

Estratégias de aprendizado ativo para melhorias no treinamento de redes neurais convolucionais

Active learning strategies to improve convolutional neural network training

RESUMO

João Marcelo Tozato
tozato@alunos.utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

Priscila Tiemi Maeda Saito
psaito@utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

O aprendizado profundo tem sido utilizado com sucesso nas mais variadas aplicações, devido à grande quantidade de dados disponíveis. No entanto, algumas delas (e.g. na área médica) apresentam escassez tanto de amostras quanto de rótulos para as mesmas. Este trabalho tem como foco apresentar a proposta de abordagens de aprendizado de máquina, incluindo métodos de aprendizado ativo no domínio de lesões mamárias, de forma a auxiliar no diagnóstico de câncer de mama. Para isto, foram realizados experimentos, considerando diversas técnicas do estado da arte para o treinamento de redes neurais convolucionais, bem como a inclusão de diferentes estratégias de seleção para o aprendizado ativo. A metodologia proposta é validada em um conjunto de dados público referente a imagens de lesões mamárias e apresenta resultados para tarefas de classificação de até 81.34% de acurácia. Foi possível avaliar o impacto das estratégias de aprendizado ativo no processo de aprendizado das redes. Tais estratégias possibilitam alcançar resultados significativos, com uma menor quantidade de dados de treinamento rotulados, minimizando o esforço dos especialistas no processo de anotação.

PALAVRAS-CHAVE: Visão por computador. Inteligência artificial. Processamento de imagens.

ABSTRACT

Deep learning has been used successfully in a variety of applications due to the large data availability. However, some of these present a shortage of both samples and labels, for instance the medical area. In this work, we focus on presenting the proposal of machine learning approaches including active learning methods in the breast lesion domain, in order to aid breast cancer diagnosis. We propose experiments using many state-of-the-art techniques for training convolutional neural networks and the introduction of distinct selection strategies for the active learning. The validation of our proposed methodology is done on a public breast lesion related dataset and our results show classification accuracies of up to 81.34%. We assessed the impact of active learning strategies in the learning process of the networks reaching solid results with a dataset containing fewer labeled images, minimizing the annotation effort of the samples for the specialist.

KEYWORDS: Computer vision. Artificial intelligence. Image processing.

Recebido: 19 ago. 2020.

Aprovado: 01 out. 2020.

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



INTRODUÇÃO

A utilização de aprendizado profundo para tarefas de classificação ou segmentação de imagens vem se tornando extremamente eficiente nos mais diversos domínios de aplicação, tal como em problemas médicos. Apesar da existência de alguns esforços na literatura, a maioria deles consideram abordagens de aprendizado que necessitam de uma quantidade significativa de dados rotulados, bem como não levam em consideração determinadas restrições exigidas por aplicações médicas (RIBLI et al., 2018; VALÉRIO et al., 2019; KOOL et al., 2017).

Segundo o Instituto Americano de Pesquisa em Câncer (2017), o câncer de mama é o tipo mais comum e a quinta principal causa de morte de câncer em mulheres no mundo. Já no Brasil, de acordo com o Instituto Nacional de Câncer (2020), este aparece em primeiro lugar tanto na taxa de incidência quanto na de mortalidade. Dessa forma, a aplicação de sistemas inteligentes neste cenário pode auxiliar na decisão do profissional responsável pelo diagnóstico, além de auxiliar no tratamento e diminuir o risco de morte.

A obtenção de imagens de lesões, referentes às massas mamárias de pacientes, não é trivial devido à manutenção da privacidade dos pacientes, à disponibilidade dos dados por parte de hospitais, entre outros. A necessidade de rotulação das amostras por um ou mais especialistas para garantir que sejam atribuídos rótulos corretos, também impacta e contribui para a menor disponibilidade de amostras rotuladas. O processo de rotulação dos dados requer tempo e esforço do especialista, além de ser altamente suscetível a erros.

