DeepPet: A Pet animal tracking system in Internet of Things using deep neural networks

Ahmed Ali Hammam
Faculty of Computers and Information
Cairo University
Scientific Research Group in Egypt
(SRGE)
ahmed.a.hammam@grad.fci-cu.edu.eg
www.egyptscience.net

Mona M.Soliman
Faculty of Computers and Information
Cairo University
Scientific Research Group in Egypt
(SRGE)
Mona.solyman@fci-cu.edu.eg
www.egyptscience.net

Aboul Ella Hassanein
Faculty of Computers and Information
Cairo University
Scientific Research Group in Egypt
(SRGE)
aboitcairo@gmail.com
www.egyptscience.net

***Abstract*—Pet animal monitoring in smart cities is a big challenge. The classical animal identification and surveillance methods like air tags, GPS, and RFID fail to provide the required level of security and management of pets. Such devices have many limitations and costly. The massive development in the world of Internet of things (IOT) in smart cities can be used to examine the ability of computation, communication, and control technologies to improve human interaction with pets by the technology of the Internet of Things. This paper will present an approach of pet animal tracking on the video stream using deep learning capabilities with the aim of recognizing and classifying the object of interest.**

Keywords- Pet animal, conventional neural network, tracking, Object detection.

I. Introduction

The goal of building a smart city is to use technology to improve the quality of life. In smart cities, technology is the first player. There we can find high-tech solutions to the significant challenges foreseen for the next years, which are expected to affect the current urban landscapes strongly. Many capabilities are considering when building smart cities to ensure providing the high level of comfort to our life. Recently, there was an interest in research to combine digital devices signals and information to enhance planning of smart cities [13]. The remarkable growth of digital devices, such as sensors, actuators, smart-phones and smart appliances leads to broad commercial objectives of the Internet of Things (IoT). It is possible to interconnect all devices and create communications between them through the Internet. Smart City tries to connect physical infrastructures, Information and Communication Technologies (ICT), knowledge resources, and social infrastructures to create a better city administration, and infrastructure management [1].

The first initiatives in Smart Cities are related with areas like Intelligent Transport and Smarter Parking, Efficient Resource Management (e.g. water, energy), Building Automation and Smart Buildings. To cope with such applications, some public transport information, e.g., real-time location and utilization, occupancy of parking spaces, traffic jams, and other data like weather conditions, air and noise pollution status, water contamination, energy consumption, etc. should be gathered continuously [13]. To this end, different technologies have been applied to address the specific features of each application. The required technologies cover a wide range and layer from the physical level to the data and application layers. According to the United Nations, half of the world's population already live in cities, and by 2050, nearly seven out of 10 people worldwide—more than 6 billion people in total—will lead urban lives. Already, mega-cities are responding to growth by becoming *" smart cities"*. Smart Cities have been further defined to six axes or dimensions: Economy; Mobility; Environment; People; Living and Governance [14]. There is a new demand for integrating Pets in the Smart City. For example, in the United States, there is roughly one dog for every four humans, and the rate of canine ownership keeps rising. In 2011, the number of dogs in the city of Seattle was greater than the number of children: 153,000 dogs to 107,000 children. New York City is home to more than 600,000 dogs [15].

In addition to putting sensors on cars and in roads and parks, We need to give more interest for pet animals in smart cities. We need to examine the ability of computation, communication, and control technologies to improve human interaction with pets either for entertainment purpose or security purpose. One of the main concern of the pet owners is keeping better track of their pet animals. They always need a way to monitor pet animal and tracking it when it's out from home. There is a need to know where it is saving them from being lost or stolen. Monitoring and recognition of pet animal are done using tools like ear-tag and GPS tools. These tools can be damage easily and can be affected easily by surrounding environment and also can be stolen from the pet and tracking. Recently, recognition systems are gaining more attention due to a variety of applications and use for registration and monitoring of pet animal in the smart city [2].

Pet animal recognition and tracking are considering a hot research area during last years. [2] Propose an algorithm that categorizes animal locomotive behavior by combining detection and tracking of animal faces in wildlife videos. The detection algorithm is based on a human face detection method, utilizing Haar-like features and AdaBoost classifiers.

