UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
CAMPUS DOIS VIZINHOS
CURSO DE ESPECIALIZAÇÃO EM CIÊNCIA DE DADOS

GUSTAVO HENRIQUE MIGLIORINI

IDENTIFICAÇÃO DE AVES DA FAUNA DO BRASIL ATRAVÉS DE APRENDIZADO PROFUNDO PARA CLASSIFICAÇÃO DE IMAGENS

TRABALHO DE CONCLUSÃO DE CURSO DE ESPECIALIZAÇÃO

DOIS VIZINHOS
2022

GUSTAVO HENRIQUE MIGLIORINI

IDENTIFICAÇÃO DE AVES DA FAUNA DO BRASIL ATRAVÉS DE APRENDIZADO PROFUNDO PARA CLASSIFICAÇÃO DE IMAGENS

Trabalho de Conclusão de Curso de Especialização apresentado ao Curso de Especialização em Ciência de Dados da Universidade Tecnológica Federal do Paraná, como requisito para a obtenção do título de Especialista em Ciência de Dados.

Orientador: Prof. Dr. Dalcimar Casanova

DOIS VIZINHOS
2022

GUSTAVO HENRIQUE MIGLIORINI

IDENTIFICAÇÃO DE AVES DA FAUNA DO BRASIL ATRAVÉS DE APRENDIZADO PROFUNDO PARA CLASSIFICAÇÃO DE IMAGENS

Trabalho de Conclusão de Curso de Especialização apresentado ao Curso de Especialização em Ciência de Dados da Universidade Tecnológica Federal do Paraná, como requisito para a obtenção do título de Especialista em Ciência de Dados.

Data de aprovação: 10/novembro/2022

Dalcimar Casanova
Doutorado
Universidade Tecnológica Federal do Paraná - Câmpus Pato Branco

Jefferson Tales Oliva
Doutorado
Universidade Tecnológica Federal do Paraná - Câmpus Pato Branco

Marcelo Teixeira
Doutorado
Universidade Tecnológica Federal do Paraná - Câmpus Pato Branco

DOIS VIZINHOS
2022

AGRADECIMENTOS

À minha companheira Mariane por apoiar desde o início a minha decisão de transição de carreira e o desafio de inicar e concluir o curso de especialização em Ciência de Dados. Aos meus pais Maria e Antonio pelo apoio incondicional e suporte durante toda a minh vida! Ao professor Dalcimar Casanova pela orientação para o desenvolvimento desse trabalho de conclusão de curso. Ao colega Luiz F Giolo, pela contribuição para o desenvolvimento de código para *webscraping* para coleta de imagens. À UTFPR e professores do curso de especialização em Ciência de Dados pela dedicação e por todo conhecimento passado durante o período do curso.

RESUMO

Aplicações baseadas em aprendizagem profunda têm se tornado frequentes em diversas áreas da biologia. Especificamente, as redes neurais convolucionais (CNNs) são amplamente utilizadas para predição de imagens e têm chamado a atenção de biólogos uma vez que podem automatizar análises que por métodos convencionais exigiriam grande esforço e tempo para execução, como a identificação de espécies. Nesse estudo utilizamos aprendizagem por transferência para a obtenção de um modelo classificador de aves da fauna do Brasil. Utilizando um banco de imagens com 1953 classes e mais de 600000 imagens desenvolvemos o modelo classificador utilizando a arquitetura ResNet50V2. O modelo alcançou 45% de acurácia no conjunto de validação, desempenho que pode ser melhorado em estudos futuros com a exploração de outras técnicas e métodos. Esse trabalho destaca a importância de se fornecer para a sociedade uma ferramenta para identificação de aves de uma das maiores biodiversidades da Terra.

Palavras-chave: redes neurais profundas; pássaros; reconhecimento de imagens, ResNet.

ABSTRACT

Deep learning use have become frequent in several areas of biology. Specifically, Convolutional Neural Networks (CNNs) are widely used for image prediction and have drawn the attention of biologists since they can automate analyzes that by conventional methods would require great effort and time to perform, such as species identification. In this study, we used transfer learning to obtain a classifier of bird species from the Brazilian fauna. Using a image database with 1953 classes and more than 600000 images, we developed our classifier model using the ResNet50V2 architecture. The model reached 45% of accuracy in the validation set, a performance that can be improved in future studies with the exploration of other techniques and methods. This work highlights the importance of providing to society a tool to identify birds from one of the greatest biodiversity on Earth.

Keywords: deep neural networks; birds; image recognition; ResNet.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

API	Application Programming Interface
CNN	Convolutional Neural Network
GAN	Generative Adversarial Network
GPU	Graphical Processing Units

SUMÁRIO

1 INTRODUÇÃO

Aplicações baseadas em aprendizagem profunda tem se tornado frequentes em diversas áreas da biologia. Esses algoritmos captam as características em bases de dados complexas, como imagens, sons ou genomas, e as utilizam para criar ferramentas preditivas baseadas em padrões detectados (XU; JACKSON, 2019). Uma vez treinado, o modelo pode ser utilizado para executar previsões em novos dados.

Redes neurais profundas tem sido aplicadas em tarefas de reconhecimento e classificação de espécies de animais e plantas à partir de bases de dados de sons e imagens (WäLDCHEN; MäDER, 2018). Especificamente, as redes neurais convolucionais (CNNs) são amplamente utilizadas para predição de imagens e tem chamado a atenção de biólogos uma vez que podem automatizar análises que por métodos convencionais exigiriam grande esforço e tempo para execução, como a identificação de espécies (WEINSTEIN, 2017). Geralmente, a identificação de espécies de animais e plantas é feita por meio de chaves taxonômicas ou por técnicas genéticas. O primeiro necessita de habilidades e connhecimentos no campo da taxonomia e zoologia, pode ser bastante trabalhoso e a conclusão sobre a identificação pode ser demorada. Já o segundo método é bastante preciso e relativamente rápido, porém pode ser bastante caro já que depende de diversos processos e materiais, como kits de extração de DNA específicos e equipamentos para sequenciamento genético, além, é claro, de conhecimento e habilidades com técnicas moleculares.

