

Uma Abordagem Multiclasse para a Caracterização de Flores Utilizando Redes Neurais Convolucionais

Diego Nunes Noceli, Alessandra Martins Coelho, Matheus de Freitas Oliveira Baffa

¹Departamento Acadêmico da Ciência da Computação
Instituto Federal de Educação, Ciência e Tecnologia do Sudeste de Minas Gerais
IF Sudeste MG, campus Rio Pomba
36.500-180 – Rio Pomba – MG – Brasil

mfreitas826@gmail.com

Abstract. *In Brazil, there are about 30 thousand flowers species, with 250 to 300 new species being discovered annually, which may contain different shapes, colors and perfumes. Due to the similar characteristics found in plants, their classification is not a trivial task. This work aims to develop a preliminary method of flower classification, based on Convolutional Neural Networks to identify patterns and automatically differentiate the types of flowers. For the development, a database of 4,135 images was used, containing five groups of flowers, being the dandelion, the daisy, the rose, the tulip and the sunflower. These images were used in the training and evaluation of two types of Convolutional Neural Networks, LeNet-5 and ResNet152. Although simpler, LeNet-5 obtained a superior result, reaching 99.69 % of categorical accuracy.*

Resumo. *No Brasil existem cerca de 30 mil espécies de plantas com flores, sendo descobertas entre 250 e 300 novas espécies anualmente podendo conter diferentes formas, cores e perfumes. Devido as características semelhantes encontradas nas plantas, a sua tipificação não é uma tarefa trivial. Este trabalho tem como objetivo, desenvolver um método preliminar de classificação de flores, baseado em Redes Neurais Convolucionais para identificar padrões e diferenciar automaticamente os tipos de flores. Para o desenvolvimento, foi utilizada uma base de dados de 4.135 imagens contendo cinco grupos de flores, sendo elas o dente de leão, a margarida, a rosa, a tulipa e o girassol. Essas imagens foram utilizadas no treinamento e avaliação de dois tipos de Redes Neurais Convolucionais, a LeNet-5 e a ResNet152. Embora mais simples, a LeNet-5 obteve um resultado superior, atingindo 99.69% de acurácia categórica.*

1. Introdução

A Visão Computacional é uma área da Ciência da Computação que tem como objetivo fazer com que a máquina interprete imagens e, através da análise e do reconhecimento de padrões, extraia novas informações sobre os elementos que a compõem [Azevedo and Conci 2003]. A literatura relacionada à Visão Computacional consiste em trabalhos que envolvem desde a captura das imagens, ao melhoramento e a segmentação das regiões de interesse a fim de melhor extrair informações e interpretar uma cena [Backes and de Mesquita Sá Junior 2019].

Essa área de estudo, que relaciona métodos de Processamento e Análise de Imagens à técnicas de Aprendizado de Máquina, consiste em três etapas fundamentais, sendo

elas (i) o pré-processamento das imagens, (ii) a análise descritiva das imagens e (iii) o reconhecimento de padrões. O processamento das imagens consiste em reforçar algumas características visuais, remover ruídos e segmentar regiões de interesse. Por outro lado, a análise da imagem visa descrevê-la sob o olhar de uma característica específica, tal como a textura, a cor ou as formas presentes na imagem. O reconhecimento de padrões, por sua vez, é realizado por um algoritmo de aprendizado de máquina que detecta padrões entre as características previamente extraídas.

Esse campo de pesquisa tem aplicações em múltiplas áreas adjacentes. Pode-se utilizar visão Computacional nas áreas de ciências agrárias [Oyama et al. 2012], na astronomia [Moen et al. 2019], na biologia [Nieto et al. 2019], nas ciências médicas [Fernandes et al. 2012] entre outras, constituindo uma vasta área multidisciplinar com várias aplicações práticas [Neves et al. 2012].

No âmbito das aplicações da Visão Computacional na botânica, o reconhecimento e classificação de flores se difere bastante de um algoritmo de classificação entre cães e gatos, por exemplo, pois nas flores existem grandes quantidades de características semelhantes, tais como várias flores diferentes compartilharem as mesmas formas, aparências e até cores. Além disso, as imagens de flores geralmente contêm muitos ruídos, como folhas, grama, entre outros [Hiary et al. 2018]. Dessa forma, é necessário um estudo que realize a classificação das imagens utilizando combinações de múltiplos tipos de características e algoritmos de aprendizado de máquina.

