Scholar Publishing Group
International Journal of Neural Network
https://doi.org/10.38007/NN.2020.010105
ISSN 2790-2692 Vol. 1, Issue 1: 31-40

Algorithm of Fusion Convolution Neural Network in Animal Recognition

Manuel Gonzalez*

Indira Gandhi Delhi Technical University for Women, India

**corresponding author*

Keywords: Convolution Neural Network, Animal Recognition, Image Recognition, Recognition Algorithm

Abstract: The application of convolutional neural network has emerged in various fields in recent years, and has been favored by more and more experts, and has gradually entered the vision of ordinary people. Convolution neural network is a kind of deep neural network. Its main feature is that the front end input uses multi-layer locally interconnected neurons to extract the input information, and can consider the translation, rotation and scaling invariance of the signal target in space. In this paper, CNN is applied to image processing, and an animal recognition algorithm based on self normalized convolution neural network is proposed. Animal recognition technology is one of the most successful applications in image analysis and understanding, and has been attached importance by researchers from all walks of life. Especially in the past few years, Internet technology and information technology have been developing continuously, and the research on animal recognition has become more and more significant.

1. Introduction

With the rapid development of computer software and hardware technology and the arrival of big data era, deep learning has become a research hotspot in the direction of computer science. As one of the typical representative models of deep learning, convolutional neural network plays an important role in the field of computer vision [1]. It has the characteristics of local receptive field, weight sharing and downsampling, and can achieve end-to-end training and testing, thus replacing the traditional machine learning algorithm, and has made remarkable achievements in the field of image processing. Image recognition task is the key and representative research direction in the field of computer vision. In the current society, to deal with thousands of image data, traditional artificial feature extraction methods can no longer meet the basic needs, so the use of convolutional neural network technology can greatly improve the efficiency of image recognition. However, the

current convolutional neural network model has the problems of difficult design, low recognition accuracy and huge consumption of computing resources, which is not conducive to the promotion of practical applications and services. Therefore, a more lightweight network model needs to be designed to greatly reduce the size of network parameters under the condition of ensuring a fairly accurate rate.

In recent years, the convolutional neural network technology has developed rapidly. In view of its strong feature extraction ability for images, some scholars have begun to apply the convolutional neural network method to the field of animal recognition. Kamminga J W studied the basic challenges faced by real-time animal activity recognition, including the change of sensor direction, many features and energy and processing limitations of animal tags caused by changes in motion data. The aim is to find small optimized feature sets that are lightweight and robust for sensor orientation. The method consists of four main steps. First, select 3D feature vectors because they are theoretically independent of direction. Second, the least interesting feature is suppressed to speed up computation and improve robustness against over fitting. Third, the function is further selected through the embedded method, which selects features by simultaneously selecting features and classifications. Finally, the feature set is optimized through 10 times cross validation. Collect real world data through multiple sensors around the necks of five goats. The results show that only accelerometer data and some lightweight functions are used. In addition, the performance is robust to sensor direction and position. This work supports embedded, real-time, energy-saving and robust animal activity recognition [2]. Petrosky A L studied that many vertebrates cry for help when predators attack, as the last attempt to survive. However, few studies have investigated whether the identification of distress calls involves learning or acoustic similarity to familiar calls. The study assessed the importance of these two factors and the phylogenetic correlation in the tropical rain forest bird recognition distress call. In a lowland tropical rainforest in Costa Rica, the response of bird communities to familiar and unfamiliar cries for help was measured by replaying a sympatric species, namely the orange billed sparrow and a closely related alien species, the white eared sparrow. In addition, in order to test whether the recognition is driven by the phylogenetic correlation with the caller, the phylogenetic distance close to the individual of Staphylococcus aureus when using different playback stimuli was compared. It is found that individuals call back the same domain and different domain calls in similar time, supporting the role of voice similarity in distress call recognition [3]. At the same time, animal behavior analysis is an important research direction of the intersection of biology and computer science, which automatically analyzes the behavior of animal body, joints and other parts through the use of computer vision and artificial intelligence technology.