Portanto, o presente projeto tem como objetivo o estudo, o desenvolvimento e a validação de estratégias de aprendizado ativo para melhorias no processo de treinamento de redes neurais convolucionais. Estas estratégias têm sido amplamente utilizadas e bem-sucedidas em outros domínios de aplicação e possibilitam obter um conjunto reduzido de amostras mais informativas ao processo de aprendizado de classificadores de padrões, obtendo classificadores mais eficazes e eficientes, atingindo acurácias elevadas mais rapidamente e minimizando o esforço do especialista no processo de anotação.

MATERIAIS E MÉTODOS

A fim de realizar os experimentos utilizando estratégias de aprendizado ativo em conjunto com CNNs, são propostos diversos cenários, em que são realizadas tarefas de classificação ponta-a-ponta com diversos tipos de arquiteturas de redes neurais convolucionais. A partir da rede que obteve melhor desempenho nestes cenários propostos, é realizada a comparação entre abordagens de aprendizado supervisionado tradicional e de aprendizado ativo por meio das características extraídas por esta rede.

No processo de treinamento de qualquer rede neural, há o procedimento de inicialização de seus pesos. Nesse caso existem dois principais tipos de abordagem. Uma delas seria inicializar os pesos de forma aleatória, seguindo uma política específica, atualizando-os “do zero”, conforme o treinamento ocorre. Outra possibilidade seria inicializá-los baseando-se nos pesos de uma outra rede neural, que já foi treinada para resolver um outro tipo de tarefa. Nesse caso, seus

parâmetros são reaproveitados para o treinamento desta nova rede, dadas as devidas modificações, esta abordagem é chamada de aprendizado por transferência.

A seleção e a modificação de hiper-parâmetros representam uma parte importante para o treinamento destas redes, principalmente no que diz respeito à taxa de aprendizado destas. Uma abordagem interessante relacionada a esta taxa é a política *One Cycle*, que se baseia na variação da taxa de aprendizado e do momento de uma dada rede neural. Tal política permite que os modelos consigam alcançar a chamada super-convergência, ou seja, é possível obter os mesmos resultados em bem menos épocas quando comparado com o treinamento realizado com estes mesmos hiper-parâmetros com um valor fixo por diversas épocas (SMITH, 2018).

Em tarefas de classificação, normalmente é utilizada uma função de *loss* chamada *Cross Entropy*, que mede o desempenho de um modelo retornando um valor de probabilidade para cada uma das classes de entrada e é a base para o processo de otimização de redes neurais artificiais. No entanto, faz com que o modelo tenha uma alta confiança em suas previsões. Segundo Szegedy (2016) a redução da confiança de um determinado modelo é eficaz e pode ser realizado por meio de uma outra função de *loss*, chamada de *Label Smoothing*. Tal função aumenta a probabilidade para as categorias que o modelo previu não serem correspondentes à entrada que lhe foi dada e, conseqüentemente, reduz a probabilidade atribuída à categoria prevista para esta.

Utilizou-se também a técnica de *MixUp*, que pode ser entendida como uma forma de *data augmentation* por meio da combinação linear de duas imagens, auxiliando no treinamento de redes neurais em que não há uma grande quantidade de dados de entrada disponíveis e nem modelos pré-treinados em domínios de dados similares. Esta também requer uma mudança no formato da saída do modelo, devido ao fato que, as imagens geradas pela combinação de outras não pertencem necessariamente a apenas uma única categoria, mas proporcionalmente a duas categorias distintas (ZHANG et al., 2017).

Antes de iniciar os experimentos relacionados ao treinamento ponta-a-ponta das CNNs, são estabelecidos hiper-parâmetros que serão utilizados em todos os modelos, pois não estão relacionados a uma técnica em específico e também definidos os cenários em que estes serão executados, de acordo com as técnicas (como *One Cycle*, *MixUp*, normalização do conjunto de entrada, *Label Smoothing*) que irão compor o treinamento das arquiteturas das redes.