Diante do exposto, este trabalho tem como objetivo geral o desenvolvimento de um classificador automático de flores utilizando aprendizado de máquina. Como objetivo específico, buscamos desenvolver e comparar a eficiência da classificação a partir de duas arquiteturas de Redes Neurais Convolucionais (*Convolutional Neural Network - CNN*), a versão mais tradicional da CNN (LeNet-5), e uma versão altamente profunda (ResNet-152).

Neste trabalho, abordamos a classificação de cinco tipos específicos de flores, sendo elas o dente de leão, a rosa, a tulipa, a margarida e o girassol. Através de uma CNN, o algoritmo processa a imagem, extraindo características descritivas, e as classifica qual espécie de flor a imagem representa.

Este trabalho está organizado como se segue. Na Seção 2 é apresentado os trabalhos relacionados ao problema tratado neste trabalho. A Seção 3 contém uma descrição da base de dados de fotografias de flores utilizados no treinamento do algoritmo proposto. A Seção 4 descreve a arquiteturas das Redes Neurais Convolucionais desenvolvidas. Já a Seção 5 descreve os experimentos, os resultados obtidos e realiza uma discussão sobre este. Por fim a Seção 6 conclui o presente trabalho.

2. Trabalhos Relacionados

Alguns trabalhos na literatura abordaram o problema de classificação de flores utilizando técnicas de Visão Computacional. Nesta Seção iremos abordar alguns destes trabalhos.

[Kumar et al. 2012] realizaram o primeiro aplicativo móvel voltado para identificação espécies de plantas usando reconhecimento visual automático. O sistema identifica espécies de árvores a partir de fotos de suas folhas. A chave para esse sistema são os componentes de Visão Computacional para descartar imagens não-folha, segmen-

tar a folha de um fundo não texturizado, extrair recursos que representam a curvatura do contorno da folha em várias escalas e identificam as espécies a partir de um conjunto de dados das 184 árvores. O aplicativo acumulava quase um milhão de usuários em maio de 2011. Concluíram que a maior taxa de reconhecimento tem cerca de 73% e um reconhecimento com taxas de 96,8% para 184 espécies de árvores.

O trabalho de [Joly et al. 2014] propôs o *Pl@ntNet* como uma abordagem da ciência cidadã para acelerar a coleta e integração dos dados de imagens botânicas observadas. Desde 2010, grande quantidade de fotografias de plantas datadas e com marcação geográfica foram coletadas e revisadas por centenas de botânicos amadores e especialistas de uma rede social especializada. Uma ferramenta de identificação baseada em imagem, disponível tanto para aplicação web quanto para aplicação móvel, sincronizada com os dados crescentes, permitindo qualquer usuário consultar ou enriquecer o sistema com novas observações. Experimentos extensivos do motor de busca visual, bem como avaliações orientadas ao sistema e ao usuário do aplicativo mostraram que ele era útil para determinar uma planta entre uma grande quantidade de espécies. Na elaboração da pesquisa, cobriram aproximadamente metade das plantas encontradas França (2.200 espécies), tornando-a a maior ferramenta de identificação automatizada existente.

Já [Wu et al. 2007] fizeram o uso da Rede Neural Probabilística no reconhecimento automatizado de folhas para classificação de plantas. Doze características de folhas são extraídas e ortogonalizadas em cinco variáveis principais, que consistem no vetor de entrada da Rede neural probabilística (*Probabilistic Neural Network* - PNN). A PNN é treinada com 1800 folhas para classificar 32 tipos de plantas com uma precisão superior a 90%. Comparado com outras abordagens, o algoritmo de Inteligência Artificial é mais preciso, mais rápido e possui uma implementação mais fácil.

3. Base de Dados Visual

O conjunto de dados elaborado por Rishit Chaudhary é uma base de dados disponibilizada gratuitamente na internet para fins de pesquisa em Visão Computacional [Chaudhary 2020]. Essa base de dados foi criada a partir da obtenção de imagens disponibilizadas na internet, em ferramentas como *Flickr*, *Google Imagens* e *Yandex Images*.

