World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:15, No:1, 2021

Improved Rare Species Identification Using Focal Loss Based Deep Learning Models

Chad Goldsworthy, B. Rajeswari Matam

***Abstract*—**The use of deep learning for species identification in camera trap images has revolutionised our ability to study, conserve and monitor species in a highly efficient and unobtrusive manner, with state-of-the-art models achieving accuracies surpassing the accuracy of manual human classification. The high imbalance of camera trap datasets, however, results in poor accuracies for minority (rare or endangered) species due to their relative insignificance to the overall model accuracy. This paper investigates the use of Focal Loss, in comparison to the traditional Cross Entropy Loss function, to improve the identification of minority species in the "255 Bird Species" dataset from Kaggle. The results show that, although Focal Loss slightly decreased the accuracy of the majority species, it was able to increase the F1-score by 0.06 and improve the identification of the bottom two, five and ten (minority) species by 37.5%, 15.7% and 10.8%, respectively, as well as resulting in an improved overall accuracy of 2.96%.

***Keywords*—**Convolutional neural networks, data imbalance, deep learning, focal loss, species classification, wildlife conservation.

I. Introduction

THE use of motion sensing camera traps to automatically capture images of animals in the wild began in the early 1990s [1], allowing an efficient and unobtrusive method for capturing large amounts of images to be used for tracking wildlife populations, observing animal behaviour and monitoring endangered species, for example. Extracting information from these images, however, has traditionally been done by a team of experts and volunteers; a very time consuming and costly process [2].

Recent developments within deep learning, specifically convolutional neural networks (CNNs), have drastically improved the speed, accuracy and resources required for processing camera trap images, reaching accuracies of over 95% in identifying species [3]. However, the datasets used for training these models are generally heavily imbalanced (capturing a higher percentage of the majority species in comparison with the minority species), resulting in poor performance by CNNs in identifying minority species due to their relative insignificance to the overall model accuracy [4].

Although these minority species may not hold much significance to a model's overall accuracy, they do hold a great deal of significance within wildlife ecology and conservation as they are often the species closest to extinction or the species with the least amount of research [5].

This paper aims to investigate the addition of focal loss, a relatively new method for addressing imbalanced data which has shown promising results in other fields such as medicine, to the current leading deep learning model [24] for species identification in an attempt to improve the identification of minority species.

Chad Goldsworthy and B. Rajeswari Matam are with the School of Computing, Arden University, Birmingham, B4 4UA, United Kingdom (email: chadgoldsworthy@gmail.com, rmatam@arden.ac.uk).

II. Background and Related Work

A. Machine Learning

A subset of artificial intelligence, machine learning (ML) is a field of study enabling computers to achieve tasks without being explicitly programmed to do so [6], inspired by the biological ability to learn through experience and achieve learned tasks with minimal, or no, external assistance.

B. Deep Learning

Today's ML abilities to recognise and classify objects in images are due to the advancements in *deep learning*, a subset of ML referring to the use of deep neural networks, defined as multilayer artificial neural networks containing two or more hidden layers, enabling multiple levels of abstraction [7]. Unlike the requirement of feature engineering in traditional ML, deep learning enables automatic feature extraction through multiple hierarchical layers [8] which could, amongst many other applications, emulate the hierarchical organisation of biological visual systems as established in the neuroscientific field [9], [10].

C. Convolutional Neural Networks

CNNs are a class of deep neural networks for processing data that have a grid-like topology [11], such as the grid of pixels in image data, by employing convolution operations to detect features, and pooling operations to reduce computation. Fig. 1 shows a basic schematic diagram of a CNN. By operating across all regions of the input, a feature map is created representing a filtering of the input with each subsequent convolutional layer detecting further sophisticated features [12].

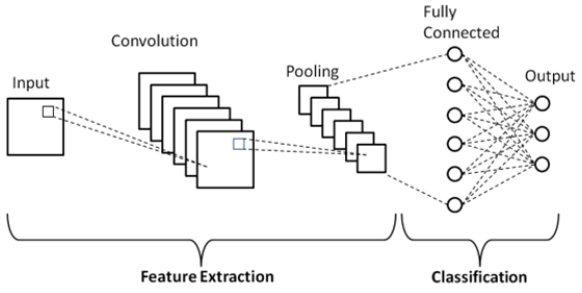

Fig. 1 Schematic diagram of a basic CNN architecture [13]

D. CNNs for Species Classification

Gomez et al. [14] compared eight different CNN models pretrained on the ImageNet dataset [15] and further trained specifically for species classification on the Snapshot Serengeti dataset [16] containing 1.5 million images at the time for 48 classes of species [2]. Their best performance was by the ResNet-101 architecture, achieving 88.9% accuracy. However, they had manually cropped the images and trained on a simplified version of the dataset containing only the 26 classes which contained the most images due to the dataset's severe measurement bias.

