

Classificação de Aves Predadoras: Fine-tuning Progressivo em Redes Neurais Convolucionais

Luan Matheus Trindade
Dalmazo

luantrindade@ufpr.br
Departamento de Informática
Universidade Federal do Paraná
Curitiba, Paraná, Brasil

Isadora Botassari

isadorabotassari@ufpr.br
Departamento de Informática
Universidade Federal do Paraná
Curitiba, Paraná, Brasil

Paulo Ricardo Lisboa de
Almeida

paulorla@ufpr.br
Departamento de Informática
Universidade Federal do Paraná
Curitiba, Paraná, Brasil

Abstract

This paper presents a process for classifying predatory birds by family and species. The motivation for this study arises from the high variability observed among birds of different species and the importance of performing classification efficiently and in a timely manner. Additionally, this work aims to analyze the impact of using RGB channels in comparison to grayscale images on classification performance, as well as the effect of applying data augmentation techniques during training. The dataset contains 42,475 images, distributed across 6 families and 41 species. The process employs fine-tuning, using the ResNet-50 model. Early stopping was applied to control overfitting and obtain the best model. The test results highlight the effectiveness of the proposed process in classification tasks, with performance varying across different input configurations. For species classification, the model trained with grayscale channels achieved an F1-Score of 0.80. Using RGB channels improved the performance significantly, resulting in an F1-Score of 0.86. Further applying data augmentation techniques to the RGB slightly improved the metrics, achieving an F1-Score of 0.87. These results demonstrate the benefits of incorporating color information and data augmentation in enhancing classification accuracy.

Palavras-Chave

Rede Neural Convolucional, Redes Neurais Profundas, Classificação, Aves

1 Introdução

Aves predadoras compõem uma parte importante nos ecossistemas nos quais estão presentes, desempenhando papéis como controle de outras espécies e dispersão de sementes. A observação desse tipo de ave é uma atividade conhecida e pode ser motivada tanto por monitoramento relacionado a ações de conservação ambiental quanto pelo interesse voluntário em *birdwatching* [1] [2]. Uma consequência dessa atividade é o acúmulo de dados a respeito desses animais, por exemplo, dados de imagens ou áudio [1]. A abundância de dados, em conjunto com técnicas computacionais, é capaz de auxiliar os esforços de monitoramento e conservação desses animais. Em particular, a abordagem de *deep learning* demonstrou sucesso em automatizar o processo de identificação de aves [3] [4].

Berno et al. [5] propôs um dataset (UTFPR-BOP) [6] composto por imagens de diversas aves predadora encontradas no Brasil. As imagens estão classificadas por família e por espécie. A fonte das imagens e de suas respectivas classificações é o *website* wikiaves¹.

¹<https://www.wikiaves.com.br/>

As imagens do conjunto de dados são coloridas e as classes estão desbalanceadas. Em análise, é perceptível que, no problema das espécies, há classes que apresentam maior dificuldade em serem reconhecidas, ou seja, com valores de métricas destoantes da média.

Este artigo propõe as seguintes contribuições:

- Avaliação do desempenho de um processo que integra a ResNet50, empregando *fine-tuning*, para a classificação de aves predadoras em níveis taxonômicos de famílias e espécies, utilizando o dataset UTFPR-BOP.
- Análise do impacto do uso de canais de cores em comparação com a escala de cinza (*grayscale*) no treinamento dos modelos.
- Análise do impacto da inclusão de técnicas de *data augmentation* no treinamento dos modelos.

Após a execução dos experimentos, foi constatada a importância do uso do sistema RGB no processo de classificação, tanto para espécies quanto para famílias. Além disso, é evidente a influência positiva da técnica de *data augmentation*. Dessa forma, espera-se que os resultados obtidos possam auxiliar tanto pesquisadores da área de computação quanto especialistas em conservação da biodiversidade.

O restante deste trabalho está organizado da seguinte forma: a Seção 2 apresenta os trabalhos relacionados; em 3, as fundamentações teóricas; em 4, os métodos utilizados para a realização dos experimentos; em 5, os resultados são discutidos tanto para o problema das famílias quanto para o problema das espécies; em 7, são disponibilizados os códigos utilizados; e, por fim, em 6, as conclusões são apresentadas.

2 Trabalhos relacionados

O problema de classificação de aves costuma envolver dois tipos de dados: áudio e imagens. No contexto de processamento de áudio, Kahl et al. (2021) [7] apresenta uma rede neural residual capaz de classificar conjuntos de dados de áudio em condições diversas, obtendo uma precisão média de 0,79 para áudios de espécies de aves. O classificador *multilayer perceptron* (MLP) em Pahuja et al. (2021) [8] é baseado em espectrogramas gerados a partir de padrões de vocalização e obteve valores de acurácia, precisão e *recall* de 0,96, 0,82 e 0,84, respectivamente. O modelo proposto por Nounmida et al. (2022) [9] utiliza mecanismos de atenção na classificação, no contexto de sinais de áudio sobrepostos. Esse classificador alcançou um *F1-score* de 0,84.

