High-Efficiency Progressive Transmission and Automatic Recognition of Wildlife Monitoring Images With WISNs

WENZHAO FENG [1,2], **WENHUA JU**[3], **ANQI LI**[1,2], **WEIDONG BAO**[4], **AND JUNGUO ZHANG** [1,2]
[1]School of Technology, Beijing Forestry University, Beijing 100083, China
[2]Key Laboratory of State Forestry Administration for Forestry Equipment and Automation, Beijing Forestry University, Beijing 100083, China
[3]Bahrain Zuoqi Forestry and Grassland Bureau, Chifeng 025450, China
[4]School of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, China

Corresponding authors: Weidong Bao (wdbao@bjfu.edu.cn) and Junguo Zhang (zhangjunguo@bjfu.edu.cn)

ABSTRACT Wireless image sensor networks (WISNs) are widely applied in wildlife monitoring, as they present a better performance in remote, real-time monitoring. However, traditional WISNs suffer from the limitations of low processing capability, power consumption restrictions and narrow transmission bandwidth. For the contradiction between the above limitations of WISNs and the wildlife monitoring images with high resolution and complex background, we propose a novel wildlife intelligent monitoring system. On the foundation of saliency object detection, the convolutional encoder-decoder network is utilized to realize the progressive compression transmission and restoration for wildlife monitoring images, which guarantees the transmission efficiency and quality of wildlife part. Moreover, to deal with the problems of high labor intensity, low efficiency and low recognition accuracy in classical manual sorting method, an improved Faster RCNN algorithm is proposed on the automatic recognition of wildlife images. The experimental results on our own wildlife dataset, show that the peak signal to noise ratio (PSNR) and structural similarity index (SSIM) are improved by 7.93%, 18.15% and 7.01%, 12.67% respectively on reconstruction image, when compared with the set partitioned in hierarchical tree (SPIHT) and embedded zerotree (EZW) algorithms. Compared with the traditional Faster RCNN algorithm, the recognition accuracy of six species wildlife is respectively improved by 1%, 18%, 5%, 17%, 2% and 19%, and the final mAP value reaches to 92.2% in test set increased by 10.9%, which demonstrates the proposed algorithm can ideally achieve the wildlife intelligent monitoring with WISNs.

INDEX TERMS Wildlife intelligent monitoring, WISNs, progressive compression transmission, image restoration, automatic recognition.

I. INTRODUCTION

Wildlife resource is abundant in China and it significantly contributes to the balance and stability of the whole ecosystem [1]. Scientific and effective monitoring of wildlife can provide necessary information on species quantity and habitat status for wildlife protection [2]. However, the precise acquisition of wildlife monitoring images in real time poses a challenge for associated practitioners. Conventional wildlife monitoring methods include crewed field survey, GPS

(Global positioning system) collar [3], infrared camera [4] and satellite remote sensing monitoring [5] approaches. However, these methods have their own limitations, such as limited monitoring range, data acquisition lag and low resolution, and so on. Recently wireless image sensor networks (WISNs) [6] are applied in wildlife image collection as they present better deployment and remote transmission ability.

For wireless image sensor networks, the collected wildlife monitoring images encompass high resolution, complex background and large information data characteristics, which can hardly be proceeded due to the limitations of WISNs with

The associate editor coordinating the review of this manuscript and approving it for publication was Abbas Jamalipour .

low processing capability, power consumption restrictions and narrow transmission bandwidth. For transmitting task through resource-constrained WISNs, image compression coding is utilized to reduce the transmission workload and improve the transmission efficiency. In this field, image compression algorithms such as static image compression coding standards: JEPG and JEPG2000 [7], [8], compressed sensing [9], discrete cosine transform [10], singular value decomposition [11] and deep convolutional neural network [12], are capable of achieving high-efficiency compression of image samples. However, these algorithms are generally applied to the encoding and decoding processes of whole images, which can not guarantee the transmission quality of regions of interest in sources, and the image data may lose in transmission process due to the environment interference.

To solve the above problem, image restoration [13] is widely utilized after image acquisition to guarantee the reliability and availability of the wildlife monitoring images. However, these algorithms have undesired algorithm complexity issues, such as edge statistics [14], total variation [15], globally and locally consistent [16], adversarial edge learning [17] and contextual attention [18]. Recently, autoencoder [19], an unsupervised learning method is utilized in WISNs thanks to its simple structure. This network architecture is only built upon standard components such as convolutional layers and skip connections [20], which can achieve the state-of-art performance in image compression and restoration tasks. However, it does not consider the difference of texture information between saliency region and background region, so we introduced the saliency detection in the image restoration process to achieve the wildlife progressive transmission and restoration.

According to the collection and analysis of the wildlife monitoring images through control center, we found that the monitoring nodes may have a serious false triggering phenomenon [21] due to the interference of external factors such as direct sunlight and wind disturbance. For the huge wildlife monitoring images with invalid images included, traditional method such as manual sorting is disadvantageous in terms of high labor intensity, low efficiency and unstable the recognition accuracy. Therefore automatic wildlife recognition [22] is demanded to improve the efficiency of wildlife monitoring. In recent years, various novel models such as AlexNet [23], VGG Net [24], Google Net [25], Deep Residual Net [26] and Dense Net [27] were proposed to improve wildlife recognition capability and achieve the automatic recognition of wildlife accurately. Main stream automatic recognition algorithms included visual vocabulary bags [28], image segmentation and deep learning technologies [29], graphical cutting algorithm [30] and YOLO algorithm [31]. However, these algorithms mainly focus on overall recognition of single image, which do not consider the relationship between the foreground and background information.

This paper proposes an intelligent wildlife monitoring system to realize the remote, real-time monitoring and transmission, image restoration and automatic recognition of wildlife.

TABLE 1. Parameters of WISNS node.

Monitoring node	Function parameters
camera	OV7725 QVGA 30fps
pixels	640×480
image format	BMP
support memory card	SD 16G
controller	STM32 contrl ship (72Mhz CPU、512K RAM)
measuring range	120°/radius 10m
transmission distance	1200m
trigger mode	infrared trigger

The key contributions of the present study are: 1) A wildlife monitoring system is constructed based on WISNs to achieve the remote, real-time monitoring of wildlife. 2) A progressive image transmission and restoration strategy is proposed based on convolution Auto-encoder for wildlife monitoring images, which addresses the WISNs bottleneck of transferring the image data. 3) The Faster RCNN algorithm is proposed to realize the automatic recognition of wildlife, which improves the sorting efficiency and overcomes the disadvantages of high labor intensity, low efficiency and low recognition accuracy.

The rest of this paper is structured as follows. Section II builds the wildlife monitoring system with WISNs. Section III introduces the proposed progressive transmission and restoration, and wildlife recognition algorithms. In Section IV, we compare the performance of our proposed algorithm with other algorithm. Finally, we give the conclusions in Section V.

II. WILDLIFE MONITORING SYSTEM WITH WISNs

WISNs is widely utilized in wildlife monitoring to capture and transmit wildlife image materials with industrial grade cameras, which consist of WISNs terminal nodes, coordination nodes, gateway nodes, control center and data processing module. The wildlife monitoring system is acknowledged to present remote, real-time, all-weather and friendly monitoring merits, and the schematic diagram with detailed configurations are shown in Fig.1.