Norouzzadeh et al. [4] tested a range of architectures similar to Gomez et al., however they trained their network on the entire 48 classes of the Snapshot Serengeti dataset without any manual cropping with their best model being the ResNet-152 achieving an accuracy of 92.1% However, their results showed extremely low accuracy for rare classes with some species being recognized 0% of the time.

Tabak et al. [17] followed Norouzzadeh et al.'s methods and trained their model on camera trap datasets across six locations within North America, with a total of 3.7 million images across 27 classes of species. Although not as severe as the Snapshot Serengeti, their dataset was also imbalanced, containing between 1,804 to 1.8 million images per class. Their accuracies for classes containing only a few thousand images were around 70-80%, an improvement on Norouzzadeh et al.'s results; however their dataset had almost half the amount of classes and was not as severely imbalanced [17].

Most recently, Schneider et al. [3] compared six modern CNNs on the Parks Canada [18] dataset containing 47,279 images across 55 classes. Their highest performing model, the DenseNet201, achieved an accuracy of 95.6%. Their dataset was also highly imbalanced, ranging between eight to 8,566 images per class. Research suggests that their overall accuracy is the highest currently achieved for a species classification task, however their dataset is not purely species based and includes a few classes for humans too. Additionally, their accuracy for rare classes was still fairly poor, averaging 63.3% for the bottom 10 classes and, as they reported, "species with fewer training images available (< 500) produce highly variable but often poor [accuracy] [3]."

E. Methods for Reducing Measurement Bias

A common issue within CNNs, as well as other deep learning networks, is measurement bias, where certain classes occur more or less frequently than others during data collection resulting in an imbalanced dataset. This leads to a bias during training, where a model will learn the features of certain classes better than others and could result in poor accuracy for the minority classes [19]. In the Snapshot Serengeti dataset, for example, the top 50% of the species classes account for over 99% of the images [20] which could negatively affect a model's performance in identifying rare species as the model would learn that they are of very little significance to the overall accuracy.

Following their main experimental results, Norouzzadeh et al. [4] applied three methods to their ResNet-152 model in an attempt to mitigate the effects of measurement bias during training, namely weighted loss, oversampling and emphasis sampling. These methods work by putting more cost on incorrectly predicting a rare class than a frequent class, repeating examples from rare classes more often, and increasing the probability of examples being fed back into the model whenever the network misclassifies them, respectively. These methods showed promise in improving the accuracy of a few rare classes; however they did not improve *all* rare classes and in some cases deteriorated the accuracy, resulting in a reduction of the overall accuracy by up to 1.54% [4].

Although their dataset contained fewer classes (27 vs. 48 in [4]) and was not as severely unbalanced, Tabak et al. [17] achieved fairly accurate results for their rare classes by applying conditional sampling, where the percentage of images used for the training set were increased for minority classes. Despite having a decent balance of species, their accuracy of classification ~80% still leaves much room for improvement.

Schneider et al. [3] applied three methods to their model in an attempt to mitigate the bias of their dataset, namely data augmentation, transfer learning and what they refer to as classification ratio training. Data augmentation repeats images from the minority classes with transformations applied, such as rotation, mirroring or colour channel modifications, to increase the number of unique images in the class. Transfer learning refers to pre-training the model on a separate dataset, such as the ImageNet dataset in Schneider et al.'s case, to initialise the model before training on the intended dataset. Classification ratio training is similar to emphasis sampling, where underrepresented classes have a higher probability of being presented to the model repeatedly [3]. Although their accuracy was 0% in classifying some rare species, there were a few rare species which were classified between 60% and 100% accuracy, suggesting that these methods may be effective in some cases.

F. Focal Loss

There are many other methods for mitigating the effects of measurement bias, some of which, research suggests, have not yet been applied to the task of species classification.

One such method is focal loss, published in a 2018 paper by Facebook AI Research (FAIR) [21], to mitigate the effects of data imbalance by adding a modulating factor to the commonly used cross entropy loss function.