No contexto de processamento de imagens, Huang et al. (2021) [10] apresenta um modelo com rede convolucional e *transfer learning* com o objetivo de classificar aves endêmicas de Taiwan. O

estudo obteve precisão, *recall* e *F1-score* de 0,98, 0,97 e 0,97. Kellenberger et al. (2021) [11] usa redes convolucionais para classificar e contar aves a partir de imagens capturadas por *drones* para fins de monitoramento de populações, atingindo precisão de 0,90 para espécies mais comuns, mas de 0,20 para espécies raras.

Alswaitti et al. (2022) [12] faz uma avaliação comparativa entre algoritmos tradicionais de *machine learning*, redes convolucionais e redes convolucionais baseadas em *transfer learning* para o problema de classificação de múltiplas espécies de ave, demonstrando que as soluções baseadas em *deep learning* são mais eficientes. Além disso, ressaltam que o uso de *transfer learning* contribui para a melhoria das métricas, especialmente em casos de conjuntos de dados reduzidos.

Berno et al. [5], além de propor o *dataset* UTFPR-BOP, também apresenta uma solução baseada na rede neural *Inception-V3* para classificação de aves em famílias e em espécies. Para avaliação, as métricas precisão, *recall* e *F1-Score* foram escolhidas. Quanto aos resultados, para famílias, uma média macro de 0,92 foi obtida para *F1-Score*, enquanto no problema de classificação de espécies, o valor foi de 0,80.

3 Fundamentação teórica

3.1 Convolutional Neural Network (CNN)

Redes Neurais Convolucionais são um tipo de arquitetura de rede neural em que, em pelo menos uma de suas camadas, a operação de convolução substitui a multiplicação geral de matrizes [13]. Soluções baseadas nessa arquitetura são comumente construídas para problemas que envolvem dados estruturados, por exemplo, imagens. Neste contexto, os *pixels* de uma imagem consistem na entrada para a primeira camada da rede. As camadas convolucionais aplicam filtros extratores de características (*features*) de imagens, tais como bordas e padrões locais. A rede é capaz de aprender estes extratores. Uma das principais vantagens de CNNs em comparação a redes *fully-connected* tradicionais é a diminuição drástica do número de parâmetros das camadas ocultas da rede [14], o que cria um sistema mais eficiente e exigindo menos dados de treinamento.

3.2 ResNet50

A ResNet50 [15] é um subtipo de rede residual (*residual network*) que apresenta 50 camadas. Redes residuais foram propostas com o objetivo principal de resolver os problemas de *vanishing gradient* e degradação de acurácia, os quais costumam aparecer em redes neurais muito profundas. He et al. [15] demonstraram que as versões mais profundas da ResNet (ResNet50, Resnet101, ResNet150) atingem maior acurácia sem incorrer nos problemas mencionados.

Redes residuais são caracterizadas por suas camadas de blocos residuais, os quais, por sua vez, consistem na combinação de uma função (associada a uma camada da rede) e uma conexão residual. Uma conexão residual é descrita pelo processo de adicionar à saída de cada função a própria entrada dessa mesma função [14].

3.3 Data augmentation

A técnica de *data augmentation* é usada para aumentar os dados de treinamento de uma classe quando estes são limitados. O processo cria amostras sintéticas ao aplicar transformações em imagens do

dataset de treino que equivalem a variações esperadas no mundo real: rotações, borrimento, cortes, etc [14].

4 Metodologia

4.1 Dataset

O *dataset* UTFPR-BOP: Birds of Prey [6] consiste em um conjunto de 42.475 imagens referentes a aves predadoras encontradas em diferentes regiões do Brasil. As imagens estão classificadas em 6 famílias e 41 espécies, nos dois casos, as classes estão desbalanceadas. Amostras de imagens para cada família podem ser visualizadas na Figura 1.

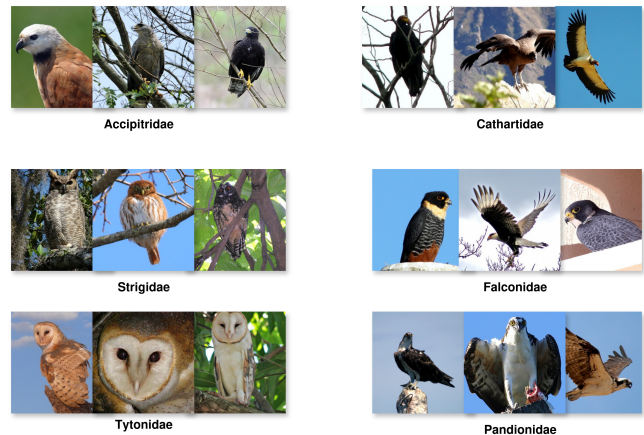


Figura 1: Amostras de imagens para cada família do *dataset*.

As imagens apresentam aves em uma variedade de posições (de frente, de costas, perfil), ações (voo, alimentação), condições de luz, distância e estágios da vida (filhote, adulto). Além disso, as imagens estão sujeitas a condições como ruído, baixa resolução e cortes. A organização das classes pode ser visualizada na Tabela 1.

4.2 Processo

A Figura 2 ilustra a arquitetura das camadas da rede deste experimento. As etapas do fluxo incluem divisão do *dataset*, pré-processamento, treinamento (que inclui *fine-tuning*), validação e teste.