This paper studies the fusion convolution neural network, image retrieval and image recognition, and animal recognition algorithms. Animal recognition methods include traditional methods based on manual features, depth learning feature extraction methods, and multimodal multi stream neural networks. In the experiment, in order to study the algorithm of fusion convolution neural network in animal recognition, this paper studies the number of iterations by setting parameters and using the convolution operation between images, and at the same time, experiments are carried out on the feature saliency map on AWA2 animal dataset.

2. Research on Algorithm of Fusion Convolution Neural Network in Animal Recognition

2.1. Fusion Convolution Neural Network

The concept of fusion initially refers to data fusion. Later, as fusion technology was increasingly valued in military applications, it became an important research content [4]. With the continuous progress of science and technology, the performance of sensors is also improving, which makes

sensors gradually used in various fields. Because the information contained in the obtained data is different due to different sensor types, data fusion technology is required to integrate these obtained data information, which is richer than the information obtained from a single data source. In other words, the data fusion technology is to process and fuse the data information obtained by different types of sensors on the same target in multiple aspects, levels and levels to obtain richer, more reliable and more accurate useful data [5].

The biggest difference between the convolutional neural network and other deep networks is that the convolutional neural network has a feature extractor, which includes a convolutional layer and a pooling layer [6]. In a convolution layer of convolutional neural network, there are many feature planes, and each feature plane contains some neurons composed of data matrix. The three basic ideas of CNN are local receptive field, shared weight and pooling. Convolution neural network is composed of several parts. The following is a brief introduction to each part of the convolution neural network: data input layer: the data input layer is different according to the different sample data. Generally, voice data needs to be converted to two-dimensional data input, and image data can be directly input. Gray scale images can be used to reduce the complexity of the model. At the same time, it also supports three channel RGB color images or more dimensional images as input [7].

2.2. Image Retrieval and Image Recognition

Image retrieval is to find one or more image data with the highest degree of relevance from the retrieval database according to the given query content. According to different query modes, image retrieval can be divided into text-based image retrieval and content-based image retrieval [8]. The input of the former is text, which can be retrieved directly by matching the text label of the image in the retrieval library; The input of the latter is an image. It is more difficult to achieve retrieval by comparing the similarity of image semantics between the query and the retrieval database. Content-based image retrieval has been widely used in pedestrian recognition, face recognition and image search. Content based image retrieval can also be used for object category recognition. In this application, the retrieval database stores the image itself and its corresponding category labels at the same time.

The core of image recognition based on image retrieval is to measure the similarity of images in category semantics. It is necessary to give higher correlation to samples of the same category and lower correlation to samples of different categories. In terms of algorithm, this process depends on feature extraction and similarity calculation [9]. The high-dimensional image input first gets its low-dimensional feature code after feature extraction, and then judges the image similarity according to the distance between its codes. Different from classification network, the prediction process of image retrieval is not end-to-end from image to tag encoding. In terms of model structure, the feature extractor of image retrieval can also use the convolutional neural network model, but its output is a fixed length feature encoding, which is different from the classification network classifier output category label encoding. Therefore, the neural network model of image retrieval does not directly give the category information, but requires subsequent matching with the feature coding of reference samples in the retrieval database to obtain the category information indirectly [10].

2.3. Animal Recognition Algorithm

(1) Traditional methods based on manual features

Local spatiotemporal features are often used to describe video representation information in action recognition before the emergence of depth learning. Local spatiotemporal features can capture the shape and motion of features in video, and provide relatively independent event features

for spatiotemporal migration and scaling of local features, as well as background confusion and motion diversity [11]. These features are directly extracted from video through local spatiotemporal detectors and descriptors, thus avoiding the possibility of other preprocessing methods such as segmentation and tracking failure in motion. After using feature detectors and descriptors to obtain local spatio-temporal features of video, these spatio-temporal features will first be quantized into visual words, and then the video features will be represented as frequency histograms of visual words [12]. The frequency histogram obtained from the number of visual words can be used as a video feature and sent to a classifier, such as a non-linear support vector machine, for classification to complete action recognition. Different local spatiotemporal features often focus on different video representation information. Feature detectors usually use maximum saliency function to obtain spatio-temporal location and scale information in video. The difference between feature detectors mainly lies in the type and sparsity of selected points of interest. The feature descriptor uses image algorithms such as spatio-temporal image gradient and optical flow algorithm to capture the shape and motion information between the neighborhood of interest points [13].

