

# *Data Augmentation* e Extração de Características utilizando Redes Neurais Convolucionais

Willian Dihanster Gomes de Oliveira, Lilian Berton  
Instituto de Ciência e Tecnologia - Universidade Federal de São Paulo

**Resumo**—O Aprendizado de Máquina (AM) é uma área de grande destaque nos dias atuais. No entanto, um dos seus problemas, é a necessidade de uma grande quantidade de dados, especialmente rotulados, que são custosos e difíceis de obter. Assim, várias técnicas têm sido propostas para obter mais dados rotulados. Neste trabalho, objetivamos estudar uma técnica de *Data Augmentation* (DA) para classificação de imagens. Além disso, quando é necessário trabalhar com as características das imagens, um bom extrator de características deve ser utilizado. Por isso, objetivou-se também o estudo de 4 redes convolucionais como extrator de características.

## I. INTRODUÇÃO

O Aprendizado de Máquina (AM) é uma área de grande destaque nos dias atuais, a qual se baseia em criar e modelar sistemas capazes de aprender a partir de exemplos e que melhorem automaticamente com a experiência. No AM há uma hierarquia do aprendizado que pode ser dividida entre: aprendizado supervisionado, semissupervisionado e não supervisionado. [1] No aprendizado supervisionado, onde é comum a classificação, há uma grande variedade de algoritmos, como *k*-vizinhos mais próximos (*k*NN), Árvores de Decisão, *Support Vector Machines*, Redes Neurais, etc. Atualmente, um classificador em destaque são as Redes Neurais Profundas, que vêm obtendo ótimos resultados na classificação de imagens, pois possuem uma alta capacidade de abstração. [2]

Dentre as maiores aplicações da classificação de imagens estão a automatização de tarefas. Um exemplo é a classificação de imagens de plantações de café que podem ajudar os agricultores a identificar mais facilmente, lugares de sua plantação que possuem um pé de café, com a ajuda de um *software* e um sensor, sem a necessidade de percorrer toda sua plantação. [3] Entretanto, para que esse reconhecimento automático funcione da melhor maneira possível é necessário uma grande quantidade de dados e em especial, rotulados para treinar um classificador. O que é uma tarefa difícil e custosa de se fazer. Assim, diversas técnicas de *Data Augmentation* (DA) vêm sendo empregadas na obtenção de mais dados rotulados.

Neste trabalho, objetivamos o estudo da técnica de Transformação de Imagens, que visa expandir o conjunto de dados aplicando efeitos como corte, girar, *blur*, etc.

Além disso, as vezes é preciso representar dados por um vetor de características, necessitando um algoritmo que extraia características dessas imagens. Uma alternativa possível é utilizar redes neurais convolucionais (CNNs), pegando seu vetor de *features*.

Os experimentos de DA, extração de características e de classificação foram realizados em Python, com a biblioteca

TensorFlow e o Keras. Foram gerados 5 conjuntos aumentados por DA e para cada conjunto gerado, foram extraído suas características pelas 4 redes, gerando 20 novos conjuntos de dados. Foi analisado os resultados de acurácia na classificação empregando a técnica DA e a variação das redes para extração.

O restante do artigo está organizado como segue: Na Seção II será detalhado alguns conceitos sobre a classificação, com foco em Redes Neurais, *Deep Learning* e *Data Augmentation*. Na Seção III são sumarizados alguns trabalhos relacionados. Na Seção IV é descrita a metodologia empregada. Na Seção V é apresentada a análise experimental e por fim, na Seção VI é apresentada as conclusões do trabalho.

## II. CLASSIFICAÇÃO

Na classificação de dados, os algoritmos recebem exemplos rotulados com um atributo classe e aprendem comparando a saída gerada pelo algoritmo com a saída correta. A partir dos erros identificados modificam o modelo para se ajustar as saídas corretas. [4] O objetivo do indutor é gerar um classificador que sirva para rotular novos elementos que não faziam parte da amostra inicial, usada no treinamento.

Há uma diversa variedade de algoritmos de classificação disponíveis em AM. Tendo as técnicas tradicionais e as CNNs, como as utilizadas neste projeto, que será comentada na seção A. Enquanto que o problema da falta de dados e uma possível solução é comentada na seção B.

