Ecological Informatics 64 (2021) 101350

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

A systematic study of the class imbalance problem: Automatically identifying empty camera trap images using convolutional neural networks

Deng-Qi Yang[a,*], Tao Li[a], Meng-Tao Liu[a], Xiao-Wei Li[b], Ben-Hui Chen[a]

[a] Department of Mathematics and Computer Science, Dali University, Dali, Yunnan 671003, China
[b] Data Security and Application Innovation Team, Dali University, Dali, Yunnan 671003, China

ARTICLE INFO

Keywords:
Camera trap images
Class imbalance
Convolutional neural networks
Deep learning
Image classification

ABSTRACT

Camera traps, which are widely used in wildlife surveys, often produce massive images, and many of them are empty images not contain animals. Using the deep learning model to automatically identify the empty camera trap images can reduce the workload of manual classification significantly. However, the performance of deep learning models is easily affected by the class imbalance problem of training datasets, which is a common problem for actual wildlife survey projects. Almost all previous studies on empty image recognition used downsampling or oversampling methods to eliminate the effect of class imbalance on the performance of deep learning classifiers. The class imbalance problem has been systematically studied in the field of traditional image recognition, yet very limited research is available in the context of identifying camera trap images taken from highly cluttered natural scenes. This study systematically studied the impact of class imbalance on model performance when using a deep learning model to identify empty camera trap images. Then we proposed the construction method of training sets of the deep learning model when the data set has different class imbalance levels. Based on results from our experiments we concluded that (i) the class imbalance showed little effect on the performance of the model when the empty image ratio (EIR) in the data set was between 10% and 70%, so the training sets can be randomly built without changing the class distribution; (ii) we recommended using oversampling to partially eliminate class imbalance to reduce omission errors when the EIR of the data set exceeded 70%; (iii) when the EIRs of the training set and the test set were close, the overall error, omission error, and commission error of the model were relatively smaller, and the model tended to achieve a better overall performance; (iv) the omission and commission errors can be adjusted by changing the percentage of empty images in the training set.

1. Introduction

Camera traps are widely used in wildlife surveys for their non-invasive and easy-to-use advantages (Duarte, 2017). Camera traps are deployed in an ecological preservation area, and the cameras are automatically triggered to take images when animals pass through the camera view (He et al., 2016). Ecologists can obtain a lot of useful information of ecosystems from the camera trap images, such as species richness (Dertien et al., 2017), species distribution and community structure (Rich et al., 2019), animal behavior (Frey et al., 2017), etc. Camera traps provide an effective tool for humans to better observe and preserve ecosystems (Kays et al., 2017; Steenweg et al., 2016; Wei et al., 2020).

However, camera traps usually produce massive images due to the complexity of the natural environment, and many of them are empty images not contain animals. For examples, the percentages of empty images in datasets from three survey projects in Africa, Snapshot Serengeti (SS; Swanson et al., 2015), Camera CATalogue (CC) (Willi et al., 2019), and Elephant Expedition (EE) (Willi et al., 2019), are 75%, 22%, and 83% respectively. The proportion of empty images in a dataset that takes from a Colombian survey project even reaches 99% (Diaz-Pulido and Payan, 2011). It is often necessary to manually remove empty images from the camera trap dataset before conducting ecological research, which is a time-consuming and tedious task. Therefore, utilizing computer technology to automatically identify and remove empty images from data sets is of great significance for the survey and preservation of wildlife.

In recent years, deep learning technologies, especially deep

* Corresponding author.
E-mail address: dqyang@dali.edu.cn (D.-Q. Yang).

https://doi.org/10.1016/j.ecoinf.2021.101350
Received 19 October 2020; Received in revised form 25 April 2021; Accepted 4 June 2021
Available online 10 June 2021

convolutional neural networks (DCNNs; Chang and Chen, 2015; Krizhevsky et al., 2012; Szegedy et al., 2015; Yousif et al., 2017; Zeiler and Fergus, 2014), have shown outstanding achievements in the field of image recognition. Deep learning technology is a supervised machine learning method, which enables computers to automatically extract hierarchical features from raw data. It uses a large number of samples with classification labels to train the model so that the model can learn how to map the input (images) to different outputs (categories). Deep Convolutional Neural Networks (DCNNs) is one of the most popular deep learning models in the field of image classification, which can output all probabilities that the image belongs to different categories. Many researchers (Norouzzadeh et al., 2018; Tabak et al., 2019; Willi et al., 2019; Yousif et al., 2019) are trying to use the DCNN model to automatically identify and filter empty images from camera traps.

A common problem is that one class in the data set have a significantly higher number of examples than another class when using DCNN models to automatically identify images. This difference is referred to as class imbalance (Chawla, 2010; Japkowicz and Stephen, 2002; Mazurowski et al., 2008). The performance of the DCNN model is often affected by the class imbalance of the training set, and this effect is considered to be detrimental to the model's performance (Buda et al., 2018; Johnson and Khoshgoftaar, 2019). Methods for solving class imbalance problems can be divided into data-level methods, algorithm-level methods, and hybrid methods (He and Garcia, 2009; Johnson and Khoshgoftaar, 2019). The essence of the data-level method is to operate on the training set to change its class distribution. The most straightforward and common data-level method is sampling, including both oversampling (Chawla et al., 2002; Guo and Viktor, 2004; Jaccard et al., 2017; Bennin et al., 2018; Marouf et al., 2020; Shen et al., 2016) and down-sampling methods (Drummond and Holte, 2003; Japkowicz and Stephen, 2002; Kubat, 1997; Ling and Li, 1998; Liu et al., 2009). The simplest oversampling method is to randomly repeat the samples in minority classes to eliminate class imbalance. Down-sampling, called random majority down-sampling, which removes a portion of examples from majority classes randomly. The idea of the algorithm-level method is to keep the class distribution of the data set unchanged and adjust the machine learning algorithm to improve the model performance (Li et al., 2017; Havaei et al., 2017; Japkowicz et al., 2000; Lee and Cho, 2006; Zhou and Liu, 2006). According to our review of the literature, the data-level methods are the most commonly used when using the deep learning model to identify empty camera trap images.