(2) Feature extraction method of deep learning

The traditional manual feature extraction method depends on the design of the feature description operator. The whole design process not only depends on the researchers' knowledge of the image field, but also is tedious and prone to errors. However, the emergence of deep learning has changed this situation. It only needs to design a loss function, and then iteratively convex optimize the loss function. The deep learning method can adaptively extract the desired features [14]. This method does not require prior knowledge or manual intervention to extract features, and is more purposeful, which is the main reason why deep learning can make achievements in the field of computer vision [15-16].

(3) Multiflow Neural Network Based on Multimode

In order to learn the static appearance information and dynamic motion information in video respectively, a dual stream neural network is proposed. They propose to use two different neural networks for action recognition, one of which is a spatial flow network to learn the static information of video frames, and the other is a time flow network to learn the dynamic information through optical flow. At the same time, a time sequence segmentation network is proposed to capture long-term dependence. Its main idea is to process multiple video clips with equal intervals at the same time, and then aggregate the output of the last layer through post fusion, so as to convert the difficulty of processing multiple frames into processing a single frame. It improves the ability of the network to learn time sequence relations without increasing network parameters, and studies the fusion effect in the middle layer, it shows that the method of middle layer fusion can effectively reduce the number of parameters and improve the accuracy [17-18].

3. Investigation and Research on Algorithm of Fusion Convolution Neural Network in Animal Recognition

3.1. Research Content

In order to study the algorithm of fusion convolutional neural network in animal recognition, this paper studies its iteration times by setting parameters, and at the same time, experiments are carried out on the SISURF feature saliency map and SUSIFT feature saliency map of AWA2 animal dataset experiment on the experimental dataset.

3.2. Convolution Operation between Images

The scale space $L(x,y,\sigma)$ of an image is represented as the convolution between the scale changing Gaussian function $G(x,y,\sigma)$ and the $I(x,y)$ image:

$$L(x,y,\sigma)=G(x,y,\sigma)*I(x,y) \tag{1}$$

Where m and n represent the dimensions of the Gaussian template, x and y represent the positions of pixels, and σ is the scale space factor. The convolution operation is:

$$G(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e \tag{2}$$

4. Analysis and Research on Algorithm of Fusion Convolution Neural Network in Animal Recognition

4.1. Parameters and Training Speed of the Network

Ten convolutional neural networks, including ResNet50, ResNet101, ResNet152, FractalNet, DenseNet121, DenseNet161, DenseNet169, MobileNet, MobileNetV2, and MobileNet Beta, are used on the AWA2 dataset for balanced processing of the dataset to conduct experiments, compare the number of network parameters and the iterative training duration of the network, so as to select a network that can be suitable for the characteristics of small mobile terminal delay, small resource consumption, etc, The experimental results are shown in Table 1 and Figure 1:

Table 1. Convolutional neural network experimental data

Experimental network	Number of Parameters (M)	Iteration time (s)
MobileNet	33.44	669
MobileNet V2	32.02	551
MobileNet- Beta	54.31	685
ResNet50	103.4	559
ResNet101	186.3	1365
ResNet152	260.71	1423
FractalNet	96.34	863
DenseNet121	90.37	874
DenseNet161	113.24	1363
DenseNet169	145.36	1535