4.2.1 Divisão do *dataset*. Do total dos dados, 70% foi destinado ao treinamento, sendo 30% desse conjunto reservado para validação. O restante do total de amostras foi destinado à etapa de teste. A separação dos dados de treinamento em conjunto de validação e conjunto de treino foi feita de forma aleatória.

4.2.2 Pré-processamento. A etapa de pré-processamento é idêntica tanto para o problema de classificação por famílias quanto para o de espécies. São três cenários para os quais o conjunto de dados e a configuração da rede devem ser preparados: uso de escala de cinza (*grayscale*), uso de canais de cores RGB e uso de canais de cores RGB com *data augmentation* aplicado.

Após a divisão do *dataset* em conjuntos disjuntos, as amostras foram normalizadas por canal (R, G e B). Para isso, foi previamente calculada a média e o desvio padrão dos valores de pixel de cada

Tabela 1: Organização das classes em famílias e espécies e as quantidades de imagens por espécie. Adaptado de Berno et al. [5]

Família	Gênero	Espécie	Amostras
Accipitridae	<i>Accipiter</i>	<i>striatus</i>	840
	<i>Busarellus</i>	<i>nigricollis</i>	1.160
	<i>Buteo</i>	<i>albicaudatus</i>	929
	<i>Chondrohierax</i>	<i>uncinatus</i>	1.120
	<i>Circus</i>	<i>bufoni</i>	840
	<i>Elanoides</i>	<i>forficatus</i>	1.140
	<i>Elanus</i>	<i>leucurus</i>	1.081
	<i>Geranoaetus</i>	<i>albicaudatus</i>	880
		<i>melanoleucus</i>	1.000
	<i>Geranospiza</i>	<i>caerulescens</i>	1000
	<i>Harpia</i>	<i>harpyja</i>	954
	<i>Heterospizias</i>	<i>meridionalis</i>	1.040
	<i>Pseudastur</i>	<i>polionotus</i>	740
Cathartidae	<i>Rupornis</i>	<i>magnirostris</i>	960
	<i>Urubitinga</i>	<i>coronata</i>	1.100
	Total		14.784
Falconidae	<i>Cathartes</i>	<i>aura</i>	1.121
		<i>burrovianus</i>	1.120
		<i>melambrotus</i>	876
	<i>Coragyps</i>	<i>atratus</i>	1.221
	<i>Sarcoramphus</i>	<i>papa</i>	1.174
Pandionidae	<i>Vultur</i>	<i>gryphus</i>	214
	Total		5.726
Strigidae	<i>Caracara</i>	<i>plancus</i>	1.180
	<i>Falco</i>	<i>femoralis</i>	1.100
		<i>peregrinus</i>	1.130
		<i>rufigularis</i>	1.140
	<i>Milvago</i>	<i>chimachima</i>	1.140
Tytonidae		<i>chimango</i>	1.120
	Total		6.810
	<i>Pandion</i>	<i>haliaetus</i>	1.260
	<i>Asio</i>	<i>clamator</i>	1.120
		<i>flammeus</i>	1.140
		<i>stygius</i>	1.120
	<i>Athene</i>	<i>cunicularia</i>	1.140
	<i>Bubo</i>	<i>virginianus</i>	1.120
	<i>Glaucidium</i>	<i>brasilianum</i>	1.140
	<i>Megascops</i>	<i>atricapilla</i>	625
Tytonidae		<i>choliba</i>	1.120
		<i>santaecatarinae</i>	1.274
	<i>Pulsatrix</i>	<i>koenigswaldiana</i>	1.120
	<i>Strix</i>	<i>hylophila</i>	1.120
		<i>virgata</i>	736
	Total		13.895
Total Geral			42.475

canal em todo o conjunto de treinamento. Os valores obtidos foram: médias de 0,4713, 0,5089 e 0,4957, e desvios padrão de 0,1775, 0,1849 e 0,200. É importante notar que, no caso do cenário com *grayscale*, esse procedimento também foi aplicado, considerando que a ResNet exige entradas com 3 canais. Na prática, os 3 canais possuem os mesmos valores, representando uma escala de cinza. Em seguida, a imagem é redimensionada para 224x224, conforme o padrão esperado pela ResNet. A ResNet utilizada foi treinada com o *dataset ImageNet*.

No cenário com *data augmentation*, é adicionada uma etapa para expandir o conjunto de dados, conforme ilustrado na Figura 3. Para a expansão do número de amostras, foram adotadas transformações de rotação, ajuste de brilho, borrramento, corte, *zoom* e espelhamento. A aplicação desses efeitos foi feita de forma aleatória.

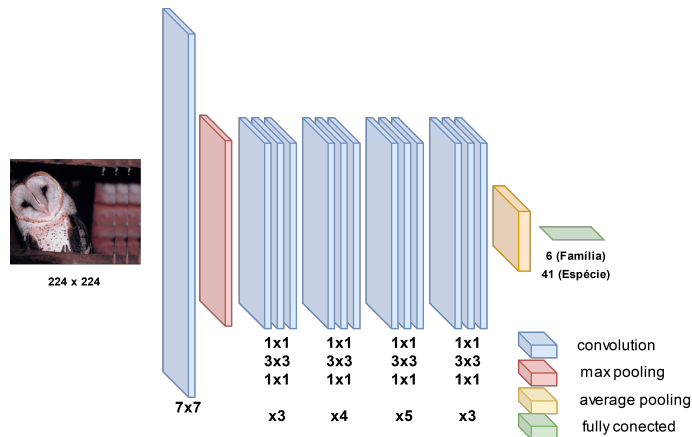


Figura 2: Esquema ilustrativo da arquitetura da rede para classificação de família e espécies.