The class imbalance problem has been systematically studied in the field of typical image recognition (Buda et al., 2018; Johnson and Khoshgoftaar, 2019), but there are very limited research is available in the context of identifying camera trap images taken from highly cluttered natural scenes. At present, the sampling method was used to eliminate the class imbalance of the dataset by almost all researchers (Norouzzadeh et al., 2018; Willi et al., 2019; Yousif et al., 2019) when trying to automatically identify empty camera trap images. Based on the SS dataset, Norouzzadeh et al. (2018) and Willi et al. (2019) used the down-sampling method to build balanced training sets and balanced test sets. Yousif et al. (2019) oversampled the SS dataset by adjusting the color content of the new images to construct a balanced dataset. All these existing researches have achieved a high accuracy of more than 95%.

It seems to be a good solution using the DCNN models to automatically filter the empty camera trap images (Norouzzadeh et al., 2018; Willi et al., 2019; Yousif et al., 2019). However, none of the previous studies answers the following questions: (1) How well the model trained on a balanced training set performs on the unbalanced data set from the actual survey project? And whether the class imbalance must be eliminated when using the DCNN model to automatically filter empty camera trap images? (2) How does class imbalance affect model performance? (3) How should the training set of a DCNN model be constructed so that the model can achieve the best performance on the actual dataset? This study designed a series of experiments to answer the above questions. Because most existing studies used the AlexNet model (Krizhevsky et al., 2012) and the Snapshot Serengeti data set, we also used them, which facilitates the comparison of experimental results.

2. Method

2.1. Data

The SS dataset is by far the world's largest publicly available camera trap images dataset, with images taken from 255 cameras deployed in the Serengeti National Park in Tanzania. The SS dataset contains 3.2 million images with an empty image ratio (EIR) of 75%. The dataset contains 48 species and each image was marked as empty or species name. The SS dataset contains six subsets of S1, S2, S3, S4, S5, and S6. The subset S1 includes 168 camera sites with a total of 411,418 images. This work took all images of the top-100 sites in the S1 as experimental data, notated as S1_100. The S1_100 dataset contains 255,020 images with an EIR of 79.54%.

Many images were of poor quality because some animal in images of the dataset S_100 were too close to the camera, some were far away from the camera, and some images only contained a part of the animal's body (Fig. 1). Moreover, image backgrounds were highly variable even for images from the same site due to complex climatic conditions and changing weather. The S_100 dataset included both day and night images. All of the images were in RGB format and were 2048 × 1536 pixels in size. All images were resized to 227 × 227 using the resize () method Python Imaging Library (v.6.2.1, python-pillow.org, accessed 5 May 2018) of the Python 3.5.4 platform.

2.2. DCNN model

The AlexNet model (Krizhevsky et al., 2012) was used in this study, which is a classic DCNN model. The optimization algorithm of the AlexNet model used stochastic gradient descent (SGD) with a momentum value of $\mu = 0.9$. The base learning ratio was $\eta = 0.01$, and the learning ratio was set to 0.001 if the accuracy of the model on the training set reached 90%. The batch size was set to 128, and the network weights were randomly initialized with the normal distribution. Except for the last fully connected layer, each layer used Local Response Normalization (LRN) processing and Rectified Linear Unit (ReLU) activation function processing.

2.3. Experimental method

2.3.1. Experiment I: Comparison of sampling methods

The goal of Experiment I was to study the impact of three sampling methods on model performance, including non-sampling, down-sampling, and oversampling. We chose 120,000 images from the S-100 data set to perform the 5-fold cross-validation. Our experiment included three phases. The first, 100,000 images were randomly selected from these images as the training set each time, and the remaining 20,000 images were used as the test set. This process was repeated five times to obtain five imbalanced training sets and five imbalanced test sets. The EIRs of these training sets and test sets were between 78.69% and 79.67%, which were comparable to the EIR of the S-100 data set. Then, we trained the AlexNet model on these five imbalanced training sets and tested their errors on five imbalanced test sets. The average errors on five test sets were taken as the model error of the non-sampling method. Second, the five imbalanced training sets were down-sampled by reducing the number of empty images to construct the balanced training sets. Then we trained the AlexNet model on these five balanced training sets and tested their errors using the above imbalance test sets. The average errors on these five tests were taken as the model errors of the down-sampling method. Finally, oversampling the five imbalanced training sets by increasing the number of animal images to build a balanced training set, and train the AlexNet model on these five balance

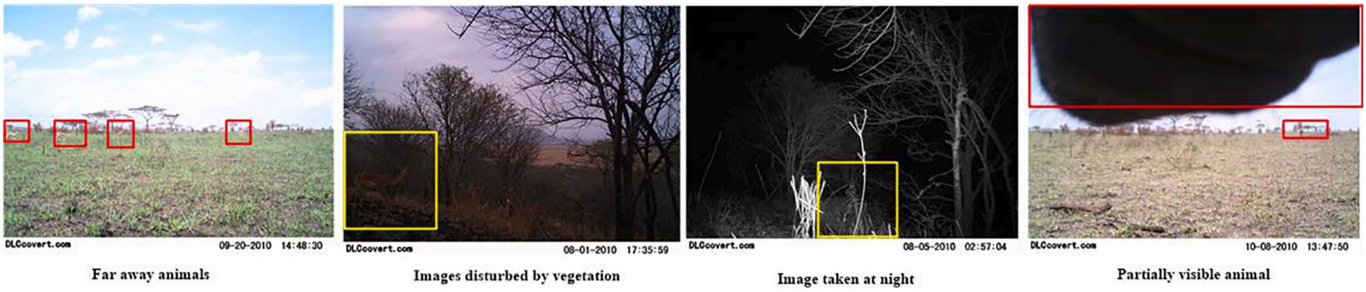

Fig. 1. Various factors make identifying animals in the wild hard even for humans. There are animals in the rectangle marked part of the images.

training sets, and then use the same test set for testing. The average errors on five test sets were taken as the model errors of the down-sampling method.