Tracking is implemented using the Kanade-Lucas-Tomasi method. Another proposed research in [3] tries to demonstrate that face recognition of dogs can be used to recognize the dog efficiently. Develop an algorithm to provide security to pet animal with the help of biometrics using face recognition. Using feature extraction algorithms such as PCA (Principal Component Analysis), LDA (Local Discriminant Analysis) and ICA (Independent Component Analysis). Authors in [4] provide a system for monitoring of pet animals (dogs) based on their primary animal biometric identifiers. The proposed recognition approach uses the one-shot similarity and distance metric based learning methods for matching and classifying the extracted features of face images for recognition of pet animals (dog). Using a Fisher Linear Projection and Preservation (FLPP) as feature extraction. And the One-Shot Similarity (OSS) via distance metric based learning with online incremental support vector machine to classify the extracted features of face image database of pet animals (dogs). Current best-performing detectors are based on the technique of finding region proposals to localize objects. [5] Provide a method R-CNN: Regions with CNN features. They try to localizing objects with a deep network and training a model with only a small quantity of annotated detection data. Apply convolutional neural networks to bottom-up region proposals to localize and segment objects. Their algorithm gives a 30% relative improvement over the best previous results on PASCAL VOC 2012. [6] Propose a framework for object localization and recognition where regions are introduced from a CNN, developing a region-based object detection framework that boosts up the classification performance and reduces the computational complexity and improves the performance in object detection. Trying to generate regions that are semantically meaningful and related to objects. Their method achieves better or similar performance when compared to R-CNN with significantly less number of region proposals. In [7] they trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. To make training faster, they used non-saturating neurons and an efficient GPU implementation of the convolution operation. Reduce the over-fitting in the fully-connected layers by using a regularization method called "dropout". Achieved an error rate of 15.3%, compared to 26.2% obtained by the second-best entry. They try to show that a large, deep convolutional neural network is capable of achieving record breaking results on a highly challenging dataset using purely supervised learning. In recent years, deep artificial neural networks (including recurrent ones) have won numerous contests in pattern recognition and machine learning. Deep learning is a new way of fitting neural nets. Traditionally a neural net is fit to labelled data all in one operation. The weights are usually started at random values near zero. Due to the non-convexity of the objective function, the final solution can get caught in a poor local minimum. In deep learning, multiple layers are first fit in an unsupervised way, and then the values at the top layer are used as starting values for supervised learning. Apparently by modeling the joint distribution of the features, this can yield better starting values for the supervised learning phase. These two stage approach also allows the use of unlabeled data (sometimes available in abundance) I There are different deep learning models-those with quantitative latent factors, which look like a form of nonlinear PCA, and those with discrete hidden factors. The latter are favored by Hinton and his group, and are much harder to fit (due to the intractability of the partition function).

This work proposes an approach to track pet animal based on deep learning solution. We will apply conventional neural network (CNN) to video streaming of pet animal. Pet animal tracking will be depended on determine its location using Convolutional Neural Network (CNN). We will use a pre-trained model of Fast R-CNN. This paper proposed an approach consisting from three steps; first one is Object localization. Second step is object recognition. Third step is tracking the object frame by frame. The rest of the work is divided as follows. Section 2 will introduce the fundamentals that support the theory exposed in this work. Section 3 introduces the proposed approach used to provide pet tracking based on deep learning. Detailed results of the proposed approach will be shown in section 4. Section 5 draws conclusions and future work to be done during next years.

II. Basics and Background

A. Deep Neural Network (DNN)

Traditional machine learning techniques were limited in their ability to access natural data in their raw form. Usually, constructing a pattern-recognition or machine-learning system needs a careful engineering and considerable domain expertise to design a feature extractor that used to transform the raw data into a suitable internal feature vector from which the learning subsystem, could detect patterns in the input [15]. Since the late 2000's, neural networks have become one of most successful machine learning technique. This happen due to o the availability of inexpensive, parallel hardware (graphics processors, cluster computer) and a massive availability of labeled data. Deep learning (also known as Deep Neural Network "DNN") is initiated within the same domain of neural networks. Although the original neural networks contain very few layers in the network, but deep networks had many more layers in the network. Networks today have five to more than a thousand of layers [16]. The ability of Deep Neural Network to extract high-level features from raw input data after using statistical learning over a large amount of data gives it a high superior performance. Deep learning is providing a major advance in solving many y challenging tasks that the artificial intelligence community fails to solve for many years [17]. Deep Neural Network Can be consider a revolution in Machine learning field. Many models are proposed to be used in different research areas. The learning methods of Deep-Neural Network consisting of multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more