Research suggests that focal loss has not yet been applied to the task of species classification, but research in the medical field has shown it to be very effective in mitigating dataset imbalance issues. Lotfy et al. [22] compared the use of a cross entropy loss function and focal loss function on a CNN used for femur fractures classification, showing an increase in accuracy of 3% and 6%, respectively, reporting that focal loss could "address scenarios with unremarkable imbalance among the classes [22]." Pasupa et al. [23] applied focal loss to a CNN used to classifying human red blood cell morphology, showing that it significantly reduced the bias towards the

majority class when compared to the use of an ordinary cross entropy loss function.

The modulating factor which is added to the cross entropy function is defined as $(1-p_t)^\gamma$, which includes a tunable focusing parameter, $\gamma \geq 0$, resulting in the focal loss function (1):

$$FL(p_t) = -(1-p_t)^\gamma \, log(p_t) \quad (1)$$

When used in a CNN, for example, as $p_t \to 1$, the model has a high confidence in its prediction as $(1-p_t)^\gamma \to 0$ and therefore down-weights $FL(p_t)$ for well classified examples. The focusing parameter, γ, adjusts the rate that these examples are down-weighted. As reported in the FAIR paper, "In practice we use an α-balanced variant of the focal loss [...] as it yields slightly improved accuracy over the non-α-balanced form" [21], resulting in (2):

$$FL(p_t) = -\alpha(1-p_t)^\gamma \, log(p_t) \quad (2)$$

The focal loss function results in reduced contribution to the loss for well classified examples, so that the network focuses more on classifying the rare, minority classes.

III. Materials and Methods

A. Camera Trap Species Classifier

The CNN model used in this study, available on GitHub under the name "Camera Trap Species Classifier [24]", was developed by Schneider et al. and is the current leading model for species classification, achieving 95.6% accuracy [3]. It uses the MobileNetV2 model from Tensorflow's Keras library as a base model, with an additional pooling layer and four additional dense layers added.

B. Sigmoid Focal Cross Entropy

The focal loss function which was applied to the model is provided from Tensorflow's Addons repository, named 'SigmoidFocalCrossEntropy', and is an implementation of the focal loss function developed by FAIR. It includes α and γ parameters and can be passed as the loss function when compiling the CNN model, just like any other tensorflow loss function [25].

C. Bird Species Dataset

The '225 Bird Species' dataset [26], available from Kaggle, includes images of 225 different bird species. 20 species (as seen in Table I) were selected, ranging from 2-248 images per species for the training set, and 20-60 images for the testing set, with a total of 1,811 and 677 images in the training and testing set, respectively.

The selected dataset was split with 80% training and 20% testing images. However, as the original dataset was fairly balanced, the training set was manually imbalanced after the split in order to properly test the possible improvements focal loss would bring in improving the identification accuracy of minority species. Fig. 2 shows the imbalance of training images across the 20 selected species. Table I shows the list of birds and the number of images used for training.

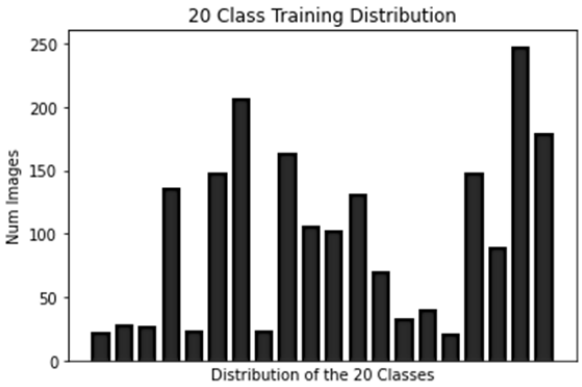

Fig. 2 Data distribution of training images for the 20 species in alphabetical order

TABLE I
Amount of Training Images per Species

Class	Species	No. Images
1	Sora	248
2	House Finch	207
3	Wood Duck	179
4	Northern Parula	164
5	Shoebill	148
6	Glossy Ibis	148
7	Common Poorwill	136
8	Peacock	132
9	Ostrich	106
10	Palila	103
11	Snowy Owl	89
12	Puffin	70
13	Black Skimmer	29
14	Javan Magpie	15
15	Carmine Bee-Eater	11
16	Razorbill	8
17	Quetzal	6
18	Crow	6
19	Robin	4
20	Araripe Manakin	2

D. Process and Performance Metrics

For this study, the model was trained on the training set using the cross entropy loss function, as was already defined within the model. The model was then trained using the focal loss function, with a range of values for α and γ applied as recommended within the FAIR paper [21]. The model was trained for a total of 20 epochs, and evaluated on the test set each time after being trained. The models w were set back to its initial, untrained state before being tested with different combinations of $\alpha$and γ to ensure a fair comparison.

