

## Treinamento de Redes Neurais Convolucionais por Meio de Estratégias de Aprendizado Ativo

### Convolutional Neural Network Training Through Active Learning Strategies

#### RESUMO

**Luigi Freitas Cruz**  
[luigicruz@alunos.utfpr.edu.br](mailto:luigicruz@alunos.utfpr.edu.br)  
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

**Pedro Henrique Bugatti**  
[pbugatti@utfpr.edu.br](mailto:pbugatti@utfpr.edu.br)  
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

**Priscila Tiemi Maeda Saito**  
[psaito@utfpr.edu.br](mailto:psaito@utfpr.edu.br)  
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

Devido ao crescimento vertiginoso da internet, o tamanho dos conjuntos de dados coletados também atingiu proporções sem precedentes. Desta forma, beneficiando os classificadores baseados em redes neurais. No entanto, redes neurais profundas necessitam de uma grande quantidade de dados rotulados para serem treinadas. A classificação manual destes dados leva tempo e recursos, portanto, a tendência é que se torne inviável. Uma das soluções propostas para melhorar a eficiência dessa classificação é utilizar abordagens de aprendizado ativo. Tais abordagens utilizam o próprio classificador para selecionar amostras mais informativas ao aprendizado. Este trabalho tem como objetivo avaliar o impacto na inclusão de abordagens de aprendizado ativo durante o processo de treinamento de redes neurais convolucionais. Para tanto, foi adotado o critério de incerteza por meio da heurística BALD para seleção das amostras. Foram realizados experimentos entre as abordagens de aprendizado supervisionado tradicional e de aprendizado ativo, utilizando a arquitetura VGG16 e o conjunto de dados CIFAR-10. De acordo com os resultados dos experimentos, os classificadores treinados com a abordagem de aprendizado ativo alcançaram uma acurácia 1.81% superior em relação ao classificador tradicional.

**PALAVRAS-CHAVE:** Aprendizado em máquina. Inteligência Artificial. Rede Neural. Visão computacional.

#### ABSTRACT

**Recebido:** 19 ago. 2020.

**Aprovado:** 01 out. 2020.

**Direito autoral:** Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



Due to the growth of the internet in recent years, the amount of data collected has also reached unprecedented proportions, thus positively benefiting classifiers based on deep neural networks. However, neural networks need a large amount of labeled data to be trained. The manual classification of these data takes time and resources. Therefore, the tendency is that it becomes impractical. One of the proposed solutions to improve the efficiency of this classification is to use active learning approaches. These approaches use the classifier itself to select more informative samples for the learning process. Therefore, this work aims to assess the impact on the inclusion of active learning approaches during the process of training convolutional neural networks. Therefore, the uncertainty criterion was adopted using the BALD heuristic for sample selection. Experiments were performed using the VGG16 architecture and the CIFAR-10 dataset. The classifiers trained with active learning achieved an accuracy 1.81% higher than the traditional classifier.

**KEYWORDS:** Machine learning. Artificial intelligence. Neural network. Computer vision.



## INTRODUÇÃO

Com o constante aumento de usuários, a internet mundial está cada vez mais atraindo empresas e serviços. Este crescimento substancial combinado com a cultura de mineração de informações tem criado uma quantidade sem precedentes de conjunto de dados de imagens, textos, livros, músicas, etc. Estes podendo estar classificados ou não. Este aumento de disponibilidade de dados tem incentivado outras áreas da indústria tecnológica a desenvolver técnicas de análise de dados cada vez mais avançadas. Estas tecnologias tem como âmbito de decifrar os conjuntos de dados a procura de comportamentos e hábitos dos usuários.

O aprendizado de máquina (*machine learning*) é um conjunto de métodos desenvolvidos com o objetivo de reconhecer padrões e quantizar um conjunto de dados em categorias específicas de acordo com suas similaridades. A abordagem de aprendizado mais amplamente estudada e utilizada atualmente é ao aprendizado supervisionado que consiste em treinar o classificador com um conjunto de dados já anotados por um especialista.

