

Sistema de Reconhecimento Automático de Pássaros da Fauna de Linhares - ES Utilizando Redes Neurais Convolucionais

Brenda Bayerl Pinheiro*. Lucas de Assis Soares**

Instituto Federal de Educação, Ciência e Tecnologia do Espírito Santo – IFES, Campus Linhares, ES

**(e-mail: bren dabayerl@outlook.com).*

*** (e-mail: lucas.soares@ifes.edu.br)*

Abstract: Faced with the threat and extinction of some of the bird species that live in the state of Espírito Santo and the lack of incentive tools that encourage bird watching, environmental education and ecotourism, this work, inspired by the idea of building a facilitating and accessible tool for bird watching, proposes the recognition of images of birds using artificial intelligence, through convolutional neural networks. For this purpose, an image database with some birds from the Espírito Santo region with 64 bird species observed in the institution where this work was developed was built, seeking the success in the accuracy of the recognition of those chosen species, which was possible to assess through the results and analysis of the accuracy. In addition, to compare and verify the results, other bird image databases were evaluated and the GradCAM method was applied to visualize the region which were more important in the classification performed by the convolutional neural network. The applied network obtained a good performance, which allows a possible application in environmental monitoring systems and assistance in studies in the area of intelligent image recognition.

Keywords: Bird species recognition; convolutional neural networks; deep learning; artificial neural networks; computer vision.

Resumo: Diante da ameaça de extinção de algumas das espécies de pássaros que habitam no estado do Espírito Santo, e da falta de ferramentas incentivadoras que estimulem a observação das aves, a educação ambiental e o ecoturismo, este trabalho, inspirado na ideia de construir uma ferramenta facilitadora e acessível para observação dos pássaros, propõe o reconhecimento de imagens de pássaros utilizando a inteligência artificial, através das redes neurais convolucionais. Com esse objetivo, foi construído um banco de imagens de alguns pássaros da região capixaba com 64 espécies avistadas na instituição onde foi desenvolvido o trabalho, buscando o sucesso na acurácia do reconhecimento das espécies escolhidas, que foi possível aferir através dos resultados e análises das acurácias. Além disso, para comparar e averiguar os resultados, foram testados outros bancos de imagens de pássaros e aplicado o método GradCAM a fim de visualizar as regiões de maior importância na classificação feita pela rede neural convolucional. A rede aplicada obteve um bom desempenho, o que permite uma possível aplicação em sistemas de monitoramento ambiental e auxílio em estudos na área do reconhecimento inteligente de imagens.

Palavras-chaves: Reconhecimento de espécies de pássaros; redes neurais convolucionais; aprendizado profundo; redes neurais artificiais; visão computacional.

1. INTRODUÇÃO

No estado do Espírito Santo, especificamente na cidade de Linhares, existem famílias de aves com as maiores riquezas de espécies na Floresta Nacional (FLONA) de Goytacazes e em sua Zona de Amortecimento. Nota-se, portanto, que a área de estudo possui uma evidente relevância para a conservação ambiental, pois abriga, segundo as fontes de dados primários e secundários com relação à fauna de pássaros, os seguintes elementos especiais da fauna de pássaros (táxons relevantes): 14 espécies ameaçadas de extinção (5 em nível nacional, 12 em nível regional e 3 em nível nacional/regional); 90 espécies dependentes do ambiente florestal; 16 espécies endêmicas da Mata Atlântica;

e; 38 espécies sob pressão de caça e captura (ICMBIO, 2013).

De acordo com esses dados, tem-se que a ameaça de espécies de pássaros, os quais compõem a diversidade e a riqueza da fauna regional, reflete a desvalorização da importância da preservação das aves do estado, intensificando o desequilíbrio ambiental e, por consequência, ameaçando a vida do ser humano no planeta, pois os pássaros possuem um papel importante na cadeia alimentar do ecossistema. Eles são polinizadores e dispersantes de sementes, participam do controle biológico das pragas, se alimentam de insetos e pequenos animais, entre outras contribuições para a manutenção saudável da vida na Terra. Por esses motivos, a

