Machine Learning for Earth and Environmental Sciences (2025), **xx**:xx 1–6

Final Report

UNIL | Université de Lausanne

Finding the spots in owls

Kinga Kaszap1[1]

[1]Department of Ecology and Evolution, Université de Lausanne, Lausanne, CH-1015, Switzerland.

Keywords: semantic segmentation, barn owl, u-net, transfer learning

Abstract

Patterns on animal plumages provide valuable information for conservation and evolutionary biology, but their manual quantification is subjective and time-consuming. Convolutional neural networks (CNNs) offer an automated alternative. In barn owls, spottiness is a useful proxy for female fitness and an indicator of population-level local adaptation. Here, we develop a CNN for semantic segmentation to delineate spots on owls' plumages using a small dataset of 270 images and corresponding masks. With the aid of data augmentation and transfer learning, the final model reliably identifies spots on spottier owls and performs adequately on lightly spotted individuals. These results demonstrate the potential of deep-learning-based animal biometrics for phenotyping and evolutionary research in this species.

1. Introduction

Animal biometrics, i.e. computational representation and quantification of phenotypes, is an emerging tool for evolutionary and conservation applications in wild animals[1]. Describing individual phenotypes from images has been used to aid species identification, population size estimation, and individual identification in species such as Grevy's zebras and seals [2, 3]. Here, we study barn owls, where plumage spottiness is a key proxy for female quality and local adaptation[4, 5]. Therefore, quantifying the number and size of spots on individuals can yield minimally invasive fitness estimates, and help detect population-level adaptive evolution. However, these spots are tiny, making manual counting undesirable, subjective and most of all time consuming. Deep learning, particularly convolutional neural networks (CNN-s) can be a promising alternative to automatically identify fine-scale patterns on animal plumage nepovinnykh2024species. To get any spot-based metric for barn owls, the first step is to detect the spots. Semantic segmentation, a CNN approach that could assign every pixel in images to either spot or background, is an ideal technique for this task [6], but has not been tested in barn owl spot detection.Here, we test two CNN architectures for semantic segmentation to locate the spots on owl plumage images.

2. Methods

2.1. Data

We used a dataset of 270 images of adult and juvenile barn owls photographed at their nest boxes located around Switzerland during their breeding season between March and October. The animals are photographed in a specific "black box" to get maximum quality images, but are not harmed in the process. The images include different body parts: belly, flank, wing and back; and different base-colours of owls (the two main colourations are brown-ish and white-ish, also correlated with sex), with various "spottiness" level. Corresponding "ground truth" masks with background and spots separated were obtained with a) pre-processing (identifying the owl, homogenising images) conducted by students and b) spots identified by the software Cellpose (https://www.cellpose.org/).

2.2. Pre-processing

Images were matched to their corresponding masks, and padded to have a uniform size of 256x256 pixels, with a default black padding colour (0). Images and the corresponding masks were normalised for pixels to have values between 0 and 1 for further operations (original images had a RGB, while masks a grayscale channel, but both with values between 0 and 255).

2.3. Data augmentation

Due to the limited dataset, and the risk of overfitting evident from preliminary analyses, we created a relatively aggressive data augmentation pipeline summarised in Table 1.

Table 1. *Data augmentation.*

Operation	Range	Images?	Masks?
random flip	1:1 chance	yes	yes
small shift	<= 5px	yes	yes
random rotation	0 / 90 / 180 / 270°	yes	yes
brightness	±0.2	yes	no
contrast	0.8–1.2×	yes	no
gaussian noise	std = 0.05	yes	no

2.4. Models

For all initial models, we used a training-validation-test set split of 0.8-0.1-0.1, resulting in sample sizes of 216-24-27, respectively (before data augmentation, which increased the training set size). We initially experimented with k-fold cross-validation with various k-values, but due to the limited dataset and the various possible bases of stratification (body part, base colour, spottiness), this proved to make models severely unstable, therefore, we chose the simpler, constant training-validation-test set split.

First, we wanted to evaluate the (potential) added power from a) transfer learning and b) data augmentation. To this end, we tested a pre-trained and a non-pre-trained CNN U-net architecture for binary semantic segmentation to identify and visualise the spots on the owls' plumage; once with and once without data augmentation each. For augmented models, data augmentation was repeated 5 times, and the augmented dataset combined with the original, resulting in a training set size of 1296. All four preliminary models were fit with binary cross-entropy as the loss metric; Adam optimiser with a learning rate of 1e-4, and ran for 10 epochs. To evaluate the performances, we used two metrics: Binary Intersection of Union (BIoU) and Dice coefficient.

The two U-nets are summarised as follows:

1) Simple U-Net trained from scratch (hereafter referred to as simple U-Net): A 'lightweight' U-Net architecture with 3 encoder blocks (32->64->128 filters) and corresponding decoder blocks with transposed convolutions, with batch normalization and dropout added for regularization. Output layer is a probability map (sigmoid activation).