Neste contexto, buscar por métodos alternativos para identificação de espécies de animais e plantas, que sejam precisos, rápidos e acessíveis pode favorecer a sociedade como um todo. Explorar o uso de técnicas de inteligência artificial, particularmente das redes neurais profundas como as CNNs, pode tornar o processo de identificação de espécies acessível a todas as pessoas, desde pesquisadores até observadores eventuais.

1.1 Problema de Pesquisa

O Brasil é um dos países com a maior diversidade de aves no mundo. Atualmente são conhecidades 1971 espécies de aves que ocorrem no Brasil, sendo 293 endêmicas e o restante está distribuido entre espécies migrantes e sazonais (PACHECO et al., 2021). A identificação das espécies de aves não é uma tarefa trivial e pode exigir habilidades que somente especialistas possuem, como ornitólogos e hobbistas. Entretanto há um interesse amplo da sociedade em conhecer e poder, de forma independente, identificar espécies de aves que encontram no dia a dia ou durante viagens, por exemplo. Sendo assim, há a necessidade de uma ferramenta que facilite essa tarefa, como por exemplo, através de uma pagina de internet ou aplicativo para celular. Uma potencial solução para essa lacuna pode ser um modelo baseado em redes neurais convolucionais treinado para o reconhecimento das diferentes espécies de aves que ocorrem

no Brasil, e que posteriormente possa ser disponibilizado de forma funcional através de uma aplicação.

1.2 Justificativa

Reconhecer e identificar espécies de aves é uma tarefa que pode ser bastante difícil até mesmo para zoólogos especialistas, uma vez que diferenciar as espécies já classificadas sem o uso de ferramentas genéticas pode exigir elevado conhecimento específico dos grupos, além de chaves taxônomicas bem resolvidas e atualizadas. Além do interesse científico por biólogos, existe grande interesse de uma parte da sociedade que pratica a observação de aves, ou mesmo daqueles que apreciam observar as aves que pousam em seus jardins. Sendo assim, fornecer um ferramenta que permita o reconhecimento de espécies de aves para a sociedade como um todo pode ser importante, tanto do ponto de vista recreativo, como científico e conservacionista. CNNs tem sido utilizadas com sucesso em diferentes campos da biologia (ANGERMUELLER et al., 2016). Na ecologia, tais técnicas de redes neurais profundas são aplicadas para identificar e contar espécies de animais e plantas a partir de imagens. Por exemplo, Norouzzadeh et al. (2018) treinaram uma CNN para identificar 48 espécies de animais africanos utilizando uma base de dados com mais de 3 milhões de imagens rotuladas. Essa CNN, devidamente inserida em uma aplicação, pode substituir a identificação manual podendo fornecer resultados mais rápidos e, até mesmo, mais precisos.

1.3 Objetivos

Os objetivos deste trabalho de conclusão de curso foram obter um banco de imagens de aves representativo da fauna brasileira e desenvolver um modelo baseado em redes neurais convolucionais para executar o reconhecimento e classificação das espécies.

1.4 Materiais e Métodos

Seguindo a lista de espécies de aves publicada por Pacheco et al. (2021), nós buscamos e coletamos imagens em diversos bancos de domínio público e através do buscador de imagens do Google. Todas as imagens foram armazenadas em disco e rotuladas com o nome popular da espécie. Após limpeza do banco de imagens (e.g., remoção de imagens indesejadas), realizamos o treinamento do modelo utilizando a arquitetura pré-treinada ResNet50V2, disponível na biblioteca Keras. Após treinado, o modelo foi avaliado quanto a acurácia e taxa de erro. Todo o trabalho foi realizado em linguagem Python utilizandos o ambiente Google Colab.

1.5 Organização do Trabalho

Esse trabalho está dividido da seguinte maneira:

No Capítulo 2 apresentamos uma revisão da literatura que aborda desde conceitos das teorias que fundamentam este trabalho até estudos relacionados que abordaram tema semelhante.

No Capítulo 3 apresentamos os materiais e métodos utilizados para o desenvolvimento do trabalho.

No Capítulo 4 apresentamos e discutimos os resultados obtidos.

Finalmente, no Capítulo 5 concluímos o estudo, apresentando as limitações encontradas durante o desenvolvimento do trabalho e direções para trabalhos futuros.

2 REVISÃO DE LITERATURA

O interesse pela possibilidade de fazer computadores "pensar" vem desde o início da ciência da computação, nos anos 50. Inteligência artificial (IA) pode ser descrita como a área da ciência que busca automatizar tarefas intelectuais que normalmente são executadas por humanos (CHOLLET, 2021). Conceitualmente, a IA engloba o aprendizado de máquina (*machine learning*) e o aprendizado profundo (*deep learning*), além de outras abordagens que não envolvem aprendizado. As primeiras aplicações de IA não envolviam aprendizado, ao invés, utilizavam conjuntos de regras codificados por programadores, como programas de jogo de xadrez. Tal abordagem da IA se mostrou eficiente para resolver problemas lógicos bem definidos, como aqueles de jogos de xadrez, entretanto, era incapaz de resolver problemas mais complexos, como reconhecimento de fala e classificação de imagens. Assim, com o avanço no conhecimento e nas tecnologias, a aprendizagem de máquina começou a se desenvolver.