Monitoring node devices are deployed in the active wildlife area according to practical needs, and the wireless sensor network is established in a self-organizing manner by using ZigBee network protocols. The detailed parameters of monitoring node devices are shown in Tab.1. After the terminal node infraredly senses that the wildlife enters the monitoring field, the camera is triggered to capture images and data are saved in the SD card. Then the coordination node transmits the image data to back-end server when the monitoring image data is received from terminal node in a multi-hop manner.

The WISNs monitoring system for wildlife is mainly deployed in northern China, and the wildlife species collected in the experimental site include Red Deer, Chinese goral, Roe Deer, Lynx, Wild boar, and Raccon Dog. Among them, Red Deer and Lynx are secondary national-protected species
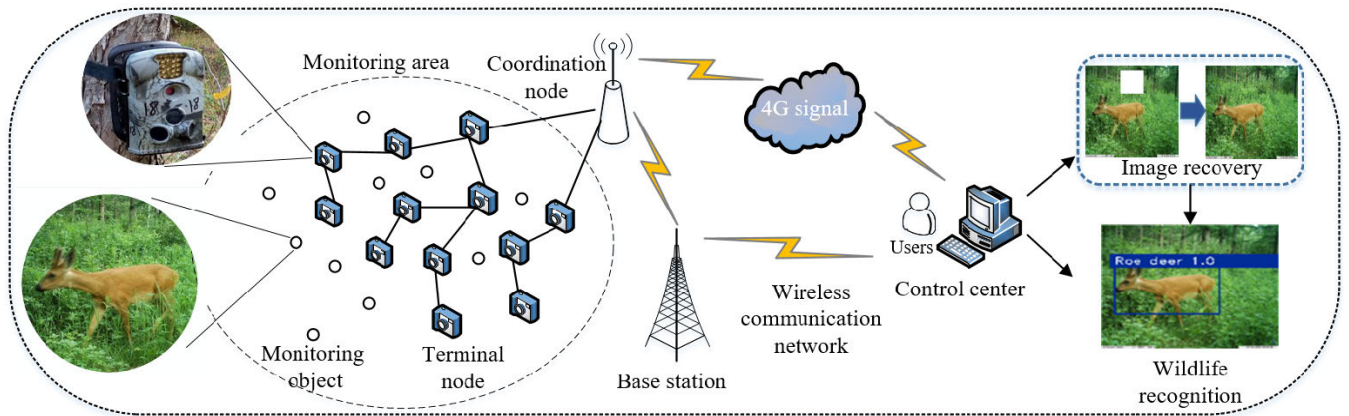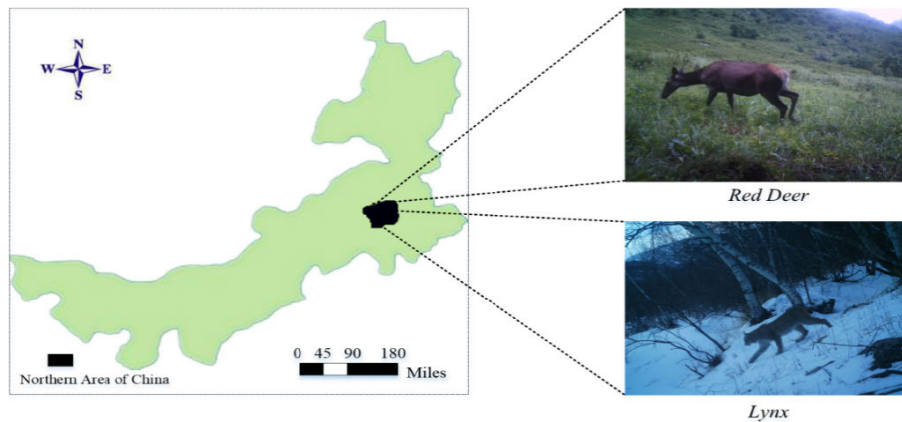
FIGURE 1. Wildlife monitoring intelligent system.

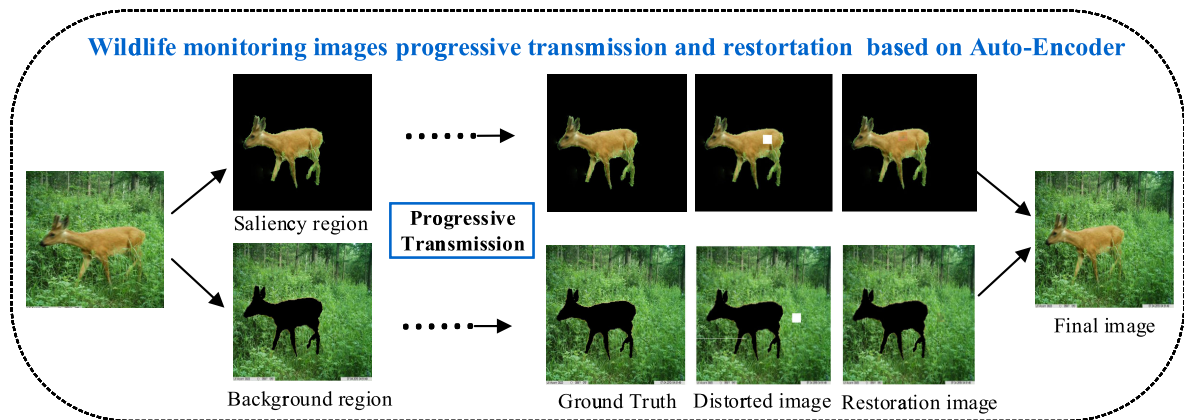**FIGURE 2.** Wildlife monitoring area Northern of China.

FIGURE 3. Image progressive transmission and restoration strategy.

(shown in the Fig.2) and the total quantity of collected images is over 10,000 of the six species.

III. THE PROPOSED ALGORITHM

A. PROGRESSIVE TRANSMISSION AND RESTORATION

For the wildlife monitoring images with high-resolution, complex background and large information data, a novel image progressive transmission and restoration method is proposed in this study as shown in Fig.3. On the foundation of wildlife saliency detection and extraction [32], the saliency region and background region are separately compressed using the Convolution encoder-decoder [33] to guarantee the saliency region. In other words, the wildlife region is placed in the highest priority of compression and transmission. Then the skip connections [20] are utilized to realize the restoration of wildlife images, which ensures the availability of the experimental material for the subsequent study of wildlife automatic recognition.

1) SALIENCY OBJECT DETECTION AND EXTRACTION

In the process of wildlife monitoring, the saliency object region (wildlife region) can provide an intuitive understanding and judgment for wildlife while the background region is only a supplementary information to the saliency region. Therefore, saliency object detection and extraction is utilized in this paper to generate the mask image for the progressive transmission strategy.

In our proposed algorithm, the input images are quantified according to the number of quantify channels CN and the main color is arranged into a color matrix by histogram statistics. After that the image pixels are recorded by color value and the terms with the same color value are grouped together. The saliency value $S(C_i)$ between different colors is calculated and expressed as shown in the (1), besides the saliency value are the same when the color value of the pixels is the same.

$$S(c_i) = \sum_{j=1}^{n} f_j D(c_i, c_j) \tag{1}$$

where c_i denotes the color value of the pixel P_k in input image and the $D(c_i, c_j)$ denotes the color distance metric between the pixel c_i and c_j in Lab space. n denotes the total color numbers of input image and f_j represents the ratio of the pixels whose color value is c_i to the total pixel numbers in the image.

