# Leak detection of gas pipelines using acoustic signals based on wavelet transform and Support Vector Machine

Gudithoti Chendra Sekhar

Department of Computer Science, Sri Venkateswara University, Tirupati

***Abstract—*** *Over the last few years, there has been a steady rise in number of reported human–animal conflicts. While there are several reasons for increase in such conflicts, foremost among them is the reduction in forest cover. Animals stray close to human settlements in search of food, and often end up raiding crops or preying on cattle. There are at times human causalities as well. Proficient, reliable, and autonomous monitoring of human settlements bordering forest areas can help reduce such animal–human conflicts. A broad range of techniques in computer vision and deep-learning has shown enormous potential to solve such problems. In this paper, a novel, efficient, and reliable system is presented which automatically detects wild-animals using computer vision. The proposed method uses the YOLO object detection model to ascertain presence of wild animals in images. The model is fine-tuned for identifying ten different types of animals (Dog, Horse, Butterfly, elephant, Hen, Cat, Cow, Sheep, Spider, and Squirrels). The proposed method achieves an accuracy of 98.8% and 99.8% to detect animals.*

## I. INTRODUCTION

In the recent few years, diverse research work happened to develop a practical approach to accelerate the development of deep learning methods. Numerous developments accomplished excellent results and followed by continuous reformations in deep learning procedures. Object localization is the identification of all the visuals in a photograph, incorporating the precise location of those visuals. By using deep learning techniques for object identification and localization, computer vision has reached a new zenith. Due to significant inconstancies in viewpoints, postures, dimensions, and lighting positions, it is challenging to succeed in the identification of objects perfectly. Accordingly, considerable concern has been given by researchers to this area in the past few years. These techniques create region proposal networks (RPN), and then the region proposals are divided into categories afterward. On the other side, object detection algorithms using regression includes SSD and YOLO, etc. These methods also generate region proposal networks (RPN) but divide these region proposals into categories at the moment of generation. All of the procedures mentioned above have significant accomplishments in object localization and recognition. YOLO consolidates labels in diverse datasets to form a tree-like arrangement, but the merged labels are not reciprocally exclusive.

Computer vision-based methods are being used increasingly as tools to assist wild animal object recognition. The ability to identify individual animals from images enables population surveys through sight-resight statistics and forms the basis for demographic studies. The pipeline of processing for animal recognition includes several stages, starting with the detection of animals in images and ending with identification decisions. By making all stages of this pipeline more reliable and automated, animal identification studies can be increased in spatial and temporal resolution, provide better conservation statistics, and – importantly – allow citizens without specialized training to participate in engaging census data collection events [1, 5, 23, 26]. Computer vision is one such technology which could potentially solve most of the associated problems. Use of deep learning methods to classify images that contain entities of in-tersest are gaining popularity. Deep Neural Networks (DNN) is known to be accurate, and outperform all other existing methods in the task of image classification. Krizhevsky et al., who submitted the winning entry for the ImageNet classification challenge, introduced a Deep Neural Network based solution for image classification. It is now considered a landmark achievement in computer vision, and has contributed to increased research in the field. The main intent of this paper is to describe the design for a computer vision system, capable of detecting animals. DNN's could be leveraged to detect the presence of animals in the captured images. In addition to detecting the presence of an animal, in order to effectively track them and monitor their actions, it is also necessary to localize the animals within the image. This is the task of object detection. Object detection systems predict regions of interest within images, and in addition classify entities within these regions. Thus, object detection is the ideal choice for the system proposed in this paper. This paper introduces a novel method of reducing human animal conflicts, through constant and automatic monitoring of vulnerable areas using a system of cameras. The proposed solution is accurate and cost effective and to an extent, can be customized specifically for a particular region. Section 2 presents notable existing research work in the area. Section 3 describes the design of the proposed system and highlights the role of various components. Section 4 summarizes the important results of the study, followed by a brief discussion and scope for future enhancements in section.

### 1.1 Artificial intelligence:

Artificial intelligence (AI) is the ability of a computer program or a machine to think and learn. It is also a field of study which tries to make computers "smart". As machines become increasingly capable, mental facilities once thought to require intelligence are removed from the definition. AI is an area of computer sciences that emphasizes the creation of intelligent machines that work and reacts like humans. Some of the activities computers with artificial intelligence are designed for include: Face recognition, Learning, Planning, Decision making etc.,

Artificial intelligence is the use of computer science programming to imitate human thought and action by analysing data and surroundings, solving or anticipating problems and learning or self-teaching to adapt to a variety of tasks.

