Recognition Of Animal Species On Camera Trap Images Using Machine Learning And Deep Learning Models

***Rajasekaran Thangarasu ,Vishnu kumar Kaliappan ,Raguvaran Surendran,Kandasamy Sellamuthu ,Jayasheelan Palanisamy**

Abstract: Wild animal movement monitoring and its distribution are essential for the conservation of animal life. Camera trap, a most commonly used technique for animal monitoring which automatically activate the camera on animal presence and obtain a huge volume of data. The present work aims to investigate various machine learning algorithms including Support Vector Machine (SVM), Random Forest (RF) and deep learning models such as Alexnet, Inception V3 for classification of animal species. Among which deep learning models outperforms than machine learning algorithms. In this paper, the overall comparison of accuracy between machine learning and deep learning models has been observed and discussed. The outcomes of the experiment suggest that InceptionV3 attains more accuracy than SVM, Random Forest, AlexNet and also results highly accurate classification is obtained with the availability of enough data and precise techniques. The experiment uses KTH dataset that composed of 19 different categories of animals among which 12 classes are selected to measure the performance of the models.

Keywords: Animal species recognition, camera trap, KTH dataset, Random Forest, SVM, AlexNet, Inception v3, fine-tuning

1. INTRODUCTION

Wildlife monitoring is crucial for tracking animal habitat utilization, population demographics, poaching incidents, and movement patterns. Numerous technology has been introduced that includes motion-sensitive camera traps, radio tracking, wireless sensor network tracking, and satellite for monitoring wild animals[1-2]. Currently, the animal espial and recognition are still an arduous challenge and there is no unique method that provides a sturdy and efficient solution at all situations. Monitoring wild animals through camera traps are prominent due to their commercial availability, equipped features, and ease deployment. The Extraction of knowledge from these camera-trap images is implemented using machine learning and deep learning models. Machine learning (ML) plays a key role in a wide range of statistical, image recognition, natural language processing, and expert systems.

Every instance in the dataset is represented using the set of features and class labels where the training data builds a predictive model. Some of the ML algorithms for feature extraction and classification are Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), and Random Forest. In this paper, Random forest and SVM has been used for the classification of wild animals using camera trap images. A proposed CNN model, Deep learning (DL) outperforms in the identification of wild animals using camera trap images without any manual intervention[15]. Most popular deep learning architectures are GoogleNet, AlexNet, VGG, ResNet, NiN, and inception-1, 2, 3 that are employed in the classification of images[3-4]. Among these architectures, Alexnet and Inception V3 are used for the classification of animal species images in KTH dataset. The level of accuracy at each method has been calculated and compared. The outline of this paper is organized as follows. Section 2 gives a brief overview of the related works in animal identification. In Section 3 the details about the datasets used for experiment purpose followed by In Section 4 the animal identification based on feature extraction and classification is discussed. Section 5 the obtained experimental results are listed. Finally, Section 6 Concludes and suggests future work.

2. RELATED WORK

Gullal Singh Cheema and Saket Anand developed an automatic wildlife monitoring system for detection and recognition of individual animal species like tigers, zebras, and Jaguars which are in different pattern. The faster Regional CNN is proposed to detect the whole body and the flank region of individual animals. The features and flanks extracted from animals are fed to the AlexNet to train the logistic regression for identification of individual[16]. Olivier Chapelle, Patrick Haffner, and Vladimir N. Vapnik developed Support Vector Machines for Histogram based Image Classification. It describes that the support vector machines (SVM's) are used to generalize well on difficult and different image classification problems where the only features which are used are high dimensional histograms[17]. Tibor TRNOVSZKY, Patrik KAMENCAY and Richard ORJESEK implemented the Animal Recognition System Based on Convolutional Neural Network. The main goal of this paper was to compare the overall recognition accuracy of the PCA, LDA, LBPH, and SVM with proposed CNN model. Mohammad Sadegh Norouzza deh demonstrated the state-of-the-art deep learning neural network method which results in automatic identification of animals on Snapshot Serengeti dataset which had of 3.2 million images with 42 different species of animals. Various deep learning architectures such as AlexNet, NiN, VGG, ResNet-18, 34, 152 were used to classify animal species. Among the six architectures, VGG provided a better accuracy of 96.8 %[7].