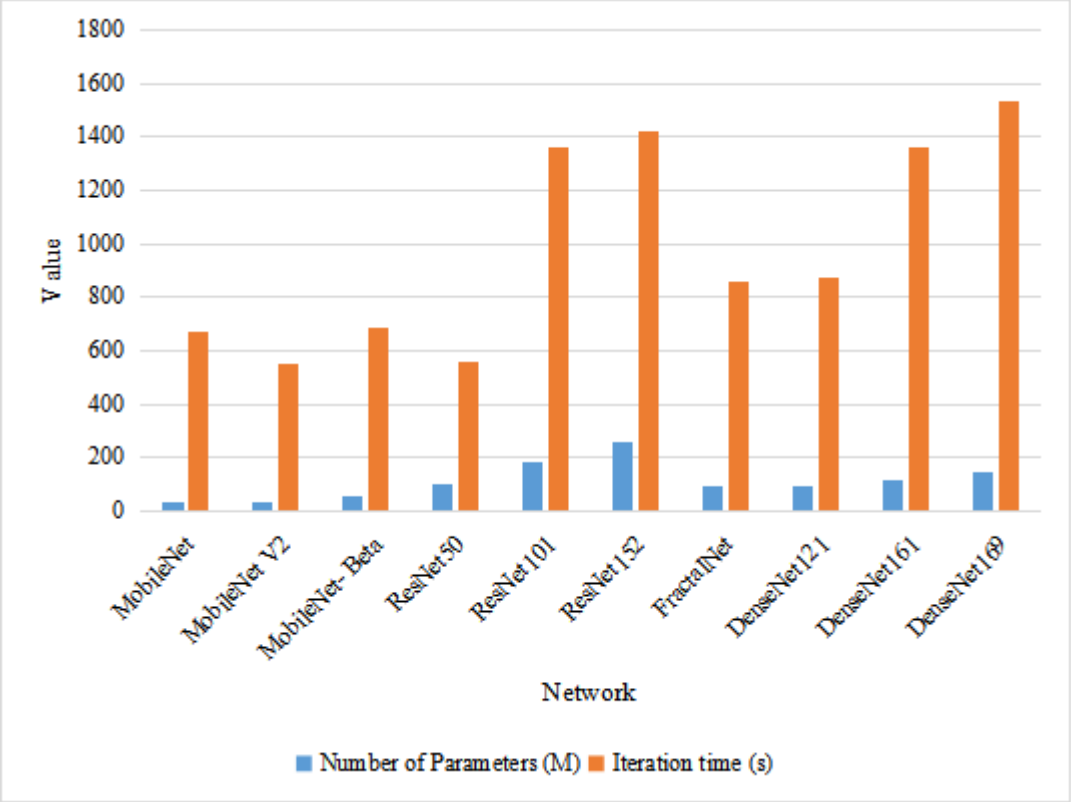

Figure 1. Number of parameters and iteration times of the networks on A W A 2

The experimental results show that among the AWA2 data sets recognized by various networks, the parameters of MobileNetV2 network are the least compared with other networks, only 32.02M in size. From the experimental results, we can see that MobileNet and MobileNetV2 networks and MobileNet Beta networks have greatly reduced the number of parameters and iteration time compared with traditional networks, and are more suitable for the development of mobile networks. In terms of training speed, the MobileNetV2 network using the linear bottleneck reciprocal residual structure is the fastest, and the required single iteration time is only 551s.

4.2. Experimental Results and Analysis

After the feature saliency map is obtained, the DenseNet121 network is pre trained using the feature saliency map as the network input to strengthen the network's local feature learning of the image. After the pre trained network is obtained, the original dataset is trained using the network. The balanced AWA2 animal data set experiment uses the SISURF feature saliency map and the SUSIFT feature saliency map with pixel values of (255255255) and (0,0,0) on the experimental data set. The experimental results are as follows:

(1) The identification accuracy of AWA2 dataset is shown in Table 2 and Figure 2:

Table 2. Data table

Iterations	10	20	30
Black significant plot of the SISURF	0.96	0.98	0.99
SISURF significant plot in white	0.85	0.90	0.95
The SUSIFT interior color developer diagram	0.73	0.84	0.86
SUSIFT significant plot in white	0.70	0.83	0.85