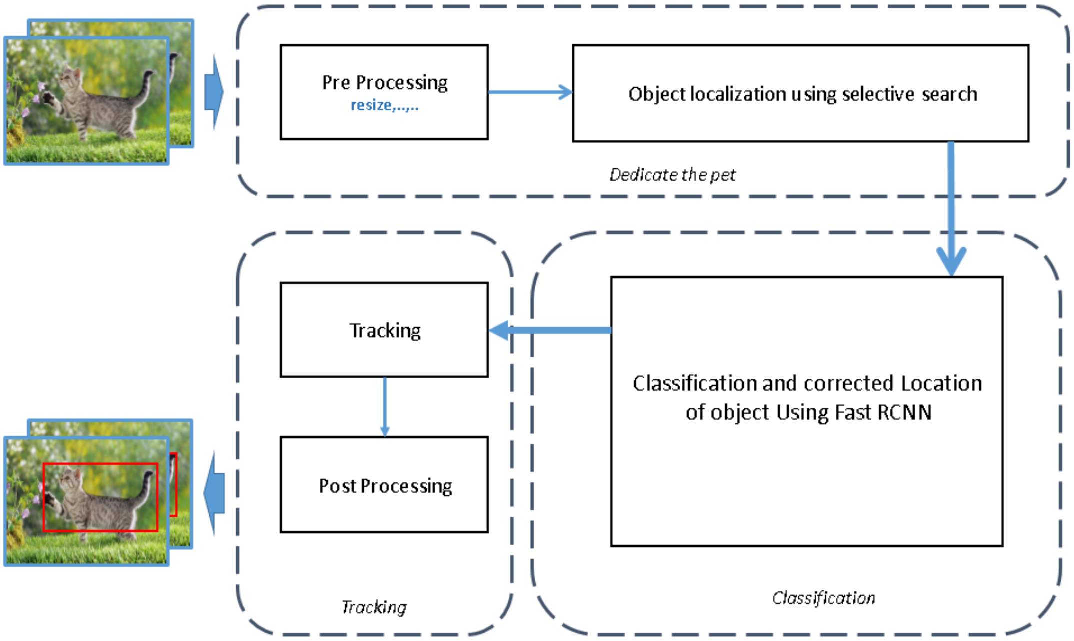
Pre Processing
resize,..,..
Object localization using selective search
Dedicate the pet
Tracking
Classification and corrected Location of object Using Fast RCNN
Post Processing
Tracking
Classification

Fig. 1: The proposed tracking PET animal approach

abstract level. Using this composition with enough number of transformations, DNN can be trained and learned very complex functions. For classification tasks, higher layers of representation amplify aspects of the input that are important for discrimination and suppress irrelevant variations [16].Deep learning architectures such as deep neural networks, deep belief networks and recurrent neural networks have been used to solve problems in many domains including: natural language processing, computer vision, bioinformatics,audio recognition, speech recognition, machine translation, and social network filtering. The produced results of DL in such applications is comparable to and in some applications superior to human experts [18].

III. THE PROPOSED SYSTEM

This Paper introduce a new approach For Tracking Pet animal as shown in figure 1, we will use a pre-trained model of Fast R-CNN. This paper proposed an approach involves three steps; first one is object localization or where is the pet. Second step is object recognition or is it a cat?. Third step is tracking the object frame by frame.

A. Pet Animal Localization

The main steps in this system is Pet Detection (or localization), This Paper use a Pre-trained approach using Faster Region-based Convolutional Neural Network (R-CNN) for localize the Pet. The Faster R-CNN consists of two Steps. The region proposal step produces a set of 2k proposed regions using selective search method where the objects of interest could localize within the image. The region classifier step then determines if the region belongs to an object class of interest [9]. Overlapping regions are producing the final bounding boxes for detected objects. In this phase as shown in Fig 2., the video is converted to frame. Each frame is entered to the fast R-CNN network takes the current frame and using region proposal method (selective search) to find all possible places where the target object can be found. In this phase, a lot of regions of interest are generate. If an object is not detected during the first phase (region proposal), in the second phase there is no way to correctly classify. To achieve a good detection result needs to generate very large numbers of proposals (2k proposed regions). Second input in this phase is a fixed-size feature map obtained from a deep convolutional network (VGG). Next stage in this phase is taking the feature map obtained from a deep convolutional network and a list of regions of interest for the input image and insert theme to a Region of interest pooling layer in Fast R-CNN architecture. This layer is used for object detection. This layer takes each region of interest from the input list, plus the input feature map that corresponds to it and make a scales to it. The Region of interest pooling layer will resize, the region proposals to the same resolution expected on the fully connected layer which is the layer take the output from the Region of interest pooling layer which is a fixed sized feature map for each Region of interest.

B. Classification of the Pet Animal

Fast R-CNN take the input image that are pre-processed and divided it to region of interest (ROI) using selective search and then compute CNN Feature and make refine to the region of interest to narrow it to the search class and determine if the region belong to a pet animal. In this paper, we interested in cat as a target pet animal. Convolutional neural networks have a fixed input size, so each image need to be resized to