2.3.2. Experiment II: Effects of class imbalance on model performance

2.3.2.1. Data splitting. The purpose of experiment II was to study the effects of class imbalance of the dataset on the model performance when using the DCNN model to automatically filter empty camera trap images. A series of training and test sets with different levels of class imbalance was constructed by randomly extracting images from the S-100. The class imbalance levels of image datasets from actual survey projects vary widely due to the significant differences in the ecological environment. Test sets with different levels of class imbalance were constructed to simulate the datasets to be identified in actual projects. The test sets consisted of type-A (Table 1) and type-B (Table 2) sequences. The EIR of the test set in type-A sequence was increased from 50% to 90%, while that in type-B sequence was decreased from 50% to 10%. Test sets in type-A and type-B sequences were notated as $TestA_i$ and $TestB_i$ ($i = 1, 2, \ldots, 5$), respectively. The test sets of type-A satisfied the Eq (1), and the test sets of type-B satisfied the Eq (2).

$$TestA_1 \subset TestA_2 \subset TestA_3 \subset TestA_4 \subset TestA_5 \tag{1}$$

$$TestB_1 \subset TestB_2 \subset TestB_3 \subset TestB_4 \subset TestB_5 \tag{2}$$

Training sets with different EIRs, denoted as $Train_i$ ($i = 1, 2, \ldots, 9$), were constructed to train the DCNN model (Table 3). These training and test sets satisfied the Eq (3) and the Eq (4).

$$Train_1 \subset Train_2 \subset \ldots \subset Train_9 \tag{3}$$

$$Train_i \cap TestA_j = Train_i \cap TestB_k = \varnothing, i = 1, 2, \ldots, 9, j, k = 1, 2, \ldots, 5 \tag{4}$$

2.3.2.2. Experiment. Datasets collected from actual ecological survey projects often have different levels of class imbalances. The purpose of experiment II was to study the impact of dataset class imbalance on the performance of the model. On this basis, we explored the construction method of training sets of the DCNNs model when the image datasets had different levels of class imbalance so that the model can achieve the best performance on the target test set. In this experiment, we used the 5-fold cross-validation to train the AlexNet model and tested its performance. Specifically, we randomly chose animal images and empty images according to the numbers of images in the training and test sets in Tables 1–3 to build the training sets and test sets for cross-validation. We tested the models, which trained on the same training sample size, on the test sets with the same sample size, and took the average errors of these models as the errors of the model under this training sample size.

Table 1
Type A test dataset sequence with different empty image ratios (EIRs)[a].

Data sets	N_A	N_E	N	EIR (%)
$TestA_1$	6000	6000	12,000	50
$TestA_2$	6000	9000	15,000	60
$TestA_3$	6000	14,000	20,000	70
$TestA_4$	6000	24,000	30,000	80
$TestA_5$	6000	54,000	60,000	90

[a] N_A was the number of animal images. N_E was the number of empty images. N was the number of images in the dataset. EIR represented the empty images ratio.

Table 2
Type B test data set sequence with different empty image ratios (EIRs)[a].

Data sets	N_A	N_E	N	EIR (%)
$Test\ B_1$	4000	4000	8000	50
$Test\ B_2$	6000	4000	10,000	40
$Test\ B_3$	9333	4000	13,333	30
$Test\ B_4$	16,000	4000	20,000	20
$Test\ B_5$	36,000	4000	40,000	10

[a] N_A was the number of animal images. N_E was the number of empty images. N was the number of images in the dataset. *EIR* represented the empty images ratio.

Table 3
Training datasets with different Empty Image Ratios (EIRs)[a].

Data sets	N_A	N_E	N	EIR (%)
$Train_1$	15,000	1667	16,667	10
$Train_2$	15,000	3750	18,750	20
$Train_3$	15,000	6429	21,429	30
$Train_4$	15,000	10,000	25,000	40
$Train_5$	15,000	15,000	30,000	50
$Train_6$	15,000	22,500	37,500	60
$Train_7$	15,000	35,000	50,000	70
$Train_8$	15,000	60,000	75,000	80
$Train_9$	15,000	135,000	150,000	90

[a] N_A was the number of animal images. N_E was the number of empty images. *N* was the number of images in the dataset. EIR represented the empty images ratio.

2.3.3. Error evaluation method

The most widely used metric to evaluate the performance of a deep learning classifier is overall accuracy, which is the proportion of test samples that were correctly classified. However, it has some significant limitations, particularly in the context of imbalanced datasets (Chawla, 2010). An example of this is a situation the classier assigns the empty label to all test images when the empty images represent 99% of all images. A misleading accuracy of 99% will be assigned to the classier, but it has limited use due to its error rate is 100% when identifying non-empty images. We used three metrics of the overall error, the commission error, and the omission error to evaluate the model performance in this study, which was defined by Eq (5), Eq (6), and Eq (7).

$$\text{Overall error} = (FN + FP)/(TP + TN + FN + FP) \tag{5}$$

$$\text{Commission error} = FP/(FP + TP) \tag{6}$$

$$\text{Omission error} = FN/(TP + FN) \tag{7}$$

Here, TP was the number of images that were labeled as animal images by both experts and models. TN was the number of images that were labeled as empty images by both experts and models. FN was the number of images that were labeled as animal images by experts but labeled as empty images by the model. FP was the number of images that were labeled as empty by experts but labeled as animals by the model. The commission error intuitively reflected the ratio that the model incorrectly assigned empty labels to non-empty images. Oppositely, the omission error explicitly reflected the ratio that the model incorrectly assigned non-empty labels to empty images.

3. Results

3.1. Results of experiment I

The experimental results of experiment I showed that the overall error and commission error of the non-sampling method were smaller than those of the oversampling and down-sampling methods; and the omission error of the non-sampling method was bigger than that of the oversampling and down-sampling method (Fig. 2a). The mean value and the standard deviation of three errors of non-sampling were much smaller than those of the oversampling and down-sampling method (Fig. 2b). The mean value and the standard deviation of the three errors of the oversampling method were small than those of the down-sampling (Fig. 2b).