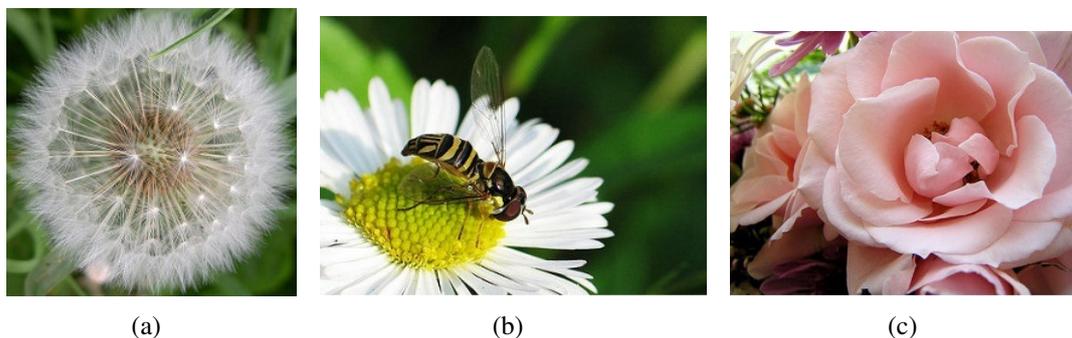
Ao todo, a base de dados contém 4.135 imagens, divididas em cinco classes distintas. A base de dados é composta por 746 imagens de Margaridas, 1028 imagens de dentes-de-leão, 728 imagens de rosas, 698 imagens de girassol e 935 imagens de tulipa. Observa-se que a quantidade média de cada classe está próxima, não havendo a predominância de uma classe específica. Na Figura 1 são apresentadas algumas imagens de flores que compõem a base de dados, tais como o dente de leão (Fig. 1(a)), a margarida (Fig. 1(b)) e a rosa (Fig. 1(c)).

Devido ao processo automático de recuperação das imagens utilizado pelo autor da base, existem imagens que estão fora do contexto de classificação. Essas imagens, consideradas *outliers*, ou ruídos, foram eliminadas da base de dados para fins de melhor qualidade da classificação.

4. Metodologia de Classificação

Algumas características físicas das flores são importantes, pois permitem distinguir uma espécie de flor de outra, como padrões de cor, forma e textura. A dificuldade de se

Figura 1. Imagens que compõem a base de dados.



(a)

(b)

(c)

Fonte: [Chaudhary 2020].

implementar sistemas de Visão Computacional para essa tarefa é encontrar descritores adequados que permitam descrever as imagens de forma que o classificador entenda as características únicas entre as classes de dados.

No método proposto, a extração de características é realizada pela própria arquitetura de Rede Neural Profunda. A CNN é composta por duas etapas base, sendo elas (i) a convolução e (ii) o reconhecimento de padrões. A convolução é a etapa que extrai características visuais das imagens. O reconhecimento de padrões utiliza de filtros como descritores da imagem para aprender a reconhecer novas entradas.

Neste trabalho foram utilizadas duas arquiteturas de CNN: a arquitetura tradicional LeNet-5 e a arquitetura de CNN altamente profunda, ResNet152. O objetivo do emprego dessas arquiteturas foi avaliar a eficiência de classificação nesse problema multiclasse e verificar se há ganhos comparado ao uso de um modelo tradicional e um modelo mais robusto, composto por mais camadas ocultas e diferentes técnicas de convolução.

A arquitetura proposta por [LeCun et al. 1995] é mais simples e menos custosa quando comparada à arquitetura altamente profunda que veremos a seguir. A LeNet-5 é composta por três camadas de convolução, seguidas por uma camada de subamostragem (*Pooling*) cada, uma camada de achatamento (*Flatten*), uma camada totalmente conectada e uma camada de saída. As camadas de convolução e de subamostragem são responsáveis pela extração de características e representação das imagens no formato de filtros. Esses filtros são convertidos posteriormente em vetores descritivos pela camada de achatamento e são enviados as camadas seguintes para o reconhecimento de padrões.