### 1.2 Object Detection and Tracking

There is a wide range of computer vision tasks benefiting society such as object classification, detection, tracking, counting, Semantic Segmentation, Captioning image, etc. Process of identifying objects in an image and finding its position is known as object detection. Various object detection tasks. With advancements in field of computer vision assisted by AI, realization of tasks was realizable along t time scale. Semantic segmentation task of clustering pixels based on similarities. Classification + Localization and object detection method of identifying class of object and drawing a bounding box around it to make it distinct. Instance segmentation is semantic segmentation applied to multi objects. The general intuition to perform the task is to apply DNN over the image. DNN works on image patches to carry out the task many such salient regions can be obtained by Region-Proposal Networks like Region Convolution Neural network (RCNN), Fast- Region Convolutional Neural Network (Fast-RCNN), Faster- Region Convolutional Neural Network (Faster-RCNN). To perform selective search for object recognition Hierarchal Grouping Algorithm is used. Few bottlenecks by these approaches are mitigated by state-of the-art algorithms like You Only Look Once (YOLO), Single shot Detector (SSD). The efficient object detection algorithm is one which assures to give bounding box to all objects of vivid size to be recognized, with great computational capabilities, faster processing. YOLO and SSD assure to render promising results, but have a tradeoff between speed and accuracy. Hence, selection of algorithm is application specific.

### 1.3 Deep Neural Networks (DNN)

A deep neural network (DNN) is an artificial neural network (ANN) with multiple layers between the input and output layers. There are different types of neural networks but they always consist of the same components: neurons, synapses, weights, biases, and functions. These components functioning similar to the human brains and can be trained like any other ML algorithm. DNNs can model complex non-linear relationships. DNN architectures generate compositional models where the object is expressed as a layered composition of primitives. The extra layers enable composition of features from lower layers, potentially modelling complex data with fewer units than a similarly performing shallow network. For instance, it was proved that sparse multivariate polynomials are exponentially easier to approximate with DNNs than with shallow networks. Deep architectures include many variants of a few basic approaches. Each architecture has found success in specific domains. It is not always possible to compare the performance of multiple architectures, unless they have been evaluated on the same data sets. DNNs are typically feedforward networks in which data flows from the input layer to the output layer without looping back. At first, the DNN creates a map of virtual neurons and assigns random numerical values, or "weights", to connections between them. The weights and inputs are multiplied and return an output between 0 and 1. If the network did not accurately recognize a particular pattern, an algorithm would adjust the weights.

**1) Convolutional layer:** Convolutional Layer encompasses filters and feature maps. Filters are processors of a particular layer. These filters are distinct from one another. They take pixel value as input and gives out feature Map. Feature map is output of one filter layer. Filter is traversed all along the image, moving one pixel at a time. Activation of few neurons takes place resulting in a feature map.

**2) Pooling layer:** Pooling layer is employed to reduce dimensionality. Pooling layers are included after one or two convolutional layers to generalize features learnt from previous feature maps. This helps in reducing chances of over fitting from training process.

**3) Fully connected layer:** Fully connected layer is used at the end to assign the feature to class probability after extracting and consolidating features from Convolutional Layer and pooling later respectively. These layers use linear activation functions or softmax activation function.

### 1.4 You Only Look Once (YOLOv3)

YOLO version 1 and 2 applies softmax functions convert score into probabilities. This approach is feasible when objects are mutually exclusive only. YOLOv3 employs multi label classification. Independent logistic classifier is used to calculate likeliness of input belong to a specific label. Loss is calculated using binary-cross entropy of each label. Since we omit the softmax function complexity is reduced.

**Optimization of Bounding Boxes:** By using Logistic, regression YOLO v3 predicts the score of presence of object. A ground truth box is defined to all objects, if anchor box overlaps the most with ground truth box, then objectness score is said to be 1. For the anchor boxes whose overlap is greater than the preselected threshold, the anchor box incurs null cost. Every ground truth box is mapped with only one anchor box. If anchor box is not selected and assigned to bounding box then no classification and localization loss is considered, only confidence loss is calculated.

