Joint Distribution and Class-based Data Augmentation for Wildlife Detection

Yunhao Pan
Physics and Electronic Information College
Yantai University
Yantai, China
panyunhao@s.ytu.edu.cn

Chenhong Sui*
Physics and Electronic Information College
Yantai University
Yantai, China
sui6662015@ytu.edu.cn

Fuhao Jiang
Physics and Electronic Information College
Yantai University
Yantai, China
jiangfuhao@s.ytu.edu.cn

Guobin Yang
Physics and Electronic Information College
Yantai University
Yantai, China
yangguobin@s.ytu.edu.cn

Ankang Zang
Physics and Electronic Information College
Yantai University
Yantai, China
zangankang@s.ytu.edu.cn

Shengwen Zhou
Physics and Electronic Information College
Yantai University
Yantai, China
zhoushengwen@s.ytu.edu.cn

2022 International Conference on Automation, Robotics and Computer Engineering (ICARCE) | 978-1-6654-7548-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICARCE55724.2022.10046567

***Abstract*—Data augmentation is of great importance to alleviate the insufficiency of training samples, and further improve wildlife detection accuracy. However, current data augmentation methods tend to augment all kinds of samples equally, ignoring the problem of uneven distribution of the number and size of all kinds of samples in wildlife detection datasets, resulting in poor generalization of the model. To address this problem, this paper proposes a joint distribution and class-based data augmentation method for wildlife detection. In this method, diverse rather than universal data augmentation methods are introduced for different classes with a small proportion. This makes the distributions of different classes more balanced. Therefore, each class even with a small number of samples gets good training as well. To evaluate the effectiveness of the proposed method, extensive comparative experiments are conducted. Experimental results show the superiority of our proposed method. Specifically, the detection accuracy of Faster RCNN with Swin Transformer as the backbone network is improved by 0.8% to 96.2% after data augmentation with our method.**

Keywords—computer vision and image processing, data augmentation, wildlife detection, neural networks and deep learning

I. INTRODUCTION

Many object detection models are data-driven, and their detection performance is closely related to the number and quality of training samples. Since the acquisition and labeling of datasets are time-consuming and labor-intensive, excellent data augmentation methods are of great value to alleviate the lack of samples and improve the detection accuracy of models.

Based on the way of sample generation, current data augmentation methods can be classified as single-sample transformation-based augmentation, multiple-sample synthesis-based augmentation, and neural network-based augmentation. Single-sample transformation-based data augmentation methods are those that generate multiple augmentation samples by performing different transformations on the original samples, such as random inversion, color channel change [1], rotation [2], cropping [3], noise [4], RandomErasing [5], Cutout [6], GridMask [7], and a combination of methods such as RandAugmentation [8]. Multi-sample synthesis-based data augmentation, on the other hand, obtain new samples by fusing multiple samples, such as Mixup [9], CutMix [10], augmentation for small objects [11], Mosaic, and Copy-Paste [12]. However, in wildlife datasets with unbalanced object classes, the single use of data augmentation ignores the differences in the size and number distribution of the various sample classes, resulting in insufficient generalization of the model.

To this end, this paper proposes a joint distribution and class-based data augmentation method for the differences in the distribution of various types of samples in the wildlife dataset.

II. OUR METHON

A. Dataset Distribution

A total of 6041 images of six types of wildlife in the dataset were divided 1:1 into train and validation sets distributed as shown in TABLE I.

As can be seen from TABLE I, the monitored animal images show the following characteristics: (1)influenced by the regional ecological environment, the number of various types of animals accounted for a seriously uneven proportion, lying between 2.23% and 46.7%. (2) Influenced by the living habits of pig badgers, grass rabbits, and squirrels, the training samples containing these three types of animals are not only small in number but also have only one object per image. (3) The small size of objects such as birds, grass rabbits, and squirrels, which account for a very low percentage of the images, has become a difficult point for recognition.

Therefore, wildlife detection faces three main challenges: the uneven distribution of samples in the training set leads to a more difficult detector training and a smaller number of class objects and contains which indicates that rare class pictures have less useful information and more background information, increasing the difficulty of object detection.