Color quantization greatly simplifies calculation procedure, but similar colors may be quantized to different values during the process. In order to reduce noisy saliency results caused by such randomness, we replace the saliency value of each color by the weighted average of the saliency value of similar colors.

$$S'(c) = \frac{1}{(m-1)T} \sum_{i=1}^{m} (T - D(c, c_i)) S(c_i) \tag{2}$$

where the equation $T = \sum_{i=1}^{m} D(c, c_i)$ denotes the distance between the color c and its nearest colors. Typically m is quarter of color numbers n in the images after quantification.

$$\sum_{i=1}^{m} (T - D(c, c_i)) = (m-1)T \tag{3}$$

Then the saliency area is obtained by comparing appearance frequency [34] of the first n kinds color. If the appearance frequency is greater than CR (Color Retention Rate), the color with low frequency is discarded subsequently and the closest color will replace it. The color appearance frequency of the first n kinds color will be increased in accordance with the number of quantify channels until it is greater than CR. In this experiment, the color value of each channel (R, G, and B) is quantified between 0-12 (CN) so that the number of colors are reduced to $12^3 = 1728$. To obtain high-frequency colors in saliency area, the CR is selected as 0.95.

In terms of saliency object detection and extraction, the corresponding coefficients in the saliency and background regions are set to 1 and 0 respectively to obtain the binary
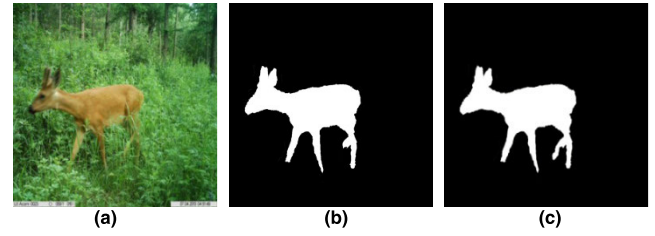

FIGURE 4. Saliency object detection and extraction. (a) Original Image. (b) Saliency object region. (c) Ground truth.

mask image, as shown in Fig.4. After obtaining the mask image of the wildlife region, progressive transmission is utilized based on the labeled wavelet coefficients through mask image. We utilize the bit plane lifting method to raise the coefficients of the saliency region to background region which ensures the transmission priority of saliency object region. Therefore, staff can prioritize obtain the important region information in the background server to know the species of wildlife.

2) IMAGE COMPRESSION CODING AND RESTORATION

After generating the mask of the wildlife region, we utilized the convolutional encoder-decoder to prioritize the transmission of the important region and ensure that the staff can receive the most interested region at the first time. Then the image restoration is applied to restore the distorted image due to the interference of the external noise, which guarantees the material validity and recognition accuracy for the subsequent research of wildlife recognition.

We selected 1,000 images as the training set, 100 images as the validation set and 200 images as the testing set. All images have a size of 128×128 pixels. In the network architecture in Fig.5, we utilized the ADAM optimizer [35] with learning rate l_r and a mini-batch size of b, and each convolutional layer has a filter with the number N and the size k, which is followed by ReLU. Additionally, the skip connections that connect the layer to its mirrored counterpart in the decoder part, are utilized to learn the feature in different level layers. Specifically, the output feature maps (obtained after ReLU) of the layer are element-wise added to the output feature maps (obtained before ReLU) of the counterpart layer.

For the training loss, the mean squared error (MSE) is used between the restored images and their ground truths as follows, which indicates the degree of inconsistency between the estimate value and the truth value. Compared with the cross-entropy function, this loss function is related to the correct and wrong prediction result to make the wrong result become average value.

$$L(\theta_D) = \frac{1}{|S|} \sum_{i=1}^{|S|} \|D(y_i; \theta_D) - x_i\|_2^2 \tag{4}$$

where S denotes the training set, x_i is a ground truth image and y_i is a distorated image.
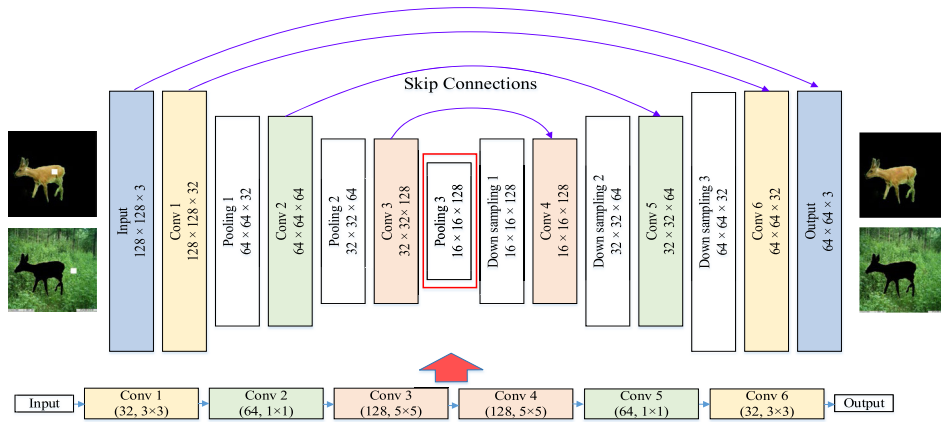
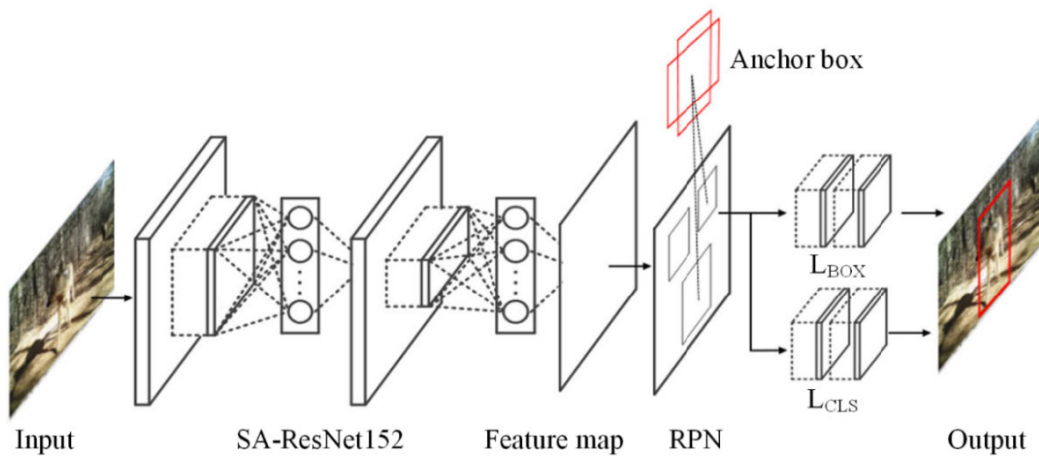**FIGURE 5.** Network architecture of image compression and restoration.