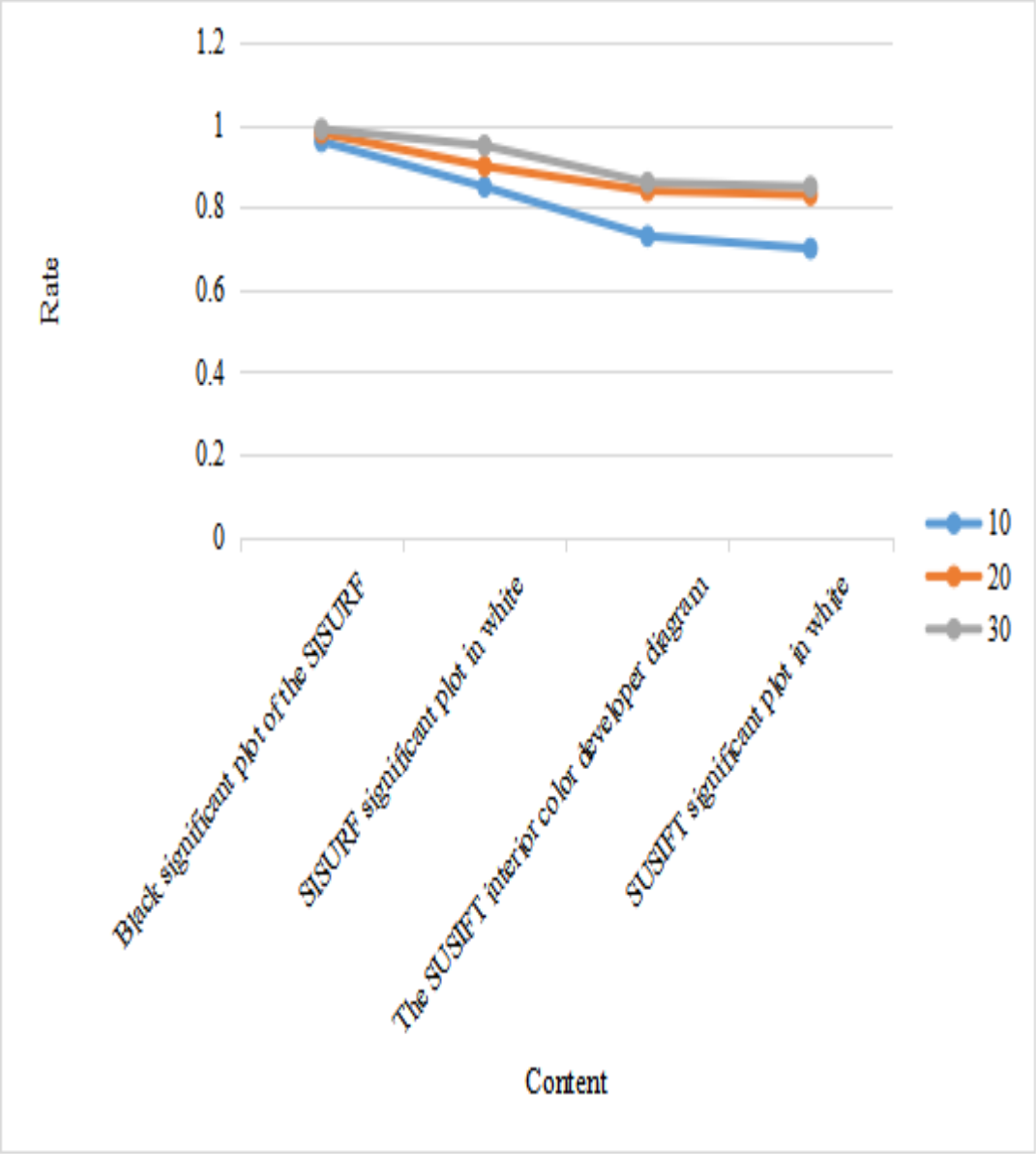

Figure 2. Training accuracy of the A W A 2 dataset

(2) The training loss function value of AWA2 dataset is shown in Table 3 and Figure 3:

Table 3. Loss function values

Iterations	10	20	30
Black significant plot of the SISURF	0.13	0.08	0.03
SISURF significant plot in white	0.27	0.12	0.07
The SUSIFT interior color developer diagram	0.34	0.28	0.26
SUSIFT significant plot in white	0.38	0.15	0.16