The comparison of total accuracy for the two, five and ten classes with the least amount of training images (the minority classes) and the five and ten classes with the most amount of training images (the majority classes) was conducted. The two classes (19 and 20 in Table I) combined, the five classes (16-20 in Table I) combined and, the ten classes (11-20 in Table I) combined account for 0.3%, 1.4% and 13.3% of the total

amount of training images, respectively, and the five classes (1-5 in Table I) and the ten classes (1-10 in Table I) account for 52.2% and 86.8% of the total amount, respectively.

The F1-score was also evaluated, which represents a balance of accuracy across all classes and is useful for evaluating the accuracy across an unbalanced dataset. The F1-score (5) is calculated with the precision (3) and recall (4) values defined as:

$$precision = \frac{true\ positives}{true\ positives + false\ positives} \quad (3)$$

$$recall = \frac{true\ positives}{true\ positives + false\ negatives} \quad (4)$$

$$F1 = \frac{precision \times recall}{precision + recall} \quad (5)$$

A reduction in the overall accuracy may still result in a higher F1-score if the accuracy is better balanced across all classes [3].

IV. Results and Discussion

Table II shows the performance metrics after evaluating the network, trained using the cross entropy function and the focal loss function with a range of values for α and γ as recommended within the FAIR paper [21], with the highest accuracy achieved for each row in bold.

TABLE II
Summary of Performance Metrics Comparing Cross Entropy and Focal Loss

Classes	Cross Entropy Loss	Focal Loss				
		α=0.25, γ=1	α=0.25, γ=2	α=0.25, γ=5	α=0.5, γ=0.5	α=0.5, γ=1
19-20	13.46	34.48	23.18	42.03	50.00	**50.93**
16-20	42.39	51.13	56.63	54.02	20.00	**58.04**
11-20	62.70	63.65	68.03	68.80	10.00	**73.48**
1-5	**95.14**	88.70	91.24	92.28	00.00	92.85
1-10	**95.41**	92.56	95.25	93.12	00.00	93.48
f1-score	00.78	00.80	00.83	00.82	00.00	**00.84**
total test accuracy	81.68	80.50	84.19	82.87	03.40	**84.64**

The results show that focal loss improved the accuracy for the minority classes in every pair of values for the focal loss function, except for the pair where α = 0.5, γ = 0.5 which, interestingly, achieved 100% accuracy for the one class (Araripe Manakin) with the least amount of images, and 0% for all of the other 19 classes. It can also be seen that all versions of the focal loss function decreased the classes 1-5 and classes 1-10 accuracy slightly, however still resulted in an increased F1-score, showing a greater balance of accuracy across all classes.

The best result came from the pair of values α = 0.5, γ = 1 which, although worsening the accuracy for the majority classes, significantly improved the accuracy of all minority classes and increased the F1-score by 0.06.

Although the accuracy for the majority classes decreased, it can be seen that the focal loss function still resulted in an increase of the overall model accuracy as the increase in accuracy for the minority classes outweighed the decrease of accuracy for the majority classes.

V. Conclusion

The use of CNNs for the identification of species in camera trap images has, in recent years, achieved accuracies surpassing the accuracy of human classification [4]. The accuracy in identifying rare, minority species still remains fairly poor due to their relative insignificance to the models' overall accuracy [4], however their identification may be of more importance in some cases within wildlife ecology and conservation efforts [5].

In this study, the application of the focal loss function to the leading CNN for species identification was investigated to improve the identification of rare species. The results showed that the focal loss function was able to increase the F1-score by 0.06, and improve the identification of classes 19-20, classes 16-20 and classes 11-20 (minority, shown in Table I) species by 37.5%, 15.7% and 10.8%, respectively.

Although the focal loss function decreased the accuracy of the majority species, and may decrease the overall accuracy in larger datasets, it has shown itself to be useful in specific applications where the identification of rare or endangered species is of greater importance than overall identification accuracy.

References

[1] T. G. O'Brien, "Abundance, Density and Relative Abundance," in *Camera Traps in Animal Ecology,* A. F. O'Connel, J. D. Nichols, K. U. Karanth. Tokyo: Springer, 2011, pp. 71-96.

[2] A. Swanson, M. Kosmala, C. Lintott, C. Packer, "A generalised approach for producing, quantifying, and validating citizen science data from wildlife images," *Conservation Biology,* vol. 30, issue 3, pp. 520-531, Apr 2016.

