Animal Intrusion Detection Using Deep Learning

Varshini Sri S, First Author

Bannari Amman Institute Of Technology, Sathyamangalam

[1]varshinisri.ad20@bitsathy.ac.in

Shakthi Sree R C, Second Author

Bannari Amman Institute Of Technology, Sathyamangalam

[2]shakthisree.ad20@bitsathy.ac.in

Abstract— **One of the researcher's interests and challenges is animal identification techniques. There are several challenges that researchers in this field confront that limit detection effectiveness and efficiency, such as picture illumination variation, animal occlusion, colour similarity of animal colours with backdrop environment, and so on. The purpose of this study is to detect and classify multi-label images of mammals, which we propose to do using the Single Shot Multi-Box Detector (SSD) and the MobileNet v1 coco_2017 model. Another objective is to locate and identify several items (animals) from the Mammal category in digital photos. Based on deep learning technology, the recommended SSD is regarded as a more accurate, rapid, and efficient technique to recognise objects of various sizes. We utilised 2000 pictures in the network taken from standard datasets (such as Caltech 101) and the net in this proposal. The SSD framework enhances Convolution Neural Network (CNN) detection and identification operations. During the prediction phase, the network assigns scores to the existence of each object class and draws a box around each object in the picture. Each box includes a name that defines the kind of object, and the score denotes the likelihood of the object's association to that category. During the procedure, boxes are changed to get the best fit to the form of the item. The experimental findings of this work demonstrated the efficacy of identifying and detecting animals even when light, position, and occlusion were varied. The detection and classification accuracy can reach 98.7%. Unlike other comparable efforts, this recommendation is more dependable and accurate, and it identifies a wide variety of Mammals species.**

Keywords— **MobileNet , Convolution Neural Network (CNN), Single Shot Multi-Box Detector (SSD)**

I. Introduction

Visual monitoring in animal settings is now one of the most prominent study topics in the field of Computer Vision (CV), yet methods for identifying and interpreting dynamic objects remain unavailable [1], [2]. Animal detection technologies can assist solve a variety of problems, such as preventing dangerous animal entry in a residential environment [3]. It is also a big part of a robot's job to recognise and categorise items in crucial scenarios such as natural catastrophes or other calamities that need diverse objects such as human, animal, and other [4]. Another example is self-driving automobiles, where successfully identifying people, animals, street signs, or other vehicles is a critical component for maximising system safety.
It also adds significant utility in biomedical laboratories by monitoring laboratory animals with a powerful following approach capable of extracting a rodent from a frame in an uncontrolled setting [5]. Furthermore, it improves laboratory worker efficiency and production by minimising the time spent directly watching animals and gaining a better knowledge of animal behaviour. Deep (CNN) or ConvNet has achieved significant success in the field of computer vision, such as target recognition, target tracking, picture classification, and semantic image segmentation [6]. It is a multi-layered neural network with a unique design for detecting complicated data properties. item detection is the prediction of the position and category of an item in static pictures, which is one of the most difficult computer vision issues [7], [8]. It frequently use extracted attributes and learning algorithms to recognise things in a static image that correspond to a certain category of objects [9]. There are several forms of CNNs; MobileNet is one of the most important applications for this sort of CNN network [10]. When compared to a regular CNN with the same depth, the design of "Depthwise Separable Convolutions" considerably reduces the number of parameters. It is a multi-layered neural network with a one-of-a-kind architecture that detects complex data attributes. One of the most difficult computer vision problems is item detection, which is the prediction of an object's position and category in static images [7], [8]. It often use extracted properties and learning algorithms to identify items in a static image that belong to a specific category of objects [9]. There are several types of CNNs; one of the most important applications for this type of CNN network is MobileNet [10]. When compared to a standard CNN of the

same depth, the design of "Depthwise Separable Convolutions" reduces the number of parameters significantly.

During prediction, the network assigns scores to the existence of each object type in each default box. It also modifies these boxes in order to acquire the best fit of the object's form. The present network then aggregates predictions from multi-feature maps of varying sizes for dealing with objects of varying volumes.