Todas as classes foram balanceadas para possuir a o dobro da quantidade de imagens que a classe mais numerosa, *Megascops santaecatarinae*, com 1274 imagens. Após a divisão descrita na Seção 4.2.1, o conjunto de treinamento dessa classe ficou com 668 imagens. O dobro desse valor, 1336, foi definido como o valor máximo para o *data augmentation*. Assim, cada conjunto de treinamento de cada espécie foi aumentado para esse valor. Esse método foi aplicado para garantir que todas as classes se beneficiassem do aumento de dados.



Figura 3: Exemplos de *data augmentation* aplicados a algumas amostras.

4.2.3 Treinamento e validação. Para o treinamento, a camada de classificação da rede foi substituída, redefinindo a saída para 6 neurônios na tarefa de classificação por família e 41 neurônios na classificação por espécie. Além disso, todos os pesos das demais camadas da rede foram congelados, preservando os parâmetros pré-treinados.

A rede é inicialmente configurada para ser treinada por até 20 épocas, com o otimizador AdamW, a função de perda (*loss function*) *Cross Entropy*, um *learning rate* de $1e-4$ e um *batch size* de 32, com aleatorização das amostras a cada iteração de treino. A técnica de *early stopping* é aplicada com um fator de paciência de 3 épocas, garantindo que o modelo com a melhor *loss* de validação seja salvo. Caso não haja melhoria após 3 épocas, o treinamento é interrompido.

Após a terceira época, os pesos das camadas restantes são descongelados, visando melhorar a eficiência da rede na identificação de padrões, introduzindo um comportamento de *fine-tuning* progressivo ao modelo. Ao final do treinamento, o gráfico de *loss* é gerado, permitindo o acompanhamento do desempenho do modelo ao longo do treino e a verificação de possíveis casos de *overfitting*.

Para cada caso analisado (sem RGB, com RGB, e com RGB e *data augmentation*), são realizados cinco treinos independentes. Ao fim de cada treino, o modelo correspondente é salvo. A etapa de teste é executada uma vez para cada modelo, e a média das métricas de precisão, *recall* e *F1-score* é calculada para os resultados de cada classe, juntamente com o desvio padrão de cada métrica.

4.2.4 Teste. Finalmente, a etapa de teste é iniciada, e o conjunto de imagens destinado a essa fase é avaliado com base nas seguintes métricas, em que NN representa a quantidade de classes, VP o número de verdadeiros positivos, FP os falsos positivos e FN os falsos negativos.

- Precisão:

$$\text{Macro-Precisão} = \frac{1}{N} \sum_{i=1}^N \frac{VP_i}{VP_i + FP_i} \quad (A)$$

- Recall:

$$\text{Macro-Recall} = \frac{1}{N} \sum_{i=1}^N \frac{VP_i}{VP_i + FN_i} \quad (B)$$

- *F1-Score*:

$$\text{Macro-}F_1 = \frac{1}{N} \sum_{i=1}^N \frac{2 \cdot \text{Precisão}_i \cdot \text{Recall}_i}{\text{Precisão}_i + \text{Recall}_i} \quad (C)$$

As três métricas são particularmente úteis em cenários com classes desbalanceadas, como é o caso deste estudo. Além disso, o conjunto de métricas permite a comparação do modelo com [5].

Para analisar a melhoria na classificação de diferentes famílias e espécies, foi utilizada principalmente a métrica *F1-Score*, visto que combina os valores de precisão e *recall*. O processo de análise consistiu em observar a métrica individualmente (para cada classe) e a média macro.

4.3 Hardware e Framework utilizados

Para execução do experimento, foi utilizada uma máquina com dois processadores Intel(R) Xeon(R) CPU E5-2620 0 @ 2.00 Gigahertz (cada um com 6 núcleos e 12 threads), 32 Gigabytes (GB) de memória e uma NVIDIA TITAN Xp - 12GB. Como *framework*, foi utilizado o PyTorch e o *torchvision* para uso da ResNet50.

5 Resultados e discussões

Nesta seção, são apresentados os resultados dos experimentos. Inicialmente, na Seção 5.1, são apresentados os experimentos considerando a classificação de famílias. Na Seção 5.2, são apresentados os experimentos considerando a classificação de espécies.

De forma geral, o modelo demonstra capacidade de convergir em até nove épocas, o mecanismo de *early stopping* foi de grande auxílio para aprimorar a velocidade de convergência.

5.1 Classificação de famílias

Nesta seção, são apresentados os resultados dos testes de classificação por família para cada cenário avaliado: uso de escala de cinza (*grayscale*), uso de canais de cores RGB e uso de canais de cores RGB com *data augmentation*. As métricas calculadas a partir desses resultados estão retratados na Tabela 2.