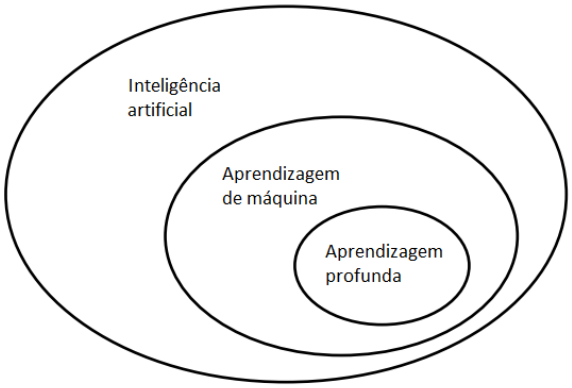

Figura 1 – Representação da relação entre inteligência artificial, aprendizagem de máquina e aprendizagem profunda. Fonte: adaptado de Chollet (2021)

2.1 Aprendizagem de máquina

Na programação clássica o programador determina regras que convertem dados de entrada em respostas apropriadas. Na abordagem em que o computador aprende, ele "olha" para um conjunto de dados de entrada e para as respostas correspondentes e descobre, através de cálculos matemáticos e estatística, quais são as regras que determinam aquelas relações entre dado e resposta.

Aprendizagem de máquina é uma forma de inteligência artificial que permite que computadores executem tarefas sem a necessidade de serem previamente programados para isso. Ao invés, eles são capazes de aprender a partir de exemplos daquela tarefa durante um processo conhecido como treinamento. Feito o treinamento, a tarefa pode ser aplicada em novos dados através de um processo conhecido como inferência (MJOLSNESS; DECOSTE,

2001). De acordo com Mitchell (2013), pode-se dizer que um programa de computador aprende com a experiência E com respeito a alguma tarefa T e alguma medida de performance P, se sua performance em T, como medida por P, melhora com experiência E.

Técnicas de aprendizado de máquina podem ser especialmente úteis para extração de informação em grandes quantidades de dados e particularmente em aplicações que envolvem dados complexos, como análises de imagens e sons.

2.2 Tipos de aprendizagem de máquina

Algoritmos de aprendizado de máquina podem ser classificados de acordo como o tipo de aprendizagem. A maior parte dos algoritmos existentes necessitam que os dados utilizados para treino e teste sejam rotulados, o que é chamado de aprendizado supervisionado. Tarefas de aprendizagem supervisionada são categorizadas como classificação e regressão, onde problemas de classificação utilizam métodos estatísticos para separar duas ou mais classes. Já os problemas de regressão utilizam métodos estatísticos de regressão para gerar previsões numéricas daquele problema. Além dos modelos supervisionados, existem os semi-supervisionados, onde somente parte dos dados possui rótulo, e os modelos não supervisionados onde todo conjunto de dados do problema não possui rótulos e o algoritmo captura as diferenças no conjunto de dados agrupando para gerar a saída.

2.3 Aprendizagem profunda

Aprendizagem profunda é uma subárea da aprendizagem de máquina que utiliza o conceito de camadas de representação para execução do aprendizado. O termo "profunda" faz referência às camadas, e a quantidade delas indica a profundidade do modelo (GOODFELLOW; BENGIO; COURVILLE, 2016). Tais modelos são também chamados de redes neurais devido a inspiração que alguns conceitos tem na neurobiologia e, em particular, no cortex visual do cérebro. Entretanto, não existem evidências de que o cérebro funcione como os modelos de redes neurais (CHOLLET, 2021).

Resumidamente, redes neurais funcionam da seguinte forma (Figura 2): é feito o mapeamento das entradas para os alvos através de sequências de representações (camadas). Cada camada aplica transformações nos dados armazenando a especificação do tipo de transformação realizada em parâmetros conhecidos como *pesos*. O objetivo principal do treinamento é, então, encontrar o conjunto ideal de pesos para todas as camadas de representação de forma que o mapeamento das entradas para seus alvos seja feito de forma correta. O controle da saída da rede neural é feito através da função de custo (ou função de perda), a qual tem a função de calcular a distância entre a predição gerada pela rede e o alvo real. Esse sinal é utilizado para ajustar os pesos nas camadas de representação de forma que diminua o distância entre a predição e alvo real ao ser calculada novamente. O responsável pela tarefa de ajustar os pesos com sinal calculado é o *otimizador* através do algoritmo de *retropropagação*. De forma ciclica,

todo o processo citado anteriormente ocorre com objetivo de minimizar a distância entre a predição e o alvo real o máximo possível e obter uma rede neural treinada (CHOLLET, 2021).

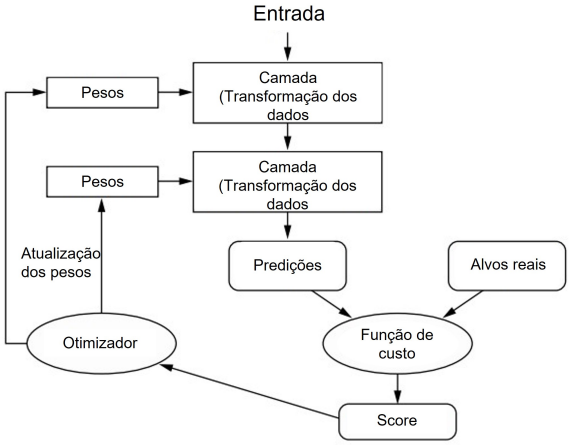

Figura 2 – Representação do funcionamento de uma rede neural. Fonte:adaptado de Chollet (2021)

Algoritmos de aprendizagem profunda vem sendo aplicados com sucesso na classificação de imagens, na transcrição de fala e escrita, na tradução de línguas, na conversão de texto para fala, em veículos autônomos, assistentes digitais como Google Assistant e Amazon Alexa, e muitos outros (CHOLLET, 2021). Classificadores de imagem baseados em aprendizado profundo têm se mostrado promissores em aplicações para medicina, parasitologia, agronomia e biologia. Além de imagens, modelos para reconhecimento e classificação de sons são aplicados para diversos problemas em ecologia (CHRISTIN; HERVET; LECOMTE, 2019).

