An Automated Classification of Mammals and Reptiles Animal Classes Using Deep Learning

Elham Mohammed Thabit A. ALSAADI[1,2], Nidhal K. El Abbadi[3]

[1]Computer Science Department, College of Science, University of Kerbala, Kerbala, Iraq
[2]College of Information Technology, University of Babylon, Babylon, Iraq
[3]Education College, University of Kufa, Kufa, Iraq

Abstract

Detection and classification of animals is a major challenge that is facing the researchers. There are five classes of vertebrate animals, namely the Mammals, Amphibians, Reptiles, Birds, and Fish, and each type includes many thousands of different animals. In this paper, we propose a new model based on the training of deep convolutional neural networks (CNN) to detect and classify two classes of vertebrate animals (Mammals and Reptiles). Deep CNNs are the state of the art in image recognition and are known for their high learning capacity, accuracy, and robustness to typical object recognition challenges. The dataset of this system contains 6000 images, including 4800 images for training. The proposed algorithm was tested by using 1200 images. The accuracy of the system's prediction for the target object was 97.5%.

Keywords: Deep Learning, Convolutional Neural Networks, Animals Classification, Mammals, Reptiles.

تصنيف آلي لفئات الثدييات والزواحف باستخدام التعلم العميق

الهام محمد ثابت عبد الأمير السعدي[1,2]، نضال خضير العبادي[3]

[1]قسم علوم الحاسبات ، كلية العلوم ، جامعة كربلاء ، كربلاء، العراق

[2]كلية تكنولوجيا المعلومات ، جامعة بابل ،بابل ، العراق

[3]كلية التربية ، جامعة الكوفة ،النجف، العراق

الخلاصة

كشف وتصنيف الحيوانات يمثل تحديا كبيرا يواجه الباحثون. حيث ان هناك خمسة أنواع من الحيوانات الفقرية، وكل نوع يشمل الالاف من الحيوانات المختلفة. في هذه الورقة، نقترح نموذجًا جديدًا يعتمد على تدريب الشبكات العصبية التلافيفية العميقة Deep Convolutional Neural Networks لغرض اكتشاف وتصنيف فئتين من هذه الحيوانات الفقرية (الثدييات والزواحف).

الشبكات العصبية التلافيفية العميقة هي أحدث ما توصلت إليه تقنية التعرف على الصور ، وهي معروفة بقدراتها التعليمية العالية، دقتها وقوتها في مواجهة تحديات التعرف على الكائنات (الحيوانات). تحتوي مجموعة بيانات هذا النظام على 6000 صورة. استخدم النظام 4800 صورة لمرحلة التدريب، بينما الخوارزمية المقترحة تم اختبارها باستخدام 1200 صورة. نتيجة دقة تنبؤ النظام للكائن الهدف (الحيوان) كانت 97.5%.

*Email: elham.thabit@uokerbala.edu.iq

1. Introduction

Animal detection methods by Computer Vision (CV) are helpful to solve various problems such as the wildlife road accidents as well as threatened and endangered species, in addition to other purposes for animal detection and classification [1 2].

Computer Vision (CV) has become widespread in our society, with applications in image understanding, medicine, mapping, self-driving cars, and drones [3, 4]. The essence of these applications is the tasks of visual recognition, such as object detection, classification, and localization [5, 6]. Some of the key difficulties that are negatively affecting the performance of object detectors in image processing include the large differences in shapes and color appearances of diverse objects which are belonging to the same class [[7, 8]. Different body postures, of the animal body, have more variations in appearance, compared to the human body and face that are almost normal and unique [9, 10]. The variations in light intensity and lighting directions have effects on animal detection [11]. Besides, the animals belonging to the same class may differ significantly in different respects, as observed in the Birds class [12]. The objects (animals) may appear truncated or partially occluded, especially in cluttered images. Another challenge is due to the phenomenon of distortion by background clutter, illumination, partial occlusion, and geometrical transformations such as skew, rotation, and scale change [2]. Recognizing many animals requires the skill to distinguish one animal from the other which belongs to the same type. These problems raised many difficulties for humans before the introduction of machines [13]. Furthermore, the detection of animal heads has apparent impediments, as the faces of animals have bigger appearance differences compared with the faces of humans. Also, textures on the faces of animals are more complex compared with those in humans, and sometimes the animals blend with their surroundings like tall grass [14]. Lastly, some animals may change their shape and colors quickly, making it difficult to detect.