- *T. Rajasekaran, K.Vishnu kumar S.Raguvaran ,S.Kandasamy are Cyber Physical System Group and Professor in Department of Computer science Engineering,KPR Institute of Engineering and Technology, Coimbatore,Tami nadu,641407,India. E-mail: phdresearchpaper@outlook.com*
- *P.Jayasheelan, is Professor in Department of Computer science,Sri Krsihna AdthiyaCollege of Arts and Science,Coimbatore, Tamilnadu.*

Image Classification using Random Forests and Ferns was prepared by Anna Bosch, Andrew Zisserman and Xavier Munoz. It demonstrates the result of random forests/ferns with an appropriate node test and reduces training and testing costs over SVM. Xiaoyuan Yu applied sparse coding spatial pyramid matching (ScSPM) and cell-structured LBP (cLBP) as descriptors of local image features to improve recognition accuracy in automatic detection of animal species. Linear support vector machine algorithm performs the classification and achieved 82 % of accuracy[15].

3. DATASET

KTH-Animal Dataset consists of 1740 images with 19 different classes of animals captured from camera traps. In Machine Learning model the level of accuracy varies in accordance with the size of the dataset. For our experiment, we acquire 12 classes of wild animals including bear, deer, elephant, giraffe, gorilla, kangaroo, leopard, lion, panda, tiger, zebra and coyote for classification. Imbalance in camera trap images is problematic for machine learning because they become strongly biased towards classes[5-7]. Whereas in deep learning model, the dataset enters the top layer of the architecture directly without any pre-processing tasks since the KTH dataset was already trained using ImageNet model. To avoid over-fitting the model undergoes Data Augmentation process. Table I. displays the different classes of animal species and a number of images in each class of KTH dataset and Table II. represents the different classes of animal species and count of images in pre-trained KTH dataset after the augmentation process. The images in the dataset result in poor in quality particularly images captured during night time which direct to misclassification. The detection and recognition of individuals in the images are challenging owing to motion blur, occlusions, and lighting changes. However, the recognition of wild animal species is achieved to a greater extent through the feature extraction in each model. Fig 1. Represent the sample images of KTH dataset.

TABLE I. KTH DATASET

Animal Species	Image count
Bear	105
Cougar	100
Cow	97
Coyote	100
Deer	100
Elephant	100
Giraffe	84
Goat	99
Gorilla	83
Horse	100
Kangaroo	90
Leopard	100
Lion	98
Panda	97
Penguin	80
Sheep	68
Skunk	62
Tiger	100
Zebra	77

***Fig.1**. Sample images from KTH dataset. Row 1: Deer Row 2: Ttiger*

TABLE II. COUNT OF AUGME NTED IMAGES IN PRE-TRAINED KTH DATASET

Animal species	Augmented images
Bear	1586
Coyote	1500
Deer	1500
Elephant	1500
Giraffe	1260
Gorilla	1245
Kangaroo	1350
Leopard	1500
Lion	1470
Panda	1455
Tiger	1500
Zebra	1155

4. METHODOLOGY

RF is a simple machine learning technique, used for both classification and regression method. RF classifier can be characterized as the assemblage of tree-structured classifiers that evaluates the best outcomes in prediction[18-19]. The response value is calculated from different decision trees to provide better performance for the classification of wild animal camera-trap images. Random forest works efficiently even on unbalanced and missing data. The key role of the RF classifier is of two steps during the random selection. The first step involves the random selection of different trees and the second sampling step determines the split conditions for each leaf node[20]. The parameters for the feature extraction includes the number of random trees, seed value, normalization of bin value, training, and testing data determines the accuracy level of the model.