2.4 Redes Neurais Convolucionais

Foi o trabalho de Hubel (1968) o grande motivador para o desenvolvimento dos principais conceitos da redes neurais convolucionais (CNN). O conhecimento do funcionamento do cortex visual de gatos, onde porções específicas do campo visual pareciam estimular neurônios específicos, motivou cientistas da computação a desenvolver a primeira arquitetura de uma CNN. Originalmente denominada *neocognitron* e mais tarde nomeada *LeNet-5* (LECUN; BENGIO et al., 1995), na arquitetura da CNN cada camada é tridimensional, possuindo extensão espacial e uma profundidade, que corresponde ao número de características (features). Além disso são utilizadas operações matemáticas de convolução, nas quais filtros são utilizados para mapear ativações de uma camada para outra (GOODFELLOW; BENGIO; COURVILLE, 2016). Classificadas como redes do tipo *feedforward*, as CNNs (Figura 3) geralmente possuem em sua estrutura camadas de convolução, camadas de subamostragem e camadas totalmente conectadas. Camadas convolucionais utilizam filtros que podem ter diferentes dimensões para extrair características específicas dos dados de entrada. As camadas de subamostragem

(*pooling*) possuem o papel de reduzir a amostragem da saída de uma camada de convolução ao longo das dimensões espaciais de altura e largura. Sendo assim, sua função principal é reduzir o número de parâmetros a serem aprendidos pela rede, o que como consequência, reduz a chance de sobreajuste e melhora o desempenho da rede.

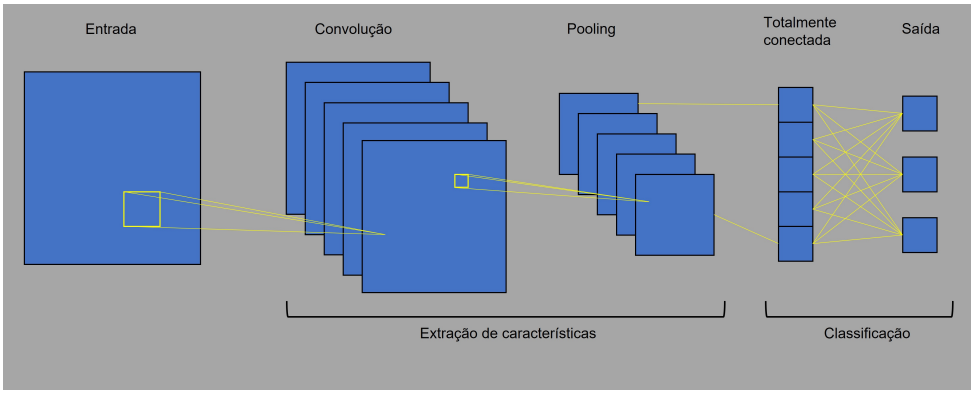

Figura 3 – Representação de uma rede neural convolucional. Fonte: autoria própria

As CNNs tem demonstrado serem muito eficientes na execução de tarefas de reconhecimento de imagens. Por conta disso, diversas arquiteturas vem sendo desenvolvidas e estão disponíveis para utilização através de bibliotecas, como o Keras. Essas arquiteturas pré-treinadas podem ser utilizadas para novas tarefas através do que se conhece como aprendizagem por transferência. Como exemplos de arquiteturas que têm mostrado resultados satisfatórios em tarefas de classificação de imagens pode-se citar a ResNet (HE et al., 2016a), VGG (SIMONYAN; ZISSERMAN, 2014), Xception (CHOLLET, 2017), entre outras.

2.5 Trabalhos Relacionados

Automatizar processos analíticos de reconhecimento e identificação de animais e plantas tem se tornado cada vez mais uma necessidade, seja para fins científicos ou apenas curiosidade. Redes neurais profundas têm sido frequentemente utilizadas para reconhecer e identificar animais (NOROUZZADEH et al., 2018; TABAK et al., 2018) e plantas (RZANNY et al., 2017), demonstrando a capacidade desses algoritmos para executar tarefas que quando executadas da maneira usual podem custar muito tempo e exigir elevado conhecimento específico.

No campo científico da ecologia, modelos baseados em CNNs tem sido utilizados em estudos com maior frequência em estudos de comportamento de animais, monitoramento de populações de animais, modelagem ecológica, e manejo e conservação de ecossistemas (CHRISTIN; HERVET; LECOMTE, 2019). Por exemplo, Norouzzadeh et al. (2018) treinaram CNNs para identificar, contar e descrever comportamentos de 48 espécies de animais africanos em uma base de dados contendo mais de 3 milhões de imagens, obtendo mais de 93% de acurácia. Wild, Sixt e Landgraf (2018) utilizaram CNNs para estudar o comportamento coletivo e interações sociais em abelhas, demonstrando a aplicabilidade desses algoritmos para uma ampla gama de animais. Dentre as espécies animais, o grupo das aves mostra elevado potencial para

uso em aplicações baseadas em CNNs uma vez que as espécies podem variar substancialmente quanto ao seu fenótipo. Mesmo assim, modelos de CNN podem ser capazes de reconhecer e diferenciar indivíduos de uma mesma espécie, onde a diferenciação a olho nu pode ser quase impossível (FERREIRA et al., 2020). Tal capacidade destaca o potencial desses algoritmos para estudos de monitoramento de espécies individuais.

Como as espécies de aves variam de acordo com a região do planeta, obter uma aplicação baseadas em redes neurais profundas para classificação das espécies pode ser uma tarefa bastante complexa, principalmente devido a dificuldade de se obter um conjunto de imagens que contemple todas as espécies desse grupo que é altamente diverso. Sendo assim, modelos treinados em bases de dados locais (e.g., por país) pode facilitar a obtenção desse tipo de aplicação. O Brasil possui uma das maiores biodiversidades de aves da Terra, porém ainda não há uma aplicação para classificar as espécies de forma automática contemplando sua biodiversidade como um todo. Assim, obter um modelo classificador de aves que ocorrem no Brasil e que posteriormente possa ser disponibilizado em forma de uma aplicação, pode contribuir para aqueles necessitam de identificações para fins científicos e para quem simplismente admira estes animais.

3 MATERIAIS E MÉTODOS

Neste capítulo descrevemos as etapas para obtenções dos resultados do estudo, incluindo a construção do banco de imagens, o processamento incial dos dados, e o treinamento e teste do algoritmo.