Modern developments in the neural networks have greatly advanced the performance of these recent visual recognition systems [15]. There is a need for a distinguishing model with a huge learning capacity to learn about thousands of animal objects in still images. Thus, this model must have masses of previous knowledge to compensate for all the datasets that are missing [16]. Consequently, compared to the normal feedforward neural networks with correspondingly sized layers, Convolutional Neural Networks have much fewer parameters and connections. Therefore, they will be easier to train and test, whereas their theoretically best performance is probable to be only slightly worse [17, 18].

2. Contributions of This Work

In this section, we summarize our contributions to this work. First, we propose to develop the necessary methodology for the automated detection and classification of generic object and concept classes in digital images. We develop an automated and fast algorithm that can assist in future robots in detecting vertebrate animals in still images and also classify the two main groups of vertebrates, i.e. Mammals and Reptiles. This algorithm should be able to detect and classify the animals in difficult cases such as the lighting/illumination conditions.

3. Review of Previous Works

There have been several attempts to automatically detect, classify, and recognize animals in still images:

Zhang *et al*. (2015) addressed the problems of current research on kangaroos, such as population tracking and monitoring of activities. A significant step in experimentation with the instruments is to help study kangaroos in the wild. A kangaroo picture dataset was developed to investigate feasibility from gathered information from various domestic parks throughout Queensland. A Multi-position strategy was studied and a framework was suggested to obtain reasonable detection precision by using the state-of-the-art Deformable Part Model (DPM) [19].

Trnovsky *et al*. (2017) proposed the use of CNN for input animal image classification. The proposed algorithms execute animal detection as a task of binary pattern classification, i.e., an input image is divided into blocks. Then every block is transformed into a feature. These features of the animal which belongs to a certain category have been used to train a certain classifier. Thus, when taken another new input image, the classifier will be capable to decide if the object is an animal or not [20].

Verma and Gupta (2018) focused on the monitoring and analysis of wildlife through the detection of the animal from natural scenes obtained by camera trap networks. They implemented a model of animal recognition using a self-learned feature of feature Deep CNN. Then this set of features was

used for the process of classification, using algorithms of state of the art Machine Learning (ML) that supported k-nearest neighbor, vector machine, and ensemble tree. Their results presented an accuracy of 91.4% [21].

Schneider *et al*. (2018) proposed a network to train and compare two classifiers of deep learning object detection, namely Faster YOLO v2.0 and R-CNN, to recognize, classify, and localize species of animals within images of camera-trap using datasets. The results were with average accuracies of 93.0% and 76.7% on the two datasets [22].

Yousif *et al*. (2019) enhanced CV procedures to detect and categorize the moving animal to support the primary step of camera-trap image filtering that is separating the animal detections from the empty frames and humans' pictures. They developed an algorithm with the segmentation of the foreground object during the subtraction process of background with the classification of deep learning to introduce an accurate scheme to the detection of human-animal objects [23].

4.Methodology

Deep Learning is a very good tool to solve complex problems in real-time which look as difficult to solve via human brains. There have been various types of research to improve the algorithms of deep learning to solve such a set of problems in which image processing plays a main role. The deep learning extracts the essential features from the raw data. The advantage of the deep learning algorithm is that it can understand the feature extraction mechanism on its own with no need for any separate algorithm. A convolutional neural network (ConvNet) is a class of deep learning algorithms that feed forward artificial neural networks and has effectively been used for analyzing visual imagery. In the proposed method, the classifier includes two stages, namely the training and testing. In the training stage, 4800 images were selected as visual images. While in the testing stage, 1200 images were selected. These test images are specified as input to the classifier. The knowledge was gained from a training aid to the test image which is accordingly classified into the most appropriate class. Three main stages were implemented in our model, which include ***receiving the input image, feature extraction, and classifying the specific categories.*** The corresponding output layer produces the probabilities of the animal detected in the image belonging to one of the potential categories.