1.1. Research Problem

Animal detection and recognition is still a difficult problem, and there is no one approach that gives a powerful and efficient answer to all scenarios. Auditory cues to create a robust system capable of identifying various animal species. To summarise some of these problems, consider the following:

1. The wide variety in the look and size of animals within a given group.
2. Lighting / Illumination Conditions: The image colour is particularly responsive to fluctuations in light intensity and light direction.
3. Detect and categorise creatures in arbitrary poses in crowded and obstructed environments, as well as in rotational states.

1.2. Paper Contributions

This section presents the following contributions of this paper:

1. Detecting single and multi-animals from the Mammals category, as well as no animal category if no mammal animal is present in the image.
2. Detecting all types of animals in the category.
3. Detecting and categorising animals in images that are fuzzy or dark due to variations in picture light/illumination.

2. Related Work

This section summarises the existing relevant efforts on animal categorization and tracking. Norouzzadeh et al. (2018) trained Deep Convolutional Neural Networks (DCNNs) on 3.2 million photos (Dataset Snapshot Serengeti) to recognise, count, and characterise the behaviours of 48 species. This neural network correctly identified animals with 93.8% accuracy. Their findings showed that Deep Learning (DL) permits the affordable, wide-scale, unobtrusive, and real-time collection of wealth information pertaining to vast numbers of wild animals. They collected millions of labelled data from the Snapshot Serengeti (SS) dataset, the latest development in (DNN) engineering, and modern supercomputing to test how well DL can automatic information elicitation from camera trap images [13].

3. SSD MobileNet Pre-Trained Model

One of the pre-trained models is the MobileNet SSD model, which combines the SSD with the MobileNet model [18]. SSD is a prominent object detecting method, and Mobilenet is a network. To generate high-level features, a convolutional neural network was employed as a features extractor [19]. It is trained on public datasets such as the Common Objects in Context (COCO) dataset to detect objects as well as multi-object classification. The SSD MobileNet model's architecture is lightweight. design that is more suited for mobile and embedded vision applications when there is little space a shortage of computer power. Google proposed this architecture. It employs depthwise separable convolutions, which means it executes a separate convolution on each colour (input) channel rather than combining and flattening all three. The depthwise convolution's outputs are then combined using an 11 convolution by the pointwise convolution. Because of a shortage of computer capacity, this factorization has the effect of dramatically lowering computation and model size.

This sort of convolution separates the output into two layers: a filtration layer and a combination layer. The combination of these two layers reduces the size of the model, lowering the computing power required. A pre-trained model transfers learnt features or weights to start the tuning process, allowing the object detector to be trained with a short amount of training. It is a highly widespread and powerful methodology for deep learning of tiny picture datasets that uses a pre-trained network

4. SSD MobileNet Architecture

SSD's CNN is completely convolutional, with MobileNet serving as the backbone network [22]. Following that, the size of numerous more CONV layers steadily decreases [23]. The SSD employs extra shoaly layers with improved precision to identify tiny things. SSD recognises several metrics for objects of varying sizes by working on multiple convolution feature maps, each of which contains the necessary bounding boxes that predict hundreds of categories and box offsets [24], [25]. In SSD, a score is calculated for the priority of each item category within each bounding box, followed by changing the bounding box to best reflect the form of the object before detection. SSD achieves its purpose by using a multi-tasking loss function that depicts the difference between expected and real values. It is designed to obtain the bare minimum of that function in order to optimise the model and improve the accuracy of forecasts. SSD is a CNN that uses feed-forward learning. It links each cell in the prediction feature maps with a set of default bounding boxes (also known as anchors) [24]. The method predicts offsets in the cell relative to a default box, as well as a confidence score expressing the existence of a target item class within the default box . To distinguish the bounding box detectors, each detector attempts to anticipate only one item, and different detectors will locate different things. SSD allocates the detector of each bounding box to a specific place in the picture. Detectors learn to specify items at certain places in this way. Figure 1 depicts a MobileNet SSD model.

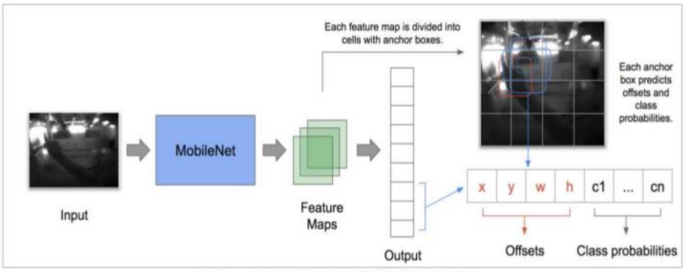
Each feature map is divided into cells with anchor boxes.
Each anchor box predicts offsets and class probabilities.
Input
MobileNet
Feature Maps
Output
x
y
w
h
c1
...
cn
Offsets
Class probabilities

Figure 1: MobileNet SSD Model [26]

5. Research Methodology

The multi-label image classification approach is presented for recognising and categorising numerous objects (animals) in static photos from the Mammal category. The SSD MobileNet v1 model is employed in this study, and Figure 2 depicts the suggested model's block diagram.