educação ambiental é uma imprescindível ferramenta para o desenvolvimento da sensibilização da sociedade sobre a importância da preservação. Por meio dela, busca-se uma mudança de atitude da população em relação às questões ambientais (Vale, 2019). A educação ambiental se dá principalmente através da informação, que é um meio facilitador de aprendizado, e pode ser proporcionada, no caso das aves, por meio do conhecimento das espécies, pois são mais fáceis de monitorar do que outras espécies de animais (Qiao et al., 2017), e da observação de pássaros, que além de contribuir com a educação ambiental, também consiste em uma forma de turismo. Nesse sentido, uma possível ferramenta para oportunizar e incentivar o conhecimento e a observação das aves são aplicativos, que mostram as diferentes espécies e suas respectivas características, sendo um instrumento acessível, rápido e gratuito. Dessa forma, torna-se interessante desenvolver um sistema de visão computacional capaz de, a partir de imagens digitais, reconhecer a espécie presente em cada imagem, além de poder ser usado em sistemas de monitoramento ambiental, e auxiliar em estudos que buscam reconhecer as espécies presentes em determinadas regiões.

Existem diversos trabalhos na literatura científica que tratam do reconhecimento de espécies de pássaros por meio de técnicas de visão computacional. Marini et al. (2013) propuseram uma técnica para a classificação de espécies de pássaros com base em características de cores extraídas de imagens naturais, com diferentes poses, tamanhos, ângulos de visão, iluminação e oclusão por outras partes do cenário. A partir disso, foi aplicado um algoritmo de segmentação de cores para delimitar regiões candidatas e, em seguida, outro algoritmo para analisar os histogramas de cores das regiões candidatas. De maneira similar, Jian et al. (2014) desenvolveram um sistema de identificação de pássaros com base na análise de imagens. O método consiste em remover o fundo da imagem e então aplicar histogramas de cada canal da imagem do pássaro isolado do fundo.

Roslan et al. (2017) propuseram um método próprio para dispositivos móveis, com baixo custo computacional e uma interface de fácil utilização para a classificação de duas espécies de pássaros. O fizeram baseados em atributos estatísticos de baixa ordem de características de cores aplicados em uma máquina de vetores de suporte após a segmentação do pássaro por um detector de bordas seguido de operadores morfológicos. Qiao et al. (2017) apresentaram um sistema de classificação de pássaros baseado em imagens no espaço de cores HSV, no qual as características utilizadas eram os momentos de cor, uma característica específica de espaço de cores a partir da distribuição das mesmas, a cor da cabeça do pássaro, os momentos invariantes e uma característica calculada a partir da identificação do bico da ave. O classificador utilizado foi uma máquina de vetores de suporte associado a uma árvore de decisão.

Nos últimos anos, as redes neurais convolucionais têm se destacado em aplicações de visão computacional, alcançando o estado da arte na classificação e o reconhecimento de objetos em imagens digitais. Neste sentido, Haobin et al. (2018) propuseram uma técnica de classificação de espécies de

pássaros usando informações extraídas de uma rede neural convolucional de arquitetura VGG-16 (Simonyan & Zisserman, 2014) aplicadas em uma análise de componentes principais para redução de dimensionalidade seguido por um modelo bilinear para representação das características. Os autores criaram um banco de dados com imagens de pássaros chineses e também avaliaram o sistema no banco de dados Caltech-UCSD Birds 200 - CUB-200 (Welinder et al., 2010). Bold et al. (2019) desenvolveram um sistema de classificação de espécies de pássaros baseados em dados de imagens, usando as imagens do banco de dados CUB-200, e de áudios dos pássaros. O treinamento é feito usando uma rede neural convolucional do tipo CaffeNet (Deng et al., 2014), sendo que o treinamento dos dados de imagens e de áudios são feitos separadamente e, posteriormente, combinados em uma técnica de Aprendizado de Kernel Múltiplo (MKL – *multiple kernel learning*).

Niemi e Tantu (2020) também apresentaram um sistema de identificação automática de espécies de pássaros, mas considerando as diferenças de quantidades de dados entre cada classe. O sistema utiliza redes neurais convolucionais em uma estrutura hierárquica, em que, primeiramente os grupos de espécies parecidas são encontrados e, posteriormente, classificadores em cascata são aplicados para encontrar as espécies dentro de cada grupo.