Tabela 2: Resultados de classificação por família. F1 (*F1-Score*) Gray (Escala de cinza) Aug (Aumentado) \pm (Desvio padrão).

Classe	F1 (Gray)	F1 (RGB)	F1 (Aug)
<i>Accipitridae</i>	0,92 \pm 0,01	0,94 \pm 0,00	0,95 \pm 0,00
<i>Cathartidae</i>	0,93 \pm 0,01	0,95 \pm 0,00	0,95 \pm 0,01
<i>Falconidae</i>	0,89 \pm 0,01	0,92 \pm 0,01	0,93 \pm 0,01
<i>Pandionidae</i>	0,83 \pm 0,01	0,87 \pm 0,01	0,88 \pm 0,01
<i>Strigidae</i>	0,98 \pm 0,00	0,98 \pm 0,00	0,99 \pm 0,00
<i>Tytonidae</i>	0,96 \pm 0,01	0,96 \pm 0,02	0,97 \pm 0,01
Média macro	0,92 \pm 0,00	0,94 \pm 0,00	0,95 \pm 0,00

Em relação aos experimentos com o uso e ausência dos canais RGB, é notável que as classes que mais se beneficiaram do uso de canais de cores foram *Pandionidae* e *Falconidae*. No caso das classes aumentadas, nenhuma classe foi particularmente beneficiada.

É possível constatar que, para a classificação de famílias, tanto a presença de canais de cores quanto de balanceamento de classes, apesar de causar melhoria, é pouco impactante. Dessa forma, verifica-se que a rede possui a capacidade de distinguir de forma eficaz as classes, utilizando estruturas e padrões físicos das aves presentes nas imagens, independentemente da coloração.

Observou-se que, em todas as famílias, as cores predominantes nas aves representadas são: branco, preto, tons de cinza e de marrom. Desse conjunto, apenas a informação da cor marrom é perdida com a ausência de canais de cores. Além disso, os cenários de fundo, quando presentes, constituem uma parte significativa da imagem. No caso de imagens capturadas durante o período diurno, esses ambientes costumam apresentar variações de azul, verde e marrom. No experimento em que o modelo usou *greyscale*, é plausível que a falta de cores tenha dificultado a distinção entre a ave e o ambiente em que ela se encontra.

Nas classes *Strigidae* e *Tytonidae*, há uma parcela considerável de imagens capturadas durante o período noturno. Na prática, esse fator resulta em variações no cenário de fundo e no contraste entre o corpo da ave e o ambiente. Nessas imagens, o fundo costuma ser majoritariamente preto, diferindo das imagens capturadas no período diurno, como pode ser visto na Figura 4. Isso indica, portanto, um padrão forte que pode ser detectado e aprendido pela rede mesmo com a ausência de canais de cores. Analisando os *F1-score*



Figura 4: Imagens pertencentes a diferentes famílias em escala de cinza. No topo, imagens com um cenário noturno. Na parte inferior, imagens com um cenário diurno.

das classes *Strigidae* e *Tytonidae*, é possível ver que foram as classes menos impactadas pelas técnicas aplicadas.

5.2 Classificação de espécies

Nesta seção, são apresentados os resultados dos testes de classificação por espécie para cada cenário avaliado: uso de escala de cinza (*grayscale*), uso de canais de cores RGB e uso do sistema RGB com *data augmentation*. As métricas calculadas a partir desses resultados estão organizadas na Tabela 3.

5.2.1 Modelo com canais grayscale e Modelo com canais RGB. Nos testes executados com o modelo configurado com escala de cinza, as médias macro de precisão, *recall* e *F1-score* obtidas foram, respectivamente: 0,82, 0,81 e 0,80.

Nos testes executados com o modelo configurado com canais de cores RGB, as médias macro de precisão, *recall* e *F1-score* obtidas foram, respectivamente: 0,86, 0,85 e 0,85. Em relação ao experimento em *grayscale*, as classes que apresentaram maior incremento no *F1-score* foram: *Megascops atricapilla*, *Vultur gryphus*, *Cathartes aura* e *Cathartes burrovianus*. Observa-se que as duas primeiras classes são as mesmas que apresentam os piores *F1-scores* nos testes do modelo com escala de cinza. Por fim, é evidente que as classes que mais foram afetadas pelo uso da escala de cinza foram *Megascops atricapilla* e *Vultur gryphus*, isso pode estar associado ao fato de que são classes pouco representadas e que utilizam da cor como um fator importante no processo de classificação. O comparativo entre as imagens dessas classes com o sistema RGB e em escala de cinza pode ser visualizado na Figura 5.

As melhorias obtidas neste experimento foram mais significativas em comparação à classificação de famílias. Uma razão plausível para esse resultado, além do evidente aumento de informações nas imagens, é o número maior de classes (41), o que deixa o problema mais complexo. As diferenças estruturais entre espécies pertencentes a uma mesma família são menos pronunciadas, portanto, as informações de coloração ganham mais relevância para classificar corretamente uma instância de espécie. Por exemplo, as aves

Tabela 3: Resultados de classificação por espécie. F1 (*F1-Score*) Gray (Escala de cinza) Aug (Aumentado) \pm (Desvio padrão).