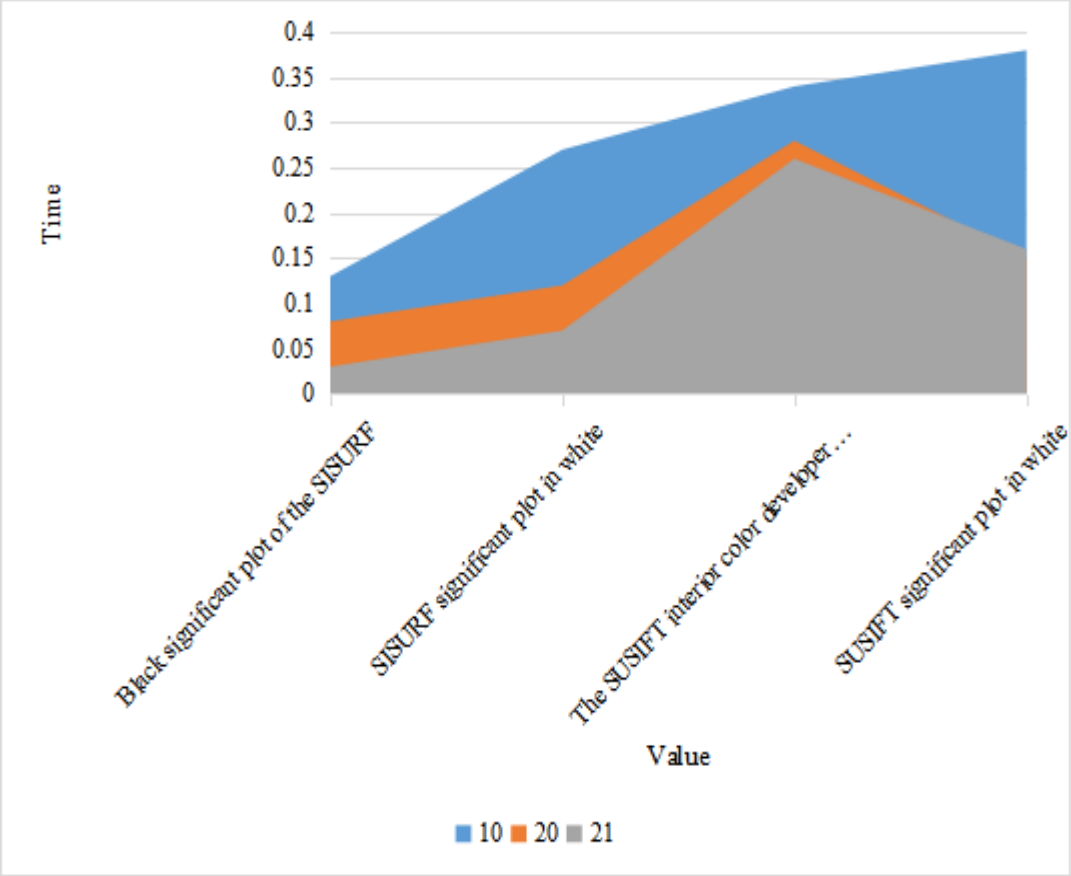

Figure 3. Training loss function values for the A W A 2 dataset

From the experimental results, it can be concluded that the performance of the network that is pre trained through the experimental black SISURF feature saliency map and then re trained with the original data set is better than that of the original data set alone, and the recognition rate on the balanced AWA2 data set is improved. The other three methods include the network that uses white SUSURF feature saliency map, black SUSIFT feature saliency map and white SUSIFT feature saliency map for pre training, and their performance is worse than that of the network that uses the original dataset for training.

5. Conclusion

The 21st century is the era of artificial intelligence. With the development of artificial intelligence technology, the rise of intelligent breeding industry and the continuous strengthening of national food safety control measures, accurate identification of animals has become an urgent task for the industry. At present, the main method of animal identification is to stick wireless identification marks through holes. This complicated identification method is easy to cause discomfort to livestock, and the labels often fall off. Convolutional neural network is the most important technical tool in the field of animal recognition, which can automatically extract more and more complex features of the image, thus greatly improving the accuracy of the algorithm. Thanks to the rapid development of computer technology and hardware, biometric recognition technology has been developed rapidly, and animal recognition has also become the focus of increasing attention. Therefore, the algorithm research of the fusion convolution neural network in animal recognition in this paper is of great significance.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Nakashima, Yoshihiro, Fukasawa, et al. Estimating animal density without individual recognition using information derivable exclusively from camera traps. The Journal of applied ecology, 2018, 55(2):735-744. https://doi.org/10.1111/1365-2664.13059