3.2. Results of experiment II

The experimental results of experiment II showed that the errors of the DCNN model on a test set were significantly affected by the levels of class imbalance of the training sets. The model always achieved the best performance when the EIR of the training set was consistent with that of the test set (Fig. 3 and Fig. 4). For example, when the EIR of the training set was 10%, the model performed best on the test set Test-B5, which also had an EIR of 10% (Fig. 4e). The model achieved the best performance on the test set $TestA_1$ and $TestB_1$ when the EIR of the training set was 50% (Fig. 3a and Fig. 4a). When the EIR of the training set was 90%, the errors of the model on the test set $TestA_5$ was the best. Furthermore, for any test set, with the increase of the EIR of the training set, the model tended to predict more non-empty images as empty labels, resulting in increased omission error and decreased of commission error.

4. Discussion

The target data sets that needed to be classified in actual ecological projects were usually imbalanced. However, we found that model trained on the balanced training set tended to produce a large commission error when the empty images in the test set account for the majority (Fig. 2a). The experiment results of experiment II also showed the model trained with the balanced training sets performed poorly on the unbalanced test set. What's more, the larger the difference between the EIR of the test set and that of the training set, the larger the commission error of the model.

Buda et al. (2018) used three benchmark datasets of increasing complexity, MNIST(Lecun et al., 1998), CIFAR-10 (Krizhevsky and Hinton, 2009), and ImageNet (Deng et al., 2009), to investigate the effects of class imbalance of training data set on classification performance of DCNN model. Their experiment results show that the oversampling method is almost always better than the down-sampling method, which is consistent with our experimental results (Fig. 2). Their experimental results also show that for simple data sets, oversampling can almost always reduce the model errors, and down-sampling usually increases the model errors. Whereas for complex data sets, these two sampling methods are worse than the non-sampling method. Our experimental results of Experiment I showed that neither oversampling nor down-sampling can reduce the model errors. So we speculate that the camera trap image dataset was a complex data set because it was greatly affected by many environmental factors, such as climate, weather, terrain, and vegetation.

Additionally, using the down-sampling method, Norouzzadeh et al. (2018) achieved a good performance, but both their training and test sets were balanced, and they have not tested their models on unbalanced test sets, which held real-world class distribution. Willi et al. (2019) and Yousif et al. (2019) respectively used down-sampling and oversampling to construct their balanced training sets. Although they did not indicate whether the test sets were balanced, their test sets should also be balanced, which can be concluded from the analysis of experimental data.

Our experiment results indicated that the omission error of the model increases as the EIR of the training set increases, and the model trended to achieve the best performance at the most time when the EIR of the training set was close to that of the test set. It meant the non-sampling method was a good alternative for identifying the empty camera trap images at most times. The mean value and standard deviation of the three errors of the model were the smallest when using the non-sampling method to build the training set. We presented the errors of the model

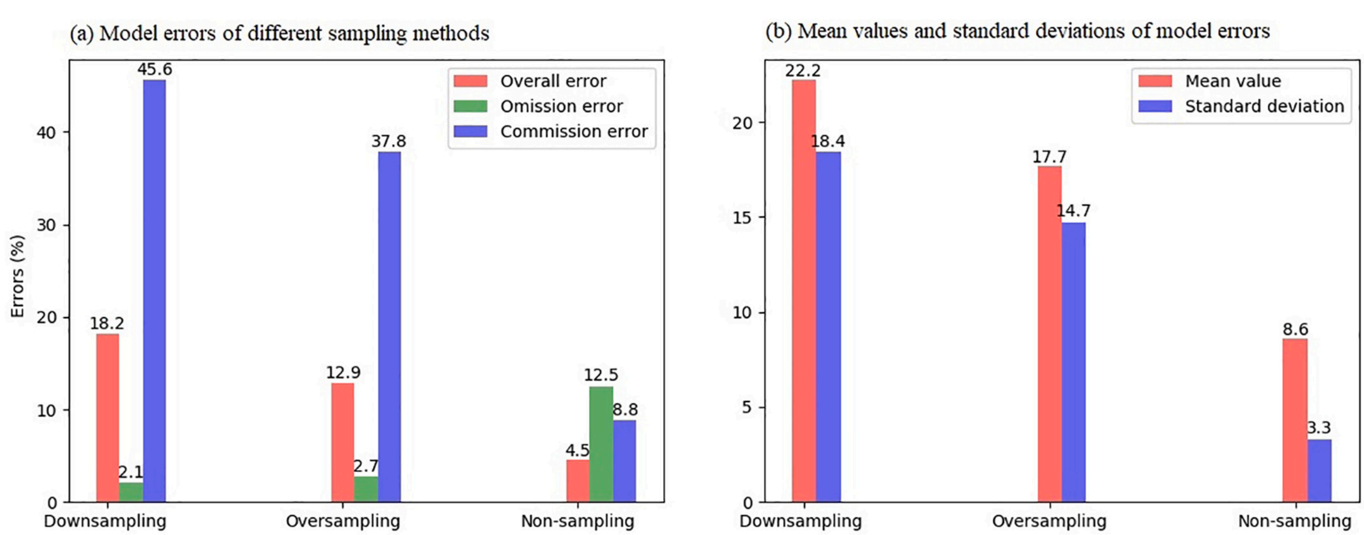

Fig. 2. The experiment results of experiment I.