FIGURE 6. Improved Faster RCNN network architecture.

B. AUTOMATIC RECOGNITION FOR WILDLIFE MONITORING IMAGES

After the progressive transmission process proposed in the last section, we could obtain large number of wildlife monitoring images, which needed an efficient, accurate recognition algorithm to solve the sorting problems of low efficiency and recognition accuracy.

In this paper, we combined the target detection (Anchor box) and wildlife recognition into a comprehensive task and the improved Faster RCNN algorithm shown in the Fig. 6, was utlized to meet the practical demands of wildlife monitoring images. Compared with traditional Faster RCNN, two contributions were made in this study: 1) The k-means algorithm is utilized to choose the better anchor size in wildlfe region detection. 2) For imbalance training data, different weights of loss function are assigned to different wildlife.

1) BACKBONE NETWORK

The deep residual network (ResNet) [39] which introduces the residual learning block could solve the problem that the recognition performance will no longer improve with deeper depth when the depth of CNN substantially increased. However, this algorithm is difficult to take into account the characteristics of large and small receptive field in the same layer of convolution network. Because of the uncertainty of wildlife target location and size, the feature extraction of detail information and global information may lead to the failure which feature map can not deal with the relationship between image details and the whole image. To solve the above problem, a novel deep ResNet network of SA-ResNet152 shown in the Fig. 7 based on self-attention mechanism is proposed in this section. And Self-attention network utilized in this paper shown in the Fig. 8, is not restricted by the receptive field in the calculation of feature graph, achieving the target location effect equivalent to weak supercision through assigning different weights to the network parameters through the softmax.

The core of the self-attention network consists of three convolution layers and the feature map extracted by CNN can be the output of the end convolution layer or any layer. First, the feature map is transformed into three feature spaces $f(x)$, $g(x)$ and $h(x)$ expressed as

$$f(x) = W_f x \tag{5}$$

$$g(x) = W_g x \tag{6}$$
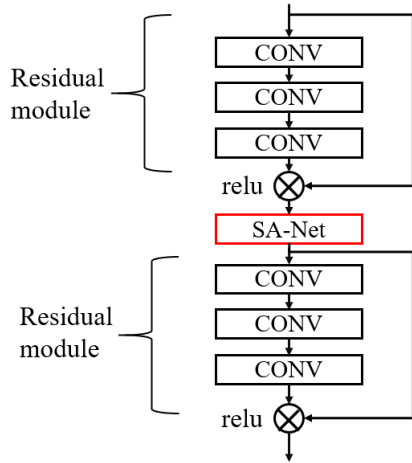
$$h(x) = W_h x \tag{7}$$

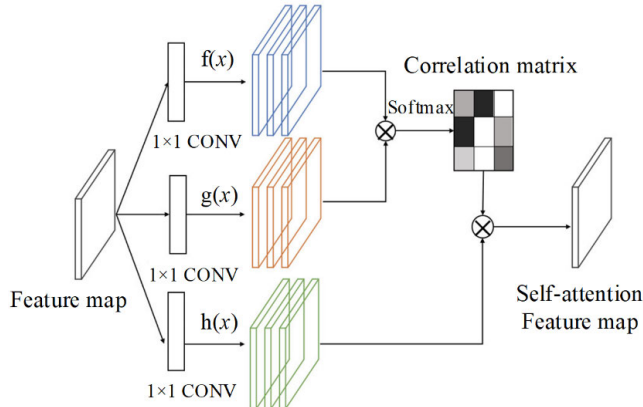**FIGURE 7.** Deep residual network architecture of SA-ResNet152.

FIGURE 8. Self-attention network architecture.

where W_f, $W_g \in R^{(c \times ratio) \times c}$, $W_h \in R^{c \times c}$ represent the three weight matrices calculated by the convolution layer. Among them, c is the dimension of the input feature graph and *ratio* is the proportional coefficient which the channel $c \times ratio$ of convolution is 1×1. The weight matrix and the feature graph are respectively convoluted by 1×1 to obtained their respective feature spaces.

Then the correlation matrix $\beta_{j,i}$ can be calculated by multiplying the matrices of $f(x)$, $g(x)$ and normalizing the attention matrix by Softmax layer, which indicates the degree of correlation between the j region and the i position in the feature graph. Finally, the output self-attention characteristic graph δ_j is calculated by the attention matrix $\beta_{j,i}$.

$$\beta_{j,i} = \frac{\exp(s_{i,j})}{\sum_{i=1}^{N} \exp(s_{i,j})}, \quad where \ s_{i,j} = f(x_i)^T g(x_j) \quad (8)$$

$$\delta_j = \sum_{i=1}^{N} \beta_{j,i} h(x_i) \quad (9)$$

In order to add the output self-attention feature matrix to the latter network architecture as a module, we need to return its reshape to [Batchsize, H, W, C], and add it to the input feature matrix as the final output y_i,
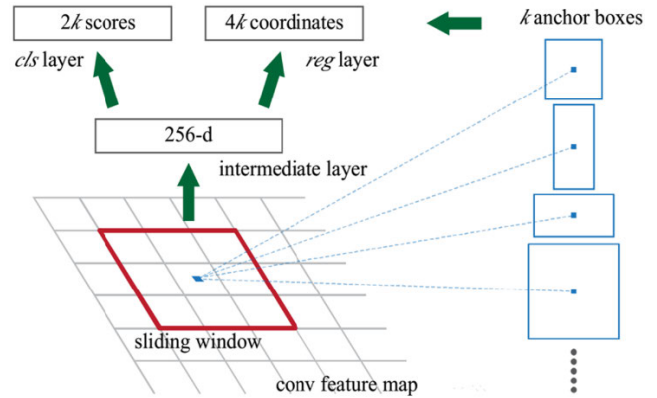
$$y_i = \lambda \delta_i + x_i \quad (10)$$

FIGURE 9. Region proposal network architecture.

where λ is the proportional coefficient of self-attention characteristics.

2) REGION PROPOSAL NETWORK

The region proposal network is used to locate candidate wildlife object, as shown in Fig.9. After image feature maps are obtained by backbone network, each feature point of feature map corresponds to the local receptive field of the orginal image. Therefore, all feature points on the map can reflect the original location features which include the candidate wildlife object and background.

Anchor boxes are utilized in the RPN to extract the features of 256-dimensional vectors at the same time, which perform a sliding scan on a feature map through convolution calculation. The size of anchor boxes are 3 scales (8,16 and 32) and 3 length/width ratios (1:2, 1:1 and 2:1) in common to deal with different scales.

And then the generated 256-dimensional vector is used to access two fully connected layers, named the border classification layer and the border regression layer respectively. We used the regression layer coordinates of the starting point (x, y). The width w and the height h were to determine the target location classification layer, which classifies the border of the target area and background area. The provisions of the same as the real area overlap is greater than 0.7 and the negative sample overlap with the real area is less than 0.3.