5.The Proposed Architecture of Network

The proposed model involves three convolutional layers; each one is followed by layers of max-pooling. This CNN used RGB image as an input. The 1st layer employed 16 filters along with a kernel size of 3 and the Rectified Linear Unit (ReLU) as an activation function. The 2nd layer has the same arrangement, except that it was with 32 filters. The 3rd convolution layer used 64 filters. After each layer, we reduced the data using max-pooling with a pool size of 2 and a stride of 1. The ANN network design was used for detecting the two main classes of Vertebrate animals.

At first, the inputs to each neuron were multiplied by the weights, then summed with the value of bias. Then the results were passed through an activation function for producing the neuron output. The layer of classification is generally the last; it is the Softmax layer that involves the Softmax activation function. Figure- 1 shows the architecture of the proposed network.
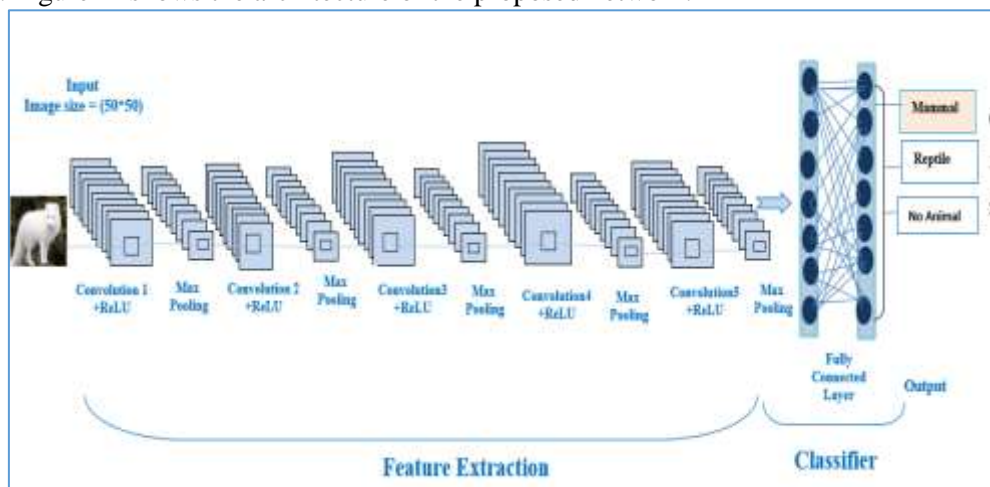

Figure 1-Proposed Architecture of CNN

The proposed model is usually trained by prediction loss minimization. Eq. (1) gives the average loss function:

$$\min_W \left\{ L(W) := \frac{1}{m} \sum_{i=1}^{m} \ell(W; x_i, y_i) + \lambda r(W) \right\}$$ …………….. Eq. (1).

where

$\{(x_i, y_i)\}$: training instances (x_i) and corresponding labels (yi)

W: network parameters to learn

$\ell(W; x_i, y_i)$: loss of network parameterized by W

r (W): regularization function

λ > 0: regularization weight

Adaptive Moment Estimation (Adam) is a stochastic optimization method designed for training deep CNN to find optimum weights during the training of the network, minimize error, and maximize accuracy.

The partial derivative of the total error about each weight was used to calculate and update weights, as shown in Eq. (2).

$$W_i = W_i - \alpha \, (\partial \, \text{Error}/\partial W_i)$$ …………………… Eq. (2).

Eq. (3) demonstrates classical gradient descent optimization as

$$W_i = W_i - \alpha. \, dW_i$$……………………………..…Eq. (3).

where ∂: a partial derivative

α > 0: learning rate (typical choices: 0.01, 0.1)

In Adam algorithm, vdw and sdw variables will be included instead of the raw partial derivatives of the error concerning the weight.

Primarily, the prime values of vdw and sdw for each weight equal 0.

Then, according to Eq. (4) and Eq. (5), vdw and sdw values will be computed for every weight.

$$vdw_i = β1. \, vdw_i + (1 - β_1). \, w_i$$………..……………. Eq. (4).

$$sdw_i = β2. \, sdw_i + (1 - β2). \, (w_i)^2$$……………………. Eq. (5).