B. Ideas for Joint Distribution and Class-Based Data Augmentation

In this paper, we will conduct a comprehensive analysis of different classes of animals in terms of the characteristics of different classes of data, the detection results of different detection models, and the effects of different data augmentation methods, and the influencing factors are shown in TABLE II as well as TABLE III. The results are as follows:

TABLE I. STATISTICS OF DIFFERENT CLASSES OF ANIMALS IN THE DATASET.

Class	PigBadger	Capreolus	Birds	Antelope	GrassRabbit	Squirrel	Total
Images	135	2842	2121	595	107	241	6041
Object	135	3544	2664	620	107	241	7311
Object share	4.6%	10.0%	1.2%	15.6%	1.9%	1.1%	-

TABLE II. EXPERIMENTAL RESULTS OF THREE DATA AUGMENTATION METHODS IN IMPROVING THE FASTER RCNN DETECTION MODEL.

Augmentation methods	$mAP_{.5}$	$mAP_{.5:.95}$	PigBadger	Capreolus	Birds	Antelope	Grass Rabbit	Squirrel
-	**0.954**	0.685	0.730	**0.817**	**0.558**	**0.826**	0.580	**0.602**
Random Erasing	0.948	**0.686**	**0.734**	0.806	0.553	0.808	**0.618**	0.596
GridMask	0.947	0.641	0.686	0.770	0.506	0.777	0.593	0.513
Mixup	0.952	0.684	0.727	0.813	0.557	0.824	0.613	0.568
Average	0.949	0.670	0.715	0.796	0.538	0.803	0.608	0.558

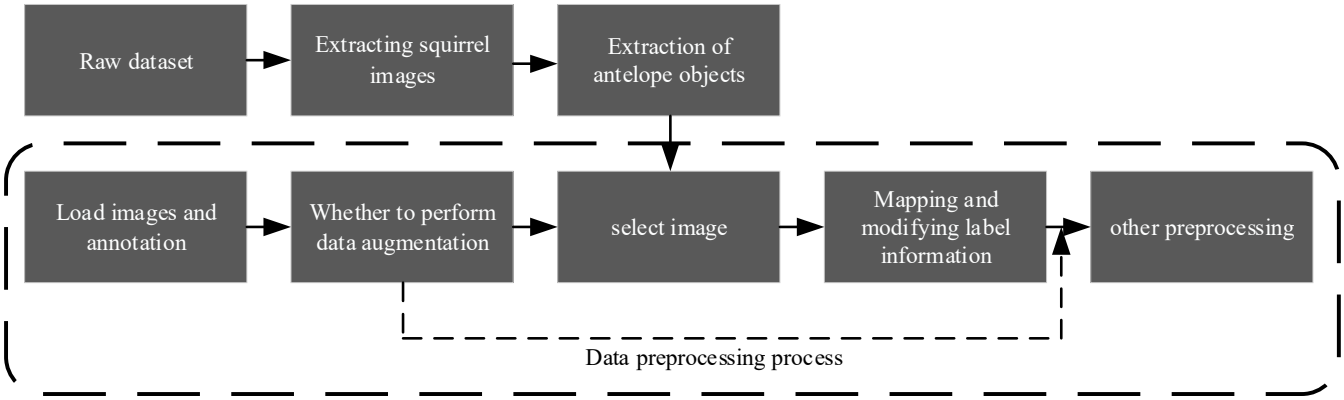

Fig. 1. Data augmentation process for Antelope.

(1) The four types of animals, namely, pig badger, antelope, grass hare, and squirrel, all accounted for less than 10% of the images. Therefore, data augmentation is of great importance to alleviate the imbalance of these classes of training samples. (2) The average accuracy of the model is the lowest for squirrels and the highest for antelopes, and the average accuracy of the model has a high correlation with the average object percentage. Usually the smaller the object, the easier the detection model is to ignore, the lower the detection accuracy and the more obvious the gain effect is if a suitable data Augmentation method is adopted. (3) The three data augmentation methods do not apply to all classes of this paper's dataset, but they can guide the selection of data augmentation classes. The highest decrease in accuracy was observed for antelope and squirrel in the mean results of the classes with improved Faster RCNN and the addition of the three data augmentation methods.