Para este trabalho, a LeNet-5 é configurada como se segue. A camada de entrada recebe as imagens da base de dados no tamanho 64x64 pixels, com 3 canais de cores. As três camadas de convolução geram 32, 64 e 128 filtros e utilizam a função de ativação *Rectifier Linear Unit* (ReLU). O tamanho da janela das camadas de *Max-Pooling* é de 2x2. Posteriormente, duas camadas densas totalmente conectadas utilizam 1048 neurônios cada, as quais possuem acoplado a operação de *dropout*, responsável por desligar aproximadamente 20% da Rede Neural, auxiliando na criação de modelos de classificação mais generalizados. Por fim, a camada de saída é composta por cinco neurônios, utilizando a função de ativação softmax.

Devido a natureza multiclasse do problema abordado neste trabalho, a mesma ta-

refa de classificação também é avaliada na arquitetura ResNet152, proposta pelos pesquisadores [Wu et al. 2019]. Esse tipo de arquitetura é mais robusta e emprega a abordagem "altamente profunda", por trazerem mais camadas ocultas e sequências de convolução, percorrendo a imagem formando filtros descritivos com alta performance para detecção de objetos e pequenos padrões na imagem. Inicialmente, o algoritmo recebe as imagens da base de dados pela camada de entrada da Rede Neural. Logo após, o algoritmo realiza as etapas de convolução, conforme a arquitetura utilizada, cria os filtros descritivos e achata a matriz em um vetor único. Por fim, esse vetor é entregue a uma camada de Rede Neural Totalmente Conectada, a qual irá buscar por padrões visuais.

A ResNet152 utilizada nesse trabalho foi configurada como se segue. A camada de entrada carrega as imagens no tamanho 224x224, em três canais de cores. São utilizadas 152 camadas residuais, compostas por camadas de convolução, de *MaxPooling* e densas. As camadas de convolução e densas usam a função de ativação *Rectifier Linear Unit* (ReLU). Por fim, a camada de saída contém cinco neurônios correspondentes as classes do problema deste trabalho.

Em ambas as arquiteturas, o tamanho do *batch* utilizado foi oito. A otimização do método foi realizada utilizando o algoritmo de *adam* e o cálculo da perda foi realizado utilizando a função *Categorical Crossentropy*.

5. Experimentos e Resultados

Para o desenvolvimento e avaliação do modelo de classificação foram utilizados um computador com um processador AMD Ryzen 7, 16 GB de memória RAM, uma GPU NVIDIA GTX 1060 e um sistema operacional Windows 10 x64. A implementação da rede neural foi feita utilizando a linguagem de programação *Python*, na versão 3.7, o *framework Keras* na versão 2.1.0 e também o *framework Tensorflow*, na versão 2.3.

Os métodos foram avaliados seguindo o protocolo de experimentação *Holdout Cross-Validation*. Neste protocolo, 80% da base de dados foi utilizada para o treinamento da base e 20% para a validação. Nesta separação, as imagens que compuseram cada *subdataset* possuíam amostras únicas e portanto, o treinamento e a validação dos métodos não foram realizados em imagens iguais.

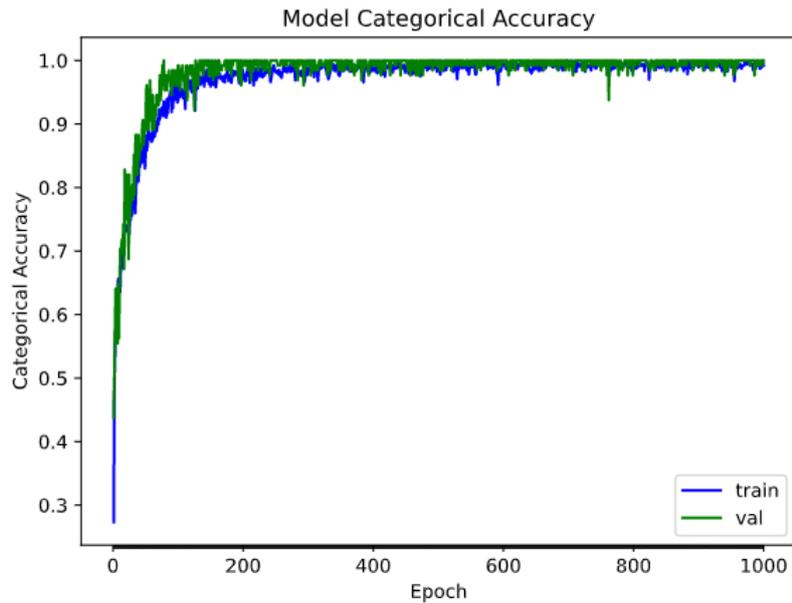
Para medir a performance da rede, nos experimentos envolvendo a LeNet-5 e a ResNet152, foram calculadas as seguintes métricas: acurácia categórica, precisão, sensibilidade e AUC. A acurácia categórica é uma medida de avaliação que nos informa sobre a taxa de acertos total do método de classificação. A precisão, por sua vez, nos diz o quão relevante são as classificações realizadas pelo algoritmo levando em consideração não apenas o número verdadeiros positivos mas também os falso negativos. Já a revocação nos informa qual a proporção de positivos reais que tiveram êxito em sua identificação, levando em conta os números verdadeiros positivos e os falsos negativos. Por fim, a AUC representa o grau de separabilidade e diz o quanto o modelo é capaz de distinguir classes. Quanto maior a AUC, melhor será o modelo de previsão. Por analogia, quanto maior a AUC, melhor o modelo distingue as espécies de flores. Os resultados obtidos em cada uma destas métricas podem ser observados na Tabela tab:resultados, abaixo.