The computed vdw and sdw will be adapted via the power of the present epoch index value.

$vdw_{i\_}$ corrected $= vdw_i / (1 - β1^{\text{epoch}})$

$sdw_{i}\_$corrected $= sdw_i / (1 - β2^{\text{epoch}})$

Lastly, according to its previous value plus vdw and sdw corrected values, each weight will be updated as:

$w_i = w_i - \alpha. \, (vdw \_\text{corrected} / (\sqrt{(sdw \_\text{corrected})} + ε))$

where: β1 = 0.9; β2 = 0.999; $ε = 10^{-8}$

The Proposed Algorithm

ConvNet was selected in this proposed algorithm, which is a distinctive type of multilayer neural networks. It aimed to identify visual patterns from pixel images with minimal pre-processing. Figure-2 illustrates the system prediction of Mammals class from three categories.

The block diagram of the proposed training and testing algorithms for detection and classification of the two categories (mammals and reptiles) is shown in Figures-(3 and 4).

The Input image is an RGB image, the dataset consists of 6000 images, including 4800 images for training and 1200 images for testing. The output of this system is to classify one of two categories. In the stage of training, the model of CNN was created, which included the following: a real image with a size of 50 x 50 was read, then filters (16, 32, 64) were used with 3 max-pooling (pool size=2) for extracting feature maps, and the rectified feature was taken as an input to produce pooled feature map. After that, all the resulted 2-D arrays from the pooled feature map were converted into a single long continuous linear vector by the flattening process to feed them as an input to the fully connected layer to classify the image. Lastly, coding of the outputs was applied, in which the animal belongs to the Mammals when the label =0, and it is a Reptile when the label =1, while a label =3 was assigned for other animals. Figure- 2 illustrates the system prediction of Mammals class from three categories. The block diagram of the proposed training and testing algorithms for the detection and classification of the two classes is shown in Figures-(3 and 4, 5).

```
Layer (type)                    Output Shape            Param #
=================================================================
conv2d_1 (Conv2D)               (None, 50, 50, 16)       208
_____
max_pooling2d_1 (MaxPooling2    (None, 25, 25, 16)       0
_____
conv2d_2 (Conv2D)               (None, 25, 25, 32)       2080
_____
max_pooling2d_2 (MaxPooling2    (None, 12, 12, 32)       0
_____
conv2d_3 (Conv2D)               (None, 12, 12, 64)       8256
_____
max_pooling2d_3 (MaxPooling2    (None, 6, 6, 64)         0
_____
dropout_1 (Dropout)             (None, 6, 6, 64)         0
_____
flatten_1 (Flatten)             (None, 2304)             0
_____
dense_1 (Dense)                 (None, 500)              1152500
_____
dropout_2 (Dropout)             (None, 500)              0
_____
dense_2 (Dense)                 (None, 3)                1503
=================================================================
Total params: 1,164,547
Trainable params: 1,164,547
Non-trainable params: 0
 Test accuracy: 0.8316239318277082
Predicting ...............................

1/1 [==============================] - 0s 165ms/step
[[9.9961841e-01 3.6404395e-04 1.7535742e-05]]
0
Reptile
```