In order to make the detection of wildlife target more accurate, k-means algorithm [40] was used to cluster the length/width ratio of anchor box in the RPN. After the regressing of the training dataset through k-means algorithm, the width w and the height h of wildlife target area were recorded. We can calculate the length/width ratio r, r = h/w, and cluster the 3 scales of r shown in the fig.10. The abscissa represents the sample number and the ordinate represents the ratio of the length/width. The red, green and blue represent the Clustering results of low, medium and high level respectively. From the clustering result we can see that the clustering values reached 1, 2 and 4 respectively which means the ratiO (1:1, 1:2 and 1:4) are used to replace the traditional ratio (1:2, 1:1 and 2:1).
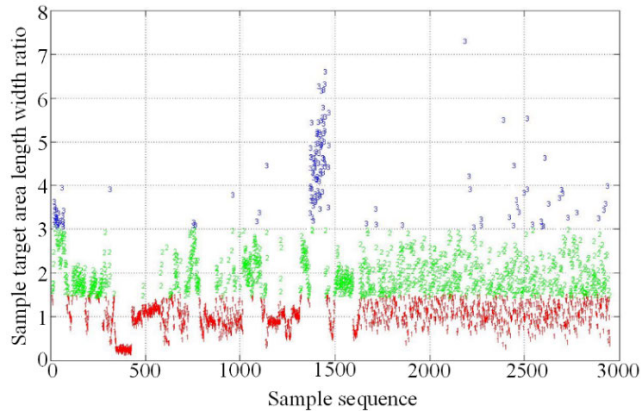
FIGURE 10. K-means clustering result graph.

TABLE 2. Wildlife image sample database.

Species of wildlfe	Number of images
red deer	4984
chinese goral	943
roe deer	2637
lynx	411
wild boar	1307
raccoon dog	438

3) LOSS FUNCTION

Due to the different quantities and habits of wildlife in test site, the number of wildlife monitoring images varies accordingly. This imbalance in data poses a challenge on the automatic wildlife recognition.

The loss function of Faster RCNN is a multi-task loss function, including the classification loss function softmax and the regression loss function. Regarding the imbalance dataset issue, different weights are assigned to different wildlife species in loss function, is expressed

$$L = -\sum_i \alpha_i \hat{y}_i log(y_i) \qquad (11)$$

$$\alpha_i = \frac{N_{total}}{n \times N_i} \qquad (12)$$

where y denotes the network prediction category result probability yielded by softmax, and $\hat{y}$ is the category label in a component of one hot vector. α is the loss function weight and i is the wildlife category. N_{total} is the total number of training images and N_i is the number of i_{th} species for wildlife training images. n is total species of wildlife.

IV. EXPERIMENTAL RESULT AND ANALYSIS
A. WILDLIFE DATASET

To evaluate the effectiveness and superiority of the proposed algorithm, we applied our proposed algorithm to the wildlife dataset, in which all monitoring images were calibrated by professional zoologists manually and eliminated the invalid monitoring images with no wildlife. The above calibration ensures the high quality of training samples.
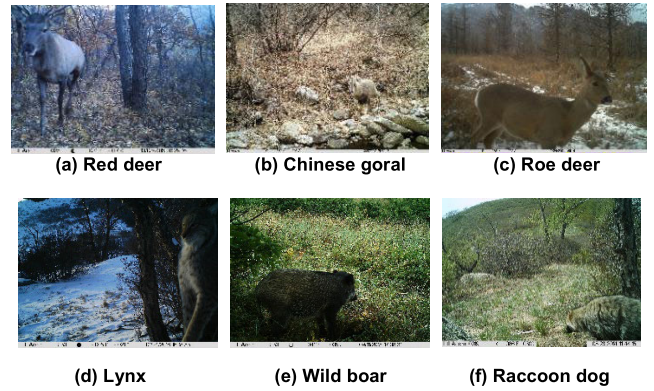

| (a) Red deer | (b) Chinese goral | (c) Roe deer |
| (d) Lynx | (e) Wild boar | (f) Raccoon dog |

FIGURE 11. Experimental samples of six species wildlife.

B. IMAGE RECONSTRUCTION QUALITY

In the process of transmission, we preferentially transmitted and restored the saliency region image, and then the background region was processed after the saliency region was completely received. We set the learning rate l_r of the ADAM optimizer at 0.001 and the mini-batch of size b at 16. For the number N and size k of filters at each layer, we selected from {32, 64, 128} and {1×1, 3×3, 5×5}, respectively.

In order to verify the effectiveness of the proposed algorithm, two traditional algorithms [36], [37] were compared with our proposed algorithm on saliency region to evaluate the quality of image reconstruction.

Both the peak signal to noise ratio (PSNR) and structural similarity index (SSIM) [38] are utilized as objective criteria to evaluate the quality of final wildlife image.

$$PSNR = 10log\left[\frac{(255)^2}{MSE}\right] dB \qquad (13)$$

$$MSE = \frac{1}{MN}\sum(g(x, y) - f(x, y))^2 \qquad (14)$$

where MN is the total number of pixels in the sample image. $g(x,y)$ is the reconstruction image and $f(x,y)$ is the orginal image.

SSIM is another criteria regarding the measurement of similarity between reconstruction and original images by calculating the image distortion degree according to the change of image structure information.

$$SSIM(x, y) = \frac{(2u_x u_y + C_1)(2\sigma_{xy} + C_2)}{(u_x^2 + u_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \qquad (15)$$

where u_x and u_y are the mean value of the luminance in the original and reconstructed image respectively. σ_x and σ_y are the standard deviation of the luminance. The constants C_1 and C_2 are used to suppress instability in structural similarity comparison.

To verify the restoration effect, the set partitioned in hierarchical tree (SPIHT) algorithm and embedded zerotree wavelets (EZW) algorithm are compared with our proposed algorithm in this section. As shown in Fig.13, the restoration image quality corresponding to each sample is improved when compared with traditional algorithms,
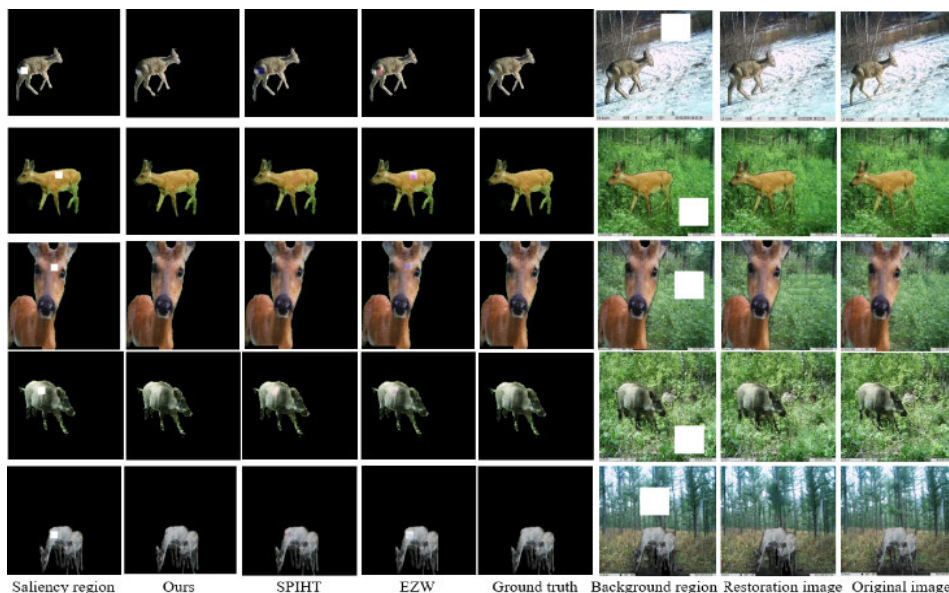
Saliency region   Ours   SPIHT   EZW   Ground truth   Background region   Restoration image   Original image