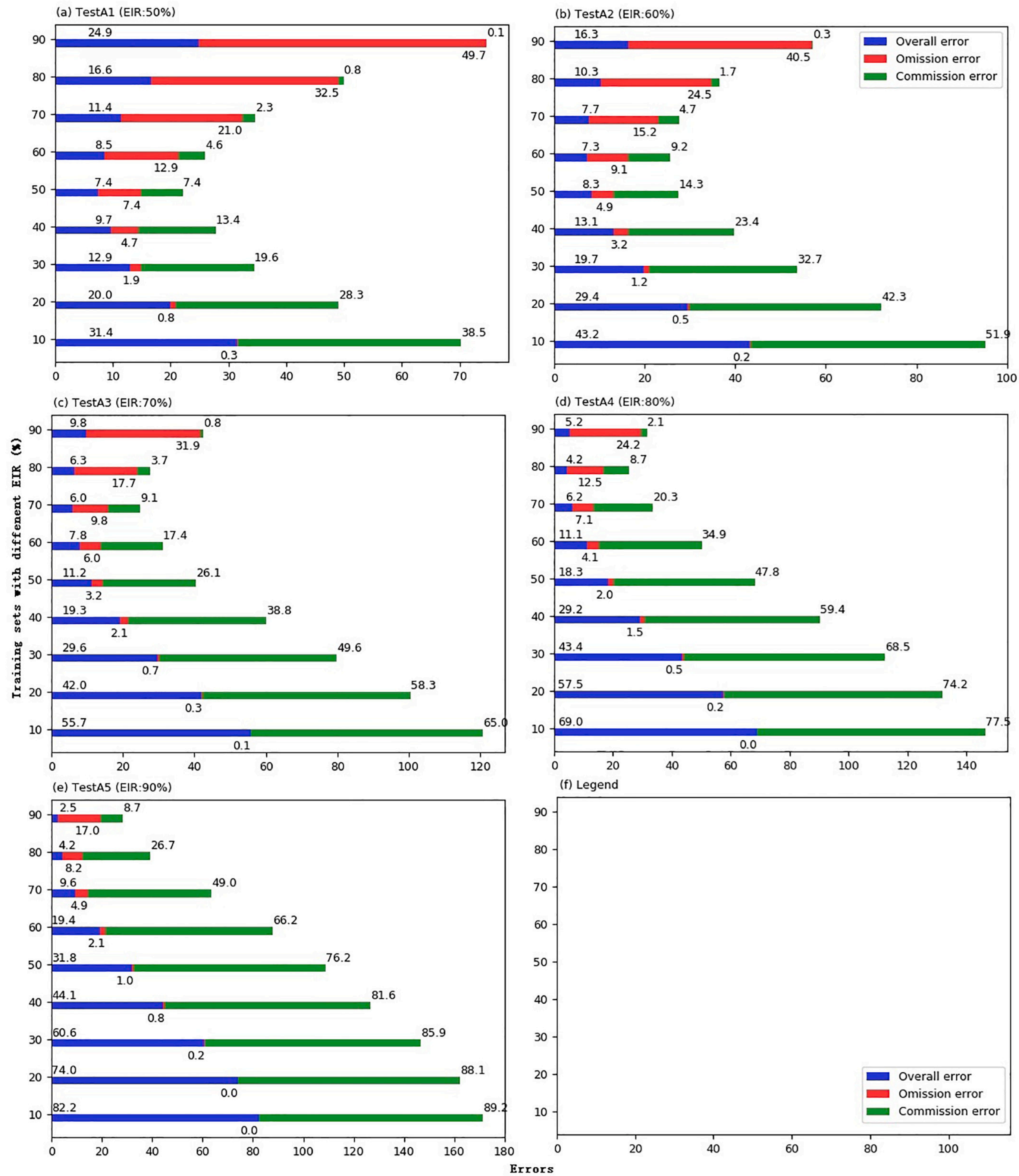

Fig. 3. Errors of different models on the type-A test sets.

built by the non-sampling method when the EIR of the data set was increased from 10% to 90% (Fig. 5). It can be found that the standard deviation of the three errors was the smallest when both the test and training sets are balanced (Fig. 5).

Buda et al. (2018) also studied the impact of training sets with different levels of class imbalance on the model performance, but all of their test sets were balanced. Their purpose of using the sampling method to completely or partially eliminate the class imbalance of training sets is to make the class imbalance level of the training and test sets closer. In this way, the model can achieve a better performance which was consistent with our conclusions.