2) Pre-trained U-Net with a ResNet50 encoder (hereafter referred to as pre-trained U-Net): a U-Net–style CNN with a ResNet50 encoder pretrained on ImageNet (i.e. using encoder weights initialized from a network trained on the large-scale ImageNet image classification dataset). Feature maps

from the intermediate encoder layers are passed as skip connections to the decoder, which consists of upsampling blocks. A final 1x1 convolutional layer with sigmoid activation then outputs the probability of each pixel belonging to a spot.

After evaluating the preliminary models during training (loss curves and performance on validation set) and on the test set with our metrics, we conducted a hyperparameter search on the architecture that yielded the best results with the fixed hyperparameters (i.e.,the pre-trained U-net with augmentation), tuning a) learning rate (tested 1e-4 and 0.001) and b) loss function (tested BCE and combined Dice + BCE with weights of 1:0.1). Finally, we fit the model with the best hyperparameters on the combined training and validation set, and evaluated its performance on the held-out test set.

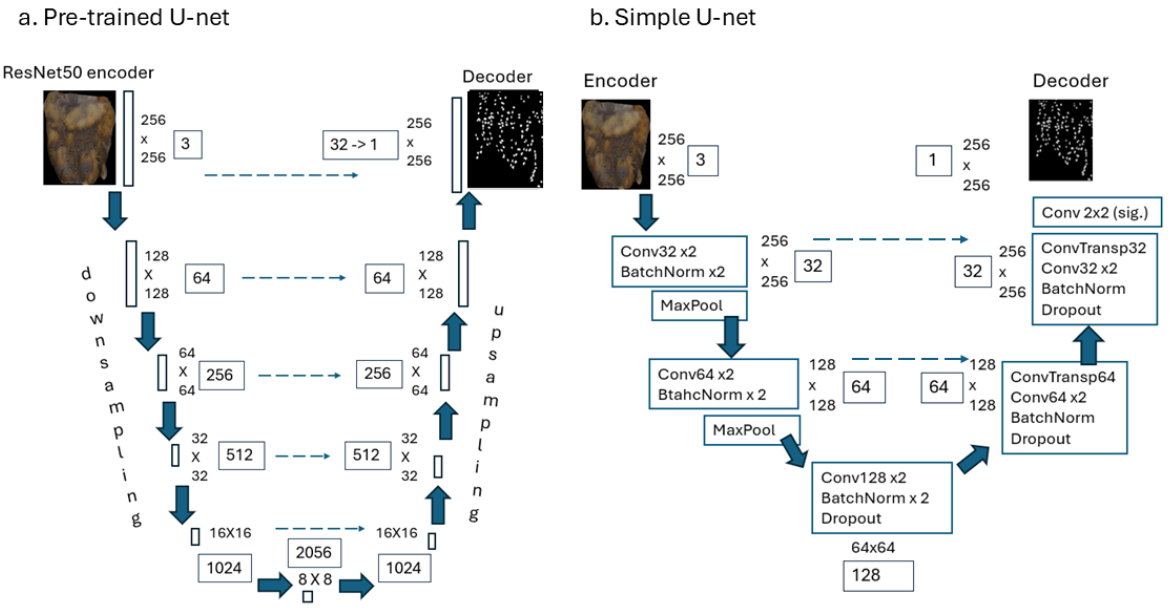

Figure 1. *Architectures of a) the pre-trained and b) the simple U-net. Numbers next to the layers indicate output dimensions, and numbers in rectangles the number of filters. Dashed lines indicate skip connections between the encoder and decoder .*

3. Results

The two primary results from fitting the initial four models was that 1) data augmentation drastically improved model performance and 2) once augmentation was applied, the pre-trained u-net outperformed the simple U-net (See Table 1). With these results in mind, we chose the pre-trained U-net with augmentation for hyperparameter tuning. The best hyperparameters, chosen as those maximising the dice coefficient on the validation set, were a learning rate of 0.001 and a loss function of combined Dice and BCE (weights 1.0:0.1), although there was not a large difference between the performances of the different hyperparameters tested. Once fit on the training + validation set, this model achieved a BIoU of 0.811 and a Dice coef. of 0.7720 on the test set (Table 2).

Learning rate	Loss	Performance (Dice) on val.set
1e-4	BCE	0.6309
1e-4	Dice+BCE	0.7467
0.001	BCE	0.6261
0.001	Dice+BCE	0.7677

Table 2. *Hyperparameter search carried out on the pre-trained UNet with data augmentation.Pink shading indicates the best-performing hyperparameter combination.*

Table 3. *Model performances. BIoU indicates binary intersection over union, and dice shows dice coefficient..*