4.2. Support Vector Machine

SVM is a supervised machine learning technique best known for pattern and image classification. SVM is Inception-v3, a convolutional neural network developed by Google and used for objects classification that consists of multiple layers. This model is one of the pre-trained models on the tensorflow environment and uses the dataset that is already trained on ImageNet model. Transfer learning is used to fine tune the top layer of the model and data augmentation technique is applied to avoid over-fitting of the model. Fine-tuning and data augmentation techniques were employed to improve the recognition accuracy for each individual category of animals. The training and AlexNet is a structure-simple proposed CNN technique that consists of eight layers where the first five layers are the convolutional layers followed by max-Pooling layers and the last three layers are fully connected layers[8-9]. The network is 8 layers in its size and is also used to classify images into 1000 object categories. In order to improve the classification accuracy and to deal with multi-scale images, a pre-trained Alexnet architecture is needed. In this model,

4.1. RF classifier

generally a binary classifier where they can be adapted to handle numerous classification tasks where it possesses the notable ability of global approximators of a multivariate function to any desired degree of accuracy[16-17]. The inadequacy of the model results from sophisticated learning that doubles the model's capacity to the data complexity assuring a good performance. The support vector machine approach is considered as a good candidate because it does not need to add prior knowledge and has high generalization performance even when the dimensions are high in the input space. The performance of the model is higher and faster for smaller dataset during the training than the testing tasks. When a set of parameter point of animal images are given, a linear SVM finds the hyper-plane leaving the largest possible fraction of points in the class than the RBF and polynomial kernel functions.

4.3.Inception v3

testing are performed with the images obtained from the selected classes of KTH dataset. This model automatically recognizes the wild animals in camera trap images without any human interventions. InceptionV3 obtains more accuracy for the larger dataset on testing data where specifically it does not require any pre-processing techniques.

4.4.Alexnet

the input data is already trained using the Imagenet technique so that it directly enters the max-pooling layer for the fine-tuning process[10]. To avoid over-fitting, the KTH dataset volume is increased using data augmentation technique. The augmentation process includes flipping, rotation, and changes in brightness, and scaling of the images.

4.5.Data Augmentation

Deep neural network is best compared to the traditional machine learning algorithms. However, over-fitting results in performance degradation of the model. Data Augmentation (DA) is the most commonly used technique to improve the accuracy of the images by increasing the number of training images and also avoids over-fitting in image classification tasks which are as important in reducing errors[11-14]. A data augmentation technique includes scaling, translation, rotation, flipping, and adding noise to the images. Varying the lighting condition of the images is widely used to improve the performance of the system. In this research paper, rotation, flipping and changes in lighting are used to multiply the numbers of images for training and testing the images.

4.6.Transfer Learning

Convolution Neural Network (CNN), the deep learning technique involves numerous parameters, which result in poor learning with a small amount of dataset. Transfer learning is employed to find a solution to the problem presented above, by training the parameter obtained from a larger dataset (source dataset) and apply the knowledge gained to retrain the smaller dataset (target dataset). In this paper, the CNN architecture with Inception v3 and AlexNet network layers are trained on ImageNet dataset that can be

The experiments are performed on an Intel i7 machine with a 3.6GHz processor and 8 GB RAM. The proposed CNN model uses tensorflow, an open source library for machine learning and deep learning developed by Google which distributes computation across multiple CPUs and GPUs. Convolution Neural Network (CNN), the deep learning technique involves numerous parameters, which result in poor learning with a small amount of dataset. Transfer learning is employed to find a solution to the problem presented above, by training the parameter obtained from a larger dataset (source dataset) and apply the knowledge gained to retrain the smaller dataset (target dataset). In this paper, the CNN architecture with Inception v3 and AlexNet network layers are trained on ImageNet dataset that can be reused to extract the feature of wild animal images for recognition.The experiments are performed on an Intel i7 machine with a 3.6GHz processor and 8 GB RAM. The proposed CNN model uses tensorflow, an open source library for machine learning and deep learning developed by Google which distributes computation across multiple CPUs and GPUs.