O conjunto de treinamento é dividido em 10 partes estratificadas mutuamente exclusivas, de forma a preservar a porcentagem de amostras das classes. Cada divisão possui um conjunto de validação distinto, de modo a verificar se há o enviesamento do modelo durante o processo de treinamento em cada uma destas partes. No final do treinamento de uma dada divisão, é realizada a inferência em um conjunto de teste que se mantém o mesmo durante todo o processo.

Após o treinamento das CNNs com as técnicas descritas anteriormente, é realizada a seleção da arquitetura de CNN que apresenta melhores resultados dentre todos os cenários propostos. As características extraídas por tal arquitetura são utilizadas pelas estratégias de aprendizado ativo, juntamente com

classificadores tradicionais como *Support Vector Machines* (SVM), *K-Nearest Neighbors* (KNN), *Random Forest* (RF) e *Naive Bayes* (NB). Estratégias de aprendizado ativo possibilitam a seleção de amostras mais informativas para o aprendizado. Sendo assim, espera-se reduzir a quantidade de imagens anotadas para o treinamento da CNN e atingir resultados significativos ou equivalentes quando comparado com a abordagem de aprendizado ponta-a-ponta, a qual requer o conjunto de treinamento completo anotado.

A cada iteração do processo de aprendizado ativo, uma quantidade de amostras (neste caso duas vezes o número de classes existentes) é selecionada pela estratégia de seleção e então incorporada gradativamente ao conjunto de treinamento. Novas instâncias do classificador são obtidas e avaliadas no conjunto de teste. Foram realizadas comparações entre as estratégias de aprendizado ativo e a seleção aleatória de amostras a cada iteração. Além disso, foi também realizada comparação entre as abordagens de aprendizado supervisionado tradicional (a qual necessita do conjunto de treinamento completamente anotado) e de aprendizado ativo. Os métodos de aprendizado ativo que foram utilizadas para selecionar as amostras mais informativas estão relacionadas ao critério de incerteza (SETTLES, 2009).

A primeira estratégia de aprendizado ativo considerada é a *Entropy* (EN), que pode ser entendida como o grau de incerteza de uma dada variável, priorizando amostras que tenham um maior valor para esta medida. Esta é calculada segundo a Equação 1, em que y é a probabilidade de um dado rótulo para uma amostra x .

Na técnica denominada *Least Confidence* (LC), o modelo seleciona a amostra que apresenta uma menor confiança para a classe mais provável, como é possível verificar na Equação 2, em que y' é a maior probabilidade dada pelo modelo para uma amostra x . Dessa forma, um valor menor para a probabilidade da classe mais provável leva a uma maior chance desta amostra ser selecionada, devido à baixa confiança atribuída a esta.

Há também a *Margin Sampling* (MS), que leva em consideração não somente o rótulo mais provável, como na estratégia anterior, mas sim baseia-se na menor diferença entre o primeiro e o segundo rótulos mais prováveis para a seleção das amostras, como é dada na Equação 3, em que y' e y'' representam as maiores probabilidades para uma amostra x .

$$EN(x) = -\sum p_i\left(\frac{y}{x}\right) \log p_i\left(\frac{y}{x}\right) \quad (1)$$

$$LC(x) = 1 - p\left(\frac{y'}{x}\right) \quad (2)$$

$$MS(x) = p\left(\frac{y''}{x}\right) - p\left(\frac{y'}{x}\right) \quad (3)$$

O conjunto de dados utilizado é um subconjunto do MAMMOSET (OLIVEIRA et al., 2017), o qual contém imagens de duas classes, incluindo dois tipos de lesões (massas e calcificações). Neste trabalho foi considerado exclusivamente o subconjunto das lesões referentes a massas mamárias, que são classificadas como malignas ou benignas, contendo, respectivamente, 635 e 746 imagens, que foram separadas na proporção 80% para treinamento e 20% para teste.

Tabela 1 Cenários propostos para o treinamento das CNNs, considerando diferentes técnicas (normalização, *loss* e *mixUp*) e hiper-parâmetros para as mesmas.