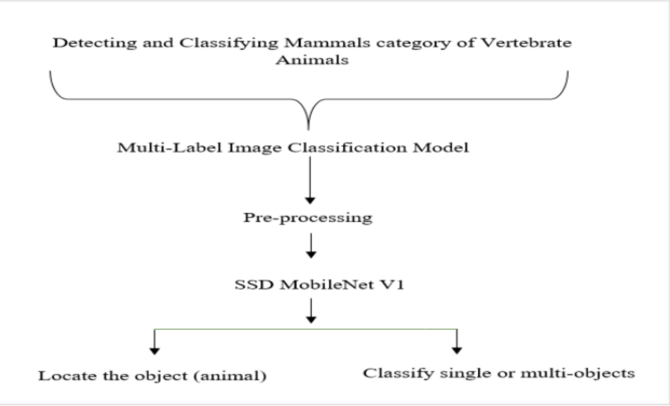
Detecting and Classifying Mammals category of Vertebrate Animals
Multi-Label Image Classification Model
Pre-processing
SSD MobileNet V1
Locate the object (animal)
Classify single or multi-objects

Figure 2: Block diagram of the proposed model

The most frequent input picture data parameters are the number of images, the number of channels, the image height, and the image width. Before using pictures in deep learning, many phases of image pre-processing need be completed:

a. Image scaling with a standard aspect ratio

b. Normalise the input picture to the range (0, 1) and lower the image dimension to fit the SSD network's specified range value and size.

c. Image Enhancements Techniques are employed in this study to address some of the issues associated with the identification and classification of animals in static photos, such as when the animal seems truncated or partially obscured, when the creatures change form, and the problem of Lighting/Illumination Conditions.

5.2 Proposed Method of Multi-Label Image Classification

5.2. Proposed Multi-Label Image Classification Method

The suggested Multi-Label Image Classification approach is utilised for the following:

1. Classification + Localization: Classifying an image as a mammal or no animal category, as well as localising the target object inside the picture to find the primary item in an image.

2. Object Detection: Detecting and drawing bounding boxes around all mammal kinds. Figure 3 depicts the method's localization, classification, and detection of two samples of our results.

The SSD MobileNet v1 model is utilised in this study to recognise and categorise several (objects) animals of the mammal class, even in tough scenarios such as detecting part of the animal under lighting/ illumination settings. For recognising and categorising numerous items in tough scenarios, MobileNets and Single Shot Multibox Detection (SSD) are integrated and pre-trained using the COCO_2017 dataset. It is an efficient CNN architecture for mobile and embedded vision applications. The SSD MobileNet v1 model is utilised in this study to recognise and categorise several (objects) animals of the mammal class, even in tough scenarios such as detecting part of the animal under lighting/ illumination settings. For recognising and categorising numerous items in tough scenarios, MobileNets and Single Shot Multibox Detection (SSD) are integrated and pre-trained using the COCO_2017 dataset. It is an efficient CNN architecture for mobile and embedded vision applications. To match the criteria of the SSD mobile net model, this model pre-processes the photos by resizing the input image sizes into 300300 pixels. This increased picture size is employed in the detecting procedure.

Localising and categorising several objects in still photos is what detection does.

The proposed model's outputs are four matrices that map to the indices 0 - 3 (Locations, Classes, Scores, and Number of detections), as shown in table 1. During network training, 20 epochs were chosen.

5.3. Algorithm of SSD Detection

The SSD algorithm stages employed in this study are as follows:

RGB picture as input.

Output: Represent four matrices (Locations, Classes, Scores, and Number of detections) that are mapped to the indices 0 - 3.