5. EXPERIMENT AND RESULTS

The experiments on machine learning and deep learning model were conducted on KTH dataset to appraise the performance of the network and training accuracy of each category with respect to training steps in the model.

5.1. Experiment on Machine Learning algorithms

5.1.1.Classification using Random Forest

Random Forests are composed of numerous individually learned random decision trees. Each node of the tree is selected randomly during the training task. To make the tree more random, additional random thresholds for each

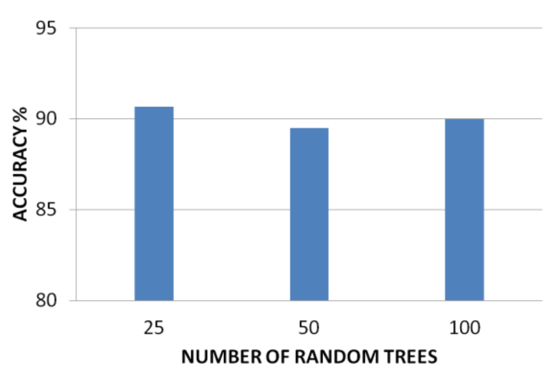

Fig. 2. *Comparison of accuracy based on number of*

The SVM classifier employs training data to create a classification model and testing data to test and evaluate trained model accuracy. SVMs are trained separately for each feature set and then it combines SVM score to generate final classification. Multi-class Support Vector Machine is even complicated and may not have significant performance. Therefore, we need a better SVM solution,

reused to extract the feature of wild animal images for recognition.

5. EXPERIMENT AND RESULTS

feature has been added. Training is achieved by identifying the node function and threshold which yields maximum information gain within the randomized search space. During training, the class posterior distributions are stored in the leaf nodes. Testing results by applying all tree to each input pixel in the test data. Table III represents the level of accuracy on varying the parameter values and by normalizing the bin value count as 8. The output for each pixel class posterior is attained as the average of the posteriors across all trees. To improve classification accuracy, each individual decision tree should differ and that difference can be accomplished through randomness where it is provided by the function parameter seed. By using the same seed, it will generate the same number sequence. Fig.2. Displays the accuracy acquired depending on the number of random trees.

TABLE II ACCURACY VALUE

Tree size	Seed value	Training size	Testing size	Accuracy %
100	1	90	10	89.97
100	6	90	10	89.56
100	9	90	10	89.42
100	9	80	20	87.97
100	9	70	30	85.57
50	9	90	10	89.49
25	9	90	10	89.82
25	1	90	10	90.66

5.1.2.Classification using Support Vector Machine

which should be memory efficient, fast and have some parallelization scheme to train the binary classifiers in parallel. Kernel, a set of mathematical functions has been used for the classification of wild animal images in KTH dataset. The functions of the kernel take data as input and transform it into the required format. Different Support Vector Machine uses different kernel functions such as linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid. The kernel functions Linear, RBF, and polynomial has been used in this work. Table IV represents accordance with the ratio of training and testing data. In linear function, the parameter value of c is taken as 0.01 to evaluate the level of accuracy. For RBF function the parameter values of c and gamma are taken as 0.01 and 0.0000001 respectively. And the polynomial function classifies the animal images using the parameter values of c as 1.0, degree as 3, and gamma as 0.1 where the accuracy varies for each kernel function depending on the training and testing data. Table V represents the animal identification rate of accuracy among the kernel functions. The experiments are performed on four different ratios of

the accuracy of kernel functions in

training and testing data. Fig 3 and 4 displays the comparison of accuracy among linear, RBF, and polynomial kernel functions and overall accuracy on Machine learning models. Among the ML algorithms, Random forest acquires more accuracy than Support vector machine. The ratio of training and testing data is divided as,
A: (90:10) 90% of training and 10% of testing data,
B: (80:20) 80% of training and 20% of testing data,
C: (70:30) 70% of training and 30% of testing data,
D: (60:40) 60% of training and 40% of testing data,
E: (50:50) 50% of training and 50% of testing data.