Mesmo apresentando um grau de acurácia elevado, a classificação manual de dados por um humano requer muito tempo e recursos. Além disso, a tendência é que os conjuntos de dados fiquem cada vez maiores, que por sua vez, irá proporcionalmente aumentar os esforços com a classificação manual tornando-a inviável.

Uma solução proposta para este problema é considerar a abordagem de aprendizado ativo, a qual possibilita selecionar um pequeno conjunto de amostras mais informativas para serem anotadas pelo especialista e consideradas no treinamento do classificador. Trata-se de um processo iterativo em que o classificador auxilia ativamente em seu processo de aprendizado, selecionando as amostras mais informativas. Diversas estratégias de seleção de amostras mais informativas têm sido desenvolvidas e bem-sucedidas, acelerando o processo de aprendizado do classificador e minimizando as intervenções do especialista nesse processo.

Portanto, o presente projeto explora o uso de abordagens de aprendizado ativo durante o processo de aprendizado de redes neurais convolucionais, de forma a obter classificadores mais robustos.

## Background

Um conjunto de dados de entrada utilizado para realizar o treinamento de um modelo de aprendizado de máquina é composto por um grande número de objetos organizados, ou não, em classes. Estes objetos podem ser imagens, áudios, textos, dados binários, etc. Como por exemplo, o *dataset* ImageNet publicado por Deng *et al.* (2009) que é composto por milhares de imagens classificadas em 21841 classes. Neste experimento, será utilizado o conjunto de dados chamado CIFAR-10 publicado por Krizhevsky *et al.* (2009). Este *dataset* é constituído por 60 mil imagens divididas entre 10 classes com 6000 imagens cada.

As redes neurais convolucionais são um dos métodos de aprendizado de máquina mais populares na academia. Este método é mais frequentemente utilizado para aprender características de uma imagem, áudio, vídeo ou até mesmo

texto. Para atingir este objetivo, os dados de entrada são abstraídos utilizando camadas de convoluções até que essa rede de camada possa identificar os padrões presentes nos dados de entrada. Após o treinamento, o classificador é capaz de reconhecer o padrão de uma entrada que não pertenceu ao conjunto de treinamento. Desta forma, é possível demonstrar que o modelo aprendeu a identificar padrões. Hoje em dia, tem-se diversas arquiteturas de redes neurais, que tem como objetivo separar imagens de acordo com classes pré-definidas. Como por exemplo, a AlexNet publicada por Krizhevsky *et al.* (2012) que ganhou a competição ImageNet em 2012, e mais recentemente a ResNext publicada por Xie *et al.* (2017). Neste experimento, será utilizada uma rede neural chamada VGG16 publicada por Simonyan *et al.* (2014).

Apesar de apresentarem resultados significativos, tais redes requerem uma quantidade grande de dados anotados, bem como recursos computacionais, considerando a abordagem de aprendizado supervisionado tradicional. Na tentativa de mitigar tais problemas, abordagens de aprendizado ativo podem ser exploradas. Tais abordagens objetivam selecionar um conjunto reduzido de amostras mais informativas para serem anotadas por um especialista e utilizadas no treinamento do classificador. Nesse sentido, é possível obter acurácias elevadas de classificação mais rapidamente, com a necessidade de menos amostras anotadas para o treinamento, consequentemente, reduzindo o esforço do especialista no processo de anotação das amostras.

Diversas estratégias de aprendizado ativo têm sido propostas, considerando diferentes critérios para seleção de amostras mais informativas. Por exemplo, o algoritmo BALD publicado por Gal *et al.* (2017) baseado no Monte-Carlo *dropout* publicado por Gal *et al.* (2016). Tal algoritmo define uma função heurística que considera o critério de incerteza para seleção das amostras.

## MATERIAL E MÉTODOS

O objetivo do trabalho é avaliar o impacto da inclusão de abordagens de aprendizado ativo durante o processo de aprendizado de redes neurais convolucionais. Para tanto, experimentos foram realizados comparando os desempenhos entre a abordagem proposta e a abordagem de aprendizado supervisionada tradicional das redes.