TABLE III. SUMMARY OF THE COMPREHENSIVE ANALYSIS OF DIFFERENT CLASSES.

Class	Pig Badger	Capreolus	Birds	Antelope	Grass Rabbit	Squirrel
Number	2.2%	47.0 %	35.1 %	9.8%	1.8%	4.0%
Object	4.6%	10.0 %	1.2%	15.6%	1.9%	1.1%
Accuracy	0.544	0.737	0.400	0.741	0.389	**0.293**

With the above analysis, this paper uses squirrels and antelopes as examples of classes of data augmentation.

C. Joint Distribution and Class-Based Augmentation Method

The main idea of joint distribution and class-based data augmentation is to augment the classes with a small proportion while remaining the rest unchanged. The following section describes the data augmentation methods and processes based on antelope and squirrel, respectively.

1) Data augmentation for antelope

In this paper, there are 595 images containing antelope in the dataset, accounting for 9.8% of the total 620 objects, of which 570 images have only one object and 25 images include two objects. The accuracy of the images including two objects is shown in TABLE IV. The reason for the low accuracy of the class antelope may be the uneven data distribution between the training set and the validation set. Therefore, this paper changes the distribution of the antelope data in the training set by replicating the online augmentation method to make the data distribution of the antelope training set more balanced.

TABLE IV. DISTRIBUTION AND ACCURACY OF IMAGES CONTAINING TWO ANTELOPES.

	Train	Val
Number	7	18
Accuracy	-	50%

This method is based on Copy-Pasting of all small objects, but the difference is that only the object antelope is extracted instead of the small object. The process of data augmentation is as follows: firstly, all the objects of antelope

are extracted according to the object segmentation and object detection annotation information, and saved on disk for selection; during the training, a judgment is made in the image pre-processing process first, and the image with only one antelope is selected with a certain probability; afterward, the arbitrarily selected object is pasted onto the selected training image, and the annotation frame and class are updated. The process is shown in Fig. 1.

When pasting the object onto the image, three guidelines need to be followed: extract the image with only the object information, excluding the background information, and the result of object extraction is shown in Fig. 2; ensure that the pasted area does not intersect with the original object, find the pasted area randomly, and give up the data augmentation of this image if the suitable area is not found even after 100 repetitions; update the annotation information after data augmentation. According to the above process and guidelines for data augmentation, the augmentation results are shown in Fig. 3.

Fig. 2. Object extraction results.

Fig. 3. Data augmentation results for Antelope.

2) Data augmentation for squirrel

Through the analysis of the squirrel dataset, we found that: there are 241 images of squirrels in the dataset of this paper, the number of images accounted for 4%, accounting and each image has only one object; there is too much background information in the images, the object size is small, the width of the object is between 256 and 559 pixels, the height is between 121 and 259, the average area of the object in the image accounted for the highest 1.2%; the color, texture and background information is similar, and the squirrel can only be photographed during daytime.

For the above characteristics of squirrels in the dataset, this paper proposes a new data augmentation method, which uses both offline and online augmentation, the GridMask method for online augmentation, and horizontal inversion for offline augmentation to solve the problems of the small number of images and too much invalid information respectively. The main purpose of offline data augmentation is to increase the number of images by horizontal inversion, and the method requires corresponding modifications to the annotation information; the online data augmentation is based on GridMask and adjusted according to the object size, by adjusting the parameters in TABLE V, the reference basis is the minimum value of the squirrel object length and width, and the parameters are restricted so that the GridMask The maximum side length of the square black block in the generated mask is 120 to avoid blocking the object and to increase the proportion of valid information in the image.

TABLE V. PARAMETER SETTINGS OF GRIDMASK METHOD.

ratio	P	d
0.5	0.5	[2, 120]

In the process of data augmentation for squirrels, the squirrel images are firstly reversed horizontally and the obtained images are added to the original dataset with the annotation information for expansion, at which the number of squirrel images is increased from 241 to 482. The square black blocks in the Mask generated by GridMask are restricted to have side lengths between 2 and 120 pixels. The overall data augmentation process is shown in Fig. 4, and the augmentation results are shown in Fig. 5.