[2] Kamminga J W, Le D V, Pieter M J, et al. Robust Sensor-Orientation-Independent Feature Selection for Animal Activity Recognition on Collar Tags. Proceedings of the Acm on Interactive Mobile Wearable & Ubiquitous Technologies, 2018, 2(1):1-27. https://doi.org/10.1145/3191747

[3] Wu Y, Petrosky A L, Hazzi N A, et al. The role of learning, acoustic similarity and phylogenetic relatedness in the recognition of distress calls in birds. Animal Behaviour, 2020, 175(9):111-121. https://doi.org/10.1016/j.anbehav.2020.02.015

[4] Day C P, E Pérez-Guijarro, A Lopès, et al. Recognition of observer effect is required for rigor and reproducibility of preclinical animal studies. Cancer Cell, 2020, 40(3):231-232. https://doi.org/10.1016/j.ccell.2020.01.015

[5] Raheja J L, Gupta H, Chaudhary A. Monitoring Animal Diseases in Remote Area. Pattern Recognition & Image Analysis, 2018, 28(1):133-141. https://doi.org/10.1134/S1054661818010145

[6] Kostuch L, Wojciechowska B, Konarska-Zimnicka S. Ancient and Medieval Animals and Self-recognition: Observations from Early European Sources. Early Science and Medicine, 2019, 24(2):117-141. https://doi.org/10.1163/15733823-00242P01

[7] Kim M, Kim J. An Effect of Tourism Motivation on Constraint Recognition and Participation Intention in Travel with Companion Animal. Journal of Tourism Management Research, 2019, 23(7):367-387. https://doi.org/10.18604/tmro.2019.23.7.17

[8] Qadr F, Prasetijo A B, Windasari I P. "Animal Introduction" Application for Children. Jurnal ULTIMATICS, 2020, 11(2):65-71. https://doi.org/10.31937/ti.v11i2.1249

[9] Guazzaloca G. 'Anyone who Abuses Animals is no Italian': Animal Protection in Fascist Italy. European History Quarterly, 2020, 50(4):669-688. https://doi.org/10.1177/0265691420960672

[10] Stoddard M C, Osorio D. Animal Coloration Patterns: Linking Spatial Vision to Quantitative Analysis. The American Naturalist, 2019, 193(2):000-000. https://doi.org/10.1086/701300

[11] Galpayage S, Chittka L. Charles H. Turner, pioneer in animal cognition. Science, 2020, 370(6516):530-531. https://doi.org/10.1126/science.abd8754

[12] Megan M S. Aerodigestive Disease in Dogs. Veterinary Clinics of North America: Small Animal Practice, 2020, 51(1):17-32. https://doi.org/10.1016/j.cvsm.2020.09.003

[13] Grume G J, Biedenbender S P, Rittschof C C. Honey robbing causes coordinated changes in foraging and nest defence in the honey bee, Apis mellifera. Animal Behaviour, 2020,

173(12):53-65. https://doi.org/10.1016/j.anbehav.2020.12.019

[14] Luang-In V, Katisart T, Nudmamud-Thanoi S, et al. Psychobiotic Effects of Multi-Strain Probiotics Originated from Thai Fermented Foods in a Rat Model. Food Science of Animal Resources, 2020, 40(6):1014-1032. https://doi.org/10.5851/kosfa.2020.e72

[15] Sarthak Y, Singh B A. Residual nets for understanding animal behavior. Journal of Animal Behaviour and Biometeorology, 2019, 7(2):97-103. https://doi.org/10.31893/2318-1265jabb.v7n2p97-103

[16] Nolan M, Dobson J. The future of radiotherapy in small animals - should the fractions be coarse or fine? The Journal of small animal practice, 2018, 59(9):521-530. https://doi.org/10.1111/jsap.12871

[17] Oguejiofor C F, Thomas C, Cheng Z, et al. Mechanisms linking bovine viral diarrhea virus (BVDV) infection with infertility in cattle. Animal Health Research Reviews, 2019, 20(1):72-85. https://doi.org/10.1017/S1466252319000057

[18] Hase K, Kutsukake N. Developmental effects on social preferences in frog tadpoles, Rana ornativentris. Animal Behaviour, 2019, 154(4):7-16. https://doi.org/10.1016/j.anbehav.2019.06.001