Cenário	Normalização	Loss	MixUp α
I	Sim	Cross-Entropy	0
II	Não	Cross-Entropy	0
III	Sim	Label-Smoothing	0
IV	Não	Label-Smoothing	0
V	Sim	Cross-Entropy	0,4
VI	Não	Cross-Entropy	0,4
VII	Sim	Label-Smoothing	0,4
VIII	Não	Label-Smoothing	0,4

Fonte: Autoria própria (2020).

Para todas as arquiteturas (DenseNet121, DenseNet161, ResNet34, ResNet50, EfficientNetB3 e EfficientNetB4) e cenários foram considerados os mesmos conjuntos de hiper-parâmetros: 25 épocas de treinamento; otimizador Adam (KINGMA, 2014) com um momento de 0,9 e ϵ de 10^{-5} ; taxa de aprendizado $3e^{-3}$, seguindo a política *One Cycle* (SMITH, 2018) para a variação desta. Em relação aos cenários que fazem uso de *MixUp*, definiu-se um valor para alpha de 0.4, ou seja, uma combinação linear de duas imagens na proporção 40% e 60%. Além disso, o número de divisões estratificadas do conjunto de dados de treinamento e validação foi de 10.

RESULTADOS E DISCUSSÃO

A Tabela 2 mostra os valores das acurácias e seus respectivos desvios padrões obtidos pelas CNNs em cada um dos cenários propostos. Em negrito estão destacados os melhores resultados para cada cenário. Observa-se que as arquiteturas EfficientNetB3 e EfficientNetB4 apresentaram resultados acima da média das demais arquiteturas. Além disso, a melhor acurácia foi obtida pela EfficientNetB3 no cenário IV, com um valor de 81.34%, indicado com um asterisco.

Tabela 2: Acurácias médias e desvios padrões das CNNs em cada um dos cenários propostos

Cenários	DenseNet 121	DenseNet 161	ResNet34	ResNet50	Efficient NetB3	Efficient NetB4
I	74,86± 1,92	77,96± 2,64	75,14± 2,04	77,32± 1,83	79,08± 1,44	79,59± 1,19
II	74,72± 2,21	79,37± 2,64	73,38± 3,18	77,46± 2,44	76,69± 2,70	78,31± 2,78
III	74,93± 2,30	78,24± 2,72	73,10± 2,92	77,39± 2,70	80,35± 1,90	80,28± 1,78
IV	73,31± 1,39	77,18± 3,09	73,80± 2,31	76,41± 1,92	81,34± 2,62*	78,38± 2,64
V	75,77± 2,50	77,75± 2,28	77,54± 2,35	76,97± 2,92	79,79± 2,50	80,21± 1,68
VI	74,72± 2,24	78,38± 2,64	75,14± 2,38	78,80± 2,77	80,92± 1,20	80,42± 2,37
VII	76,13± 2,45	76,83± 2,86	76,83± 2,64	77,39± 1,49	78,80± 3,01	76,48± 1,52

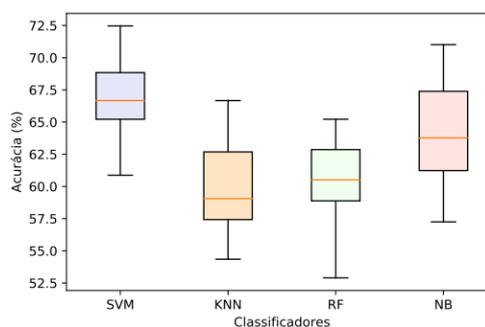
VIII	75,42± 1,56	79,08± 1,99	74,58± 2,68	76,97± 3,05	79,30± 2,00	80,70± 2,28
------	----------------	------------------------------	----------------	------------------------------	------------------------------	------------------------------

Fonte: Autoria própria (2020).