3.1 Banco de imagens

O conjunto de dados utilizado neste trabalho foi criado de acordo com a lista de espécies de aves fornecida por Pacheco et al. (2021). Foram feitas buscas pelo nome popular de cada espécies de ave em diversos bancos de domínio público e através do buscador de imagens do Google. As imagens foram armazenadas em diretórios individuais e rotulados para cada classe de ave. Das 1971 espécies registradas para o Brasil, 1953 foram incluídas no conjunto de dados, sendo as espécies faltantes não incluídas por não terem sido encontradas imagens disponíveis. No total, foram obtidas 624014 imagens. Para otimizar a obtenção das imagens foi utilizado um código para raspagem (*webscraping*). Após a criação do conjunto de dados, o mesmo foi armazenado no Google Drive para posterior utilização na plataforma Google Colab. Uma vez que buscamos classificar o máximo possível da fauna de aves que ocorre no Brasil, optamos por manter no banco as classes que tiveram poucas imagens encontradas. Tal decisão resultou em um banco com alto desbalanceamento entre as classes, onde algumas continham mais de 800 imagens e outras, menos do que 10.

3.2 Tratamento dos dados

Devido ao uso de técnicas para otimizar a obtenção das imagens, diversas imagens que não representavam aves foram inicialmente incluídas no banco. Sendo assim foi executado um processo de limpeza do banco de forma manual para remoção dessas imagens que incluíram principalmente imagens de ninhos e ovos. Posteriormente, foram utilizados códigos em Python para explorar o banco de imagens, quantificar diretórios e imagens, e detectar e remover imagens com erro de leitura.

3.3 Definição da arquitetura e implementação da CNN

Para obter um modelo de CNN para classificação de aves utilizamos a biblioteca Keras versão 2.8.0. Keras é uma API para Python que utiliza a biblioteca Tensorflow para a implementação de algoritmos de aprendizagem profunda de maneira rápida e intuitiva (CHOLLET, 2021). Para implementação da CNN, utilizamos o Google Colab Pro uma vez que disponibiliza ambientes de alta performance para esse tipo de tarefa, além de permitir conectividade com a nuvem Drive e suporte para linguagem Python.

A seguinte configuração de *hardware* foi utilizadas para o treinamento do modelo: processador Xeon Processors 2.3 Ghz, 27.3 GB de memória RAM, 166.77 GB de armazenamento, e GPU Tesla P100 de 16 GB GDDR5 VRAM.

Utilizamos a arquitetura pré-treinada ResNet50V2 (HE et al., 2016a), disponível na biblioteca Keras. Essa arquitetura obteve baixa taxa de erro no desafio ImageNet de classificação de imagens (RUSSAKOVSKY et al., 2015), além de possuir tamanho pequeno comparado a outras arquiteturas conhecidas. As arquiteturas ResNet são classificada como algoritmos de aprendizado residual. A característica principal dessa arquitetura é a utilização de atalhos (*skip connections*; Figura 4) para pular uma ou mais camadas, com o objetivo de evitar que ocorra o gradiente de fuga (*vanishing gradient*), condição em que o treinamento da rede se torna difícil conforme são adicionadas mais camadas e funções de ativação na estrutura da rede.

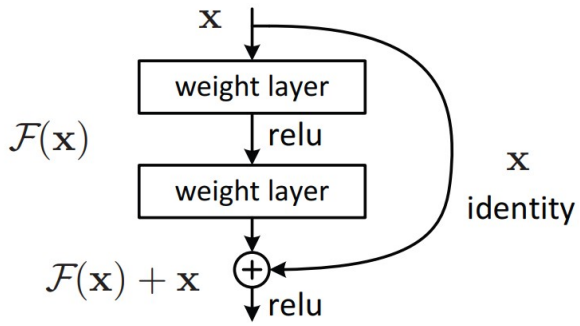

Figura 4 – Representação de um bloco da arquitetura de aprendizado residual. Fonte:He et al. (2016a)

Inicializamos o modelo Resnet com os pesos obtidos da rede pré-treinada no conjunto de imagens ImageNet. Esse banco possui mais de 14 milhões de imagens distribuídas em 20000 classes, e o propósito de se utilizar a rede pré-treinada em outro conjunto de imagens, como o ImageNet, é que características chave, como cor e textura, que são fundamentais para diferenciar diversos objetos, podem ajudar na distinção das aves.

Para evitar o treinamento da rede completa e evitar a atualização dos pesos da rede pré-treinada, todas as camadas foram *congeladas* e somente as camada adicionadas foram treinadas. Para obter um modelo com desempenho satisfatório para a execução da tarefa e considerando a complexidade do conjunto de dados, diversos testes foram executados alterando hiperparâmetros da CNN. Foram testados modelos variando tamanho do lote de imagens, inserção de camadas totalmente conectadas, número de épocas, taxa de aprendizagem, normalização de lote e de camadas, *Dropout*, etc. Após múltiplos testes, verificamos que a melhor configuração de camadas extras a serem adicionadas à arquitetura da CNN para tentar resolver o problema consistiu na estrutura apresentada na tabela 1.

A primeira camada foi uma camada *pooling*, seguida por uma camada de normalização (*batch normalization*). Na sequência foi adicionada uma camada de achatamento (*flatten*) e uma camada totalmente conectada contendo 256 neurônios. Em seguida foi adicionada outra

Tabela 1 – Estrutura adicionada à arquitetura ResNet50V2.

Camadas
GlobalAveragePooling2D()
BatchNormalization()
Flatten()
Dense(256)
BatchNormalization()
Activation(ReLU)
Dropout(0.2)
Dense(1953,activation='softmax')

camada de normalização e uma camada de *dropout*. *Dropout* contribuem para que não ocorra o sobre-ajuste do modelo através da desativação aleatória de neurônios da CNN durante o treinamento. Na última camada totalmente conectada utilizamos a função de ativação softmax. O modelo foi compilado utilizando a função de perda *CategoricalCrossEntropy* em conjunto com o otimizador ADAM. A taxa de aprendizagem do modelo foi inicializada em 0.001 e programada para reduzir 50% a cada 10 épocas. O tamanho de lote utilizado foi 32.