Figure 2-System prediction of Reptiles class

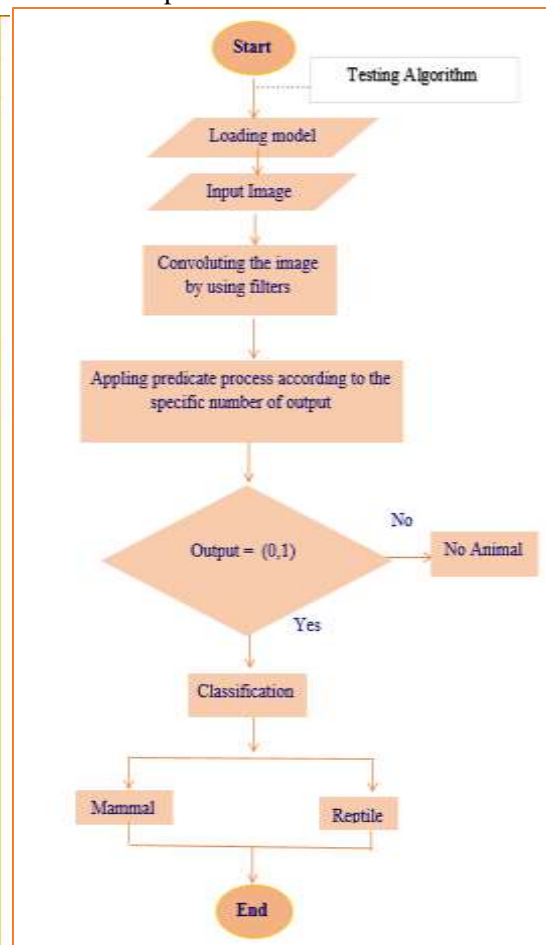**Figure 3**-Block diagram of the training algorithm

Figure 4-Block diagram of the testing algorithm

Figure 5-Sample of images of the dataset [24]

The Results and Performance Measurements

The results of our proposed network for the detection and classification of the two classes of vertebrates, when using an images size of 50x50 and several epochs that is equal to 100, are as follows:

When the test images were for Reptiles, the system predicted all these images successfully. Figure- 6 illustrates samples of testing images for Reptiles and their results.
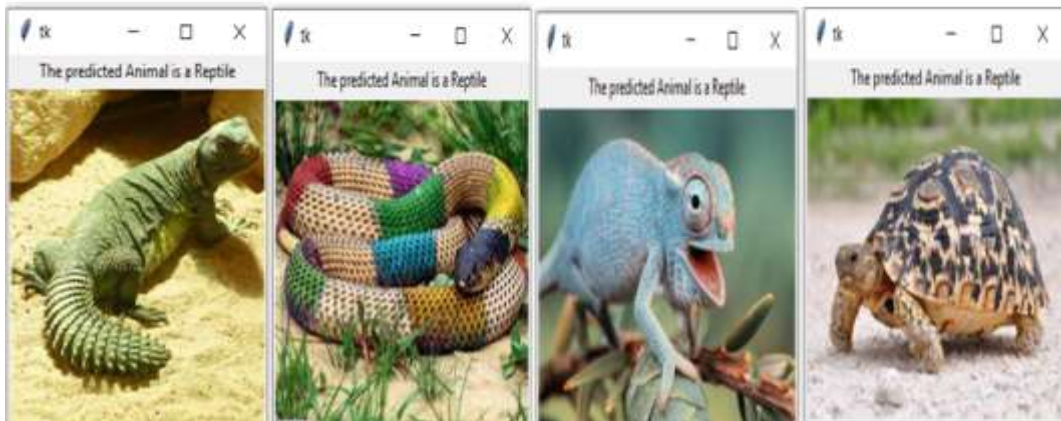

Figure 6-Samples of testing images for Reptiles.

When the test images were for Mammals, the system predicted all these images successfully. Figure- 7 illustrates samples of testing images for Mammals and their results.

Figure 7-Sample of testing images for Mammals

When the test images were neither for Reptiles nor for Mammals, the system predicted all of these images successfully as others. Figure- 8 illustrates samples of testing images of the category "Other "and their results.

Figure 8-Sample of testing images for the "Other" category

In rare cases, there were errors in object prediction in the system, as shown in Figure-9. This is because the ConvNet has to take a low trusted decision based on object size, image quality, and the number of iteration of training.