Com base nos avanços das redes neurais convolucionais, este trabalho tem por objetivo desenvolver um sistema de classificação automática de pássaros a partir de imagens digitais utilizando essas estruturas. A partir da criação de um banco de dados de imagens, chamado IFESBirds, em que foram utilizadas imagens de alguns dos pássaros existentes no município de Linhares, no Espírito Santo, localidade onde esse trabalho foi desenvolvido, a acurácia no reconhecimento das espécies é avaliada e comparada com a acurácia obtida no banco de imagens Caltech-UCSD Birds 200 (Welinder et al., 2010), chamada CUB-200-2011, e em uma base de imagens com 265 diferentes espécies, disponível na plataforma Kaggle (Kaggle, 2021), que aqui será chamado de KaggleBirds.

O restante do texto está estruturado da seguinte forma: na Seção 2 são apresentados como procedeu-se a criação do banco de dados, e as características e desenvolvimento da rede; em sequência, na Seção 3, são mostrados os resultados dos treinamentos, comparações com outros bancos de dados e a aplicação da rede em um método que apresenta os resultados de forma a serem vistos na prática; por fim, na Seção 4, são apresentadas as conclusões do trabalho e indicações para trabalhos futuros.

2. METODOLOGIA

2.1 Levantamento das espécies de pássaros regionais e a criação do banco de imagens

Para conhecer e quantizar algumas das espécies de pássaros presentes no Espírito Santo, uma lista das espécies avistadas no Instituto Federal do Espírito Santo em Linhares, campus no qual esse trabalho foi desenvolvido, foi validada a partir

de diversos documentos com levantamento de espécies presentes em reservas e florestas próximas, a saber, a Floresta Nacional dos Goytacazes, a Reserva Natural da Vale e a Reserva de Sooretama.

O banco de imagens IFESBirds foi formado por 64 espécies e, para cada uma, foram obtidas 64 imagens, em média, sendo a maioria delas de repositórios de imagens gratuitos na internet e outras cedidas por um colaborador. As Figuras 1, 2, e 3, a seguir, apresentam alguns exemplos dessas imagens. Nessas figuras, é possível observar as grandes variações que podem existir em termos de cor, iluminação e escala, além da presença de mais do que um único pássaro e outros objetos na imagem.



Figura 1: Tesourinha (*Tyrannus savana*)

A partir do banco de dados completo, as imagens foram separadas em um conjunto de treinamento, com 3584 imagens utilizadas para treinar o algoritmo e criar um modelo preditivo, e outro conjunto de teste, com 512 imagens para validar a performance modelo já treinado, juntamente com os nomes das respectivas espécies, pois trata-se de um processo de aprendizado supervisionado.



Figura 2: Corrupião (*Icterus jamaicensis*)

2.2 Arquitetura da rede neural convolucional

Para desenvolver o sistema e fazer com que a diferenciação das imagens ocorra de forma satisfatória, diante da

diversidade de cores presente nos pássaros e no fundo das imagens, posições aleatórias das aves e as variações de iluminação ambiente, as redes neurais convolucionais foram escolhidas para que o sistema pudesse, diretamente a partir dos dados, extrair as melhores características. Além disso, destaca-se a robustez dessas estruturas frente a variações de translação, rotação, escala ou mesmo iluminação (LeCun et al., 2015, apud da Silva et al.), sendo de uso promissor quando se trata de reconhecimento de imagens com as mais distintas características e formatações.



Figura 3: Beija-flor-de-banda-branca (*Amazilia versicolor*)

Redes neurais convolucionais utilizam operações de convolução para extrair diferentes características das imagens que contribuem para a discriminação das diferentes classes de acordo com o conjunto de treinamento (Data Science Academy, 2021). Essas redes são normalmente constituídas de diversas camadas, de modo a extrair características em diferentes níveis de complexidade. As operações de convolução são combinadas com funções de ativação não-lineares, fazendo com que essas camadas funcionem como bancos de filtros não-lineares. Além das camadas que realizam as operações de convolução, chamadas de camadas convolucionais, as redes neurais convolucionais também possuem camadas de *pooling*, ou agrupamento, utilizadas para redução do volume de saída, acelerando o treinamento (Masci et al., 2013, apud da Silva et al., p.3). Após a extração das características, normalmente são aplicadas camadas totalmente conectadas que utilizam as características extraídas pelas camadas convolucionais e camadas de *pooling* para fazer o reconhecimento das classes propriamente dito.