### 1.5 Single Shot Detector (SSD) algorithm

SSD is a popular object detection algorithm that was developed in Google Inc. [1]. It is based on the VGG-16 architecture. Hence SSD is simple and easier to implement. A set of default boxes is made to pass over several feature maps in a convolutional manner. If an object detected is one among the object classifiers during prediction, then a score is generated. The object shape is adjusted to match the localization box. For each box, shape offsets and confidence level are predicted. During training, default boxes are matched to the ground truth boxes. The fully connected layers are discarded by SSD architecture. The model loss is computed as a weighted sum of confidence loss and localization loss. Measure of the deviation of the predicted box from the ground truth box is localization loss. Confidence is a measure of in which manner confidence the system is that a predicted object is the actual object. Elimination of feature resampling and encapsulation of all computation in a single network by SSD makes it simple to train with MobileNets. Compared to YOLO, SSD is faster and a method it performs explicit region proposals and pooling (including Faster R-CNN).

### 1.6 Computer vision-based technique

Computer vision tasks include methods for acquiring, processing, analyzing and understanding digital images, and extraction of high-dimensional data from the real world in order to produce numerical or symbolic information, e.g. in the forms of decisions. Understanding in this context means the transformation of visual images (the input of the retina) into descriptions of the world that make sense to thought processes and can elicit appropriate action. This image understanding can be seen as the disentangling of symbolic information from image data using models constructed with the aid of geometry, physics, statistics, and learning theory. The scientific discipline of computer vision is concerned with the theory behind artificial systems that extract information from images. The image data can take many forms, such as video sequences, views from multiple cameras, multi-dimensional data from a 3D scanner, or medical scanning device. The technological discipline of computer vision seeks to apply its theories and models to the construction of computer vision systems.

## II. LITERATURE SURVEY

**[1]Title:** Large-Scale Object Detection of Images from Network Cameras in Variable Ambient Lighting Conditions.

**Author:** Caleb Tung, Matthew R. Kelleher, Ryan J. Schlueter, Binhan Xu,

Yung-Hsiang Lu, George K. Thiruvathukal, Yen-Kuang Chen, Yang Lu

**Description:** Computer vision relies on labeled datasets for training and evaluation in detecting and recognizing objects. The popular computer vision program, YOLO ("You Only Look Once"), has been shown to accurately detect objects in many major image datasets. However, the images found in those datasets, are independent of one another and cannot be used to test YOLO's consistency at detecting the same object as its environment (e.g. ambient lighting) changes. This paper describes a novel effort to evaluate YOLO's consistency for large-scale applications. It does so by working (a) at large scale and (b) by using consecutive images from a curated network of public video cameras deployed in a variety of real-world situations, including traffic intersections, national parks, shopping malls, university campuses, etc. We specifically examine YOLO's ability to detect objects in different scenarios (e.g., daytime vs. night), leveraging the cameras' ability to rapidly retrieve many successive images for evaluating detection consistency. Using our camera network and advanced computing resources (supercomputers), we analyzed more than 5 million images captured by 140 network cameras in 24 hours. Compared with labels marked by humans (considered as "ground truth"), YOLO struggles to consistently detect the same humans and cars as their positions change from one frame to the next; it also struggles to detect objects at night time. Our findings suggest that state of-the art vision solutions should be trained by data from network camera with contextual information before they can be deployed in applications that demand high consistency on object detection.

**[2]Title:** Object Detection Based on YOLO Network

**Author:** Chengji Liu, Yufan Tao, Jiawei Liang, Kai Li, Yihang Chen

**Description:** Object detection based on the deep learning has achieved very good performances. However, there are many problems with images in real-world shooting such as noise, blurring and rotating jitter, etc. These problems have an important impact on object detection. Using traffic signs as an example, we established image degradation models which are based on YOLO network and combined traditional image processing methods to simulate the problems existing in real-world shooting. After establishing the different degradation models, we compared the effects of different degradation models on object detection. We used the YOLO network to train a robust model to improve the average precision (AP) of traffic signs detection in real scenes.

**[3]Title:** A New Method of Image Detection for Small Datasets under the Framework of YOLO Network.

**Author:** Guanqing Li; Zhiyong Song; Qiang Fu

**Description:** For the image detection problems under small-scale datasets, the detection rate of deep learning method is usually very low. This paper presents a new image detection method based on transfer learning and sample enhancement under the framework of YOLO network. This method takes full advantages of the real-time feature of YOLO, as well as the enhancement of the generalization ability brought by transfer learning and sample enhancements. The detection rate of 87.4% of the 6 targets under the small-scale datasets was achieved, and this method is more than 6 times faster than the Faster R-CNN at the same detection rate. The measured data verified the method. Furthermore, this paper quantitatively analyzes the relationship between sample scale and detection performance under small-scale datasets.