### A. Redes Neurais (RN) e Deep Learning

Um dos classificadores mais populares em AM são as Redes Neurais. As Redes Neurais são baseadas em um modelo matemático de como o cérebro realiza suas tarefas e como aprende. O *Multilayer Perceptron* (MLP) simples é um modelo de Rede Neural que até então, era o mais utilizado. Nesse modelo, em geral, há a presença de uma camada de entrada, camada de saída e uma camada oculta. [5]

Com o crescimento do poder de processamento dos computadores atuais, houve a ascensão do *Deep Learning* ou Aprendizado Profundo. Essa forma de aprendizado vem obtendo ótimos resultados, pois permitem uma maior capacidade de abstração dos problemas. [2] No *Deep Learning* há também uma grande variedade de modelos de Redes Neurais, por exemplo, o modelo mais simples é basicamente uma MLP com mais camadas ocultas, as Redes Neurais Convolucionais (possuem camadas de convolução), dentre outras.

Os modelos de CNNs utilizados neste projeto, são descritas a seguir.

- ResNet50: É uma rede de 50 camadas que utiliza a técnica de *Residual Learning* [6]. O vetor de características gerado da última camada da rede (antes da classificação) tem tamanho (1, 2048). [7]
- VGG16: É uma CNN com 16 camadas de convolução. Seu vetor de características da última camada da rede tem tamanho (1, 4096). [8]
- VGG19: Parecida com a VGG16, com a diferença da adição de 3 camadas convolucionais a mais. Seu vetor de características da última camada da rede também tem tamanho (1, 4096). [8]
- Xception: Ou X-treme Inception, é uma CNN com uma técnica especial, chamada *Depthwise Separable Convolutions*. Seu vetor de características da última camada da rede também tem tamanho (1, 4096). [9]

### B. Data Augmentation (DA)

No aprendizado supervisionado, o desempenho dos algoritmos pode ser estendido de acordo com o aumento no número de instâncias numa base de dados. No entanto, em situações da vida real, nem sempre temos muitos exemplos disponíveis ou muitas vezes, não são rotulados.

Dessa forma, criou-se o conceito de *Data Augmentation* (DA), que em tradução direta, significa aumento de dados. O DA tem como objetivo utilizar técnicas computacionais para aumentar a base de dados original e como consequência, obter uma melhor acurácia na classificação.

Uma abordagem possível é a de transformação de imagens. Nessa técnica, a partir de um conjunto de imagens determinadas são aplicadas edições de imagem como corte, rotação, *blur* na imagem original, formando novas imagens. Assim, é gerado uma base aumentada, composta pelas imagens originais mais as imagens editadas.

## III. TRABALHOS RELACIONADOS

### A. Deep Features Generalize from Everyday Objects to Remote Sensing and Aerial Scenes Domains?

A ideia dos autores [3] era estimar o poder de generalização das *Convolutional Neural Networks* (CNNs) aplicado na classificação de imagens de sensores remotos e imagens aéreas. O domínio utilizado é o mesmo domínio de imagens de plantações de café do trabalho aqui realizado. No entanto, no nosso trabalho, utilizamos técnicas de DA para as bases utilizadas, o que tenta aproximar os resultados obtidos da base completa.

### B. Research on Data Augmentation for Image Classification Based on Convolutional Neural Networks

A proposta dos autores [10] era fazer DA a partir de transformação de imagens, utilizando técnicas de girar, rotacionar, cortar, etc e/ou uma combinação desses métodos. Após a geração da base aumentada, classificar os dados utilizando as CNNs e tentar descobrir dentre essas técnicas de transformação de imagens, quais são as melhores dentre os conjuntos de dados.

### C. Effectiveness of Data Augmentation in Image Classification using Deep Learning

Neste trabalho os autores [11] propuseram o estudo de outras técnicas de DA para expandir o conjunto de imagens. Nesse caso, além das técnicas tradicionais como corte, girar, etc, consideraram também filtros de estilização das imagens. O domínio do problema foi baseado em imagens de animais, como peixes, cachorros, gatos e também, dígitos.