Neste trabalho, foi usado um processo de transferência de aprendizado, no qual o modelo ResNet101V2 (He et al., 2016a, b; Tensorflow, 2021), que tem como característica a utilização de aprendizado residual em que as informações de diversos níveis da rede são combinadas visando um melhor desempenho na classificação, utilizado como um extrator de características, foi pré-treinado com as imagens da base de dados ImageNet (Deng et al., 2009). Essa arquitetura foi escolhida por sua acurácia elevada nessa base de imagens, porém futuramente outras arquiteturas podem ser avaliadas. Conectadas à estrutura convolucional do modelo ResNet101V2, camadas totalmente conectadas combinadas com operações de abandono de neutônios (Srivastava, Nitish et al., 2014) e normalização em lotes (Ioffe, 2015; Szegedy,

2015), mais conhecidas na literatura científica pelos termos em inglês, *dropout* e *batch normalization*, respectivamente, foram sintonizadas para fazer o reconhecimento das espécies de pássaros. A função de ativação utilizada em todas as camadas totalmente conectadas, com exceção da camada de saída, foi a função de retificação linear, ReLU (He et al., 2015). Para a camada de saída, que contém uma quantidade de neurônios igual à quantidade de classes, foi escolhida a função de ativação Softmax, que fornece um valor de ativação percentual, de acordo com as ativações de todos os neurônios dessa camada (Goodfellow, 2016; Bengio, 2016; Courville, 2016). Todos os pesos foram inicializados utilizando a inicialização uniforme de He (He et al., 2015). A Figura 4 apresenta a sequência de camadas da arquitetura, juntamente com a quantidade de neurônios em cada camada totalmente conectada. Para a função de abandono, uma taxa de eliminação de neurônios de 50% foi adotada.

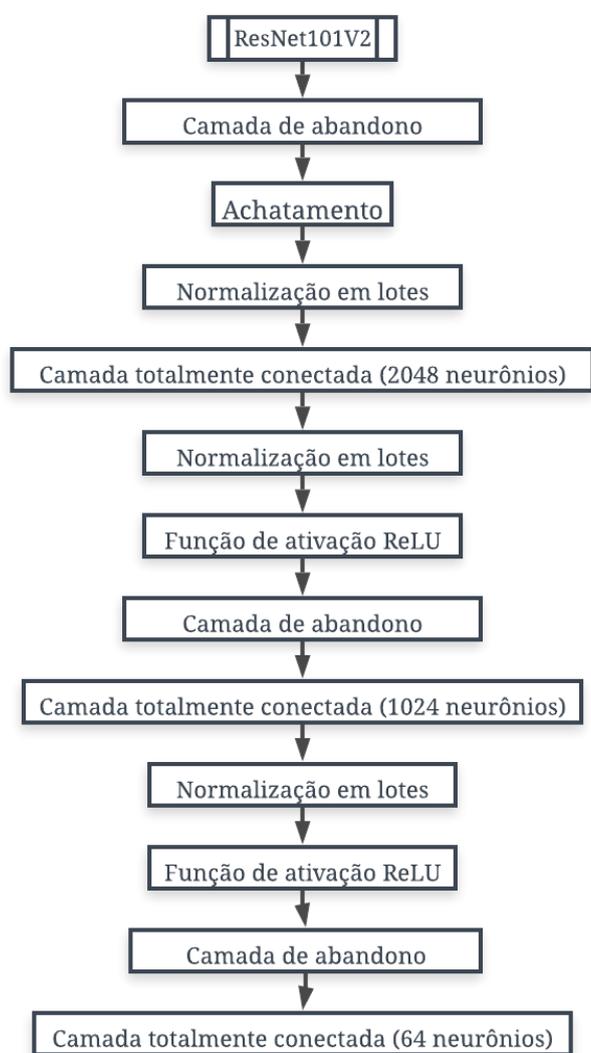


Figura 4: Fluxograma da arquitetura.