O banco de imagens foi dividido em conjunto de treinamento e conjunto de validação, sendo o primeiro composto por 80% do banco e o conjunto de teste contendo 20% do banco. Para fazer essa divisão utilizamos o processador de imagens *Image Data Generator* disponível na biblioteca Keras, e do método *flow from directory*, o qual permite ler imagens a partir de diretórios. Para tentar reduzir o impacto do desbalancemanto entre as classes do nosso banco de imagens, utilizamos o procedimento de aumento de dados (*data augmentation*). Esse procedimento consiste em aumentar de forma artificial o tamanho amostral aplicando transformações em amostras existes. Para isso utilizamos processador de imagens *ImageDataGenerator*, através da qual aplicamos nas imagens rotações aleatórias de até 40°, alteração no comprimento da imagem (0.2), alteração na altura da imagem (0.2), aproximação de até 0.2, cisalhamento (0.2) e, rotação horizontal. Essas transformações são aplicadas aleatóriamente nas imagens do banco e as imagens originais não são diretamente utilizadas no treinamento do modelo.

4 RESULTADOS e DISCUSSSÃO

Neste capítulo apresentamos os resultados obtidos no estudo, detalhando métricas do modelo no treinamento e validação. Além disso, discutimos os resultados encontrados comparando com estudos semelhantes.

4.1 Treinamento e validação do modelo

Utilizamos a técnica de aprendizagem por transferência através do modelo pré-treinado ResNet50V2, congelando as camadas desse modelo para que o treinamento e atualização dos pesos ocorressem somente nas camadas adicionais. O conjunto de imagens para treinamento do modelo conteve 500006 imagens e o conjunto de validação conteve 124008 imagens.

Após 56 épocas, o modelo alcançou 39% de acurácia no conjunto de treinamento e 45.1% no conjunto de validação. Embora esperava-se treinar o modelo durante 100 épocas, após cinco épocas sem aumentar a acurácia no conjunto de validação o modelo automaticamente encerrou o treinamento do modelo, conforme programação prévia. Apesar de o modelo não ter alcançado 50% de acurácia, foi capaz de realizar acertos em imagens que não estavam nem mesmo no conjunto de validação (Figura 5).

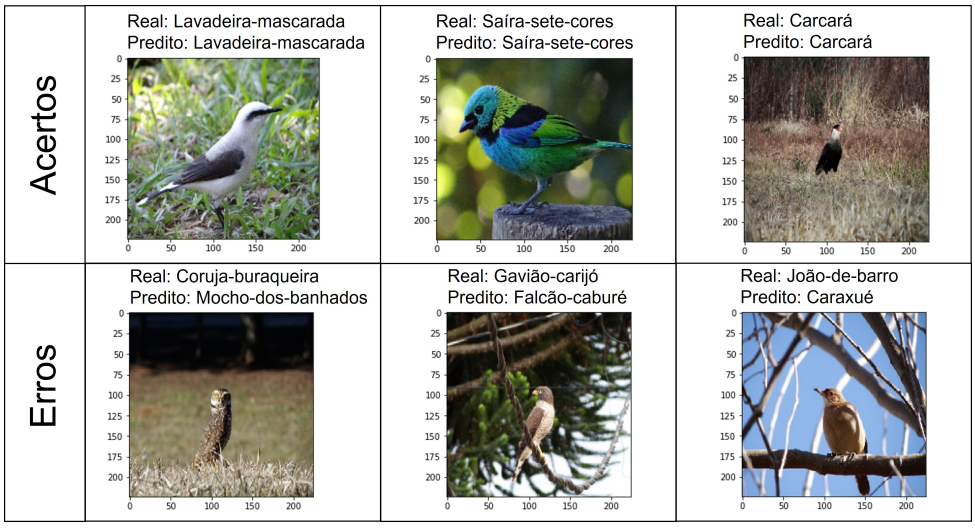

Figura 5 – Exemplos de erros e acertos realizados pelo modelo após o treinamento do algoritmo. Fonte: imagens de autoria própria

Provavelmente, o baixo desempenho do modelo se deve a complexidade do banco de imagens, e embora utilizamos técnicas para aumentar os dados, tal procedimento parece não ter sido suficientemente eficiente para resultar em um modelo com maior acurácia. Entretanto, é importante ressaltar que mesmo após diversos testes com variações em hiperparâmetros e configuração de camadas adicionadas à rede pré-treinada é possível que outras configurações possam resultar em melhor desempenho. Além disso, o foco do estudo aqui apresentado foi

utilizar a arquitetura ResNet50V2 e, portanto, não foram realizados testes utilizando outras arquiteturas pré-treinadas. Nossa escolha pela arquitetura ResNet foi motivada principalmente por ser uma arquitetura relativamente leve (98 MB; VGG19 possui 549 MB de tamanho) e por ter alcançado resultados satisfatórios no conjunto de dados ImageNet (HE et al., 2016b; RUSSAKOVSKY et al., 2015). Diversos trabalhos focados em classificação de imagens de animais e aprendizado por transferência tem alcançado resultados satisfatórios (e.g., >90% de acurácia) utilizando arquiteturas como VGG19 (FERREIRA et al., 2020), ResNet (NOROUZ-ZADEH et al., 2018), MobileNET (VISALLI; BONACCI; BORGHESE, 2021), entre outros. Portanto, explorar outras arquiteturas pré-treinadas pode ser uma alternativa para tentar obter um melhor desempenho do modelo.