### D. Operational Data Augmentation in Classifying Single Aerial Images of Animals

Baseado em técnicas de DA para classificação de imagens aéreas. A ideia dos autores [12] foi aplicar técnicas de transformação de imagens na base de dados original, com o diferencial de usar uma técnica que cria uma nova imagem a partir de diversos ângulos de rotação de uma só imagem. Isto é, forma uma espécie de grade com a mesma imagem, em vários ângulos de rotação. A aplicação é parecida com a do nosso trabalho por se tratar de imagens aéreas, mas em uma situação mais simples, onde as classes para diferenciação são com presença de vaca ou sem presença de vaca.

## IV. MATERIAIS E MÉTODOS

Nessa seção é descrito a metodologia abordada para realização dos testes e as ferramentas utilizadas.

### A. Base de dados

A base de dados original utilizada (*Brazilian Coffee Scenes Dataset* [3]) é composta por imagens de satélite de plantações de café, em cidades de Minas Gerais coletadas pelo sensor SPOT.

Cada uma das imagens possuem 64x64 pixels e foi classificada manualmente por pesquisadores agrícolas como “coffee” e “non coffee”. Assim, o conjunto de dados é composto por 2876 exemplos, sendo 50% exemplos da classe “coffee” e 50% da classe “non coffee”. Nas Figuras 1 e 2, temos alguns exemplos das imagens capturadas pelo sensor.

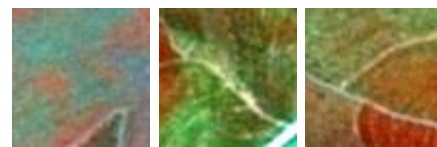


Figura 1: Exemplo de imagens rotuladas como “coffee”. Fonte: Rubens Lamparelli, Cooxupé “Brazilian Coffee Scenes Dataset”.

### B. Expansão das Bases - Data Augmentation

Para a expansão dos conjuntos de dados rotulados, primeiramente, pegou-se um subconjunto do *dataset* original, gerando 5 conjuntos de dados diferentes, com 200 imagens cada (100 exemplos de cada classe).



Figura 2: Exemplo de imagens rotuladas como “non-coffee”. Fonte: Rubens Lamparelli, Cooxupé "Brazilian Coffee Scenes Dataset".

Esses conjuntos passam pela aplicação da técnica DA, com vários efeitos aplicado aleatoriamente, até que cada conjunto atinja 2010 exemplos, ao total.

Em seguida, as 4 redes convolucionais são utilizadas para extrair as características de cada um desses conjuntos. Para isso, cada rede pré-treinada é carregada na aplicação com o pesos da *ImageNet* [13].

Assim, são gerados 20 conjuntos de dados no formato atributo-valor. Que serão utilizadas para os experimentos de classificação e avaliação dos resultados.

Também, há 4 bases de testes, uma para cada CNN, geradas a partir de um outro subconjunto da base original (30% - 866 exemplos). Sendo assim, é formado os conjuntos de treino e teste.

## V. ANÁLISE EXPERIMENTAL

Após gerar todos as bases desejadas. Essas bases passam por experimentos de classificação.

Estes experimentos foram realizados também com a biblioteca Keras em Python, utilizado uma Rede MLP com 5 camadas.

Para as bases obtidas a partir da extração de características (*features*) pela rede ResNet50, temos os resultados da Figura 3. Nota-se bons resultados, chegando a uma acurácia de 88%.

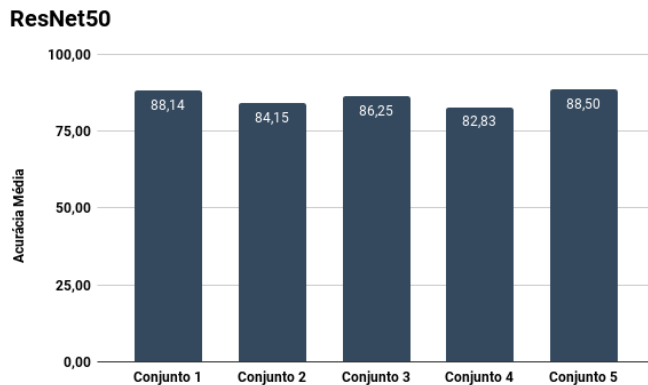


Figura 3: Resultado para as bases com *features* da ResNet50.