Ao comparar o desempenho das arquiteturas em todos os cenários, as arquiteturas DenseNet161, EfficientNetB3 e EfficientNetB4 apresentarem resultados equivalentes, considerando os seus respectivos desvios-padrões. No entanto, a EfficientNetB3 destacou-se por apresentar as melhores acurácias médias em cinco dos cenários propostos, com um número reduzido de parâmetros ao ser comparada com uma ResNet50, por exemplo. Além disso, pode-se observar que a DenseNet121 e a ResNet34 apresentaram as menores acurácias para a maioria dos cenários propostos. Em relação às técnicas utilizadas nos cenários, percebe-se que a introdução de *Label Smoothing* como função de *loss* das redes possibilitam a obtenção de resultados ligeiramente melhores e consideráveis, cerca de 4% ao comparar os cenários II e IV da EfficientNetB3.

Tendo isto em vista, selecionou-se a rede EfficientNetB3 em conjunto com o cenário IV para dar continuidade aos experimentos envolvendo as abordagens de aprendizado ativo. Entretanto, primeiro foi realizado um experimento considerando a abordagem de aprendizado supervisionado tradicional (i.e. utilizando todas as amostras do conjunto), com os mesmos classificadores e parâmetros, para servir de base de comparação com as abordagens de aprendizado ativo. Os resultados podem ser observados na Figura 1, onde pode-se verificar que o SVM e o RF são os classificadores com um menor desvio padrão, porém diferenciam-se nos valores de suas acurácias, em que os do primeiro são visivelmente maiores em relação ao segundo.

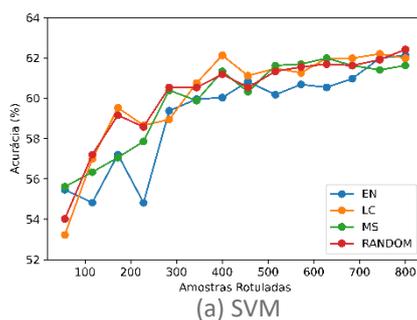
Figura 1 – Acurácias médias para o experimento utilizando aprendizado supervisionado tradicional com as características da rede EfficientNetB3.



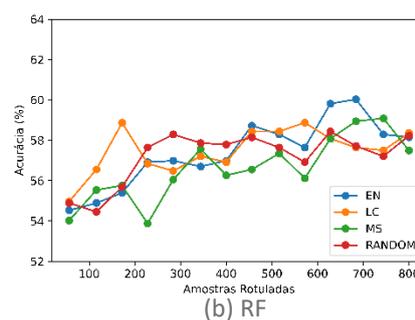
Fonte: Autoria Própria (2020)

Em relação aos experimentos considerando a abordagem de aprendizado ativo, a Figura 2 apresenta as acurácias médias obtidas pelas estratégias de seleção (EN, LC, MS e RANDOM) ao longo das iterações do processo de aprendizado, considerando cada um dos classificadores (SVM, KNN, RF e NB).

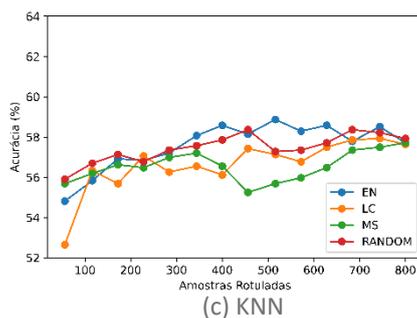
Figura 2 - Acurácias médias obtidas pelas estratégias de seleção (EN, LC, MS e RANDOM) ao longo das iterações do processo de aprendizado ativo, considerando diferentes classificadores: (a) SVM. (b) RF, (c) KNN e (d) NB.



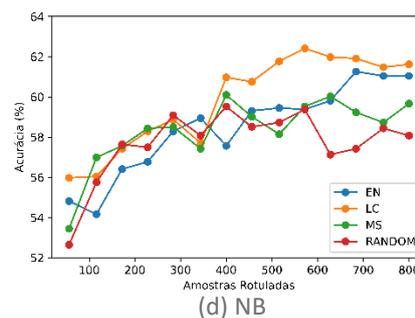
(a) SVM



(b) RF



(c) KNN



(d) NB

Fonte: Autoria Própria (2020)

Constata-se que nestes experimentos que a abordagem de aprendizado ativo proposta obtém os melhores resultados em relação à abordagem de aprendizado supervisionado tradicional, a qual considera todas as amostras (100%) do conjunto de dados. Enquanto que a abordagem de aprendizado ativo possibilita reduções (60%, 44%, 61%, 68%) significativas nas amostras de treinamento rotuladas necessárias para obter acurácias equivalentes, utilizando os classificadores SVM, KNN, RF e NB, respectivamente.