O conjunto de dados utilizado nesse trabalho representa quase que toda a fauna conhecida atualmente para o Brasil no que diz respeito às classes representadas (PACHECO et al., 2021). Porém, esse conjunto é altamente desbalanceado o que se deve, naturalmente, às diferenças de raridade das espécies e, portanto, espécies mais comuns possuem maior número de imagens devido a facilidade de observá-las em seu ambiente natural. Para tentar reduzir os efeitos do desblanceamento no número de imagens das classes pode-se utilizar métodos de *data augmentation* para aumentar artificialmente a quantidade de imagens nas classes pouco representadas. Entretanto, diversos métodos podem ser utilizados para tal tarefa (SHORTEN; KHOSHGOFTAAR, 2019) e nesse estudo não abordamos de forma exaustiva as diferentes técnicas conhecidas, focando apenas em métodos que criam variações das imagens existentes. Assim, acreditamos que explorar outras técnicas de *data augmentation* podem contribuir para a obtenção de melhores resultados.

5 CONCLUSÃO

Neste trabalho apresentamos uma solução baseada em CNN e aprendizadem por transferência para automatizar a classificação de aves que ocorrem no Brasil. Nosso modelo foi capaz de classificar corretamente 45% das espécies de aves do conjunto de dados. Embora o modelo não tenha alcançado acurácia alta consideramos que através da exploração de novas estratégias e metodologias é possível obter melhores resultados. Nesse capítulo discutimos as principais limitações do estudo e apresentamos sugestões para trabalhos futuros.

5.1 Limitações

Um dos principais problemas a se considerar quando o objetivo é treinar um modelo para classificação de imagens é a qualidade do banco de imagens quanto ao balanceamento das classes. Se o banco de imagens está desbalanceado algumas medidas podem ser tomadas para contornar esse problema que possivelmente afetará o desempenho da rede. A primeira alternativa é obter mais imagens para equilibrar as classes. Embora pareça uma tarefa simples, em muitos casos, não é uma tarefa trivial e pode, simplismente, ser impossível de fazê-la. Outra alternativa é aplicar métodos de *data augmentation* para aumentar artificialmente a quantidade de imagens contendo variações em multiplas formas, como cores, tamanhos, angulações, etc. Nesse estudo utilizamos técnicas de *data augmentation* para tentar contornar o problema de desbalanceamento dos dados. Foram utilizadas técnicas disponíveis na biblioteca Keras, porém não foram eficientes o suficiente para resultar em um modelo com elevada acurácia. Ainda, poderíamos pensar que o baixo desempenho do modelo poderia estar relacionado ao elevado número de classes (espécies de aves) que o conjunto de dados possui, entretanto, vale ressaltar que não só a arquitetura ResNet quanto as outras disponíveis no Keras, são treinadas no conjunto de dados ImageNet, o qual possui 1000 classes.

Outra limitação encontrada durante o desenvolvimento desse trabalho está relacionado à forma como o modelo foi treinado. Foi utilizado o ambiente do Google Colab e o mesmo limita a utilização por longas horas. Embora o modelo pôde ser salvo para continuar o treinamento após as seções interrompidas, tal limitação resultou em atrasos no desenvolvimento do estudo. Para contornar tal problema a única solução seria a utilização de uma máquina local com GPU de alta perfomance para poder executar o treinamento do modelo ininterruptamente.

5.2 Trabalhos Futuros

Os resultados obtidos nesse trabalho sugerem que a proposta de se obter uma aplicação para identificação automática de espécies de aves da fauna brasileira é promissora. Entretanto, há a necessidade de se explorar mais exasutivamente diferentes técnicas e abordagens de forma comparativa para se obter um modelo que melhore generalize nas classificações. Por

exemplo, podem ser melhor exploradas outras técnicas de *data augmentation* além daquelas disponibilizadas na biblioteca Keras, as quais não introduzem novas imagens para o modelo, mas apenas apresentam as mesmas imagens em diferentes estados. Como exemplo, pode-se citar o uso de *Generative Adversarial Networks* (GANs), as quais podem ser utilizadas para gerar novas imagens à partir das imagens originais (GOODFELLOW et al., 2020; SHORTEN; KHOSHGOFTAAR, 2019).

Outra abordagem bastante promissora para contornar o problema do desbalanceamento das classes é o uso de redes siamesas (BROMLEY et al., 1993), as quais consistem de duas ou mais sub-redes idênticas acopladas pelas suas saídas. Durante o treinamento, essas sub-redes recebem as entradas e cada uma extrai as características, enquanto que o neurônio que liga as duas redes calcula a similaridade entre os vetores de características. Portanto, o treinamento desse tipo de rede tem como objetivo minimizar a distância entre as classes iguais e aumentar a distância entre classes diferentes. Para isso, utiliza-se alguma função de similaridade que pode ser do tipo contrastiva para comparação de pares de imagens ou triplet, através da qual uma das três entradas será a referência para a comparação das outras duas entradas. Logo, como essa metodologia trabalha com comparações e similaridade, o desbalanceamento das classes se torna menos influente no desempenho final do modelo.

5.3 Considerações Finais

Obter um algoritmo para classificação automática de aves do Brasil pode contribuir com o desenvolvimento de estudos em diversas áreas da biologia, como zoologia, ecologia e conservação. Além disse, aplicações que utilizem estes algoritmos podem ser utilizadas por pessoas além daquelas do mundo acadêmico, que admiram e praticam o hobby de observação de aves. Os resultados obtidos nesta monografia destacam o potencial do uso de algoritmos de redes neurais profundas para tarefas de classificação a partir de imagens de dados biológicos. Sendo assim, podemos concluir que o uso de CNNs para classificação de aves é uma técnica promissora que pode ser aplicada para diversos grupos de animais, favorecendo desde a pesquisa até o lazer da sociedade.

Referências

ANGERMUELLER, C. et al. Deep learning for computational biology. **Molecular Systems Biology**, EMBO, v. 12, n. 7, p. 878, jul 2016. Disponível em: <https://doi.org/10.15252%2Fmsb.20156651>. Citado na página 11.

BROMLEY, J. et al. Signature verification using a"siamese"time delay neural network. **Advances in neural information processing systems**, v. 6, 1993. Citado na página 24.