FIGURE 12. The flow chart of image restoration effect for wildlife images.
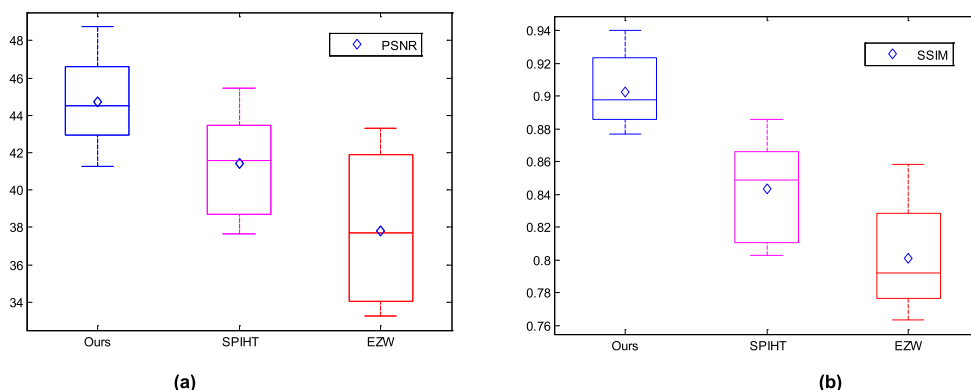

(a)

(b)

FIGURE 13. Comparison result of different algorithms. (a) Average result of peak signal to noise ratio (PSNR). (b) Average result of structural similarity index (SSIM).

which indicates the availability of sample images for wildlife recognition. Besides the average restoration results of PSNR and SSIM by our algorithm are 44.7035 dB and 0.9025, which increased by 7.93%, 18.15% and 7.01%, 12.67% respectively when compared with the SPHIT and EZW algorithms.

C. WILDLIFE AUTOMATIC RECOGNITION ACCURACY

All experiments were conducted on an Ubuntu 16.04 Linux server with an E5-2620 CPU and GTX 1080ti GPU, and the model implementation was based on the open source deep learning framework Keras. The network input materials were RGB color images and all images were re-sized to 224×224 pixels for further processing work.

We used ImageNet pre-training model and image data enhancement strategy, in which the dataset of wildlife monitoring images was randomly divided into training set, verification set and test set by the ratio of 7:1.5:1.5. In the process

of wildlife recognition, we set the learning rate of the ADAM optimizer to be 0.001 and the epochs is selected by 500.

To evaluate the effectiveness and superiority of the proposed algorithm, the accuracy P, mean average precision (mAP) [40], the sensitivity TPR and the specificity TNR are utilized as objective criteria to compare with different models for wildlife recognition.

$$P = \frac{TP}{TP + FP} \qquad (16)$$

where TP (True positives) is the number of cases which are correctly divided into positive cases, and the FP (False positives) represents the case number that is incorrectly divided into positive cases.

$$mAP = \frac{\sum_{q=1}^{n} AveP(q)}{n} \qquad (17)$$

(a) Traditional Faster RCNN algorithm

(b) Anchor box size improved algorithm

(c) Loss function improved algorithm
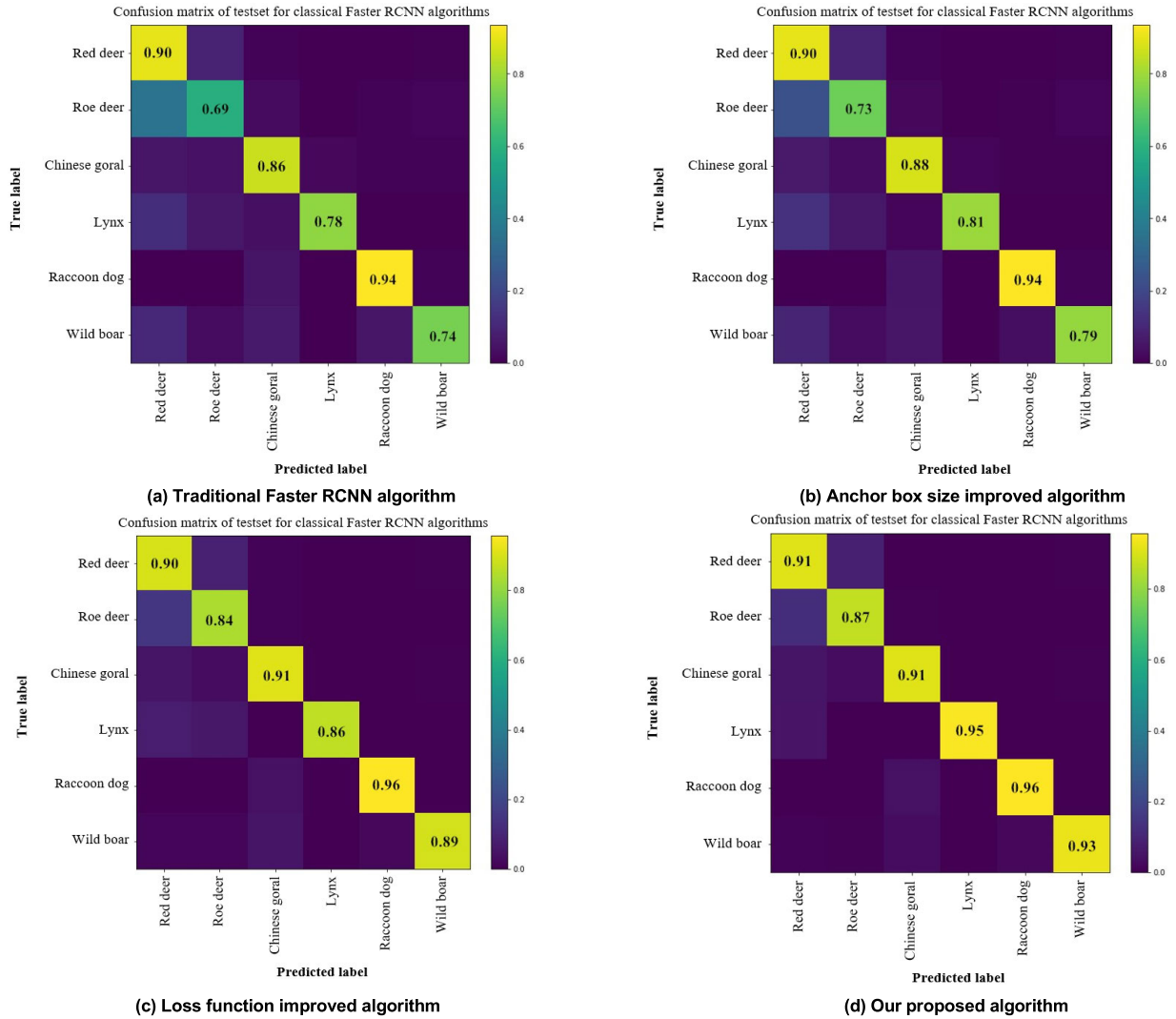

(d) Our proposed algorithm

FIGURE 14. The confusion matrix of compared algorithms.

where n is the number of wildlife species and it is set as 6 in this paper. $P(q)$ is the qth specie recognition accuracy of a certain wildlife monitoring image and the $AveP(q)$ is the average recognition accuracy of the qth specie under different thresholds.

$$TPR = \frac{TP}{TP + FN} \qquad (18)$$

where TP (True positives) is the number of cases which are correctly divided into positive cases, and the FN (False negatives) represents the case number that is incorrectly divided into negative cases.

$$TNR = \frac{TN}{TN + FP} \qquad (19)$$

where TN (True negatives) is the number of cases which are correctly divided into negative cases.