2.3 Treinamento da rede neural convolucional

Para o treinamento da rede neural convolucional, foi implementado um procedimento de aumento de imagens (Perez, 2017; Wang, 2017) que, com o objetivo de tornar o conjunto de imagens de treinamento mais diversificado e a rede mais robusta, executa pequenas deformações nas imagens, idealmente sem distorcer o objeto principal, que nesse caso são os pássaros, a ponto de torná-lo irreconhecível, acrescentando variabilidade e ruído nos dados e, assim, permitindo um maior desempenho na classificação das imagens.

As distorções aplicadas são:

- Inversão horizontal aleatória;
- Inversão vertical aleatória;
- Aplicação de um *zoom* aleatório;
- Aplicação de um cisalhamento aleatório; e
- Aplicação de uma rotação aleatória.

Além do procedimento de aumento, um redimensionamento das imagens para o tamanho 224 x 224, utilizado na ResNet101V2, também foi aplicado. As imagens também foram normalizadas para que o menor e maior valores da intensidade de cada canal de cor fosse zero e um, respectivamente.

Com a arquitetura e o processo de aumento de imagens definidos, iniciou-se o treinamento da rede, que é o processo em que os dados são apresentados ao algoritmo de aprendizado supervisionado, a fim de gerar o ajuste dos parâmetros. O algoritmo de treinamento utilizado foi o Adam (Kingma, 2014; Ba, 2014), escolhido por sua popularidade em aplicações envolvendo redes com muitas camadas, com os parâmetros padrões do TensorFlow. A função de custo utilizada foi o erro de entropia cruzada categórico, que considera a classificação multiclasse, e os resultados obtidos foram avaliados por meio de sua acurácia, que representa o percentual de acertos de classificação correta dentre todas as amostras.

Como mencionado na Seção 1, além do banco de imagens construído, IFESBirds, também foram utilizados, para fins de comparação, os bancos de imagens de pássaros CUB-200-2011 e uma base de imagens com 265 diferentes espécies, disponível na plataforma Kaggle (Kaggle, 2021), chamado aqui de KaggleBirds, sendo feito um treinamento separado para cada um desses bancos de imagens.

3. RESULTADOS E DISCUSSÕES

3.1 Acurácia das redes neurais convolucionais

O treinamento da rede neural convolucional para a classificação das espécies do banco de imagens IFESBirds foi feito em 50 épocas. A Figura 5 apresenta a evolução da acurácia para os conjuntos de treinamento e validação durante as épocas de treinamento.

Observa-se, na Figura 5, que a acurácia do conjunto de validação é menor do que aquela do conjunto de treinamento, apesar do procedimento de aumento das imagens sobre esse conjunto, e também que o treinamento ainda não parece ter

atingido um ponto de equilíbrio. Todavia, o mesmo foi interrompido na 50ª época, pois há indícios de início do sobreajuste dos parâmetros, uma vez que a acurácia do conjunto de validação começa a diminuir.

A Tabela 1 apresenta as acurácias das redes neurais convolucionais dos bancos de imagens IFESBirds, CUB-200-2011 e KaggleBirds. O treinamento da rede utilizando o banco de imagens CUB-200-2011 foi feito em 75 épocas, enquanto o treinamento da rede utilizando o banco de imagens KaggleBirds foi feito em 80 épocas. É importante lembrar que esses bancos de imagens possuem mais imagens do que o banco IFESBirds, o que justifica a utilização de uma maior quantidade de épocas de treinamento.

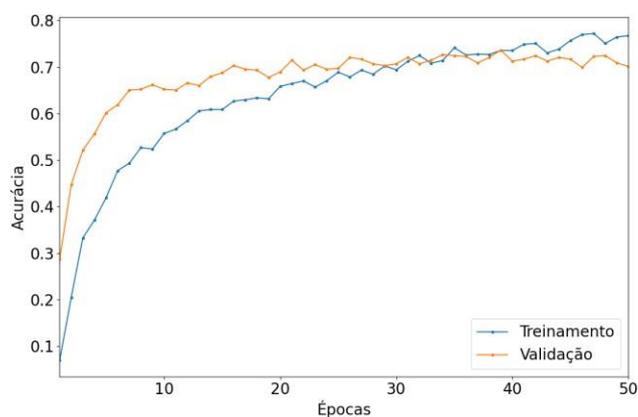


Figura 5: Acurácias de treinamento e validação em função das épocas.