Para utilizar o método de heurística BALD (baseado em Monte-Carlo *dropout*) é necessário utilizar uma rede neural que tenha camadas de abandono. Como discutido em Gal *et al.* (2017), estas camadas são utilizadas para determinar se o classificador está com dificuldades de determinar uma classe para o objeto sendo avaliado. Sendo assim, foi decidido utilizar a VGG16 como rede neural padrão que possui duas camadas de abandono no último estágio. Além disso, é um modelo simples e bastante utilizado na academia, estes fatores aumentam a confiabilidade nos resultados.

Para o treinamento, foi utilizado um tamanho de lote de dados de oito objetos. Esta é a quantidade compatível com o tamanho da memória da placa de vídeo GTX 1070 Ti que foi utilizada durante todo o processo de treinamento e testes. As imagens do conjunto de dados CIFAR-10 também foram

redimensionadas de 30 pixels para 56 pixels de altura e largura. Além disso, o processo de normalização padrão da rede VGG16 foi aplicado em cada imagem antes do treinamento ou inferência. Todos os experimentos foram executados até um limite de 200 épocas com a taxa de aprendizado fixa de 0.001. A acurácia final de um classificador foi considerada ser a época com a maior acurácia, desta forma, desconsiderando possíveis degradações após o resultado recorde. Todos os experimentos levaram em torno de 24 horas para serem processados pelo computador em questão.

### Conjunto de Dados

Para este experimento foi considerado o conjunto de dados CIFAR-10. O mesmo é composto por 60mil imagens categorizadas entre 10 classes, totalizando 6000 imagens para cada uma delas. Para os experimentos foram utilizados 90% para treinamento e 10% para testes. Tal conjunto foi considerado por ser balanceado, reduzindo a probabilidade de favorecer alguma técnica e também pela sua ampla utilização.

Para evitar inconsistências, a semente de aleatoriedade dos algoritmos foi congelada no mesmo valor durante todos os experimentos. Como este conjunto de dados foi criado originalmente para o aprendizado supervisionado, já está completamente classificado. Portanto, a ausência de anotações para o treinamento ativo foi simulada de forma sintética. A quantidade de dados disponíveis para o treinamento foi suplementada utilizando as técnicas de aumento de conjunto, sendo estas, giro com ângulo aleatório e inversão horizontal aleatória. Estes procedimentos só foram aplicados no conjunto de dados de treinamento.

### Aprendizado Ativo

Os experimentos com aprendizado ativo foram subdivididos em três testes que começam com 1000 objetos retirados aleatoriamente do conjunto de treinamento. A única diferença entre eles é a quantidade de objetos adicionados após cada avaliação heurística, que acontece a cada 20 épocas. Esta quantidade foi de 250, 500 e 1000 imagens neste experimento. O valor do gerador de ruído presente no método BALD e o número de interações recursivas do Monte-Carlo *dropout* permaneceram, respectivamente, em 0.5 e 20 durante todos os experimentos. Cada teste desta categoria seguiram os seguintes passos:

- a. O conjunto de dados CIFAR-10 é carregado do disco.
- b. Um número pré-definido de objetos é aleatoriamente classificado.
- c. A rede neural é treinada por um número pré-definido de épocas.
- d. É adicionado um número de objetos pré-definido dando prioridade aos maiores graus de incerteza de acordo com a função de heurística.
- e. Volta para o passo C até que não haja mais itens a serem classificados.