The confusion matrix was utilized in this section through the number of correct recognition result and error result

TABLE 3. The evaluation values of compared algorithms.

Compared algorithms	mAP values	TPR	TNR	Time
Faster RCNN algorithm	0.813	0.796	0.827	0.543
improved algorithm 1	0.842	0.824	0.853	0.538
improved algorithm 2	0.893	0.876	0.897	0.541
our proposed algorithm	**0.922**	**0.917**	**0.926**	**0.545**

of each wildlife specie. It is usually expressed by $n \times n$ matrix and the n represents the number of target categories. As shown in the fig.14, the total number of values in each column/row represents the total number of real/predicted results, and the diagonal line of the matrix represents the number of targets correctly classified.

To directly reflect the recognition effect of the training model, we selected the models of traditional Faster RCNN algorithm, anchor box size improved algorithm (Improved algorithm 1) and loss function improved algorithm (Improved

FIGURE 15. Results of wildlife recognition in complex cases.

algorithm 2) to compare with our proposed algorithm in this section. Compared with the traditional Faster RCNN algorithm, the accuracy of the six wildlife species has been improved by 1%, 18%, 5%, 17%, 2% and 19% respectively, and the final mAP value reaches to 92.2% which is increased by 10.9%. For a more comprehensive results to demonstrate the proposed algorithm, the TPR and TNR rates are increased by 12.1% and 9.9% respectively compared with traditional algorithm, and the running time is only added 0.002 second.

In the actual monitoring process, the monitoring camera is triggered by infrared sensor through sensing the temperature changes when the wildlife appear in monitoring range. However, there is a great uncertainty in locating the wildlife position for fixed monitoring nodes due to the habits and behaviors of wildlife, which makes the recognition accuracy not ideally high. The wildlife object may be covered by background when the object is distant from the camera and the animal is too tiny to see. Conversely, the monitoring images would not include the whole body of wildlife object if the distance is too close.

Therefore, we added the experiments to verify the effectiveness of the proposed algorithm for different types of wildlife monitoring images. As shown in the Fig. 15, the first row experiment samples are the standard wildlife monitoring images which are basically appear in the positive center of the image, and the background information is simple. Even in the extreme cases, such as the images in second row are occluded images and the third row images are all locally captured images, the algorithm can achieve precise result. The last row images include a part of the body region because the wildlife is too close to the camera. In conclusion, the propsed algorithm can deal with the influence of various complex situations and presents an excellent recognition accuracy and strong detection robustness.

V. CONCLUSION

In this paper, we proposed a wildlife intelligent monitoring system for actual wildlife images, which can achieve the remote, real-time transmission and recognition of wildlife monitoring images. Saliency object detection based on

histogram statistics was utilized to ensure the priority of wildlife region in the transmission process. For the image information lost caused by external interference during the transmission process, convolution encoder-decoder network was utilized to realize the progressive transmission and restoration of wildlife images. Finally, an improved Faster RCNN algorithm was proposed to achieve the automatic recognition of six species wildlife, which improved the efficiency and accuracy of wildlife classification. To demonstrate the effectiveness of our proposed algorithm, the experimental results on wildlife dataset show that the PSNR and SSIM value are improved by 7.93%, 18.15% and 7.01%, 12.67% respectively when compared with SPIHT and EZW algorithms, which guarantee the reliability of the experiment samples for subsequent wildlife recognition. Compared with the traditional Fast RCNN algorithm, the recognition accuracy of the six species wildlife, has been respectively improved by 1%, 18%, 5%, 17%, 2% and 19%, and the final mAP value reaches to 92.2% which is increased by 10.9%.

In the future, we will collect more wildlife sample images for validation purpose and focus on the individual classification of wildlife to improve the recognition accuracy.

REFERENCES

[1] B. Hori, R. J. Petrell, G. Fernlund, and A. Trites, "Mechanical reliability of devices subdermally implanted into the young of long-lived and endangered wildlife," *J. Mater. Eng. Perform.*, vol. 21, no. 9, pp. 1924–1931, Sep. 2012.

[2] M. Gor, J. Vora, S. Tanwar, S. Tyagi, N. Kumar, M. S. Obaidat, and B. Sadoun, "GATA: GPS-Arduino based tracking and alarm system for protection of wildlife animals," in *Proc. IEEE Int. Conf. Comput. Inf. Teleco. Syst.*, Dalian, China, Jul. 2017, pp. 166–170.

[3] J. M. Pérez , M. E. da la Varga, J. J. Garcia, and V. R. Lacasa, "Monitoring lidia cattle with GPS-GPRS technology; A study on grazing behaviour and spatial distribution," *Vet. Mex.*, vol. 4, no. 4, pp. 11–28, Oct. 2017.

[4] A. Fernández-Caballero, M. T. López, and J. Serrano-Cuerda, "Thermal-infrared pedestrian ROI extraction through thermal and motion information fusion," *Sensors*, vol. 14, pp. 6666–6676, Apr. 2014.

[5] R. N. Handcock, D. L. Swain, G. J. Bishop-Hurley, K. P. Patison, T. Wark, P. Valencia, P. Corke, and C. J. O.'Neill, "Monitoring animal behaviour and environmental interactions using wireless sensor networks, GPS collars and satellite remote sensing," *Sensors*, vol. 9, no. 5, pp. 3586–3603, May 2009.

[6] H. Wang, A. O. Fapojuwo, and R. J. Davies, "A wireless sensor network for feedlot animal health monitoring," *IEEE Sens. J.*, vol. 16, no. 16, pp. 6433–6446, Aug. 2016.

[7] A. J. Pinho, A. R. C. Paiva, and A. J. R. Neves, "On the use of standards for microarray lossless image compression," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 3, pp. 563–566, Mar. 2006.

[8] D. S. Taubman and M. W. Marcellin, "JPEG2000: Standard for interactive imaging," *Proc. IEEE*, vol. 90, no. 8, pp. 1336–1357, Aug. 2002.

[9] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[10] S. Poornachandra, V. Ravichandran, and N. Kumaravel, "Mapping of discrete cosine transform (DCT) and discrete sine transform (DST) based on symmetries," *IEEE J. Res.*, vol. 49, no. 1, pp. 35–42, Jan. 2003.

[11] R. Kumar, A. Kumar, and G. K. Singh, "Electrocardiogram signal compression based on singular value decomposition (SVD) and adaptive scanning wavelet difference reduction (ASWDR) technique," *Aeu-Int. J. Electron. C.*, vol. 69, no. 12, pp. 80–92, Dec. 2015.