Tabela 1. Acurácias dos conjuntos de treinamento e validação para os diferentes bancos de imagens.

Banco de imagens	Acurácia do conjunto de treinamento	Acurácia do conjunto de validação
IFESBirds	76,74%	70,12%
CUB-200-2011	65,79%	56,80%
KaggleBirds	92,76%	96,64%

Os resultados da Tabela 1 mostram que o conjunto de imagens KaggleBirds obteve uma acurácia superior à dos demais bancos de imagens. Conjectura-se que esse resultado se deve a três fatores. Primeiramente, esse banco de imagens contém um número maior de imagens do que os demais, com 36.465 imagens no momento de sua utilização nesse trabalho, em comparação às 11.788 imagens do banco CUB-200-2011 e 4.096 imagens do banco IFESBirds. Sabe-se que bancos de imagens mais extensos permitem melhores resultados às técnicas de aprendizado profundo e uma menor possibilidade de sobreajuste em redes com uma grande quantidade de parâmetros. Além disso, essas imagens contêm um único pássaro em cada imagem, diferentemente do banco de dados IFESBirds, nas quais mais do que um único indivíduo pode aparecer na imagem. Finalmente, nesse banco de imagens os pássaros aparecem no centro das imagens e, para uma determinada espécie, não há grande variação nas poses e escalas na dimensão das imagens dos pássaros. Isso não

acontece nos bancos de imagens IFESBirds e CUB-200-2011, os quais apresentam grande variabilidade de escala, com imagens de pássaros em movimento, voando ou parados, e de poses, sendo que, para uma mesma espécie, podem aparecer diferentes partes dos pássaros e os mesmos aparecem em diferentes ângulos.

Dessa forma, é esperada uma acurácia menor para os bancos de imagens IFESBirds e CUB-200-2011. Adicionalmente, o melhor desempenho no banco KaggleBirds reforça a capacidade das redes neurais convolucionais na classificação das espécies de pássaros de forma satisfatória, porém também mostra que um sistema de detecção dos pássaros na imagem, anterior ao seu reconhecimento, é capaz de melhorar o desempenho de classificação no caso de imagens com grandes variações nas posições e escalas dos pássaros.

3.2 Aplicação da técnica de GradCAM sobre a rede neural convolucional

Muito embora a utilização de redes neurais convolucionais em problemas de visão computacional representem o estado da arte em diferentes aplicações, essas estruturas, assim como as redes neurais artificiais e outras estruturas desse tipo, apresentam a desvantagem de funcionarem como modelos do tipo caixa-preta, nos quais não é possível saber com precisão os processos internos que levaram a rede a classificar uma imagem como uma determinada classe.

Nesse sentido, diversas técnicas têm sido propostas para tornar essas estruturas menos opacas, permitindo o seu entendimento por seres humanos, bem como o diagnóstico de possíveis falhas e apontamento de soluções. Para aplicações com redes neurais convolucionais, uma das técnicas mais populares chama-se Grad-CAM (Selvaraju, 2017), que tem o objetivo de produzir explicações visuais para as decisões dessas redes.

A técnica consiste em utilizar os gradientes de uma determinada classe na rede, que se direcionam para a última camada convolucional, a fim de produzir um mapa de localização que destaca as regiões mais importantes para a escolha da classe na imagem original.

Dessa forma, a técnica de Grad-CAM foi aplicada à rede neural convolucional desenvolvida para fazer o reconhecimento das espécies de pássaros do banco de imagens de espécies locais, IFESBirds, e os resultados sobre as imagens das Figuras 1, 2 e 3 são apresentados nas Figuras 6, 7 e 8, respectivamente. É importante ressaltar que, para as espécies das Figuras 2 e 3, a classificação foi correta, enquanto para a espécie da Figura 1, houve erro de classificação. Nessas imagens, a importância de cada região da imagem é representada por um mapa de calor, em que o vermelho e o azul significam, respectivamente, maior e menor importância.