Para as bases obtidas pela rede VGG16, temos os resultados da Figura 4. Também nota-se bons resultados, embora inferiores aos obtidos pela ResNet50.

Já para a rede VGG19, temos os resultados da Figura 5. Outra vez, observa-se bons resultados, chegando a uma acurácia de 87%.

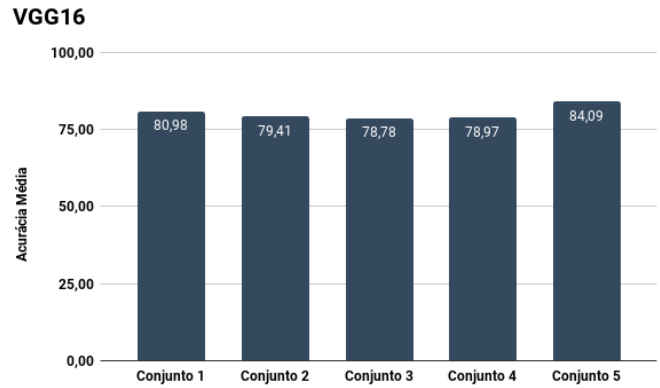


Figura 4: Resultado para as bases com *features* da VGG16.

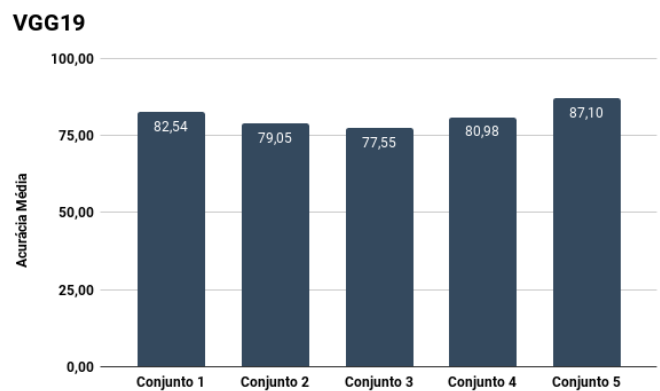


Figura 5: Resultado para as bases com *features* da VGG19.

Por fim, para a rede Xception, temos os resultados da Figura 6. Aqui, notou-se bons resultados também, com uma média de 78% de acurácia. No entanto, resultados menores que os encontrados com as outras redes.

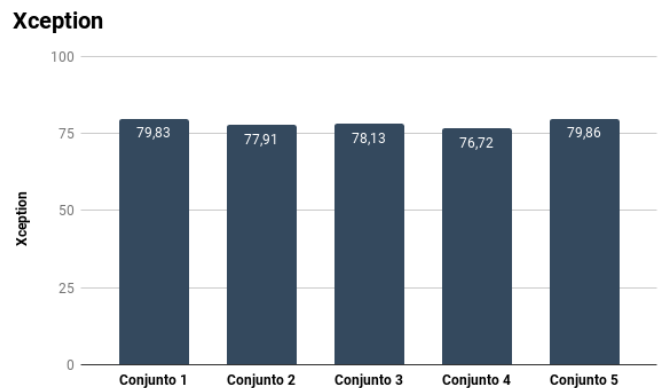


Figura 6: Resultado para as bases com *features* da Xception.

Ainda, temos uma média dos resultados obtidos por cada rede, levando em consideração todos os conjuntos gerados, que podem ser visualizados na Figura 7. Isto é, a média da acurácia de classificação dos 5 conjuntos, gerados pelas respectivas redes. Nota-se que a ResNet50 foi a rede com melhor resultado

(85,97%), em geral. Já a Xception, a com o pior resultado (78,48%), em geral.