TABLE IV. REPRESENTS ACCURACY IN THE RATIO OF TRAINING AND TESTING DATA FOR EACH KERNEL FUNCTION

Kernel	A	B	C	D	E
Linear	84.72	74.20	70.33	66.12	64.46
RBF	82.98	71.87	68.40	63.44	60.67
polynomial	83.46	72.70	69.00	64.87	62.00

TABLE V. COMPARISON OF ACCURACY AMONG THE KERNEL FUNCTIONS USING TRAINING AND TESTING DATA

Kernel	Training size	Testing size	Accuracy %
Linear	90	10	84.72
RBF	90	10	82.98
Polynomial	90	10	83.46

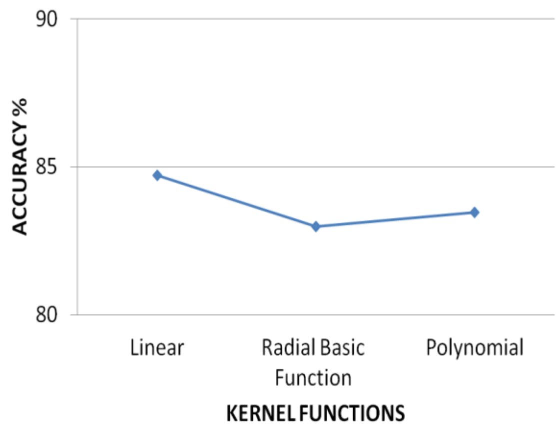

Fig. 3. *Comparison of accuracy on linear, RBF, and polynomial kernel functions.*

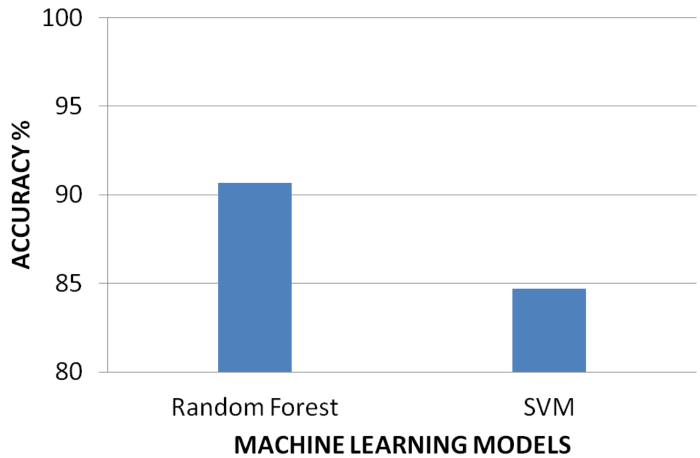

Fig. 4*. Comparison of overall accuracy of machine learning algorithms*

TABLE VI. REPRESENTS THE LEVEL OF ACCURACY OBTAINED BY ALEXNET FOR 12 CLASSES OF KTH DATASET.

Animal species	Augmented Images Accuracy %
Bear	85.12
Coyote	96.61
Deer	95.00
Elephant	93.00
Giraffe	95.23
Gorilla	95.12
Kangaroo	91.10
Leopard	93.00
Lion	87.69
Panda	98.93
Tiger	95.38
Zebra	99.20

5.2.Experiments on Deep Learning Models

5.2.1. Classification using AlexNet

AlexNet is primarily trained for semantic level image classification and tends to excerpt high-level deep depiction about images. The algorithm works on KTH dataset that is pre-trained on ImageNet. Data Augmentation is used to avoid over-fitting conflicts in the algorithm. The dimensionality of training and testing data were scaled to 256*256. As the algorithm uses pre-trained dataset, the input data directly enters the pooling layer and the output function of the AlexNet model uses Rectified linear unit(ReLU) which belongs to an unsaturated non-linear function. Also, ReLU reduces the calculations that speed up the processing time. The AlexNet model uses a learning rate of about 0.001 and the softmax layer(top-most layer) helps in fine-tuning the whole layers to increase the level of accuracy. Table VI represents the accuracy obtained by AlexNet on 12 classes of KTH dataset.