[12] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

[13] V. Papyan and M. Elad, "Multi-scale patch-based image restoration," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 249–261, Jan. 2016.

[14] R. Fattal, "Image upsampling via imposed edge statistics," *ACM Trans. Graph.*, vol. 26, no. 3, Jul. 2007, Art. no. 95.

[15] D. Perrone and P. Favaro, "Total variation blind deconvolution: The devil is in the details," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2909–2916.

[16] I. Satoshi, S. Edgar, and I. Hiroshi, "Globally and locally consistent image completion," *ACM Trans. Graph.*, vol. 36, no. 4, Jul. 2017, Art. no. 107.

[17] N. Kamyar, N. Eric, J. Tony, Z. Q. Faisal, and E. Mehran, "EdgeConnect: Generative image inpainting with adversarial edge learning," Jan. 2019, *arXiv:1901.00212*. [Online]. Available: https://arxiv.org/abs/1901.00212

[18] J. Yu, Z. Lin, J. Yang, X. Lu, and T. Huang, "Generative image inpainting with contextual attention," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5505–5514.

[19] J. Liu, C. Li, and W. Yang, "Supervised learning via unsupervised sparse autoencoder," *IEEE Access*, vol. 6, pp. 73802–73814, 2018.

[20] F. Qi, C. Lin, G. Shi, and H. Li, "A convolutional encoder-decoder network with skip connections for saliency prediction," *IEEE Access*, vol. 7, pp. 60428–60438, 2019.

[21] D. E. Swann, C. C. Hass, D. C. Dalton, and S. A. Wolf, "Infrared-triggered cameras for detecting wildlife: An evaluation and review," *Wilflife Soc. B.*, vol. 32, no. 2, pp. 357–365, Jun. 2004.

[22] A. N. Venkitasubramanian, T. Tuytelaars, and M. F. Moens, "Wildlife recognition in nature documentaries with weak supervision from subtitles and external data," *Pattern Recogn. Lett.*, vol. 81, pp. 63–70, Oct. 2016.

[23] X. Han, Y. Zhong, L. Cao, and L. Zhang, "Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification," *Remote Sens.*, vol. 9, no. 8, p. 848, Aug. 2017.

[24] S. Karen and Z. Andrew, "Very deep convolutional networks for large-scale image recognition," Sep. 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

[25] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[27] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2261–2269.

[28] E. Okafor, P. Pawara, F. Karaaba, O. Surinta, V. Codreanu, L. Schomaker, and M. Wiering, "Comparative study between deep learning and bag of visual words for wild-animal recognition," in *Proc. IEEE Conf. Sympos. Series Comput. Intelligen.*, Dec. 2016, pp. 1–8.

[29] X. Yu, J. Wang, R. Kays, P. A. Jansen, T. Wang, and T. Huang, "Automated identification of animal species in camera trap images," *Eurasip J. Image Vide.*, vol. 2013, Sep. 2013, Art. no. 52.

[30] G. Chen, T. Han, Z. He, R. Kays, and T. Forrester, "Deep convolutional neural network based species recognition for wild animal monitoring," in *Proc. IEEE Conf. Image Peocess.*, Oct. 2015, pp. 858–862.

[31] Hendry and R.-C. Chen, "Automatic license plate recognition via sliding-window darknet-YOLO deep learning," *Image Vis. Comput.*, vol. 87, pp. 47–56, Jul. 2019.

[32] T. N. Vikram, M. Tscherepanow, and B. Wrede, "A saliency map based on sampling an image into random rectangular regions of interest," *Pattern Recognit.*, vol. 45, no. 9, pp. 3114–3124, Sep. 2012.

[33] X.-L. Liu, M. Wang, Z.-J. Zha, and R. Hong, "Cross-modality feature learning via convolutional autoencoder," *ACM Trans. Multim. Comput.*, vol. 15, no. 1, Feb. 2019, Art. no. 7.

[34] K. Smet, W. R. Ryckaert, M. R. Pointer, G. Deconinck, and P. Hanselaer, "Colour appearance rating of familiar real objects," *Color Res. Appl.*, vol. 36, no. 3, pp. 192–200, Jun. 2011.

[35] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," Dec. 2017, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

[36] T. Xiang, J. Qu, and D. Xiao, "Joint SPIHT compression and selective encryption," *Appl. Soft Comput.*, vol. 21, pp. 159–170, Aug. 2014.

[37] E. Christophe, C. Mailhes, and P. Duhamel, "Hyperspectral image compression: Adapting SPIHT and EZW to anisotropic 3-D wavelet coding," *IEEE Trans. Image Process.*, vol. 17, no. 12, pp. 2334–2346, Dec. 2008.

[38] C. S. Rawat and S. Meher, "A hybrid image compression scheme using DCT and fractal image compression," *Int. Arab J. Inf. Techn.*, vol. 10, no. 6, pp. 553–562, Nov. 2013.

[39] B. Li and Y. He, "An improved ResNet based on the adjustable shortcut connections," *IEEE ACCESS*, vol. 6, pp. 18967–18974, 2018.

[40] G. Tzortzis and A. Likas, "The minmax *k*-means clustering algorithm," *Pattern Recognit.*, vol. 47, no. 7, pp. 2505–2516, 2014.

ANQI LI received the B.E. degree in automation from Beijing Forestry University, Beijing, China, in 2017, where she is currently pursuing the M.E. degree in control theory and control engineering. Her research interests include wildlife automatic recognition and convolution neural network model.

WENZHAO FENG received the B.E. degree in electronic information science and technology from Nanjing Agricultural University, Nanjing, China, in 2015. He is currently pursuing the Ph.D. degree in mechanical engineering with Beijing Forestry University, Beijing. His research interests include image compression coding and progressive transmission, image saliency object detection, image restoration, and wildlife automatic recognition.
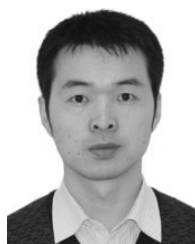
WEIDONG BAO received the B.S. degree from Inner Mongolian University, in 1988, the M.S. degree from the Chinese Academy of Agriculture Sciences, in 1993, and the Ph.D. degree from the Chinese Academy of Sciences, Beijing, China, in 1998. He is currently an Associate Professor in zoology with the College of Biological Science and Technology, Beijing Forestry University. His research interests include wild animal behavioral ecology and diversity monitoring by camera trapping and non-invasive sampling DNA analysis.

WENHUA JU received the B.S. degree in forestry from Northeast Forestry University, Harbin, China, in 1999. He is currently a Forestry Senior Engineer with Bahrain Zuoqi Forestry and Grassland Bureau, Chifeng, China. His research interests include wild animal monitoring and protection and classification of wild animal species.

JUNGUO ZHANG received the B.S. and M.S. degrees from the China University of Mining and Technology, in 2000 and 2003, respectively, and the D.E. degree from Beijing Forestry University, Beijing, China, in 2010. He is currently a Professor with Beijing Forestry University. His research interests include forestry information collection and intelligent progressing, wireless image sensor networks, artificial intelligence, and pattern recognition.

• • •