When the animal images in the data set were of the majority class

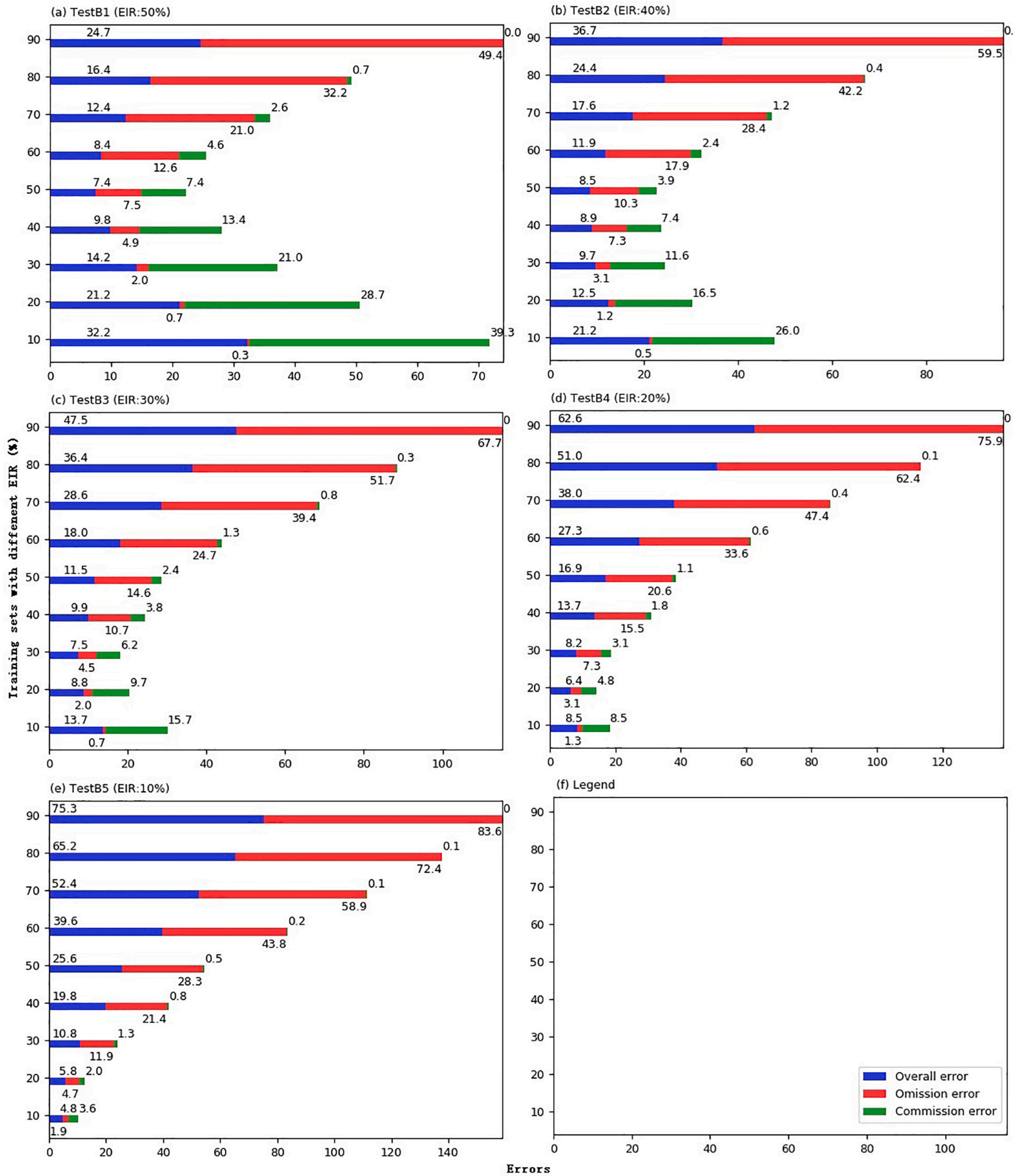

Fig. 4. Errors of different models on the type-B test sets.

(EIR < 50%), although the class distribution of the data set was imbalanced, its impact on the model performance was small. All errors of the model built by the non-sampling method did not exceed ~10% when the EIR of the data set did not exceed 70%, and as the number of training samples increases, these errors will be further reduced. As a result, we suggested that when the EIR of the camera trap image data set was less than 70%, the non-sampling method was used to construct the training set. That was, to randomly select a part of the samples directly from the original data set as the training set. In this way, the model would achieve the best performance because the EIR of the training set and that of the target test set to be classified were consistent.

When the EIR of the data set was more than 70%, the mean value and

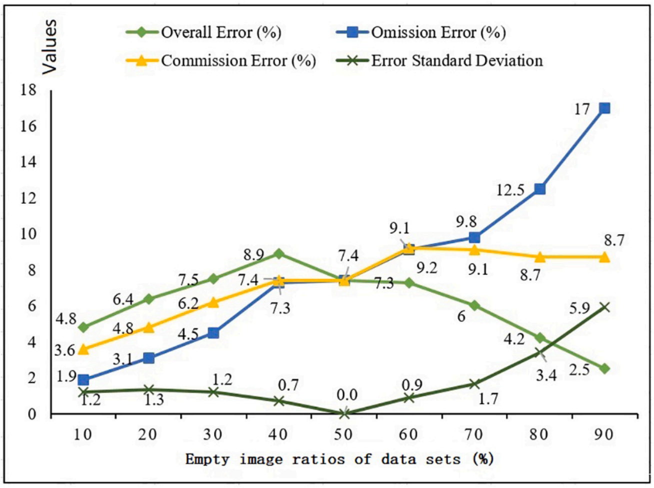

Fig. 5. Errors and error standard deviations of the model built by the non-sampling method when the EIR of datasets from 10% to 90%.

standard deviation of the model built by the non-sampling method were still the smallest (Fig. 6). However, the model produced a large omission error and may lead to many animal images be predicted as empty images, which could miss the opportunity to discover and observe the species.

We found that the omission error and commission error were opposites, which can be adjusted by changing the EIR of the training set. That was, increasing the animal images in training can reduce the omission error of the model, and increasing the empty images in the training set can reduce the commission error. If users hope to minimize the possibility of omitting the animal images, they can reduce the EIR of the training set. If users want to minimize the workload of manually recognizing empty images, they can increase the EIR of the training set. To reduce the omission error, we attempted to eliminate the class imbalance partly by the oversampling method. We constructed three training sets with different EIRs by repeating all non-empty images in the dataset *Train*$_9$ multiple times (Table 4). Then the AlexNet model was trained on these training sets and tested on the test set *TestA*$_5$. The experimental results showed that using the oversampling method to decrease the EIR of the training set, the overall and commission errors were increased, but the omission error can be reduced (Table 5). It can reduce the probability of missing species.

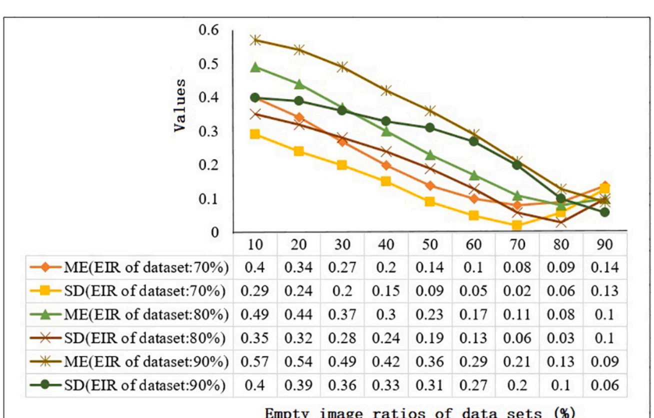

Fig. 6. Error mean values and error standard deviations of the model built by the non-sampling method when the EIR of datasets increased from 70% to 90%. The ME and SD were the mean value and standard deviation of three errors respectively.

Table 4
Training sets built by oversampling the Train$_9$ multiple times[a]

Data sets	N_A	N_E	N	EIR (%)
Train$_9$	15,000	135,000	150,000	90
Train$_{9\_1}$	30,000	135,000	165,000	82
Train$_{9\_2}$	45,000	135,000	180,000	75
Train$_{9\_3}$	60,000	135,000	195,000	69

[a] N_A was the number of animal images. N_E was the number of empty images. *N* was the number of images in the dataset. *EIR* represented the empty images ratio.

Table 5
Model errors under partly oversampling.