Figura 6: Técnica de Grad-CAM sobre a espécie da Figura 1 (*Tyrannus savana*).

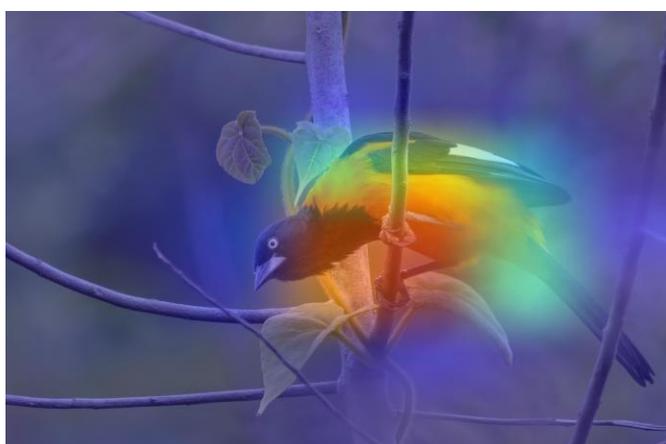


Figura 7: Técnica de Grad-CAM sobre a espécie da Figura 2 (*Icterus jamacaii*).



Figura 8: Técnica de Grad-CAM sobre a espécie da Figura 3 (*Amazilia versicolor*).

A análise das Figuras 6, 7 e 8 mostra que a rede neural convolucional está, de fato, localizando o pássaro na imagem para fazer sua detecção. Todavia, na presença de mais pássaros em uma mesma imagem, como mostrado na Figura 6, a rede foi incapaz de encontrar as outras aves. Esse resultado também é observado em outras imagens do banco

de imagens que contêm mais do que um pássaro, o que reforça a discussão feita anteriormente sobre os resultados da Tabela 1, a respeito do resultado pior em comparação ao banco de imagens KaggleBirds.

4. CONCLUSÃO

Este trabalho apresentou uma rede neural convolucional utilizada para fazer o reconhecimento automático de espécies de pássaros em imagens digitais. Para isso, um banco de imagens contendo 64 espécies locais, chamado IFESBirds, foi criado e a rede foi avaliada nesse banco de imagens, além de um banco de imagens de pássaros disponíveis na plataforma Kaggle e o banco de imagens Caltech-UCSD Birds-200- 2011. A rede neural convolucional utilizada apresentou uma acurácia de 70,12%, 96,64% e 56,80% para os conjuntos de validação dos bancos de imagens IFESBirds, KaggleBirds e CUB-200-2011.

Conjectura-se que a diferença de desempenho se deve principalmente à forma como as aves se apresentam em cada um dos bancos de imagens. Dessa forma, como indicação para trabalhos futuros, propõe-se a utilização de um sistema de detecção dos pássaros na imagem, previamente, antes do seu reconhecimento, de modo a reduzir essas influências.

AGRADECIMENTOS

Os autores agradecem ao Gabriel Scaldaferrero Bonfa, pela disponibilização de algumas imagens de espécies de pássaros utilizadas na construção do banco de dados, e ao professor Claudinei Gonçalves Pereira, do IFES Campus Linhares, pela indicação das espécies observadas na instituição.

REFERÊNCIAS

- Araújo, Aline Moura. Classificação e detecção de pessoas em ambientes não controlados utilizando redes neurais convolucionais. 2019. 46f. Monografia (Graduação em Ciência da Computação) – Universidade Federal da Paraíba, João Pessoa, 2019. Disponível em: <https://repositorio.ufpb.br/jspui/handle/123456789/16098>. Acesso em: 27 de maio de 2021.
- Bold, Naranchimeg; Zhang, Chao; Akashi, Takuya. Bird Species Classification with Audio-Visual Data using CNN and Multiple Kernel Learning. In: 2019 International Conference on Cyberworlds (CW). IEEE, 2019. p. 85-88.
- Da Silva, Luiz Augusto Zillmann et al. Reconhecimento de componentes em linhas férreas utilizando redes neurais convolucionais. Juiz de Fora, v. 1, n. 1, 2020. Disponível em: https://www.sba.org.br/open_journal_systems/index.php/sba/article/view/905/847>. Acesso em: 12 de maio de 2021.
- Data Science Academy. Deep Learning Book, 2021. Disponível em: <https://www.deeplearningbook.com.br/reconhecimento->