### Aprendizado Tradicional

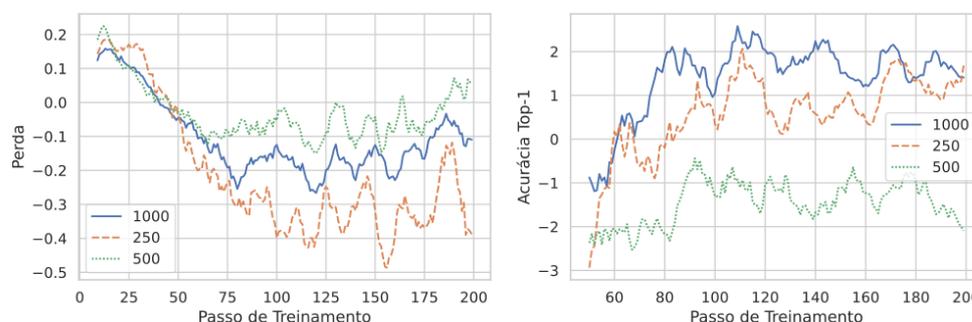
Já na categoria de experimentos com o aprendizado supervisionado tradicional foram realizados outros três testes. Assim como o aprendizado ativo, cada teste começou com 1000 objetos aleatórios classificados com um número pré-definido de objetos sendo adicionados a cada 20 épocas de treinamento. Estes números sendo 250, 500 e 1000 imagens. A única diferença entre eles é que ao invés de utilizar uma função de heurística para determinar os objetos com mais incerteza, é utilizada uma função aleatória. Com está sendo a única diferença, é possível isolar e comparar de forma justa o desempenho das duas abordagens sem que haja um viés. Em resumo, o processo de treinamento se deu da mesma forma do aprendizado ativo, com exceção do passo D em que um número pré-definido de objetos aleatórios é selecionado e classificado.

### RESULTADOS E DISCUSSÃO

A Figura 1 apresenta os resultados obtidos com o conjunto de teste durante o treinamento dos seis classificadores (três tradicionais e três ativos) dos experimentos. Para melhorar a interpretação dos resultados, os dados da figura estão normalizados em relação aos valores do treinamento tradicional. Portanto, valores positivos significam que o treinamento ativo superou os resultados do treinamento tradicional. Além disso, os dados foram filtrados com uma média móvel de 10 passos de treinamento.

No gráfico da esquerda é possível observar que a perda no treinamento ativo permaneceu inferior à do treinamento tradicional durante a maior parte do tempo. Significando que o classificador da abordagem de aprendizado ativo obteve previsões com uma melhor exatidão. Já no gráfico da direita, fica evidente que as acurácias dos classificadores treinados com o método de aprendizado ativo foram superiores, com exceção do teste de 500 unidades por turno de heurística. Em alguns momentos, esta diferença chegou até +2,54%. No fim do treinamento, a maior acurácia Top-1 de todos os testes foi de 69.93% atingida pela abordagem de aprendizado ativo de 1000 unidades por turno. Já a maior acurácia registrada nos testes tradicionais foi de 68.12%, resultando em uma diferença final de -1.81%.

Figura 1 – Comparação dos Resultados de Teste (Normalizados Em Relação Ao Aprendizado Tradicional)



Fonte: Autores

## CONCLUSÃO

Foi observado que os classificadores treinados com o método de aprendizado ativo, na maioria das ocasiões, apresentaram resultados de acurácia Top-1 de até 2.5% melhores. Estes resultados são compatíveis aos da literatura. Em futuros trabalhos, pretende-se aplicar as técnicas exploradas neste artigo em conjunto de dados maiores. Além disso, pretende-se investigar outras funções heurísticas, adotando diferentes critérios para seleção de amostras mais informativas ao aprendizado.

## AGRADECIMENTOS

Esta pesquisa foi patrocinada pelo CNPQ (bolsas #431668/2016-7, #422811/2016-5), CAPES, Fundação Araucária, SETI e UTFPR.

## REFERÊNCIAS

DENG, J. et al. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION. **Anais IEEE**, 2009. p. 248-255.

GAL, Y.; ISLAM, R.; GHAHRAMANI, Z. Deep bayesian active learning with image data. **arXiv preprint arXiv:1703.02910**, 2017. Disponível em: <https://arxiv.org/abs/1703.02910> Acesso em: 7 out. 2020.

GAL, Y.; GHAHRAMANI, Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING. New York, NY: **Anais JMLR: W&CP**, 2016. p. 1050-1059.

KRIZHEVSKY, A. **Learning multiple layers of features from tiny images**. 2009. 58p

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS. **Anais NIPS**, 2012. p. 1097-1105.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014. Disponível em: <https://arxiv.org/abs/1409.1556> Acesso em: 7 out. 2020.

XIE, S. et al. Aggregated residual transformations for deep neural networks. In: PROCEEDINGS OF THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION. **Anais IEEE CVPR**, 2017. p. 1492-1500.