The accuracy of the system was gradually increased according to the iteration; in this architecture, it reached up to 97% as a final accuracy.
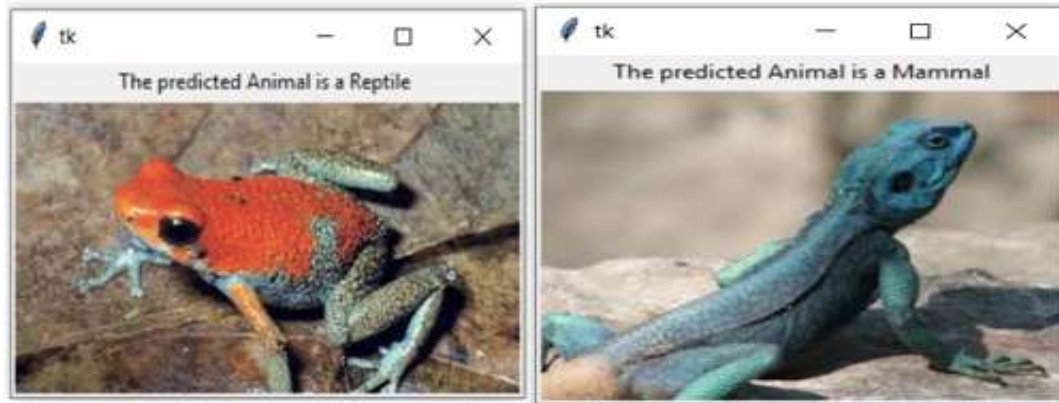

Figure 9-Sample of the wrong prediction.

We determined the best size which can be used on CNN to give the best results. Different image sizes (20x20, 40x40, 50x50, 70x70, 100x100) Pixels were used. Figure-10 illustrates the relationship between accuracy, image size, and the number of epochs. We conclude that the size of the image of 50x50 produces the best results, with many epochs of not less than 100. Increasing the number of epochs will increase accuracy, but this will also increase the testing time.
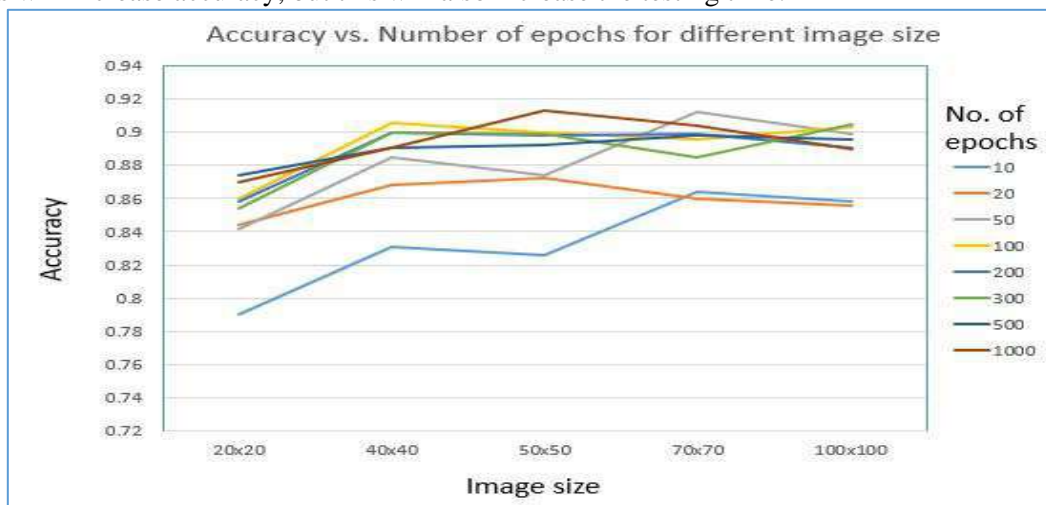

Figure 10-The relation between accuracy, image size, and number of epochs

The accuracy and efficiency of the proposed method, as shown in Table-1, is higher compared with many other similar works.

