Data Augmentation for Wildlife Animal Recognition Using Style Transfer

*Seungjin Jung[1], Yuyeon Lee[2], Inbok Lee[2], Joohyun Kang[3], Seungwook Lim[4], Jongwon Choi[1]**
[1]Department of Advanced Imaging, Chung-Ang University
[2]Republic of Korea Army
[3]Republic of Korea Air Force Air Defense Control Command
[4]Republic of Korea Navy Data System Management Group
**Correspondence to: Jongwon Choi (jwchoi@cau.ac.kr)*

Abstract

Many military facilities are located in the mountain and the forests, so the wildlife animals easily pass through the facilities. To prevent their invasion, military facilities developed an advanced surveillance system to detect wildlife animals, but the insufficient data for nighttime animals has been a challenging problem. To solve the issue, we design two methods to augment and utilize the training data for nighttime wildlife animals by using a style transfer. The first method is designed to transfer the daytime data to be a style of nighttime data, and the second method exchanges the style of nighttime data with that of daytime data. Through the experiments, we show the effectiveness of the two methods, analyzing the augmented training data.

Keywords: *Data Augmentation, Wildlife Classification, Style Transfer*

I. Introduction

Recently, army units have built an automatic boundary surveillance system to identify people, animals, and diverse objects through an AI surveillance system. While the intelligent surveillance system distinguishes objects well in days, we suffer high error rates when the capturing environment changes from day to night. This problem happens due to the insufficient variation of training data in the nighttime.

In this paper, we propose a method using data augmentation through style transfer [1] networks to remedy this issue. With the style transfer, we can supplement the insufficient data by transferring the style of sufficient data to that of insufficient ones. We design two methods to verify the effectiveness of the style-transferred training data. The first method is to transfer the style of the daytime dataset to be like the nighttime dataset, and the second method utilizes the predictions of the daytime-like datasets transferred from the nighttime datasets. From the experimental results, we discover that the usage of augmented datasets is more effective than the predictions of the transferred data.

II. Proposed Method

The goal of this paper is that our model classifies the animals well, whether daytime or nighttime. For this, first, we define the problem to be solved. Then, we introduce the data augmentation method and learning method as our solutions.

A. Problem Setting

Typically, daytime datasets are more abundant and are easier to label than other datasets, such as nighttime, snow, and rain. We define the source and target domains by the daytime and nighttime domains, respectively. The source domain defines as image and label pairs $\mathcal{D}_S = \left\{\left(\mathrm{x}_{S_i}, y_i\right)\right\}_i^n$*, while the target domain contains only images* $\mathcal{D}_T = \left\{\left(\mathrm{x}_{T_i}\right)\right\}_i^n$*. Then, our goal is to classify both the source and the target dataset using labeled and unlabeled datasets.*

B. Data Augmentation

To augment data, we use Cycle-GANs [1]. Cycle-GANs ignore the necessity of either the label information or the paired images containing the same contents. Thus, we train Cycle-GANs with daytime and nighttime datasets, and we can get two generators: $G_{S\to T}$ *changes the image style from source to target, and* $G_{T\to S}$ *changes the image style from target to source. We follow the hyperparameter and the architecture offered at public Cycle-GANs.*

C. Learning method for Classifiers

Now we train the classifier with datasets. However, since only the daytime datasets have labels, we can make a classifier only for daytime datasets. Then, we can define the training loss for the classifier as:

$$\hat{\theta}_{c_1} = \arg\min_{\theta_{c_1}} \sum_{i=1}^{n} CLE\left(\left(C_1(x_{S_i}; \theta_{c_1}), y_{S_i}\right)\right), \qquad (1)$$

where CLE *is cross-entropy loss,* θ_{c_1} *is network parameters,* $\hat{\theta}_{c_1}$ *is the optimized* θ_{c_1}*, and* i *is the index of samples.*

Unfortunately, the classifier has a limitation that does not work on the nighttime domain. To solve the problem, we use an augmented dataset through two methods. The first method uses the augmented nighttime datasets generated from daytime datasets using $G_{S\to T}$. Thus, the training loss for the first method can be denoted as:

$$\arg\min_{\theta_{c_2}} \sum_{i=1}^{n} \left(CLE\left(\left(C_2\left(\mathrm{x}_{S_i}; \theta_{c_2}\right), y_{S_i} \right) \right) + CLE\left(\left(C_2\left(G_{S\to T}\left(\mathrm{x}_{S_i}\right); \theta_{c_2} \right), y_{S_i} \right) \right) \right), \quad (2)$$

where C_2 is the classifier trained by the daytime dataset and the nighttime-like dataset from the daytime dataset.

The second method uses the pseudo-labels predicted from the classifier C_1 fed by the daytime-like dataset transferred from the nighttime dataset by $G_{T\to S}$. The training loss of the second method is designed as:

$$\arg\min_{\theta_{c_3}} \sum_{i=1}^{n} \left(CLE\left(C_3\left(\mathrm{x}_{S_i}; \theta_{c_3}\right), y_{S_i} \right) + CLE\left(C_3\left(\mathrm{x}_{T_i}; \theta_{c_3}\right), C_1\left(G_{T\to S}\left(\mathrm{x}_{T_i}\right) \right) \right) \right), \quad (3)$$

where the classifier C_3 is trained by the daytime dataset with the label and the nighttime dataset with the pseudo-label.