5.2.2.Classification InceptionV3

The InceptionV3 model consists of multiple layers mounted on top of each other and these layers are pre-trained on the ImageNet dataset which consists of 1.2 million images of 1000 different classes. The images within the dataset were scaled to 299*299 to fit the inception v3 model. Over-fitting is the challenging task which involves model fits too well to the training set and results in poor performance on new data arrives at the model. Data Augmentation averts over-fitting that takes place on the training set. The augmented images help to improve the recognition accuracy by providing more number of images for training. Then, the network is fine-tuned for the task of animal identification for selected species using KTH dataset. Network fine-tuning, a transfer learning concept allows transferring the network parameters and learned features from a certain domain to the specific domain. The softmax layer(top-most layer) computes the probability of various classes on the ImageNet dataset. Table VII represents the accuracy acquired by InceptionV3 algorithm for 12 classes of augmented images in KTH dataset. Figure 5 and 6 displays the comparison of accuracy between Alexnet and InceptionV3 models and the comparison overall accuracy between Machine learning and Deep learning models respectively. InceptionV3 algorithm attains a higher accuracy than SVM, RBF, and AlexNet algorithms.

TABLE VII. REPRESENTS ACCURACY OBTAINED BY INCEPTIONV3 ALGORITHM FOR 12 CLASSES OF KTH DATASET

Animal species	Augmented Images Accuracy %
Bear	97.70
Coyote	95.38
Deer	93.30
Elephant	97.00
Giraffe	99.60
Gorilla	99.10
Kangaroo	96.20
Leopard	96.38
Lion	97.69
Panda	97.93
Tiger	94.38
Zebra	98.20

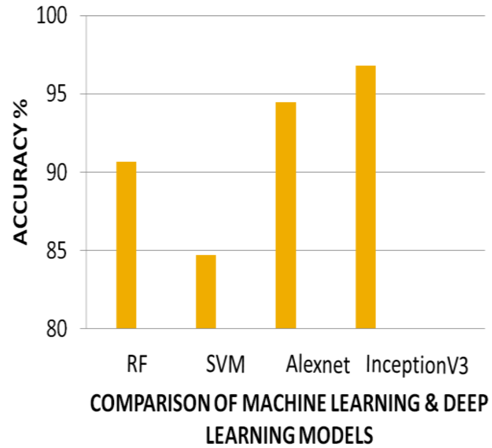

***Fig. 5**. Comparison of accuracy in animal identification between AlexNet and InceptionV3 models.*

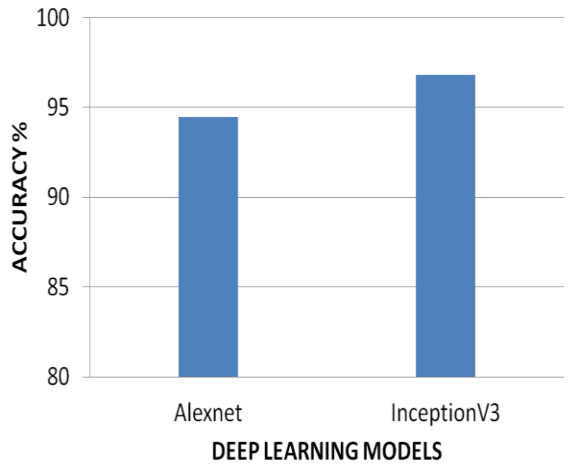

***Fig. 6**. Overall accuracy comparsion of machine learning and deep learning models*

6. CONCLUSION

In this paper, the well-known algorithms of machine learning such as SVM, Random Forest and deep learning models including Alexnet, inception V3 are compared for classification of wild animal species from KTH dataset. Among which deep learning model, inception V3 outperforms than other models and achieves better accuracy. The experiment uses KTH dataset that composed of 19 different categories of animals among which 12 classes are selected to measure the performance of the models. From the the experiment, random forest produce better results compared to SVM. Among, deep learning models inception V3 confered excellent result. However, machine learning algorithm provides good accuracy for the small dataset rather than the large dataset.