Table 1-Comparing the proposed method with other methods

No.	Research Goal	Method	Test Accuracy %
1.	Evolving an algorithm that can determine whether an individual radar image includes at least one "Tree Swallow roost" or "Purple Martin" [25].	They utilized a classical Artificial Neural Network (ANN), and a sophisticated pre-existing CNN named Inception_v3 and a shallow CNN which was constructed from scratch.	90
2.	Detecting the images which involved animals, identifying species, and counting animals [18].	They applied deep neural networks to automatically extract features from images in the largest existing labeled dataset of the wild animals called Snapshot Serengeti (SS) dataset.	93.8
3.	Autonomous detecting and monitoring of cattle with the usage of a drone [26].	Using (CNNs) to recognize the objects captured in the images.	87
4.	Detecting the large animals in the open savannah of Maasai Mara National Reserve, Kenya from very altitude resolution GeoEye -1 satellite images using hybrid image classification using [27].	Method of a hybrid image classification was utilized by incorporating the characteristics of pixels based and object-based approaches of image classification. This was executed by using an artificial neural network.	90-96
5.	Evolving an accurate system for the detection of human-animal from camera trap images through a cluttered environment as (moving objects) [28].	Introducing a new model, called Minimum Feature Difference, to model the deviation of the background of the camera trap sequences and produce the foreground object proposals. They employed then a fast scheme of deep learning classification to categorize these region proposals into three catenaries: (human, animals, and the background patches).	83.55
6.	Proposing a semi-automatic multi-object tracking method for tracking a group of unmarked zebrafish. This procedure can deal with states of partial occlusion, maintaining the correct precognition of every individual [29].	Proposing an algorithm of a multiple fish tracking with occlusion handling. Kalman Filter has been utilized to keep track of every object (animal) and to produce estimates about the region where the animal could be set in the next frame.	87.85
7.	Estimating test animal images (bear, hog, fox, deer, and wolf) using the Convolutional Neural Network (CNN) [25].	Proposing using the Convolutional Neural Network (CNN) to classify the input animal images. This approach is compared with high image recognition approaches, such as Principal Component Analysis, Local Binary Patterns Histograms, Linear Discriminant Analysis, and Support Vector Machine.	97
8.	The Proposed method detects and classifies the two categories of animals (Mammals and Reptiles).	A new model based on the training of the deep convolutional neural networks (CNN) to classify two categories of vertebrate animals (Mammals and Reptiles).	97.5

6.Conclusions

In this paper, we built an artificial convolutional neural network by deep learning to detect and classify images into two classes (Reptiles and Mammals), each class having thousands of different animals. Our network achieved very good results, where the accuracy was 97.5% and the iteration equal to 100. The system was trained by using 4800 images and tested with 1200 images.

The proposed system showed high efficiency of animal detection even in the highly cluttered natural scenes, within both day and night time. The proposed method is more efficient in animal detection when compared with other algorithms. To the best of our knowledge, this is the first work that detects and classifies animals according to the two vertebrate classes, while most of the other works focus on detecting the animals in images, and in some cases classifying few specific animals (no more than 10 different animals).

References
1. Sharma, S.U. and Shah, D. **2017**. "Design and Development of Animal Detection Algorithm Using Image Processing," 2017.
2. Hughey, L.F. and Hein, A.M. **2018**. A. Strandburg-Peshkin, and F. H. Jensen, "Challenges and solutions for studying collective animal behavior in the wild," *Philos. Trans. R. Soc. B Biol. Sci.*, **373**(1746): 20170005.
3. Liu, W., Wang, Z., Liu, N., Zeng, N. , Liu, Y. and Alsaadi, F.E. **2017**. "A survey of deep neural network architectures and their applications," *Neurocomputing*, **234**: 11–26.
4. Chen, X., Ma, H., Wan, J., Li, B. and Xia, T. **2017**. "Multi-view 3d object detection network for autonomous driving," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1907–1915.
5. Zoph, B., Vasudevan, V., Shlens, J. and Le, Q.V. **2018**. "Learning transferable architectures for scalable image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8697–8710.
6. Zhou, B., Lapedriza, A., Khosla, A., Oliva, A. and Torralba, A. **2018**. "Places: A 10 million image database for scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, **40**(6): 1452–1464.
7. Stoddard, M.C. and Osorio, D. **2019**. "Animal Coloration Patterns: Linking Spatial Vision to Quantitative Analysis," *Am. Nat.*, **193**(2):.
8. Karpestam, E., Merilaita, S. and Forsman, A. **2018**. "Size variability effects on visual detection are influenced by colour pattern and perceived size," *Anim. Behav.*, **143**: 131–138, 2018.
9. Kumar, S. and Singh, S.K. **2018**. "Monitoring of pet animal in smart cities using animal biometrics," *Futur. Gener. Comput. Syst.*, **83**: 553–563.
10. and describing wild animals in camera-trap images with deep learning Automatically identifying, counting, "Wild Animal Detection from Highly Cluttered Images Using Deep Convolutional Neural Network," *Int. J. Comput. Intell. Appl.*, **17**(04): 1850021, 2018.
11. Hollings, T., Burgman, M., van Andel, M., Gilbert, M., Robinson, T. and Robinson, A. **2018**. "How do you find the green sheep? A critical review of the use of remotely sensed imagery to detect and count animals," *Methods Ecol. Evol.*, **9**(4): 881–892.
12. Schwartz, E. *et al.* **2018**. "RepMet: Representative-based metric learning for classification and one-shot object detection," *arXiv Prepr. arXiv1806.04728*.
13. Joly *et al.* A. **2017**. "Lifeclef 2017 lab overview: multimedia species identification challenges," in *International Conference of the Cross-Language Evaluation Forum for European Languages*, 2017, pp. 255–274.
14. Deslauriers, J., Toth, M., Der-Avakian, A. and Risbrough, V.B. **2018**. "Current status of animal models of posttraumatic stress disorder: behavioral and biological phenotypes, and future challenges in improving translation," *Biol. Psychiatry*, **83**(10): 895–907, 2018.
15. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H. **2018**. "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 801–818.
16. Kheradpisheh, S.R., Ganjtabesh, M., Thorpe, S.J. and Masquelier, T. **2018**. "STDP-based spiking deep convolutional neural networks for object recognition," *Neural Networks*, **99**: 56–67.
17. Badre, P., Bandiwadekar, S., Chandanshive, P., Chaudhari, A. and Jadhav, M.S. **2018**. "Automatically Identifying Animals Using Deep Learning," *Int. J. Recent Innov. Trends Comput.*