Fazendo uma análise aprofundada no comportamento da curva da taxa de acurácia e perda do algoritmo LeNet-5, podemos observar que a acurácia da validação se mantém

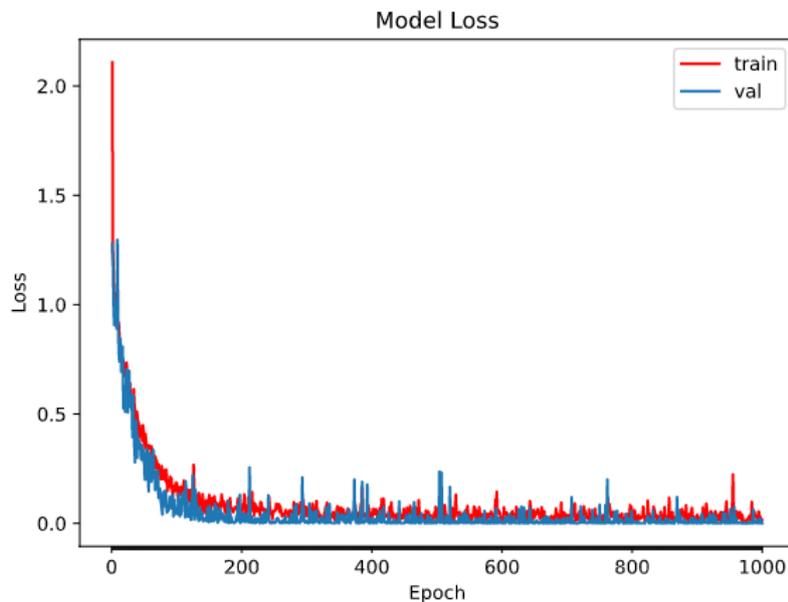
Tabela 1. Resultados obtidos na validação dos métodos.

Método	Acurácia Categórica	Precisão	Revocação	AUC
LeNet-5	99,70%	99,00%	99,10%	99,00%
ResNet152	80,45%	84,39%	78,51%	96,04%

Figura 2. Taxas de acurácia e de erro ao longo do experimento.



(a)



(b)

alta e próxima à curva do treinamento enquanto a taxa de erro na validação se mantém baixa e igualmente estável. Esse comportamento é tipicamente visto quando o modelo de classificação obteve padrões genéricos que permitem avaliar com um elevado grau de

certeza entradas não vistas anteriormente. Dessa forma, é possível dizer que não houve sobreajuste (ou *overfitting*) do modelo à base de dados. A Figura 2 contém ambos os gráficos da taxa de acurácia (Figura 2(a)) e da taxa de erro ou perda (Figura 2(b)) durante o experimento.

Com os resultados apresentados, podemos constatar que, embora mais simples, a Rede Neural Convolutiva seguindo a arquitetura de LeNet-5 é suficiente para detectar padrões e construir um classificador de flores, em baixa escala, como o problema de cinco classes proposto nesse artigo. Devido a baixa quantidade de amostras e a complexa estrutura da ResNet152, a LeNet-5 foi mais eficiente em detectar padrões e construir um modelo de classificação.

