoria na classificação de desenhos cuja classe não estava no treinamento de 2,5% quando comparado com a mesma rede sem treinamento posterior. Isso mostra que o conhecimento armazenado das 29 classes pode ter melhorado a utilização da informação que envolvia as outras 28.

4. CONCLUSÕES

Foi apresentado um método de classificação de imagens em diferentes formas de representação utilizando transferência de habilidades. Concluiu-se que é possível através da reutilização das características aprendidas nas fotos, aprender uma diferente representação de um objeto. Os resultados demonstraram que com poucas iterações a rede converge para resultados muito superiores do que utilizando uma rede sem o treinamento posterior. Além disso, também foi muito superior ao treinamento convencional da mesma rede, já que com a inicialização de pesos aleatórios, em apenas uma época, não havia obtido nenhuma melhora na classificação.

É possível que através da modificação dos parâmetros utilizados na rede e no otimizador os resultados possam ser melhorados, embora nenhum outro conjunto de valores testado tenha apresentado resultados superiores.

5. REFERÊNCIAS BIBLIOGRÁFICAS

EITZ, M.; HAYS, J.; ALEXA, M. How do Humans Sketch Objects?. **ACM Trans. Graph. (Proc. SIGGRAPH)**: 44:1-44:10, 2012.

YU, Q.; YANG, Y.; SONG, Y. -Z.; XIANG, T.; HOSPEDALES, T. Sketch-a-Net that Beats Humans. **ArXiv e-prints**: 2015.

KRIZHEVSKY, A.; SUTSKEVEER, I.; HINTON, G. E. ImageNet Classification with Deep Convolutional Neural Networks. **Advances in Neural Information Processing Systems 25 (NIPS 2012)**: 1097—1105, 2012.

RUSSAKOVSKY, O.; DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S.; MA, S.; HUANG, Z.; KARPATHY, A.; KHOSLA, A.; BERNSTEIN, M.; BERG, A.; FEI-FEI, L. ImageNet Large Scale Visual Recognition Challenge. **International Journal of Computer Vision. (IJCV)** 2015.

MILLER, G., A. Wordnet: A Lexical Database for English. **Communications of the ACM Vol. 38, No. 11**: 39-41, 1995

YOSINSKI, J.; CLUNE, J.; BENGIO, Y.; LIPSON, H. How transferable are features in deep neural networks?. **Arxiv e-prints**: 2014

BALLESTER, P.; ARAUJO, R. On the Performance of GoogLeNet and AlexNet Applied to Sketches. **AAAI Conference on Artificial Intelligence**: 2016