Model	Augmentation?	Performance	
		Val. set	Test set
Simple U-net	No	bIoU = 0.4836 dice = 0.0552	bIoU = 0.4832 dice = 0.0448
Simple U-net	Yes (5x)	bIoU = 0.7653 dice = 0.3331	bIoU = 0.7601 dice = 0.2734
Pre-trained U-net	No	bIoU = 0.4836 dice = 5.82×10^{-8}	bIoU = 0.4832 dice = 3.61×10^{-7}
Pre-trained U-net before hyperparameter tuning	Yes (5x)	bIoU = 0.7540 dice = 0.6249	bIoU = 0.7576 dice = 0.6135
Final pre-trained U-net after hyperparameter tuning	Yes (5x)	Fit on training + val. set	bIoU = 0.8111 dice = 0.7720

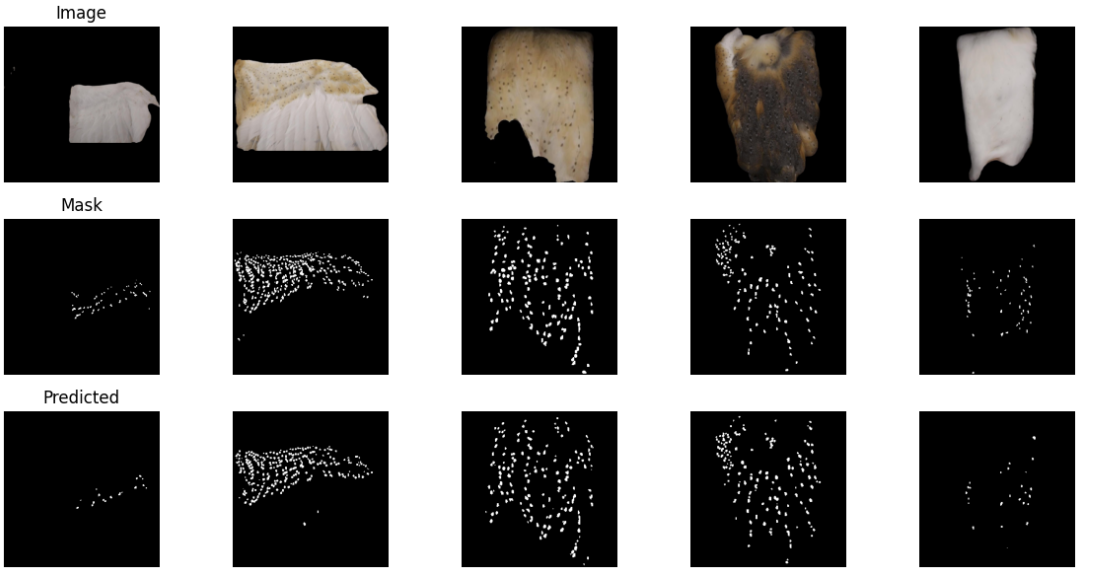

Figure 2. *Predicted masks from the final model.*

4. Discussion and conclusions

The results indicate that U-nets are appropriate to carry out semantic segmentation with the goal to locate spots on barn owls' plumages, despite the severe limitations arising from the small (250 images) and variable (various body parts, base colourations, and spot densities) nature of the dataset. Importantly, aggressive data augmentation improved training and models from practically zero performance (predicting no spots) to relatively accurate spot prediction. While a simple U-net was also able to detect spots relatively well, performance was greatly improved by using a U-net with pre-trained weights, and to some extent, by tuning the hyperparameters. This highlights the added value of data augmentation to fitting CNN-s on small datasets, a problem extremely common in wildlife studies; as well as that of transfer learning. From the predicted masks displayed, it can be observed that the architecture could still benefit from improvements: while it predicts spots on images with many spots, it under-predicts images that have fewer spots (first and last image in Fig.2.). Given that whiter owls generally have fewer spots, and spot density also often correlates with body parts (e.g. few spots on the inside of the wing), this might be improved by attaining more data from less spotty owls or body parts, or by over-sampling these combinations.

A. Appendix

The dataset I used is available here, and my code is available here.

References

[1] Hjalmar S Kühl and Tilo Burghardt. "Animal biometrics: quantifying and detecting phenotypic appearance". In: *Trends in ecology & evolution* 28.7 (2013), pp. 432–441.

[2] Ekaterina Nepovinnykh et al. "Species-agnostic patterned animal re-identification by aggregating deep local features". In: *International Journal of Computer Vision* 132.9 (2024), pp. 4003–4018.

[3] Maria Stennett, Daniel I Rubenstein, and Tilo Burghardt. "Towards Individual Grevy's Zebra Identification via Deep 3D Fitting and Metric Learning". In: *arXiv preprint arXiv:2206.02261* (2022).

[4] Alexandre Roulin et al. "Female-and male-specific signals of quality in the barn owl". In: *Journal of Evolutionary Biology* 14.2 (2001), pp. 255–266.

[5] Alexandre Roulin and Christophe Randin. "Gloger's rule in North American barn owls". In: *The Auk: Ornithological Advances* 132.2 (2015), pp. 321–332.

[6] Suet-Peng Yong, Jeremiah D Deng, and Martin K Purvis. “Novelty detection in wildlife scenes through semantic context modelling”. In: *Pattern Recognition* 45.9 (2012), pp. 3439–3450.