1. Use CONV Layers to extract features at various sizes using various filters to generate Feature Maps.
2. Output numerous feature maps of varying sizes to the supplied SSD.
3. Create MultiClass Classification and Bounding Box Regression from each spatial point in Feature Maps.
4. Continue the detection stages by adding CONV layers to produce smaller feature maps.
5. Use IoU metrics and hard negative mining to filter the bounding boxes.
6. Calculate the loss by combining classification (softmax) with detection (smooth L1).

5.4. Proposed Model Architecture

The Single Shot MultiBox Detector MobileNet model's architecture is built on (Depthwise Separable Convolutions), which is separated into two CONV layers, one for filtering and the other for merging. To begin feature extraction, the MobileNet model implements a single default filter for each neural input channel. A (1 1) Pointwise Convolution follows the Depthwise Convolution to integrate the Depthwise Convolution result. The batch norm follows all of these separable layers, and ReLU nonlinearity expects the final (FC) layer that feeds into a softmax layer to be nonlinear.

Unlike classic CNN, Mobilenet filters function for each colour channel individually and then aggregate the three outputs into one value. This factorization results in significant reductions in processing and model size.

In this suggested model, (32, 64) filters of size (5 x 5) are utilised to extract features from input pictures, followed by 2 max-pooling (pool size=2). Figure 4 depicts the suggested SSD model. MobileNet

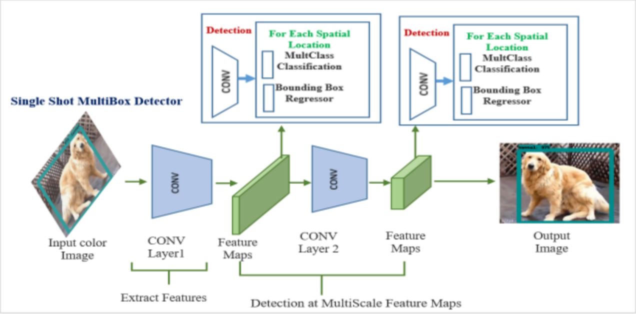

Figure 4: Proposed model SSD MobileNet V1

Many convolutional layers have been added to the model. The priority of building a bounding box in SSD is to the target item with a high score. The bounding box is then adjusted based on the position, size, and aspect ratios to obtain the best fit to the forms of the object (to the ground truth boxes).

During training, each feature map is utilised to forecast bounding boxes, and the variation in the size of the feature map allows for object recognition with varying accuracy. IoU metrics and Hard Negative Mining are used to filter boxes. IoU is an excellent statistic for measuring the overlap between the predicted box and the ground truth, as demonstrated in figure 5. The ideal value between them has a 100% IoU, yet a result greater than 50% is typically regarded a correct forecast. Finally, the projected box will be found having the greatest overlap with the ground truth.

$$IoU = \frac{area\ of\ overlap}{area\ of\ union} \text{...(1)}$$

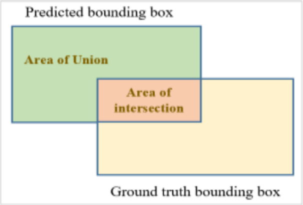

Figure 5: Intersection over union (IOU)

Loss Function is used in this model which is the weighted sum of the loss of confidence "L_{conf}" and loss of localization "L_{loc}", as shown in equation 2.

$$L(x, c, l, g) = \frac{1}{N}(L_{conf}\ (x, c) + aL_{loc}\ (x, l, g)) \text{.. (2)}$$

Where:

L_{conf}: loss of confidence
L_{loc}: loss of localization
N: number of matched default bounding boxes.
α: the weight for the localization loss.
l: predicted box
g: ground truth box
c: class score
X: Input image

6. Experiment Results

This model's input images are RGB, and the dataset contains 2000 distinct images. Each category (mammals and no animal) received 1000 photos. The follow 6.1. Detection and

6.1. Classification Single Animal from Different Type of Animals:

In this exam, 1000 photographs of animals were chosen, including different sorts of mammals such as a dog, cat, deer, bear, rabbit, and so on. This suggested model effectively recognised and categorised all sorts of mammal groups in photos and bordered them with a box, as illustrated in figure 6:

6.2. Detection and Classification Single Object of no animal Category:

In this test, 1000 photos from the no animal category were utilised, which included various sorts of things from the no animal category. As demonstrated in figure 7, our model correctly recognised and categorised items of this category in photos and bordered them with a box.