- de-imagens-com-redes-neurais-convolucionais-em-python-parte-1/. Acesso em: 12 de maio de 2021.
- Deng, Jia et al. Imagenet: A large-scale hierarchical image database. Em: 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009. p. 248-255.
- Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron. 6.2.2.3 Softmax Units For Multinoulli Output Distributions. Deep Learning. MIT Press. pp. 180–184. 2016. ISBN 978-0-26203561-3.
- Haobin, Shi; Renyu, Zhang; GANG, Sun. Fine-grained bird classification based on low-dimensional bilinear model. In: 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC). IEEE, 2018. p. 424-428.
- He, Kaiming et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision. 2015. p. 1026-1034.
- He, Kaiming et al. Deep residual learning for image recognition. Em: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016a. p. 770-778.
- He, Kaiming et al. Identity mappings in deep residual networks. In: European conference on computer vision. Springer, Cham, 2016b. p. 630-645.
- ICMBIO. Plano de Manejo da Floresta Nacional de Goytacazes: Volume I - Diagnóstico. Vitória, 2013. 143p. Disponível em: https://www.icmbio.gov.br/portal/images/stories/docs-planos-de-manejo/volumeI-diagnostico_flon_a_Goytacazes.pdf. Acesso em: 12 de maio de 2021.
- Ioffe, Sergey; Szegedy, Christian. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning. PMLR, 2015. p. 448-456.
- Jia, Yangqing et al. Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM international conference on Multimedia. 2014. p. 675-678.
- Jian, Li; Lei, Zhang; Baoping, Yan. Research and application of bird species identification algorithm based on image features. In: 2014 International Symposium on Computer, Consumer and Control. IEEE, 2014. p. 139-142.
- Kaggle. 270 Bird Species. Disponível em: <https://www.kaggle.com/gpiosenka/100-bird-species>. Acesso em 12 de maio de 2021.
- Kingma, Diederik P.; BA, Jimmy. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
- Marini, Andréia; Facon, Jacques; Koerich, Alessandro L. Bird species classification based on color features. In: 2013 IEEE International Conference on Systems, Man, and Cybernetics. IEEE, 2013. p. 4336-4341.
- Niemi, Juha; Tantt, Juha T. Deep Learning Case Study on Imbalanced Training Data for Automatic Bird Identification. In: Deep Learning: Algorithms and Applications. Springer, Cham, 2020. p. 231-262.
- Perez, Luis; Wang, Jason. The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621, 2017.
- Qiao, Baowen et al. Bird species recognition based on SVM classifier and decision tree. In: 2017 First International Conference on Electronics Instrumentation & Information Systems (EIIS). IEEE, 2017. p. 1-4.
- Roslan, Rosniza et al. Color-based bird image classification using Support Vector Machine. In: 2017 IEEE 6th Global Conference on Consumer Electronics (GCCE). IEEE, 2017. p. 1-5.
- Selvaraju, Ramprasaath R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. 2017. p. 618-626.
- Simonyan, Karen; ZISSERMAN, Andrew. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
- Srivastava, Nitish et al. Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, v. 15, n. 1, p. 1929-1958, 2014.
- TensorFlow. ResNet101V2. Disponível em: https://www.tensorflow.org/api_docs/python/tf/keras/applications/ResNet101V2. Acesso em 20 de abr. de 2020.
- Vale. *Relatório de atividades 2019*: Reserva Natural Vale. Linhares, 2019. 17p. Disponível em: http://www.vale.com/brasil/PT/initiatives/environmental-social/natural-reserve/Documents/novaReserva/relatorio-reserva-natural_2019.pdf. Acesso em: 12 de maio de 2021.
- Welinder, P.; Brandon, S.; Mita, T.; Wah, C.; Schroff, F.; Belongie, S.; Perona, P. Caltech-UCSD Birds 200. California Institute of Technology. CNS-TR-2010-001. 2010.