CHOLLET, F. Xception: Deep learning with depthwise separable convolutions. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2017. p. 1251–1258. Citado na página 16.

CHOLLET, F. **Deep learning with Python**. Second edition. Shelter Island: Manning Publications, 2021. OCLC: on1289290141. ISBN 9781617296864. Citado 5 vezes nas páginas 6, 13, 14, 15 e 18.

CHRISTIN, S.; HERVET, É.; LECOMTE, N. Applications for deep learning in ecology. **Methods in Ecology and Evolution**, Wiley, v. 10, n. 10, p. 1632–1644, jul 2019. Disponível em: <https://doi.org/10.1111%2F2041-210x.13256>. Citado 2 vezes nas páginas 15 e 16.

FERREIRA, A. C. et al. Deep learning-based methods for individual recognition in small birds. **Methods in Ecology and Evolution**, v. 11, n. 9, p. 1072–1085, set. 2020. ISSN 2041-210X, 2041-210X. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/2041-210X.13436>. Citado 2 vezes nas páginas 17 e 22.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>. Citado 2 vezes nas páginas 14 e 15.

GOODFELLOW, I. et al. Generative adversarial networks. **Communications of the ACM**, ACM New York, NY, USA, v. 63, n. 11, p. 139–144, 2020. Citado na página 24.

HE, K. et al. Deep Residual Learning for Image Recognition. In: **2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Las Vegas, NV, USA: IEEE, 2016. p. 770–778. ISBN 9781467388511. Disponível em: <http://ieeexplore.ieee.org/document/7780459/>. Citado 3 vezes nas páginas 6, 16 e 19.

HE, K. et al. Identity mappings in deep residual networks. In: SPRINGER. **European conference on computer vision**. [S.l.], 2016. p. 630–645. Citado na página 22.

HUBEL, D. H. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. **J. Physiol**, v. 195, p. 215–244, 1968. Citado na página 15.

LECUN, Y.; BENGIO, Y. et al. Convolutional networks for images, speech, and time series. **The handbook of brain theory and neural networks**, Cambridge, MA USA, v. 3361, n. 10, p. 1995, 1995. Citado na página 15.

MITCHELL, T. M. **Machine learning**. Nachdr. New York: McGraw-Hill, 2013. (McGraw-Hill series in Computer Science). ISBN 9780071154673 9780070428072. Citado na página 14.

MJOLSNESS, E.; DECOSTE, D. Machine Learning for Science: State of the Art and Future Prospects. **Science**, v. 293, n. 5537, p. 2051–2055, set. 2001. ISSN 0036-8075, 1095-9203. Disponível em: <https://www.science.org/doi/10.1126/science.293.5537.2051>. Citado na página 14.

NOROUZZADEH, M. S. et al. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. **Proceedings of the National Academy of Sciences**, Proceedings of the National Academy of Sciences, v. 115, n. 25, jun 2018. Disponível em: <https://doi.org/10.1073%2Fpnas.1719367115>. Citado 3 vezes nas páginas 11, 16 e 22.

PACHECO, J. F. et al. Annotated checklist of the birds of Brazil by the Brazilian Ornithological Records Committee—second edition. **Ornithology Research**, v. 29, n. 2, p. 94–105, jun. 2021. ISSN 2662-673X. Disponível em: <https://link.springer.com/10.1007/s43388-021-00058-x>. Citado 4 vezes nas páginas 10, 11, 18 e 22.

RUSSAKOVSKY, O. et al. ImageNet large scale visual recognition challenge. **International Journal of Computer Vision**, Springer Science and Business Media LLC, v. 115, n. 3, p. 211–252, apr 2015. Disponível em: <https://doi.org/10.1007%2Fs11263-015-0816-y>. Citado 2 vezes nas páginas 19 e 22.

RZANNY, M. et al. Acquiring and preprocessing leaf images for automated plant identification: understanding the tradeoff between effort and information gain. **Plant Methods**, Springer Science and Business Media LLC, v. 13, n. 1, nov 2017. Disponível em: <https://doi.org/10.1186%2Fs13007-017-0245-8>. Citado na página 16.

SHORTEN, C.; KHOSHGOFTAAR, T. M. A survey on image data augmentation for deep learning. **Journal of Big Data**, Springer Science and Business Media LLC, v. 6, n. 1, jul 2019. Disponível em: <https://doi.org/10.1186%2Fs40537-019-0197-0>. Citado 2 vezes nas páginas 22 e 24.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014. Citado na página 16.

TABAK, M. A. et al. Machine learning to classify animal species in camera trap images: Applications in ecology. **Methods in Ecology and Evolution**, Wiley, v. 10, n. 4, p. 585–590, nov 2018. Disponível em: <https://doi.org/10.1111%2F2041-210x.13120>. Citado na página 16.

VISALLI, F.; BONACCI, T.; BORGHESE, N. A. Insects image classification through deep convolutional neural networks. In: **Progresses in Artificial Intelligence and Neural Systems**. [S.l.]: Springer, 2021. p. 217–228. Citado na página 22.

WEINSTEIN, B. G. A computer vision for animal ecology. **Journal of Animal Ecology**, Wiley, v. 87, n. 3, p. 533–545, nov 2017. Disponível em: <https://doi.org/10.1111%2F1365-2656.12780>. Citado na página 10.

WILD, B.; SIXT, L.; LANDGRAF, T. Automatic localization and decoding of honeybee markers using deep convolutional neural networks. **arXiv preprint arXiv:1802.04557**, 2018. Citado na página 16.

WäLDCHEN, J.; MäDER, P. Machine learning for image based species identification. **Methods in Ecology and Evolution**, v. 9, n. 11, p. 2216–2225, nov. 2018. ISSN 2041-210X, 2041-210X.

Disponível em: <https://onlinelibrary.wiley.com/doi/10.1111/2041-210X.13075>. Citado na página 10.

XU, C.; JACKSON, S. A. **Machine learning and complex biological data**. [S.l.]: Springer, 2019. 1–4 p. Citado na página 10.