6.3.Detection and Classification Multi animals of Mammals Category

In this test, the model correctly recognised and categorised many similar and dissimilar mammals in one image and bounded them with a box, as seen in figure 8.

6.4. Detection and Classification Parts of Animals in Image

In this test, the model has been detected and classified parts of animals displayed in the image successfully, figure 10 shows samples of the detected cut parts of animals.

6.5. Detection and Classification the Animals from Dark Images

7. Comparison Between the Proposed Work Against the Others

Most studies only identify certain sorts of animals within a given group, rather than all types of creatures within that category. These investigations also do not detect a section of a (object) animal in a picture, nor do they detect objects in dark photos or under different lighting settings. The majority of them did not acquire photographs; instead, they used the available ImageNet collection. depicts a comparison of the proposed work to the other works.

Creating a detector that evaluates video images from camera traps in real time. These photos depict "rhinoceros, humans, and a group of six common large animals in the African savannah."CNN " SSD MobileNet V2" is being used. The precision is 90%.

Animal detection, localization, and classification in images. Also, tackle the challenges of detecting animals in the event of a portion of an animal in an image, as well as diverse lighting environments in an image. The suggested technique employed "CNNs training in conjunction with pre-trained Single Shot Detector (SSD) MobileNet v1 architecture." The precision is 97.8%.

8. Conclusions:

The SSD and Mobilenet-v1 are demonstrated in this paper as an accurate and quick approach for detecting, classifying, and localising animals in static photos. Most other studies only detect particular forms of a single category but do not detect As we did in the present proposal, we included all species of animals in the Mammals group. Several tests have been performed on the SSD MobileNet v1 model, and the model has performed well. The object (animal) is detected and classified as a result. The results were really effective, and The detection accuracy utilising 20 epochs increased to 98.7%, which is encouraging when compared to previous efforts.

In static photos, our model correctly spotted and categorised single and numerous items. Furthermore, this model correctly spotted and categorised the animal in a variety of lighting conditions, including dark and fuzzy photos. When there is a portion of an animal in the image, it can detect and categorise it. In addition to labelling each object in the image, it was recognised and delimited by boxes.

References :

[1] Z. He et al., “Visual informatics tools for supporting large-scale collaborative wildlife monitoring with citizen scientists,” IEEE Circuits Syst. Mag., vol. 16, no. 1, pp. 73–86, 2016.

[2] A. J. Shepley, G. Falzon, P. Meek, and P. Kwan, “Location Invariant Animal Recognition Using Mixed Source Datasets and Deep Learning,” bioRxiv, 2020.

[3] C. Zhu, T. H. Li, and G. Li, “Towards automatic wild animal detection in low-quality cameratrap images using two-channeled perceiving residual pyramid networks,” in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2860–2864.

[4] H. Dong, G. Sun, W.-C. Pang, E. Asadi, D. K. Prasad, and I.-M. Chen, “Fast ellipse detection via gradient information for robotic manipulation of cylindrical objects,” IEEE Robot. Autom. Lett., vol. 3, no. 4, pp. 2754–2761, 2018.

[5] A. C. Coelho and J. García Díez, “Biological risks and laboratory-acquired infections: a reality that cannot be ignored in health biotechnology,” Front. Bioeng. Biotechnol., vol. 3, p. 56, 2015.

[6] W. Wang, Y. Li, T. Zou, X. Wang, J. You, and Y. Luo, “A Novel Image Classification Approach via Dense-MobileNet Models,” Mob. Inf. Syst., vol. 2020, 2020.

[7] G. Rogez, P. Weinzaepfel, and C. Schmid, “Lcr-net: Localization-classification-regression for human pose,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3433–3441.

[8] S. Tang and Y. Yuan, “Object detection based on convolutional neural network,” in International Conference-IEEE–2016, 2015.

[9] E. M. T. A. ALSAADI, “Auto Animal Detection and Classification among (Fish, Reptiles and Amphibians Categories) Using Deep Learning.”

[10] A. G. Howard and M. Zhu, “Mobilenets: Open-source models for efficient on-device vision,” Google AI blog, 2017.

[11] L. Leal-Taixé and S. Roth, Computer Vision–ECCV 2018 Workshops: Munich, Germany, September 8-14, 2018, Proceedings, Part VI, vol. 11134. Springer, 2019.