III. Dataset

Table 1 : Wildlife dataset

Animal	*Train*		*Test*	
	#daytime	*#nighttime*	*#daytime*	*#nighttime*
Elk	*16,761*	*13,573*	*2,101*	*1,666*
Half-moon bear	*12,471*	*12,018*	*1,529*	*1,466*
Raccoon	*13,549*	*17,064*	*1,683*	*2,060*
Roe deer	*7,202*	*6,535*	*903*	*626*
Weasel	*6,843*	*9,911*	*1,934*	*1,302*
Wild boar	*8,421*	*6,556*	*1,060*	*808*
Wild rabbit	*7,989*	*6,865*	*1,029*	*831*

The wildlife dataset [2] consists of 11 animals captured by diverse camera sensors, such as CCTV cameras, DSLRs, and infrared cameras. Since four animals are only active during the daytime, and our setting considers daytime and nighttime, we use seven animals' dataset from wildlife datasets. Then we crop the image patches for the respective wildlife animals using the bounding box annotations. In Table 1, we present the number of class-wise data and observe the data imbalance in terms of categories and domains.

IV. Experiments

We implement it on RTX 3070 laptops. The classifier architecture is ResNet18 [3] and optimizer is SGD [4] with learning rate 0.1, momentum 0.9, and weight decay=0.0002. We set the batch size to 128, the image resolution to 36×36, and the crop size to 32.

Table 2 : Comparison experiments

Experiments	*epochs*	*10*	*20*	*30*	*40*	*50*
#1	*Day*	*87.49*	*86.74*	*89.28*	*89.69*	*88.76*
	Night	*31.30*	*26.71*	*36.70*	*34.04*	*32.56*
#2	*Day*	*88.55*	*93.46*	*88.94*	*90.05*	*89.83*
	Night	*43.04*	*43.80*	*42.00*	*42.65*	*46.86*
#3	*Day*	*89.25*	*89.75*	*89.80*	*89.61*	*89.70*
	Night	*42.08*	*42.61*	*42.45*	*41.17*	*43.14*

In Table 2, we perform the three experiments, #1, #2, and #3, respectively. The table indicates accuracy as the interval of 10 epochs to daytime and nighttime, respectively. We conduct experiments twice to show the accuracies as the average of experiments.

#1 is the result in the classifier C_1 trained only on daytime datasets mentioned in equation (1). As the classifier C_1 is the baseline, our method needs to improve this.

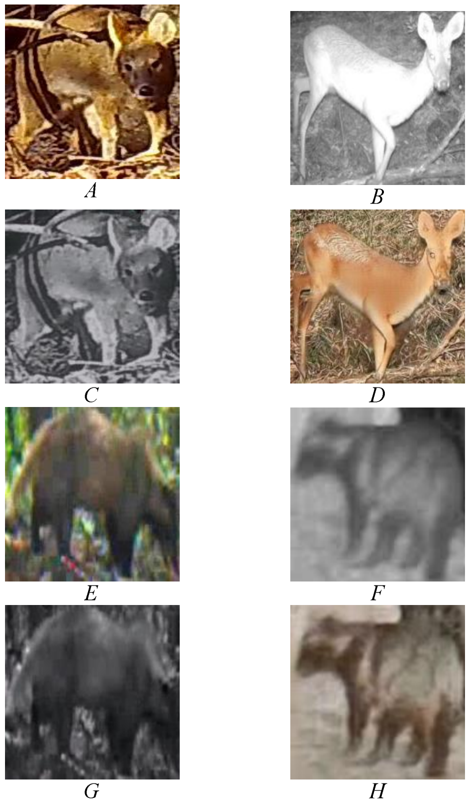

Figure 1 : Augmented Image

#2 is the result of the classifier C_2 mentioned in equation (2). The classifier trained daytime datasets with the augmented dataset transferred style from daytime to nighttime. #2 shows that the classifier mentioned in the equation performs better than the baseline classifier. In both daytime and nighttime datasets, the classifier C_2 performs better than the baseline classifier C_1.

#3 is the result of the classifier C_3 mentioned in equation (3). The classifier trained daytime with the nighttime dataset, which consists of images and pseudo-labels pairs. We can get pseudo-labels using C_3 as a function and augmented datasets transferred style from nighttime to daytime as inputs. In both daytime and nighttime datasets, the classifier C_3 also performs better than the baseline classifier C_1. However, C_3 has lower performance than C_2 because Cycle-GANs decrease performance when translating an image from insufficient to sufficient information.

In Figure 1, we visualize original images and generated images. A, B, E, and F are images from the dataset, and C, D, G, and H are images generated by Cycle-GANs. A and E are daytime images, and B and F are nighttime images.

C and G are images generated nighttime style from A and E. Because the daytime datasets have more information than nighttime datasets, C and G look like nighttime images, such as B and F.

D and H are images generated daytime style from B and F. D is a well-generated image because the image is large. However, H is a badly generated image because the resolution is low and grayscale.

V. Conclusion

Through our experiments, we demonstrated that the data augmentation based on the style transfer does improve the accuracy of wildlife animal recognition. Among the tested methods, we discovered that we achieve improved effectiveness when the style transfer is employed to transfer the domain with rich information into the one with poor information. In the future, we are planning to make our model work well in other environments, such as rain, snow, and foggy. Furthermore, we will also perform experiments with the military surveillance dataset.

Acknowledgements.

It was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (2022-0-00601, military-specific AI training courses (Chung-Ang University), 90%; 2021-0-01341, Artificial Intelligence Graduate School Program (Chung-Ang University), 10%).

References

[1] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros. "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," in 2017 IEEE International Conference on Computer Vision (ICCV)

[2] NIA. "Wildlife activity video data". NOAA SNC. Updated in September 2022. Download at August 2022. &topMenu=100&aihubDataSe=realm&dataSetSn=645

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[4] Bottou, Léon. "Online Algorithms and Stochastic Approximations". In 1998 Online Learning and Neural Networks.