**[4]Title:** Poly-YOLO: higher speed, more precise detection and instance segmentation for YOLOv3

**Author:** Petr Hurtik, Vojtech Molek, Jan Hula, Marek Vajgl, Pavel Vlasanek & Tomas Nejezchleba

**Description:** We present a new version of YOLO with better performance and extended with instance segmentation called Poly-YOLO. Poly-YOLO builds on the original ideas of YOLOv3 and removes two of its weaknesses: a large amount of rewritten labels and an inefficient distribution of anchors. Poly-YOLO reduces the issues by aggregating features from a light SE-Darknet-53 backbone with a hypercolumn technique, using stairstep upsampling, and produces a single scale output with high resolution. In comparison with YOLOv3, Poly-YOLO has only 60% of its trainable parameters but improves the mean average precision by a relative 40%. We also present Poly-YOLO lite with fewer parameters and a lower output resolution. It has the same precision as YOLOv3, but it is three times smaller and twice as fast, thus suitable for embedded devices. Finally, Poly-YOLO performs instance segmentation by bounding polygons. The network is trained to detect size-independent polygons defined on a polar grid. Vertices of each polygon are being predicted with their confidence, and therefore, Poly-YOLO produces polygons with a varying number of vertices.

**[5]Title:** Broken Corn Detection Based on an Adjusted YOLO With Focal Loss

**Author:** Zechuan Liu; Song Wang

**Description:** Corns may be broken during corn mechanical harvesting. The ratio of broken corns measures the quality of mechanical harvesting and should be monitored in real time. This paper presents a method of detecting both broken and non-broken corns at the conveyor belt of a corn harvester based on the YOLO. The network structure of the YOLO is adjusted here to obtain more robust features so that it can work well in the open working space of the corn harvesting. Moreover, we improve the loss function to ensure that the hard examples can catch more attention during training. As it is difficult to obtain many training data of broken corns, the simulated broken corn images are generated from the real images of corns by a simple synthetic method. The concerned corn detection network of the proposed YOLO-based method is first trained with plenty of simulated samples and then fine-tuned with the real corn images. The experiments on real corn data confirm that the proposed YOLO-based method can achieve good accuracy and fast speed on the NVIDIA TX2.

## III. PROBLEM STATEMENT

Animal images captured in a field represent a challenging task while classifying animals since they appear with a different pose, cluttered background, different light and climate conditions, different viewpoints, and occlusions. Additionally, animals of different classes look similar. All these challenges necessitate an efficient algorithm for classification.

In this challenge, you will be given 12,000 images of 10 different animal species. Given the image of the animal, your task is to predict the probability for every animal class. The animal class with the highest probability means that the image belongs to that animal class.

For several years, animal detection in the wildlife has been an area of great interest among biologists. They often study the behaviour of the animals to predict their actions. Since there are a large number of different animals, manually identifying them can be a daunting task. So, an algorithm that can classify animals based on their images can help researchers monitor them more efficiently. Also, animal detection and classification can help prevent animal-vehicle accidents, trace animal facility, prevent theft, and ensure the security of animals in zoos.

The application of deep learning is rapidly growing in the field of computer vision and is helping in building powerful classification and identification models. We can leverage this power of deep learning to build models that can classify and differentiate between different species of animals as well.

In this dataset, we provide 12,000 images of 10 different species of animals. In the next 90 days, we challenge you to build models such that given an image, the model will predict the probability of every animal class. The animal class with the highest probability will signify that the image belongs to that animal class.

**Existing System**

- Our Existing system describes a system for segmentation of animals from images.
- The procedure employed uses a multi-level iterative graph cut to generate object region proposals and accurately recognize regions of interest.
- This is especially useful when the animal blends together with the background and is difficult to identify.
- These proposals segmented into background and foreground in the second stage. Feature vectors are extracted from each image using AlexNet.

**Disadvantage**

1. It cannot give an accuracy value.
2. Less storage
3. Less performance

## IV. PROPOSED SYSTEM

- ✓ The proposed method uses the YOLO object detection model to ascertain presence of wild animals in images.
- ✓ The model is fine-tuned for identifying Ten different types of animals (Dog, Horse, Butterfly, elephant, Hen, Cat, Cow, Sheep, Spyder, Squirrels.)
- ✓ The proposed method achieves an accuracy of 98.8% and 99.8% to detect animals respectively.