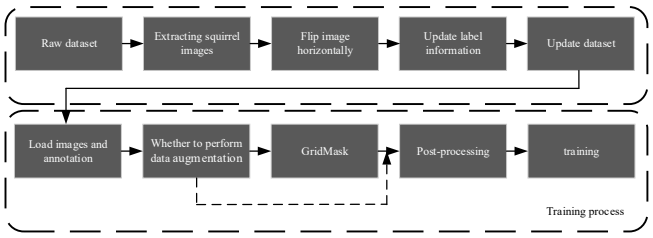

Fig. 4. Data augmentation process for squirrel.

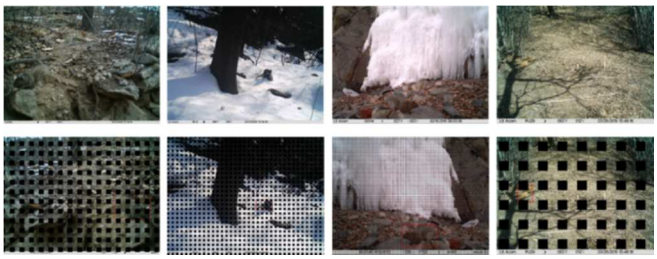

Fig. 5. Data augmentation results for squirrel.

III. EXPERIMENT

A. Experimental Environment and Parameter Settings

The Faster RCNN object detection model used in this study, the training of the model is all based on the MMDetection [13] platform, and the hardware and software configuration of the experimental environment used in this study is shown in TABLE VI.

TABLE VI. SOFTWARE AND HARDWARE ENVIRONMENT CONFIGURATION.

Equipment	Type
Operating System	Ubuntu18.04
cuDNN	7.6.5
Python	3.7.11
Pytorch	1.8.1
GPU	Tesla V100
VideoMemory	32GB
MMDetection	2.22.0

When the object recognition model is trained on the wildlife dataset constructed in this paper, it is initialized using pre-training weights to save time. The pre-training weights are all officially provided, and the image scaling size (Resize), the number of cycles of the training set (epoch), the learning rate (Lr) decay coefficient (lr-gamma), and the trained image batch size (Batch-Size, BS) are set as shown in TABLE VII.

TABLE VII. PARAMETER SETTING TABLE.

Methods	Backbone	Resize	epoch	Lr	Lr-steps	Lr-ga	BS
SSD300	Resnet50	(300,300,3)	15	0.0005	3	0.33	32
Mask RCNN	Resnet50	(224,224,3)	12	0.02	[8, 11]	0.1	16
Faster RCNN	Resnet50	(224,224,3)	12	0.02	[8, 11]	0.1	16
Faster RCNN	Swin-Transformer	(224,224,3)	12	0.0001	[27, 33]	0.1	8

B. *Experimental Environment and Parameter Settings*

1) Comparative experiments of distribution data augmentation in different detection models

Data augmentation based on the distribution of antelope and squirrels was added to the selected model, with the parameters of the augmentation method set as before and the remaining four wildlife classes kept unchanged. The models selected for the experiments are SSD [14], Mask RCNN [15], and Faster RCNN [16] three object detection models, the backbone network is used Resnet50 [17], the experimental results are shown in TABLE VIII and Fig. 6.

TABLE VIII. OVERALL ACCURACY RESULTS FOR THREE MODELS USING JOINT DISTRIBUTION AND CLASS-BASED DATA AUGMENTATION.