Classe	F1 (Gray)	F1 (RGB)	F1 (Aug)
<i>Glaucidium brasilianum</i>	0,96 \pm 0,01	0,98 \pm 0,01	0,97 \pm 0,00
<i>Caracara plancus</i>	0,89 \pm 0,00	0,92 \pm 0,01	0,92 \pm 0,01
<i>Milvago chimachima</i>	0,82 \pm 0,02	0,88 \pm 0,01	0,87 \pm 0,02
<i>Falco rufigularis</i>	0,87 \pm 0,01	0,93 \pm 0,00	0,92 \pm 0,01
<i>Milvago chimango</i>	0,85 \pm 0,01	0,90 \pm 0,02	0,92 \pm 0,01
<i>Vultur gryphus</i>	0,13 \pm 0,15	0,34 \pm 0,12	0,48 \pm 0,04
<i>Athene cunicularia</i>	0,94 \pm 0,01	0,95 \pm 0,01	0,96 \pm 0,01
<i>Strix virgata</i>	0,93 \pm 0,02	0,96 \pm 0,01	0,96 \pm 0,01
<i>Strix hylophila</i>	0,95 \pm 0,03	0,97 \pm 0,01	0,98 \pm 0,00
<i>Asio clamator</i>	0,96 \pm 0,01	0,96 \pm 0,01	0,96 \pm 0,01
<i>Megascops choliba</i>	0,85 \pm 0,02	0,87 \pm 0,02	0,89 \pm 0,03
<i>Megascops sanctaecatarinae</i>	0,70 \pm 0,02	0,67 \pm 0,05	0,70 \pm 0,03
<i>Megascops atricapilla</i>	0,12 \pm 0,14	0,35 \pm 0,15	0,34 \pm 0,12
<i>Falco femoralis</i>	0,87 \pm 0,02	0,89 \pm 0,00	0,89 \pm 0,01
<i>Falco peregrinus</i>	0,82 \pm 0,01	0,84 \pm 0,02	0,86 \pm 0,01
<i>Busarellus nigricollis</i>	0,84 \pm 0,02	0,94 \pm 0,01	0,95 \pm 0,01
<i>Heterospizias meridionalis</i>	0,80 \pm 0,01	0,90 \pm 0,02	0,91 \pm 0,01
<i>Chondrohierax uncinatus</i>	0,76 \pm 0,01	0,80 \pm 0,01	0,82 \pm 0,01
<i>Rupornis magnirostris</i>	0,81 \pm 0,02	0,87 \pm 0,02	0,89 \pm 0,01
<i>Geranoaetus albicaudatus</i>	0,75 \pm 0,02	0,78 \pm 0,01	0,81 \pm 0,01
<i>Circus buffoni</i>	0,83 \pm 0,03	0,82 \pm 0,02	0,85 \pm 0,02
<i>Accipiter striatus</i>	0,74 \pm 0,01	0,79 \pm 0,03	0,81 \pm 0,02
<i>Elanus leucurus</i>	0,88 \pm 0,01	0,89 \pm 0,02	0,91 \pm 0,01
<i>Geranoospiza caerulescens</i>	0,77 \pm 0,02	0,87 \pm 0,01	0,89 \pm 0,01
<i>Pseudastur polionotus</i>	0,81 \pm 0,02	0,82 \pm 0,02	0,84 \pm 0,03
<i>Harpia harpyja</i>	0,82 \pm 0,00	0,84 \pm 0,03	0,87 \pm 0,02
<i>Elanoides forficatus</i>	0,91 \pm 0,02	0,93 \pm 0,02	0,93 \pm 0,02
<i>Buteo albonotatus</i>	0,76 \pm 0,02	0,78 \pm 0,03	0,83 \pm 0,02
<i>Bubo virginianus</i>	0,94 \pm 0,01	0,95 \pm 0,01	0,95 \pm 0,01
<i>Asio stygius</i>	0,93 \pm 0,01	0,95 \pm 0,01	0,96 \pm 0,00
<i>Asio flammeus</i>	0,93 \pm 0,00	0,94 \pm 0,01	0,94 \pm 0,01
<i>Pulsatrix koeniswaldiana</i>	0,95 \pm 0,01	0,97 \pm 0,02	0,97 \pm 0,00
<i>Tyto alba</i>	0,97 \pm 0,01	0,98 \pm 0,00	0,98 \pm 0,00
<i>Cathartes melambrotus</i>	0,71 \pm 0,02	0,81 \pm 0,01	0,79 \pm 0,04
<i>Cathartes burrovianus</i>	0,70 \pm 0,03	0,84 \pm 0,01	0,83 \pm 0,02
<i>Coragyps atratus</i>	0,85 \pm 0,01	0,89 \pm 0,01	0,91 \pm 0,01
<i>Cathartes aura</i>	0,75 \pm 0,01	0,91 \pm 0,01	0,90 \pm 0,01
<i>Sarcorampus papa</i>	0,81 \pm 0,01	0,86 \pm 0,01	0,85 \pm 0,01
<i>Urubitinga coronata</i>	0,68 \pm 0,03	0,75 \pm 0,02	0,77 \pm 0,02
<i>Pandion haliaetus</i>	0,83 \pm 0,02	0,87 \pm 0,01	0,90 \pm 0,01
<i>Geranoaetus melanoleucus</i>	0,74 \pm 0,01	0,74 \pm 0,03	0,77 \pm 0,01
Média macro	0,80	0,85	0,87

Cathartes aura e *Cathartes burrovianus* pertencem à mesma família e gênero e compartilham padrões corporais nas penas e na cabeça. Contudo, as espécies possuem colorações diferentes no rosto, o que pode ser um forte indicador de distinção entre as classes baseado na cor.