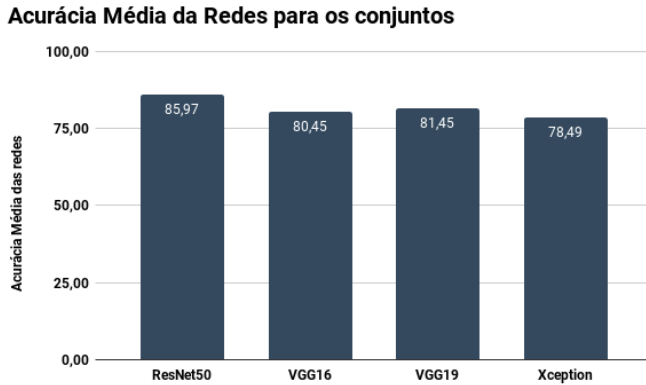


Figura 7: Resultado medio das redes.

#### A. Discussão

Pode-se concluir que a aplicação da técnica de DA foi sucedida neste caso, pois foram obtidos bons resultados em geral e usando apenas 10% do conjunto original, como treino, um acurácia menor foi atingida (cerca de 70% para todas as redes).

Verificou-se também que há a possibilidade de haver uma variância entre os resultados dos conjuntos gerados, mas não é uma discrepância tão alta.

Além disso, as redes utilizadas também influenciaram no resultado. Tendo a ResNet50 obtido melhores resultados em relação as outras redes, enquanto que a Xception obteve piores resultados.

A explicação encontrada pode ser devido ao fato da ResNet50 utilizar 50 camadas e o aprendizado residual, de sua arquitetura. Relação também encontrada comparando as rede VGG16 e VGG19, onde a rede VGG19 obteve melhores resultados em relação a rede VGG16, tendo a rede VGG19, três camadas a mais que a rede VGG16.

### VI. CONCLUSÕES

Pode-se concluir que a técnica de DA pode ajudar no aumento da acurácia. Podendo ser utilizada no caso de se possuir poucos exemplos em um conjunto de dados.

Dentre as vantagens da técnica de Transformação de Imagens é a possibilidade de utilização em qualquer aplicação que tenha imagens. Como desvantagem, deve-se tomar cuidado com a redundância (caso o efeito aplicado seja mínimo e as imagens fiquem praticamente iguais, pode não resultar em uma melhora na acurácia) ou na perda de sentido (caso o efeito aplicado seja excessivo e a imagem não pareça mais com a classe desejada).

Além disso, redes convolucionais podem ser também ser utilizadas como extratores de características. Em especial, a ResNet50, que apresentou melhores resultados.

Para trabalhos futuros, o objetivo é testar outros conjuntos de dados, além de outras redes convolucionais para a extração de características.

### REFERÊNCIAS

- [1] Olivier Chapelle, Bernhard Scholkopf, and Alexander Zien. *Semi-Supervised Learning*. The MIT Press, 1st edition, 2010.
- [2] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- [3] Otávio AB Penatti, Keiller Nogueira, and Jefersson A dos Santos. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2015 *IEEE Conference on*, pages 44–51. IEEE, 2015.
- [4] Maria Carolina Monard and José Augusto Baranauskas. Conceitos sobre aprendizado de máquina. *Sistemas inteligentes-Fundamentos e aplicações*, 1(1):32, 2003.
- [5] Simon Haykin. *Redes neurais: princípios e prática*. Bookman Editora, 2007.
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [8] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [9] François Chollet. Xception: Deep learning with depthwise separable convolutions. *arXiv preprint*, pages 1610–02357, 2017.
- [10] Jia Shijie, Wang Ping, Jia Peiyi, and Hu Siping. Research on data augmentation for image classification based on convolution neural networks. In *Chinese Automation Congress (CAC)*, 2017, pages 4165–4170. IEEE, 2017.
- [11] Jason Wang and Luis Perez. The effectiveness of data augmentation in image classification using deep learning. Technical report, Technical report, 2017.
- [12] Emmanuel Okafor, Rik Smit, Lambert Schomaker, and Marco Wiering. Operational data augmentation in classifying single aerial images of animals. In *Innovations in Intelligent Systems and Applications (INISTA)*, 2017 *IEEE International Conference on*, pages 354–360. IEEE, 2017.
- [13] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.