	Train$_9$	*Train*$_{9\_1}$	*Train*$_{9\_2}$	*Train*$_{9\_3}$
Overall error	2.50	2.57	3.76	4.84
Omission error	17.00	11.59	7.76	6.41
Commission error	8.70	13.77	24.44	30.93

5. Conclusion

The class imbalance problem of datasets is a common problem when using the DCNN model to identify the empty camera trap images, which is generally considered to affect the performance of the classifier. We designed two experiments to study the impact of the class imbalance problem on the performance of DCNN models in this work. Experiment I systematically compared the DCNN model performance using different sampling methods. Experiment II artificially constructed multiple training and test sets with various levels of the class imbalance and studied the impact of the class imbalance problem on model performance. Then we explored the construction method of training sets for DCNN models when the camera trap image datasets have different levels of class imbalance. We drew the following conclusions based on the experimental results:

(1) The model based on the non-sampling method achieved a better overall performance than the model based on the sampling method (including over-sampling and down-sampling) when using the AlexNet model identifying empty camera images.
(2) When the animal images in the data set were of the majority class (EIR $<$ 50%), the class imbalance impact of the data set on the model performance was little. Compared with animal images, empty images have a significant effect on the performance of the model.
(3) When the EIR of the data set was not more than 70%, we suggested randomly divide the data set to build the training set rather than use the sampling methods to eliminate the class imbalance of the training set.
(4) When the EIR of the data set was more than 70%, although the overall performance of the model built using the non-sampling method was better than the overall performance of the model built using the sampling method, the previous will produce larger omission errors. The larger the EIR of the training set, the bigger the omission error of the model.
(5) Increasing the animal images in training can reduce the omission error of the model, and increasing the empty images in the training set can reduce the commission error.

Declaration of Competing Interest

None.

Acknowledgments

We appreciate the support of the National Natural Science

Foundation of China (31960119) and the Yunnan Provincial Science and Technology Department University Joint Project (2017FH001-027) and the Innovative Project of Dali University (ZKLX2020308).

References

Bennin, K.E., Keung, J., Phannachitta, P., Monden, A., Mensah, S., 2018. [journal first] MAHAKIL: Diversity based oversampling approach to alleviate the class imbalance issue in software defect prediction. In: 2018 IEEE/ACM 40th international conference on software engineering (ICSE), p. 699. https://doi.org/10.1145/3180155.3182520. Presented at the 2018 IEEE/ACM 40th International Conference on Software Engineering (ICSE).

Buda, M., Maki, A., Mazurowski, M.A., 2018. A systematic study of the class imbalance problem in convolutional neural networks. Neural Netw. 106, 249–259. https://doi.org/10.1016/j.neunet.2018.07.011.

Chang, J., Chen, Y., 2015. Batch-normalized Maxout Network in Network. arXiv. arXiv:1511.02583.

Chawla, N.V., 2010. Data mining for imbalanced datasets: an overview. In: Maimon, O., Rokach, L. (Eds.), Data Mining and Knowledge Discovery Handbook. Springer US, Boston, MA, pp. 875–886. https://doi.org/10.1007/978-0-387-09823-4_45.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357.

Deng, J., Dong, W., Socher, R., Li, L., Li, K., Fei-Fei, Li, 2009. ImageNet: A large-scale hierarchical image database. In: Presented at the 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. https://doi.org/10.1109/CVPR.2009.5206848.

Dertien, J.S., Doherty Jr., P.F., Bagley, C.F., Haddix, J.A., Brinkman, A.R., Neipert, E.S., 2017. Evaluating dall's sheep habitat use via camera traps. J. Wildl. Manag. 81, 1457–1467. https://doi.org/10.1002/jwmg.21308.

Diaz-Pulido, A., Payan, E., 2011. Densidad de ocelotes (*Leopardus pardalis*) en los llanos colombianos. Mastozoología Neotropical 18, 63–71.

Drummond, C., Holte, R., 2003. C4.5, class imbalance, and cost sensitivity: why under-sampling beats OverSampling. In: Proceedings of the ICML'03 Workshop on Learning from Imbalanced Datasets, vol. 11, pp. 1–8.

Duarte, A., 2017. Candid creatures: how camera traps reveal the mysteries of nature. Roland Kays. 2016. The Johns Hopkins University press, Baltimore, USA. 280 pp. $39.95 hardcover. ISBN: 978-1-421-41888-9. J. Wildl. Manag. 81, 182. https://doi.org/10.1002/jwmg.21146.

Frey, S., Fisher, J.T., Burton, A.C., Volpe, J.P., 2017. Investigating animal activity patterns and temporal niche partitioning using camera-trap data: challenges and opportunities. Remote Sens Ecol Conserv. 3, 123–132. https://doi.org/10.1002/rse2.60.

Guo, H., Viktor, H.L., 2004. Learning from imbalanced data sets with boosting and data generation: the DataBoost-IM approach. SIGKDD Explor. Newsl. 6, 30–39. https://doi.org/10.1145/1007730.1007736.

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M., Larochelle, H., 2017. Brain tumor segmentation with deep neural networks. Med. Image Anal. 35, 18–31. https://doi.org/10.1016/j.media.2016.05.004.

He, H., Garcia, E.A., 2009. Learning from imbalanced data. IEEE Trans. Knowl. Data Eng. 21, 1263–1284. https://doi.org/10.1109/TKDE.2008.239.

He, Z., Kays, R., Zhang, Z., Ning, G., Huang, C., Han, T.X., Millspaugh, J., Forrester, T., McShea, W., 2016. Visual informatics tools for supporting large-scale collaborative wildlife monitoring with citizen scientists. IEEE Circ. Syst. Mag. 16, 73–86. https://doi.org/10.1109/MCAS.2015.2510200.

Jaccard, N., Rogers, T.W., Morton, E.J., Griffin, L.D., 2017. Detection of concealed cars in complex cargo X-ray imagery using deep learning. J. X-ray Sci. Technol. 25, 323–339. https://doi.org/10.3233/XST-16199.

Japkowicz, N., Stephen, S., 2002. The class imbalance problem: a systematic study. Intell. Data Anal. 6, 429–449. https://doi.org/10.3233/IDA-2002-6504.

Japkowicz, N., Jose Hanson, S., Gluck, M.A., 2000. Nonlinear autoassociation is not equivalent to PCA. Neural Comput. 12, 531–545. https://doi.org/10.1162/089976600300015691.