[3] S. Schneider, S. Greenberg, G. W. Taylor, S. C. Kremer, "Three critical factors affecting automated image species recognition performance for camera traps," *Ecology and Evolution*, vol. 10, issue 7, pp. 3503-3517, Mar 2020.

[4] M. S. Norouzzadeh, A. Nguyen, M. Kosmala, "Automatically identifying wild animals in camera trap images with deep learning," *Proceedings of the National Academy of Sciences*, vol. 115, issue 25, pp. E5716-E5725, Jun 2018.

[5] J. Martin, W. M. Kitchens, J. E. Hines, "Importance of well-designed monitoring programs for the conservation of endangered species," *Conservation Biology,* vol. 21, issue 2, pp. 472-481, Apr 2007.

[6] A. Panesar, *Machine Learning and AI for Healthcare.* Coventry: Apress, 2019.

[7] M. Mitchell, *Artificial Intelligence: A Guide for Thinking Humans.* London: Penguin, 2019.

[8] Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, et al., "The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches," Post-Doctoral research, Dept. Comp. Sci., Univ. of Dayton, OH, 2018.

[9] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," *The Journal of Physiology*, vol. 148, issue 3, pp. 574-591, Oct 1959.

[10] S. Sutherland, "The vision of David Marr," *Nature*, vol. 298, pp. 691-692, Aug 1982.

[11] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning (Adaptive Computation and Machine Learning series).* Cambridge: MIT Press, 2016.

[12] Z. Yang, T. Dan, Y. Yang, "Multi-temporal Remote Sensing Image Registration Using Deep Convolutional Features," *IEEE Access*, vol. 6, pp 38544-38555, Jul 2018.

[13] V. H. Phung and E. J. Rhee, "A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets," *Applied Sciences*, vol. 9, issue 21, pp. 4500, Nov 2019.
[14] A. Gomez, G. Diez, A. Salazar, A. Diaz, "Animal Identification in Low Quality Camera-Trap Images Using Very Deep Convolutional Neural Networks and Confidence Thresholds," in *International Symposium on Visual Computing*, Las Vegas, NV, 2016, pp. 747-756.
[15] ImageNet. (2016, May 31). *Large Scale Visual Recognition Challenge 2016 (ILSVRC2016)* (Online). Available: http://image-net.org/challenges/LSVRC/2016/
[16] Zooniverse. (2020). *Snapshot Serengeti* (Online). Available: https://www.zooniverse.org/projects/zooniverse/snapshot-serengeti
[17] M. A. Tabak, M. S. Norouzzadeh, D. W. Wolfson, S. K. Sweeney, et al., "Machine learning to classify animal species in camera trap images: Applications in ecology," *Methods in Ecology and Evolution*, vol. 10, issue 4, pp. 585-590, Nov 2018.
[18] Parks Canada. (2019, Oct. 16). *Wildlife webcams and remote cameras* (Online). Available: https://www.pc.gc.ca/en/nature/science/controle-monitoring/cameras
[19] O. J. Robinson, V. R. Gutierrez, D. Fink, "Correcting for bias in distribution modelling for rare species using citizen science data," *Diversity and Distributions*, vol. 24, issue 4, pp. 460-472, Dec 2017.
[20] A. Swanson, M. Kosmala, C. Lintott, R. Simpson, A. Smith, C. Packer, "Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna," *Scientific Data*, vol. 2, Article 150026, Jun 2015.
[21] T. Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, "Focal Loss for Dense Object Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, issue 2, pp. 318-327, Feb 2020.
[22] M. Lofty, R. Shubair, S. Albarqouni, "Investigation of Focal Loss in Deep Learning Models For Femur Fractures Classification" in *The 2019 IEEE International Conference on Electrical and Computing Technologies and Applications*, UAE, 2019.
[23] K. Pasupa, S. Vatathanavaro, S. Tungjitnob, "Convolutional neural networks based focal loss for class imbalance problem: a case study of canine red blood cells morphology classification," *Journal of Ambient Intelligence and Humanized Computing*, Feb 2020.
[24] S. Schneider. (2019, Sep. 9) *Camera Trap Species Classifier* (Source Code). Available: https://github.com/Schnei1811/Camera_Trap_Species_Classifier
[25] Tensorflow. (2020, Aug. 5). *Focal_loss.py* (Source Code).Available: https://github.com/tensorflow/addons/blob/v0.11.2/tensorflow_addons/losses/focal_loss.py
[26] G. Piosenka. (2020, Jul. 27). *225 Bird Species* (Online). Available: https://www.kaggle.com/gpiosenka/100-bird-species