CONCLUSÃO

Neste trabalho realizou-se experimentos extensivos seguindo a metodologia proposta, de forma a comparar os resultados do treinamento de CNNs através de aprendizado profundo ponta-a-ponta, aprendizado tradicional supervisionado com classificadores clássicos e aprendizado ativo utilizando as características extraídas destas. Em relação ao treinamento das CNNs ponta-a-ponta, observou-se os melhores resultados para esta tarefa de classificação ao utilizar a arquitetura EfficientNetB3 em conjunto com *Label Smoothing* e sem a normalização dos dados de entrada, alcançando cerca de 81% de acurácia neste cenário.

Avaliou-se também a importância da utilização de estratégias de aprendizado ativo para o processo de classificação de imagens com CNNs, demonstrando que com um conjunto de dados rotulado reduzido obteve-se resultados expressivos em relação ao aprendizado supervisionado tradicional. Nos resultados obtidos pelos experimentos realizados verificou-se o quão interessante é o uso destas estratégias no desenvolvimento de um classificador, visto que neste contexto médico em que está sendo aplicada nem sempre todas as imagens disponíveis estão anotadas e, ao selecionar as amostras mais informativas para o treinamento, portanto há a diminuição do esforço e do tempo de anotação de um especialista.

REFERÊNCIAS

AKKUS, Z. et al. Deep learning for brain mri segmentation: state of the art and future directions. **Journal of digital imaging, Springer**, v. 30, n. 4, p. 449–459, 2017.

INCA. **Estatísticas de câncer**. Disponível em: <https://www.inca.gov.br/numeros-de-cancer>. Acesso em: 03 jul. 2020

GREWAL, M. et al. Radnet: Radiologist level accuracy using deep learning for hemorrhage detection in ct scans. *In: IEEE.2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. 2018. p. 281–284.

HUYNH, B. Q.; LI, H.; GIGER, M. L. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. **Journal of Medical Imaging**, International Society for Optics and Photonics, v. 3, n. 3, p. 034501, 2016.

KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. **arXiv preprint arXiv:1412.6980**, 2014.

KOOI, T. et al. Large scale deep learning for computer aided detection of mammographic lesions. **Medical image analysis**, Elsevier, v. 35, p. 303–312, 2017.

OLIVEIRA, P. et al. Mammaset: An enhanced dataset of mammograms. *In: Brazilian Symposium on Databases - Dataset Showcase Workshop*. Uberlândia, 2017.

AMERICAN INSTITUTE FOR CANCER RESEARCH. **Breast cancer**: How diet, nutrition and physical activity affect breast cancer risk. Disponível em: <https://www.wcrf.org/dietandcancer/breast-cancer>. Acesso em: 03 jul. 2020

RIBLI, D. et al. Detecting and classifying lesions in mammograms with deep learning. **Scientific reports**, Nature Publishing Group, v. 8, n. 1, p. 1–7, 2018.

SETTLES, B. **Active learning literature survey**. University of Wisconsin-Madison Department of Computer Sciences, 2009.

SMITH, L. N. A disciplined approach to neural network hyper-parameters: Part 1-- learning rate, batch size, momentum, and weight decay. **arXiv preprint arXiv:1803.09820**, 2018.

SZEGEDY, Christian et al. Rethinking the inception architecture for computer vision. *In: Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. p. 2818-2826.

TAN, M.; LE, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*, 2019.

VALÉRIO, L. M. et al. Deepmammo: Deep transfer learning for lesion classification of mammographic images. *In: IEEE.2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*. 2019. p. 447–452.

ZHANG, Hongyi et al. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017.