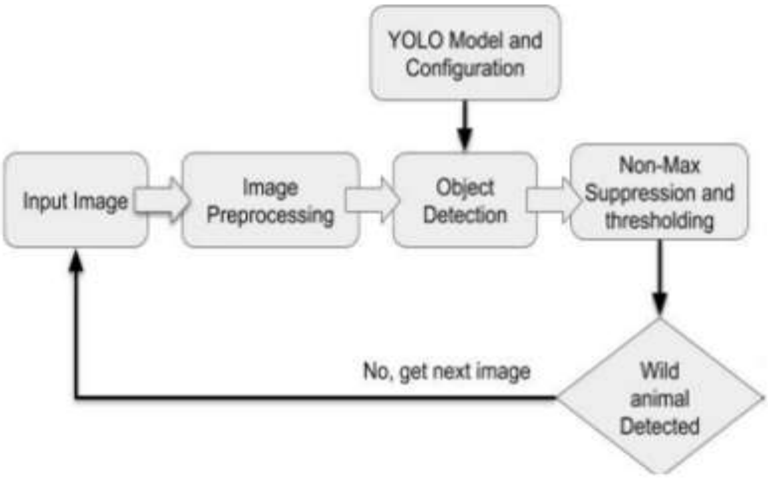
The system proposed in this paper uses a network of cameras, connected to PIR motion sensors, so that image capture is triggered only when some movement is detected. This enables power conservation. The images captured through these cameras are processed to detect presence of wild animals, and if an animal is found, identify the species. Once identified, the animals are tracked for a suitable time in order to determine their intent – such as to find whether they are moving across the village, or into it. In the latter case, alerts are generated and local authorities are notified through proper channels. Understanding the intent goes a long way to reduce false positives, either due to a false detection or when there is no actual threat posed due to presence of the animal.

**Advantages**

- Easily detection with the help of Yolov3 and SSD
- High Accuracy
- Good Performance

| Variable | Description |
|---|---|
| Image_id | Image name |
| Animal | Name of the Animal |

## V. SYSTEM ARCHITECTURE




### 5.1 System Modules

- Module 1: Dataset Collection
- Module 2: Pre-processing
- Module 3: Tracking Object
- Module 4: Detection
- Module 5: Pre-processing

**Module 1: Dataset Collection**

You're given two types of files (CSV and Images) to download. The train data consists of 12,000 images and the test data consists of 6,000 images of 10 different species of animals. The image ID and the corresponding animal name are stored in .csv format, while the image files are sorted into separate train and test image folders. Data in the .csv file is in the following format:

Following are the 10 different species of animals in the dataset:

- ✓ Dog
- ✓ Horse
- ✓ Butterfly
- ✓ Elephant
- ✓ Hen
- ✓ Cat

- ✓ Cow
- ✓ Sheep
- ✓ Spyder
- ✓ Squirrels

**Module 2: Pre-processing**

Once the data is extracted from the twitter source as the datasets, this information has to be passed to the classifier. The classifier cleans the dataset by removing redundant data like stop words, emoticons in order to make sure that non textual content is identified and removed before the analysis.

The sklearn.preprocessing package provides several common utility functions and transformer classes to change raw feature vectors into a representation that is more suitable for the downstream estimators.

In general, learning algorithms benefit from standardization of the data set. If some outliers are present in the set, robust scalers or transformers are more appropriate. The behaviors of the different scalers, transformers, and normalizers on a dataset containing marginal outliers is highlighted in Compare the effect of different scalers on data with outliers.

**Module 3: Tracking Object**

Internet is the main network connecting millions of people in world. Main entertainment factor and the source of greater knowledge is image. Video is collection of frames. The negligible time gap between frames makes the stream of photos looks like flow of scenes. When designing algorithm for video processing. Videos are classified into two classes. Video stream is an ongoing process for video analysis. The processor is not aware of future frames. Video sequence is video of fixed length. All the consecutive frames are obtained prior to processing of current frame. Motion is distinct factor that differentiates video form frame. Motion is a powerful visual Que. Object properties and action can be realized by noticing only sparse points in the image.