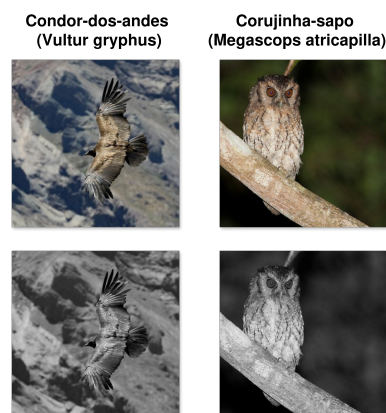


Figura 5: Comparativo visual entre as classes que foram mais afetadas pela ausência do sistema RGB. No topo imagens em RGB e na parte inferior em escala de cinza.

38 das 41 classes obtiveram algum grau de melhoria, 2 permaneceram com o mesmo *F1-score*, e 2 tiveram um resultado inferior.

5.2.2 Modelo com canais de cores e data augmentation. As médias macro de precisão, *recall* e *F1-score* obtidas foram, respectivamente: 0,87, 0,87 e 0,87.

Em relação ao experimento sem *data augmentation*, o aumento de *F1-score* mais expressivo se deu na classe *Vultur gryphus*.

25 das 41 classes obtiveram algum grau de melhoria, 9 permaneceram com o mesmo *F1-score*, e 8 tiveram um resultado inferior.

Essa técnica mostrou um impacto menor em comparação ao uso de RGB. Como apresentado na Tabela 3, mesmo as classes que se beneficiaram do aumento de dados apresentaram melhorias pequenas.

As três classes com os piores *F1-scores* neste experimento foram: *Megascops atricapilla* (0,34), *Vultur gryphus* (0,48) e *Megascops sanctaecatarinae* (0,70). A classe *Vultur gryphus* é a mais subrepresentada no *dataset*, caracterizando um cenário em que o uso de *data augmentation* pode mitigar os efeitos negativos decorrentes do número reduzido de amostras. Entretanto, esse efeito positivo não foi reproduzido para a classe *Megascops atricapilla*, a classe com a segunda menor quantidade de imagens. Por fim, o *data augmentation* também não produziu impacto significativo nas métricas associadas à classe *Megascops sanctaecatarinae*, a qual apresenta o maior número de amostras no *dataset*. Dessa forma, observa-se que a técnica foi mais útil para a classe com o caso mais crítico de poucas amostras de treinamento.

5.2.3 Testes qualitativos. Para avaliar o desempenho do modelo de classificação por espécies em cenários alternativos, foram selecionadas duas imagens externas ao conjunto de dados. A primeira é uma fotografia de uma coruja-buraqueira, capturada à distância e com qualidade reduzida. A segunda é a imagem de um chaveiro representando a mesma espécie. Ambas imagens foram coletadas por um dos autores e foram utilizadas com o objetivo de testar a capacidade de generalização do modelo. Para classificar, foi escolhido aleatoriamente um dos modelos treinados com canais de cores e classes aumentadas. Os resultados estão ilustrados na Figura 6.

Os resultados demonstram o potencial do modelo ao lidar com imagens externas ao *dataset* e de qualidade variada.

5.2.4 Comparação com trabalhos similares. Esta seção comenta e compara sucintamente os resultados obtidos nestes experimentos com os resultados obtidos por um trabalho que utiliza o mesmo conjunto de dados, Berno et al. [5]. Tal como mencionado na Seção 2, o *F1-score* registrado em [5] para o problema de famílias foi 0,92, enquanto para o problema de espécies, foi 0,80. Observa-se que esses valores de *F1-score* são iguais aos valores obtidos no experimento feito com o modelo sem canais de cores, ao passo que os experimentos subsequentes obtiveram métricas superiores. Os comparativos para as famílias podem ser visualizados na Tabela 4 e para as espécies na Tabela 5.

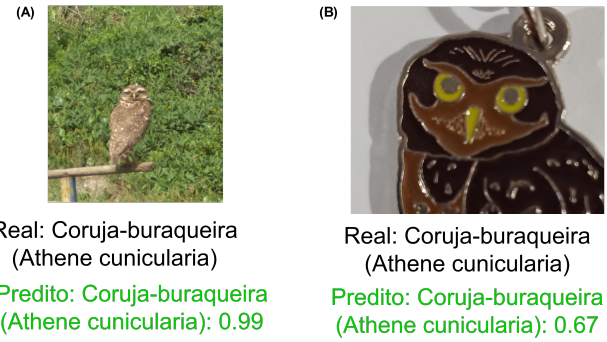


Figura 6: Testes de classificação com fotografias coletadas por um dos autores em visita ao Parque Das Aves - Foz do Iguaçu - Paraná, (A) é uma fotografia e (B) um chaveiro. Em verde, a probabilidade obtida pela inferência do modelo.