[12] A. Tydén and S. Olsson, “Edge Machine Learning for Animal Detection, Classification, and Tracking.” 2020.

[13] M. S. Norouzzadeh et al., “Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning,” Proc. Natl. Acad. Sci., vol. 115, no. 25, pp. E5716– E5725, 2018.

[14] A. Joly et al., “Lifeclef 2017 lab overview: multimedia species identification challenges,” in International Conference of the Cross-Language Evaluation Forum for European Languages, 2017, pp. 255–274.

[15] P. Badre, S. Bandiwadekar, P. Chandanshive, A. Chaudhari, and M. S. Jadhav, “Automatically Identifying Animals Using Deep Learning,” Int. J. Recent Innov. Trends Comput. Commun., vol. 6, no. 4, pp. 194–197, 2018.

[16] X. Liu, Z. Jia, X. Hou, M. Fu, L. Ma, and Q. Sun, “Real-time Marine Animal Images Classification by Embedded System Based on Mobilenet and Transfer Learning,” in OCEANS 2019-Marseille, 2019, pp. 1–5.

[17] D. Roy, P. Panda, and K. Roy, “Tree-CNN: a hierarchical deep convolutional neural network for incremental learning,” Neural Networks, vol. 121, pp. 148–160, 2020.

[18] A. G. Howard et al., “Efficient convolutional neural networks for mobile vision applications,” arXiv Prepr. ArXiv1704.0486, 2017.

[19] Z.-Q. Zhao, P. Zheng, S. Xu, and X. Wu, “Object detection with deep learning: A review,” IEEE Trans. neural networks Learn. Syst., vol. 30, no. 11, pp. 3212–3232, 2019.

[20] E. Suharto, A. P. Widodo, and E. A. Sarwoko, “The use of mobilenet v1 for identifying various types of freshwater

fish," in Journal of Physics: Conference Series, 2020, vol. 1524, no. 1, p. 12105.

[21] S. GHOURY, C. SUNGUR, and A. DURDU, "Real-Time Diseases Detection of Grape and Grape Leaves using Faster R-CNN and SSD MobileNet Architectures."

[22] R. Verma and C. Arora, "Modeling and implementation of real-time animal detection module for mobility assistant for visually impaired (MAVI) system." 2017.

[23] W. Liu et al., "Ssd: Single shot multibox detector," in European conference on computer vision, 2016, pp. 21–37
Using Mixed Source Datasets and Deep Learning," bioRxiv, 2020.

[24] J. L. Masache Narvaez, "Adaptation of a Deep Learning Algorithm for Traffic Sign Detection," 2019.

[25] A. Younis, L. Shixin, S. Jn, and Z. Hai, "Real-Time Object Detection Using Pre-Trained Deep Learning Models MobileNet-SSD," in Proceedings of 2020 the 6th International Conference on Computing and Data Engineering, 2020, pp. 44–48.

[26] R. Pandey, M. White, P. Pidlypenskyi, X. Wang, and C. Kaeser-Chen, "Real-time Egocentric Gesture Recognition on Mobile Head-Mounted Displays," arXiv Prepr. arXiv1712.04961, 2017.

[27] Z. Yang, T. Wang, A. K. Skidmore, J. de Leeuw, M. Y. Said, and J. Freer, "Spotting east African mammals in open savannah from space," PLoS One, vol. 9, no. 12, p. e115989, 2014.

[28] H. Nguyen et al., "Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring," in 2017 IEEE international conference on data science and advanced analytics (DSAA), 2017, pp. 40–49.

[29] S. Matuska, R. Hudec, P. Kamencay, M. Benco, and M. Zachariasova, "Classification of wild animals based on SVM and local descriptors," AASRI Procedia, vol. 9, pp. 25–30, 2014.

[30] A. Rivas, P. Chamoso, A. González-Briones, and J. Corchado, "Detection of cattle using drones and convolutional neural networks," Sensors, vol. 18, no. 7, p. 2048, 2018.

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

[32] T. Trnovszký, P. Kamencay, R. Orješek, M. Benčo, and P. Sýkora, "Animal recognition system based on convolutional neural network," 2017. [33] S. Kumar and S. K. Singh, "Monitoring of pet animal in smart cities using animal biometrics," Futur. Gener. Comput. Syst., vol. 83, pp. 553–563, 2018