Johnson, J.M., Khoshgoftaar, T.M., 2019. Survey on deep learning with class imbalance. J. Big Data 6, 27.

Kays, R., Parsons, A.W., Baker, M.C., Kalies, E.L., Forrester, T., Costello, R., Rota, C.T., Millspaugh, J.J., McShea, W.J., 2017. Does hunting or hiking affect wildlife communities in protected areas? J. Appl. Ecol. 54, 242–252. https://doi.org/10.1111/1365-2664.12700.

Krizhevsky, A., Hinton, G., 2009. Learning multiple layers of features from tiny images. Computer Science Department, University of Toronto, Tech. Rep 1.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc., Lake Tahoe, Nevada, USA, pp. 1097–1105.

Kubat, M., 1997. Addressing the curse of imbalanced training sets: One-sided sampling. In: Proc of the 14th International Conference on Machine Learning 97, pp. 179–186.

Lecun, Y., Bengio, Y., Haffner, P., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324. https://doi.org/10.1109/5.726791.

Lee, H., Cho, S., 2006. The novelty detection approach for different degrees of class imbalance. In: King, I., Wang, J., Chan, L.-W., Wang, D. (Eds.), Neural Information Processing. Springer, Berlin, Heidelberg, pp. 21–30.

Li, Feimo, Li, Shuxiao, Zhu, Chengfei, Lan, Xiaosong, Chang, Hongxing, 2017. Class-imbalance aware CNN extension for high resolution aerial image based vehicle localization and categorization. In: 2017 2nd international conference on image, vision and computing (ICIVC), pp. 761–765. https://doi.org/10.1109/ICIVC.2017.7984657. Presented at the 2017 2nd International Conference on Image, Vision and Computing (ICIVC).

Ling, C., Li, C., 1998. Data Mining for Direct Marketing: Specific Problems and Solutions. KDD.

Liu, X., Wu, J., Zhou, Z., 2009. Exploratory undersampling for class-imbalance learning. IEEE Trans. Syst. Man Cybernetics,B Cybernetics 39, 539–550. https://doi.org/10.1109/TSMCB.2008.2007853.

Marouf, M., Siddiqi, R., Bashir, F., Vohra, B., 2020. Automated hand X-ray based gender classification and bone age assessment using convolutional neural network. In: 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (ICoMET), pp. 1–5. https://doi.org/10.1109/iCoMET48670.2020.9073878. Presented at the 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET).

Mazurowski, M.A., Habas, P.A., Zurada, J.M., Lo, J.Y., Baker, J.A., Tourassi, G.D., 2008. Training neural network classifiers for medical decision making: the effects of imbalanced datasets on classification performance. Neural Netw. 21, 427–436. https://doi.org/10.1016/j.neunet.2007.12.031.

Norouzzadeh, M.S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M.S., Packer, C., Clune, Jeff, 2018. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proc. Natl. Acad. Sci. U. S. A. 115, E5716–E5725. https://doi.org/10.1073/pnas.1719367115.

Rich, L.N., Beissinger, S.R., Brashares, J.S., Furnas, B.J., 2019. Artificial water catchments influence wildlife distribution in the Mojave Desert. J. Wildl. Manag. 83, 855–865. https://doi.org/10.1002/jwmg.21654.

Shen, L., Lin, Z., Huang, Q., 2016. Relay Backpropagation for effective learning of deep convolutional neural networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision – ECCV 2016. Springer International Publishing, Cham, pp. 467–482.

Steenweg, R., Whittington, J., Hebblewhite, M., Forshner, A., Johnston, B., Petersen, D., Shepherd, B., Lukacs, P.M., 2016. Camera-based occupancy monitoring at large scales: power to detect trends in grizzly bears across the Canadian Rockies. Biol. Conserv. 201, 192–200.

Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A., Packer, C., 2015. Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. Sci. Data 2, 150026. https://doi.org/10.1038/sdata.2015.26.

Szegedy, C., Liu, Wei, Yangqing, Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Boston, MA, USA, pp. 1–9. https://doi.org/10.1109/CVPR.2015.7298594. Presented at the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Tabak, M., Norouzzadeh, M.S., Wolfson, D., Sweeney, S., Vercauteren, K., Snow, N., Halseth, J., Salvo, P., Lewis, J., White, M., Teton, B., Beasley, J., Schlichting, P., Boughton, R., Wight, B., Newkirk, E., Ivan, J., Odell, E., Brook, R., Miller, R., 2019. Machine Learning to Classify Animal Species In Camera Trap Images: Applications in Ecology. https://doi.org/10.1101/346809.

Wei, W., Luo, G., Ran, J., Li, J., 2020. Zilong: a tool to identify empty images in camera-trap data. Ecol. Inform. 55, 101021. https://doi.org/10.1016/j.ecoinf.2019.101021.

Willi, M., Pitman, R.T., Cardoso, A.W., Locke, C., Swanson, A., Boyer, A., Veldthuis, M., Fortson, L., 2019. Identifying animal species in camera trap images using deep learning and citizen science. Methods Ecol. Evol. 10, 80–91. https://doi.org/10.1111/2041-210X.13099.

Yousif, H., Yuan, J., Kays, R., He, Z., 2017. Fast human-animal detection from highly cluttered camera-trap images using joint background modeling and deep learning classification. In: 2017 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1–4. https://doi.org/10.1109/ISCAS.2017.8050762. Presented at the 2017 IEEE International Symposium on Circuits and Systems (ISCAS).

Yousif, H., Yuan, J., Kays, R., He, Z., 2019. Animal scanner: software for classifying humans, animals, and empty frames in camera trap images. Ecol. Evol. 9, 1578–1589. https://doi.org/10.1002/ece3.4747.

Zeiler, M.D., Fergus, R., 2014. Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014. Springer International Publishing, Cham, pp. 818–833.

Zhou, Zhi-Hua, Liu, Xu-Ying, 2006. Training cost-sensitive neural networks with methods addressing the class imbalance problem. IEEE Trans. Knowl. Data Eng. 18, 63–77. https://doi.org/10.1109/TKDE.2006.17.