6. Conclusão

A classificação de espécies de flores é um trabalho árduo, porém não necessita de uma rede neural complexa para ter um ótimo resultado. A utilização da base de dados com imagens sem segmentação, nos permitiu usar a Visão Computacional para extrair os dados e através do Aprendizado Profundo detectar padrões nesses dados e assim gerar conhecimento para uma Inteligência Artificial fosse capaz de classificar espécies de flores.

Neste trabalho apresentamos dois algoritmos de classificação que obtiveram resultados bem distintos um do outro, sendo eles o ResNet152 e o LeNet-5. Comparamos o tempo de treinamento com suas respectivas métricas e concluímos que a Rede Neural LeNet-5 é a mais adequada para ser utilizada na tarefa de classificação de espécies de flores.

Como trabalhos futuros, sugerimos o desenvolvimento de um algoritmo de busca de flores pela internet, de espécies distintas das utilizadas neste trabalho, com o intuito de alimentar o conhecimento da rede neural, ampliando a sua utilização.

Agradecimentos

Os autores agradecem ao IF SudesteMG e ao Programa de Educação Tutorial (PET - Conexões de Saberes Ciência da Computação) pelo apoio técnico e financeiro para o desenvolvimento do projeto.

Referências

- Azevedo, E. and Conci, A. (2003). *Computação Gráfica: Teoria e Prática*, volume 2. Elsevier.
- Backes, A. R. and de Mesquita Sá Junior, J. J. (2019). *Introdução à visão computacional usando Matlab*. Alta Books Editora.
- Chaudhary, R. (2020). Final flowers course project dataset. Disponível em: <<https://www.kaggle.com/rishitchs/final-flowers-course-project-dataset/metadata>>. Acesso em: 21 nov. 2020.
- Fernandes, M. P., Sato, J. R., Busatto Filho, G., and Thomaz, C. E. (2012). Classificação estatística e predição da doença de alzheimer por meio de imagens médicas do encéfalo humano. *Avanços da Visão Computacional*. Omnipax Editora.
- Hiary, H., Saadeh, H., Saadeh, M., and Yaqub, M. (2018). Flower classification using deep convolutional neural networks. *IET Computer Vision*, 12(6):855–862.

- Joly, A., Goëau, H., Bonnet, P., Bakić, V., Barbe, J., Selmi, S., Yahiaoui, I., Carré, J., Mouysset, E., Molino, J.-F., et al. (2014). Interactive plant identification based on social image data. *Ecological Informatics*, 23:22–34.
- Kumar, N., Belhumeur, P. N., Biswas, A., Jacobs, D. W., Kress, W. J., Lopez, I. C., and Soares, J. V. (2012). Leafsnap: A computer vision system for automatic plant species identification. In *European Conference on Computer Vision*, pages 502–516. Springer.
- LeCun, Y., Bengio, Y., et al. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995.
- Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M., and Van Valen, D. (2019). Deep learning for cellular image analysis. *Nature methods*, pages 1–14.
- Neves, L. A. P., VIEIRA NETO, H., and Gonzaga, A., editors (2012). *Avanços em Visão Computacional*. Omnipax, Curitiba, PR, 1 edition. 406 p.
- Nieto, D., Brill, A., Feng, Q., Humensky, T., Kim, B., Miener, T., Mukherjee, R., and Sevilla, J. (2019). Ctlearn: Deep learning for gamma-ray astronomy. *arXiv preprint arXiv:1912.09877*.
- Oyama, P. d. C., JORGE, L. d. C., Rodrigues, E. L. L., and Gomes, C. C. (2012). Sistema para classificação automática de café em grãos por cor e forma através de imagens digitais. *Embrapa Instrumentação-Capítulo em livro científico (ALICE)*.
- Wu, S. G., Bao, F. S., Xu, E. Y., Wang, Y.-X., Chang, Y.-F., and Xiang, Q.-L. (2007). A leaf recognition algorithm for plant classification using probabilistic neural network. In *2007 IEEE international symposium on signal processing and information technology*, pages 11–16. IEEE.
- Wu, Z., Shen, C., and Van Den Hengel, A. (2019). Wider or deeper: Revisiting the resnet model for visual recognition. *Pattern Recognition*, 90:119–133.