Commun., **6**(4): 194–197.

18. M. S. Norouzzadeh *et al.*, M.S. **2018**. "Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning," *Proc. Natl. Acad. Sci.*, **115**(25): E5716–E5725.

19. Zhang, T., Wiliem, A., Hemsony, G. and Lovell, B.C. **2015**. "Detecting kangaroos in the wild: the first step towards automated animal surveillance," in *2015 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2015, pp. 1961–1965.

20. Trnovský, T., Kamencay, P., Orješek, R., Benčo, M. and Sýkora, P. **2017**. "Animal recognition system based on convolutional neural network,".

21. Verma, G.K. and Gupta, P. **2018**. "Wild animal detection using deep convolutional neural network," in *Proceedings of 2nd international conference on computer vision & image processing*, 2018, pp. 327–338.

22. Schneider, S., Taylor, G.W. and Kremer, S. **2018**. "Deep learning object detection methods for ecological camera trap data," in *2018 15th Conference on Computer and Robot Vision (CRV)*, 2018, pp. 321–328.

23. Yousif, H., Yuan, J., Kays, R. and He, Z. **2019**. "Animal Scanner: Software for classifying humans, animals, and empty frames in camera trap images," *Ecol. Evol.*.

24. http://www.vision.caltech.edu/Image_Datasets/Caltech101/ (Accessed 11-10-2019).

25. McGovern, A., Chilson, C., Avery, K., Bridge, E., Sheldon, D. and Kelly, J. **2018**. "Automated detection of bird roosts using NEXRAD radar data and Convolutional neural networks,".

26. Rivas, A., Chamoso, P., González-Briones, A. and Corchado, J. **2018**. "Detection of cattle using drones and convolutional neural networks," *Sensors*, **18**(7): 2048, 2018.

27. Yang, Z., Wang, T., Skidmore, A.K., de Leeuw, J., Said, M.Y. and Freer, J. **2014**. "Spotting east African mammals in open savannah from space," *PLoS One*, **9**(12): e115989.

28. Yousif, H., Yuan, J., Kays, R. and He, Z. **2018**. "Object detection from a dynamic scene using joint background modeling and fast deep learning classification," *J. Vis. Commun. Image Represent.*, **55**: 802–815.

29. de Oliveira Feijó, G., Sangalli, V.A., da Silva, I.N.L. and Pinho, M.S. **2018**. "An algorithm to track laboratory zebrafish shoals," *Comput. Biol. Med.*, **96**: 79–90.