Methods	$mAP_{.5}$	$mAP_{.75}$	$mAP_{.5\text{-}.95}$
SSD	0.706	0.499	0.458
SSD^{+}	**0.714**	**0.511**	**0.480**
Mask RCNN	0.880	0.656	0.583
Mask RCNN^{+}	**0.887**	**0.675**	**0.597**
Faster RCNN	0.775	0.541	0.504
Faster RCNN^{+}	**0.779**	**0.565**	**0.517**

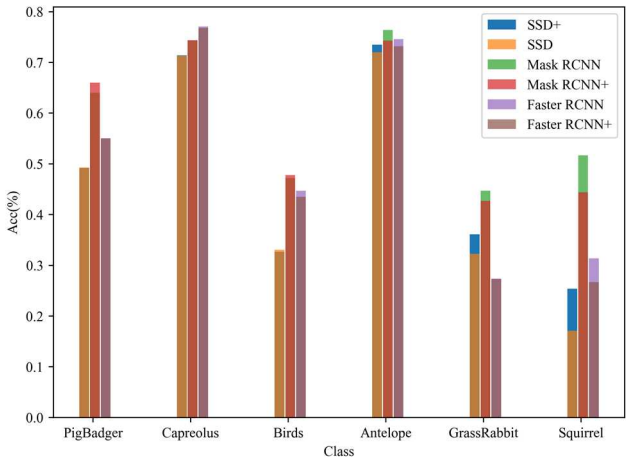

Fig. 6. The overall accuracy results for the three models were augmented using joint distribution and class-based data.

TABLE VIII indicates the overall accuracy of the three models before and after improvement using joint distribution and class-based data augmentation, with + indicating the use of the joint distribution and class-based data augmentation proposed in this paper, the improved object detection model improves the overall accuracy in all three different scales. mAP.5, mAP.75, and mAP.5-.95 improve by 2.83%, 1.90%, and 1.37% on average, which experimentally proves that the joint distribution and class-based data augmentation method proposed in this paper is effective in different models.

Fig. 6 shows the comparison of the accuracy of the three models for six wildlife classes before and after using joint distribution and class-based data augmentation. In the single-class detection, the class accuracy of hog badger, bird, and roe deer is basically the same as that without the augmentation method, and the class accuracy of grass hare is improved in all three improved models, and the AP.5-.95 of antelope and squirrel are also improved, and the class detection accuracy of antelope is improved by 1.63% on average, and that of the squirrel is improved by 6.7% on average, which is very obvious, which is closely related to the augmentation method proposed in this paper for squirrel. This is closely related to the proposed augmentation method for squirrels, which effectively solves the problem of less number of class pictures and a smaller object size for squirrels.

2) Comparative experiments of distribution data augmentation in different detection models

The above experimental results show that the joint distribution and class-based data augmentation methods are effective for different object detection models as well as different backbone networks. To further demonstrate the effectiveness of the results, the object detection model selected for the experiments is Faster RCNN and the backbone network is Swin Transformer [18], by using the remaining three online data augmentation methods as control experiments, which are GridMask, Random Erasing, and Mixup, respectively, and the code implementation is based on MMdetection platform.

TABLE IX represents the overall results of the Faster RCNN object detection model with Swin Transformer as the backbone network and after data augmentation by fusing different methods. The evaluation metrics are mAP.5, mAP.75 and mAP.5-.95, where the proposed joint distribution and class-based data augmentation method is the best in all three metrics, improving 0.8, 1.2%, and 0.4% compared to the original Swin Transformer model. In contrast, the Swin Transformer incorporating Random Erasing, GridMask, and Mixup data augmentation methods all performed poorly, with a decrease in overall accuracy, which is related to the variability of different classes of data in the wildlife dataset in this paper, while the distribution-based data augmentation methods can be augmentation according to the characteristics of different datasets. Where mAP is the value that calculates the AP of all images in each class and then averages over all classes, mAP.5 is the mAP when the IoU is set to 0.5, mAP.75 is the mAP when IoU is set to 0.75, mAP.5-.95 is the average over different IoU thresholds (from 0.5 to 0.95, in steps of 0.05).

TABLE IX. OVERALL ACCURACY USING DIFFERENT DATA AUGMENTATION METHODS.