Tabela 4: Comparativo para famílias entre o trabalho [5] e os resultados obtidos neste estudo. Valores em negrito representam melhoria. O comparativo foi realizado para a configuração que obteve os melhores resultados: com *data augmentation*. Todas as métricas estão apresentadas como a média macro.

Métrica	Berno et al. [5]	Obtida
Precisão	0,93	0,95
Recall	0,91	0,94
<i>F1-Score</i>	0,92	0,94

Tabela 5: Comparativo para espécies entre o trabalho [5] e os resultados obtidos neste estudo. Valores em negrito representam melhoria. O comparativo foi realizado para a configuração que obteve os melhores resultados: com *data augmentation*. Todas as métricas estão apresentadas como a média macro.

Métrica	Berno et al. [5]	Obtida
Precisão	0,81	0,87
Recall	0,81	0,87
<i>F1-Score</i>	0,80	0,86

6 Conclusão

Este estudo propõe a utilização da ResNet50 em conjunto com técnicas de *fine-tuning* para a tarefa de classificação de aves predadoras.

Os resultados alcançados destacam o impacto positivo dos canais de cores nos dados de imagens no problema de classificação. No entanto, em geral, a ausência desses canais não compromete drasticamente o desempenho do modelo, considerando as métricas obtidas e o número de classes nos problemas de classificação por família e por espécie.

Por outro lado, a técnica de *data augmentation* foi pouco impactante nos resultados, beneficiando majoritariamente a classe mais subrepresentada do conjunto de dados no problema de espécies.

Em geral, os modelos apresentados são eficientes na identificação de diferentes famílias e espécies, mesmo diante da complexidade do problema, que se deve principalmente às similaridades entre espécies de determinadas famílias e ao número limitado de imagens para classes específicas.

Para trabalhos futuros, sugerem-se algumas possibilidades, como a exploração de modelos neurais multimodais capazes de combinar dados de áudio e imagem para aprimorar a classificação, o enriquecimento do dataset com mais imagens para classes com métricas baixas (especialmente *F1-score*), além do uso de abordagens mais avançadas de *data augmentation*, como Redes Geradoras Adversárias para a geração de novas imagens. Além disso, é importante a avaliação dos procedimentos desenvolvidos neste trabalho a outros conjuntos de dados públicos, ampliando sua validação e potencial de generalização.

Acknowledgments

Agradecimentos ao Laboratório Visão Robótica e Imagem (VRI) por fornecer o ambiente para realização dos experimentos.

7 Material complementar

Os códigos utilizados neste trabalho podem ser consultados em github.com/luandalmazot/classificacao-aves-predadoras.

Referências

- [1] K.V.C. Barbosa, P.F. Develey, and M.C. et al. Ribeiro. The contribution of citizen science to research on migratory and urban birds in Brazil. In *Ornithology Research*, 2021.
- [2] P. F. Develey. Bird conservation in Brazil: Challenges and practical solutions for a key megadiverse country. In *Perspectives in Ecology and Conservation*, 2021.
- [3] J. J. Lahoz-Monfort and M. J. L. Magrath. A comprehensive overview of technologies for species and habitat monitoring and conservation. In *BioScience*, 2021.
- [4] S. Christin, E. Hervet, and N. Lecomte. Applications for deep learning in ecology. In *Methods in Ecology and Evolution*, 2019.
- [5] B.C.S. Berno, L. Schneider, L.A. Albin, and H.S. Lopes. Brazilian birds of prey - a new dataset and classification with deep neural networks. In *Proc. Congresso Brasileiro de Inteligência Computacional - CBIC*, 2019.
- [6] LABIC Bioinformatics and Computational Intelligence Laboratory. Utfpr-bop: Birds of prey. URL <https://labic.utfpr.edu.br/datasets/UTFPR-BOP.html>. Accessed: 2024-11-10.
- [7] S. Kahl, C. M. Wood, M. Eibl, and H. Klinck. Birdnet: A deep learning solution for avian diversity monitoring. In *Ecological Informatics*, 2021.
- [8] R. Pahuja and A. Kumar. Sound-spectrogram based automatic bird species recognition using mlp classifier. In *Applied Acoustics v. 180*, 2021.
- [9] A. Noumida and Rajan R. Multi-label bird species classification from audio recordings using attention framework. In *Applied Acoustics v. 197*, 2022.
- [10] Y. Huang and Basanta H. Recognition of endemic bird species using deep learning models. In *IEEE Access*, 2021.
- [11] B. Kellenberger, T. Veen, E. Folmer, and D. Tula. 21 000 birds in 4.5 h: efficient large-scale seabird detection with machine learning. In *Remote Sensing in Ecology and Conservation v. 7*, 2021.
- [12] M. Alswaitti1, L. Zihao1, W. Alomoush, A. Alrosan, and K. Alissa. Effective classification of birds' species based on transfer learning. In *International Journal of Electrical and Computer Engineering (IJECE)*, 2022.
- [13] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.
- [14] C. M. Bishop and H. Bishop. *Deep Learning: Foundations and Concepts*. Springer, 2024.
- [15] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition, 2015. URL <https://arxiv.org/abs/1512.03385>.