**Module 4: Detection**

Frames are captured from camera at regular intervals of time. Difference is estimated from the consecutive frames. Optical Flow This technique estimates and calculates the optical flow field with algorithm used for optical flow. A local mean algorithm is used then to enhance it. To filter noise a self-adaptive algorithm takes place. It contains a wide adaptation to the number and size of the objects and helpful in avoiding time consuming and complicated pre-processing methods. Background Subtraction Background subtraction (BS) method is a rapid method of localizing objects in motion from a video captured by a stationary camera. This forms the primary step of a multi-stage vision system. This type of process separates out background from the foreground object in sequence in images.

**Module 5: Accuracy score evaluated**

The trained model using deep learning must be evaluated for its performance on unseen data called as test dataset. The choice of performance metrics will influence the analysis of algorithms. This helps in identifying reasons for misclassifications so that it can be corrected by taking necessary measures.

1) **Accuracy and Loss:** Accuracy measure is calculated by using formula. The accuracy measure, as a stand-alone measure is not reliable since it gives equal costs for both type of errors and works well for a well-balanced dataset. The loss is calculated by loss functions of used for training, and average of the loss is calculated when used batch learning that computes loss after each training each batch.

2) **Precision, Recall and F1-** score: Precision is the percentage of classification results that are relevant. Recall is the percentage of total relevant results that are classified correctly by algorithm. F-1 score considers both precision and recall values hence must be maximized to make the model better.

## VI. CONCLUSION

The proposed system detects the animals using computer vision. The proposed system is cost-effective and highly efficient, with an average accuracy of 98.8% in detecting and identifying animals in images. Although the prototype described in this paper is trained to recognize ten different species of animals, it is easily extendable to detect other types of animals with sufficient training data. The choice of species can also be region specific, thereby providing a unique edge over other existing solutions. Such a system if implemented on a large scale, has potential to detect large data of animal categories.

## REFERENCES

[1] Caja, Gerardo, J. J. Ghirardi, M. Hernández-Jover, and D. Garín. "Diversity of animal identification techniques: From ‘fire age’to ‘electronic age’." *ICAR Technical Series* 9 (2004): 21-39.

[2] Donovan, John, and Patricia Brown. "Animal identification." *Current protocols in immunology* 73, no. 1 (2018): 1-5.

[3] Eradus, Wim J., and Mans B. Jansen. "Animal identification and monitoring." *Computers and Electronics in Agriculture* 24, no. 1-2 (2019): 91-98.

[4] Voulodimos, Athanasios S., Charalampos Z. Patrikakis, Alexander B. Sideridis, Vasileios A. Ntafis, and Eftychia M. Xylouri. "A complete farm management system based on animal identification using RFID technology." *Computers and electronics in agriculture* 70, no. 2 (2020): 380-388.

[5] Kellenberger, B., Volpi, M., Tuia, D.: Fast animal detection in uav images usingconvolutional neural networks. In: 2017 IEEE International Geoscience and RemoteSensing Symposium (IGARSS), pp. 866–869 (2017)

[6] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep con-volutional neural networks. In: F. Pereira, C.J.C. Burges, L. Bottou, K.Q. Wein-berger (eds.) Advances in Neural Information Processing Systems 25, pp. 1097–1105. Curran Associates, Inc. (2012)

[7] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.:Ssd: Single shot multibox detector. Lecture Notes in Computer Science pp. 21–37(2016)7. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Interna-tional Journal of Computer Vision 60(2), 91–110 (2004)

[8] Lukeˇziˇc, A., Voj´ıˇr, T., ˇCehovin Zajc, L., Matas, J., Kristan, M.: Discriminativecorrelation filter tracker with channel and spatial reliability. International Journalof Computer Vision 126(7), 671–688 (2018)

[9] Matuska, S., Hudec, R., Benco, M., Kamencay, P., Zachariasova, M.: A novel sys-tem for automatic detection and classification of animal. In: 2014 ELEKTRO, pp.76–80 (2014)1

[10] Norouzzadeh, M.S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M.S., Packer,C., Clune, J.: Automatically identifying, counting, and describing wild animals incamera-trap images with deep learning. Proceedings of the National Academy ofSciences 115(25), E5716–E5725 (2018)

[11] Parham, J., Stewart, C., Crall, J., Rubenstein, D., Holmberg, J., Berger-Wolf, T.:An animal detection pipeline for identification. In: 2018 IEEE Winter Conferenceon Applications of Computer Vision (WACV), pp. 1075–1083 (2018)

[12] Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified,real-time object detection. In: 2016 IEEE Conference on Computer Vision andPattern Recognition (CVPR), pp. 779–788 (2016).