Augmentation methods	$mAP_{.5}$	$mAP_{.75}$	$mAP_{.5\text{-}.95}$
-	0.954	0.799	0.685
Random Erasing	0.948	0.793	0.686
GridMask	0.947	0.727	0.641
Mixup	0.952	0.788	0.684
Ours	**0.962**	**0.804**	**0.697**

IV. Conclusions

In this paper, we first analyze the characteristics of wildlife datasets and the results of commonly used online data augmentation methods, propose a joint distribution and class-based data augmentation method, and select classes for improvement using data augmentation; then the data augmentation process and methods for different classes are introduced separately; finally, different object detection models and different data augmentation methods are experimented with to demonstrate the effectiveness of the proposed data augmentation methods.

References

[1] J. Aranzazu, P. Miguel, G. Mikel , L. M. Carlos and P. Daniel, "A comparison study of different color spaces in clustering based image segmentation," in International conference on information processing and management of uncertainty in knowledge-based systems, 2010, pp. 532-541.

[2] E. Logan, T. Brandon, and T. Dimitris, "A rotation and a translation suffice: Fooling cnns with simple transformations," 2018, https://arxiv.org/pdf/1712.02779v2.pdf.

[3] T. Ryo, M. Takashi, and U. Kuniaki, "Data augmentation using random image cropping and patching for deep CNNs," IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 9, pp. 2917-2931, 2019.

[4] M. Barea, and J. Francisco, "Forward noise adjustment scheme for data augmentation." 2018 IEEE symposium series on computational intelligence, IEEE, 2018, pp. 728-734

[5] Z. Zhun, Z. Liang, K. Guoliang, L. Shaozi, and Y. Yi, "Random erasing data augmentation," in Proceedings of the AAAI conference on artificial intelligence, 2020, pp. 13001-13008.

[6] T. DeVries, and G. W. Taylor, "Improved regularization of convolutional neural networks with cutout," arXiv preprint arXiv:1708.04552, 2017.

[7] C. Pengguang, L. Shu , Z. Hengshuang and J. Jiaya, "Gridmask data augmentation," arXiv preprint arXiv:2001.04086, 2020.

[8] C. D. Ekin, Z. Barret, S. Jonathon, and L. V. Quoc, "Randaugment: Practical automated data augmentation with a reduced search space," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, 2020, pp. 702-703.

[9] Z. Hongyi, C. Moustapha, D. N. Yann and L. David, "mixup: Beyond empirical risk minimization," arXiv preprint arXiv:1710.09412, 2017.

[10] Y. Sangdoo, H. Dongyoon, O. S. Joon, C. Sanghyuk, C. Junsuk, and Y. Youngjoon, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 6023-6032.

[11] K. Mate, W. Zbigniew, M. Jakub, N. Jacek, and C. Kyunghyun, "Augmentation for small object detection," arXiv preprint arXiv:1902.07296, 2019.

[12] G. Golnaz, C. Yin and S. Aravind, Q. Rui, L. Y. Tsung, C. D. Ekin, L. V. Quoc, and Z. Barret, "Simple copy-paste is a strong data augmentation method for instance segmentation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 2918-2928.

[13] C. Kai, W. Jiaqi, P. Jiangmiao, C. Yuhang, X. Yu, L. Xiaoxiao, S. Shuyang, F. Wansen, L. Ziwei, X. Jiarui, et al., "MMDetection: Open mmlab detection toolbox and benchmark," arXiv preprint arXiv:1906.07155, 2019.

[14] L. Wei, A. Dragomir, E. Dumitru, S. Christian, R. Scott, F. C.Yang, and B. C. Alexander, "Ssd: Single shot multibox detector," in European conference on computer vision, 2016, pp. 21-37.

[15] H. Kaiming, G. Georgia, D. Piotr, and G. Ross, "Mask r-cnn," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961-2969.

[16] R. Shaoqing, H. Kaiming, G. Ross, and S. Jian, "Faster r-cnn: Towards real-time object detection with region proposal networks," Advances in neural information processing systems, vol. 28, pp. 2961-2969, 2015.

[17] H. Kaiming, Z. Xiangyu, R. Shaoqing, and S. Jian, "Deep residual learning for image recognition,",in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.

[18] L. Ze, L. Yutong, C. Yue, H. Han, W. Yixuan, Z. Zheng, L. Stephen, and G. Baining, "Swin transformer: Hierarchical vision transformer using shifted windows," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10012-10022.