REFERENCES

[1] R.Kayas, S.Tilak, B.Kranstauber, P.A.Jansen, C.Carbone, M.J.Rowcliffe, T.Fountain"Monitoring wildlife communities with arrays of motion sensitive camera traps,"arXiv:1009.5718,2010.
[2] G.C. White and R.A.Garrot, Analysis of wildlife radio tracking data. Elseviefr, 2012.
[3] Animal Recognition System Based on Convolutional Neural Network was implemented by Tibor TRNOVSZKY, Patrik KAMENCAY and Richard ORJESEK, September 2017.
[4] G. Chen, T. X. Han, Z. He, R. Kays, and T. Forrester, "Deep convolutional neural network based species recognition for wild animal monitoring," in Proceedings of the IEEE International Conference on Image Processing (ICIP), 2014.
[5] A. Gómez, A. Salazar, and F. Vargas, "Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks," arXiv:1603.06169, 2016.
[6] A. Gomez, G. Diez, A. Salazar, and A. Diaz, "Animal identification in low quality camera-trap images using very deep convolutional neural networks and confidence thresholds," in International Symposium on Visual Computing, 2016, pp. 747–756.
[7] Mohammad Sadegh Norouzzadeh, Anh Nguyen, Margaret Kosmala, Craig Packer, Automatically identifying wild animals in camera trap images with deep learning, March 2017.
[8] Han, X.; Zhong, Y.; Cao, L.; Zhang, L. Pre-Trained AlexNet Architecture with Pyramid Pooling and Supervision for High Spatial Resolution Remote Sensing Image Scene Classification. Remote Sens. 2017, 9, 848.

[9] Szegedy, Christian, et al., "Going deeper with convolutions", Proceedings of the IEEE Conference on Computer Vision and PatternRecognition, 2015.
[10] Olga Russakovsky, Jia Deng, Hao Su, ImageNet Large Scale Visual Recognition Challenge, International Journal of Computer Vision 115(3).
[11] PornntiwaPawara, Emmanuel Okafor, Lambert Schomaker, Marco Wiering, Data Augmentation for Plant Classification, International Conference on Advanced Concepts.
[12] K.Simnoyan and A.Zisserman,"very deep convolutional networks for large scale image recognition,"arXiv: 1409.1556,2014.
[13] Zhang, W., Sun, J., Tang, X.: From tiger to panda: Animal head detection. IEEE Transactions on Image Processing 20(6), 1696–1708 (2011)
[14] Agnieszka Mikołajczyk, Michał Grochowski, Data augmentation for improving deep learning in an image classification problem, 2018 International Interdisciplinary Ph.D. Workshop (IIPhDW).
[15] Xiaoyuan Yu, Jianping Wang, Roland Kays, Patrick A Jansen, Tianjiang Wang, Thomas Huang, Automated identification of animal species in camera trap images, EURASIP Journal on Image and Video Processing,2013.
[16] Gullal Singh Cheema and Saket Anand, Automatic Detection and Recognition of Individuals in Patterned Species, Machine Learning and Knowledge Discovery in Databases, 2017.
[17] Olivier Chapelle, Patrick Haffner, and Vladimir N. Vapnik developed Support Vector Machines for Histogram based Image Classification, September 1999.
[18] Image Classification using Random Forests and Ferns was prepared by Anna Bosch, Andrew Zisserman and Xavier Munoz.
[19] D. Lowe. Distinctive image features from scale-invariant keypoints,IJCV 2004.
[20] F.Moosmann, B. Triggs, and F. Jurie. Fast discriminative visual codebooks using randomized clustering forests, NIPS2006.