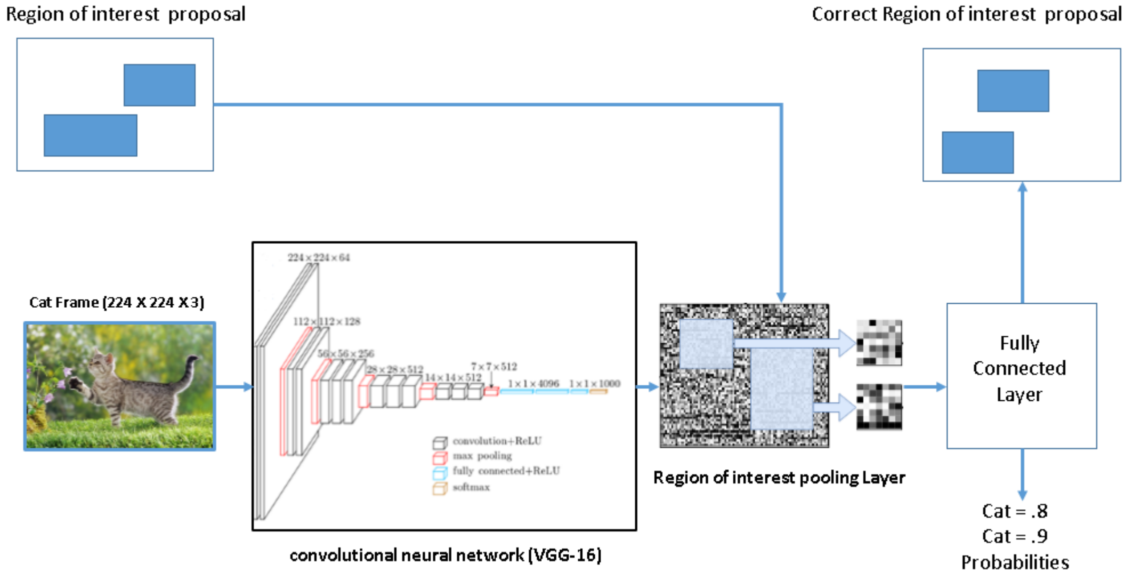

Fig. 2: Fast RCNN Architecture

a pre-defined size that the network expects. First the Input image need to pre-processed (resized and cropped) before being submitted to the fast R-CNN network. As shown this paper used what is called a Transfer learning which commonly used in deep learning applications. Here take a pre-trained network and use it as start point to learn a new task which her is tracking pet animal. The input to this stage is a max-pooled region of interest proposals. The region of interest pooling layer do the following to its inputs. Make the region proposal into equal-sized sections, finding the largest and finally insert these max values to the output buffer.

The final result bounding boxes is produced for our target pet animal. We can describe the final step as the system is branch into two sibling output layers: one that produces soft max probability estimates over K object classes plus a special “background” class in case no object was recognized. And another layer that outputs four real-valued numbers for each of the K objects classes. For each group of 4 values is a representation for bounding-box positions for one of the K classes [9].

C. *Pet Animal Trcaking*

To track the pet animal this paper uses Convolutional Neural Network (CNN) and region proposals for each frame. We can say that tracking is an integrated step. We will track the pet animal by localize it in each frame. Convolutional Neural Network (CNN) have proven itself as a main player in object detection and recognition. Here the tracking which refers to localize the pet animal in frame by frame and classify it is done in two steps. First step is Calculate the candidate ROIs for detection. Using methods like Selective Search to produce identical ROIs for each frame. The Convolutional Neural Network (CNN) model that we are used in The Faster R-CNN is the very deep VGG16 [14]. This Model as show in Fig 2 take an input image 224*224 it consist of 16 CONV/FC layers. Each convNet has filters of size 3x3. The main algorithm used in this work is illustrated as follow:

```
input : Frame F, Region of interest
output: bounding boxes for detected Pet B
Procedure:
Frame=0
while hasFrame () do
    1.Resize Frame to compatible with the
      Convolutional neural network
    2.Run the resized Frame on CNN and extract Row
      of feature.
    3.Take the feature together with the candidate ROIs
      through ROI pooling Layer.
    $Frame \leftarrow Frame + 1$
end
```

Algorithm 1: Pet animal tracking

IV. EXPERMENTAL RESULTS

In this experiment, we are using a virtual machine provided by google cloud with 12 gigabytes of ram and one Intel Xeon CPU and NVIDIA Tesla P100 and K80 GPU run on windows server 2012 and using matlab to implement the Deep Learning model. We test this model against cat videos in different shots. In Fig. 3 we can see the result of the first phase for the introduced model which is pet animal localization. The final result for the introduced model is tracking and detected the pet animal (cat) which can be shown in Fig. 4. To compare the model introduced in this paper we have test two different model based on architecture provide in [11] and [12] ,this model depended on deep neural network as classifier or feature extractor and not used it as a method for localization of the object. In [11,12] model they are using MatConvNet, which are a pretrained CNN classifier that has been trained on the ImageNet dataset. It is a CNN package for MATLAB that uses the NVIDIA cuDNN. They are using CNN as Feature extractor and use these feature to train an SVM classifier. We compare the result of our detection and classification model against both models proposed in 11 and 12 (e.g. GoogleNet ,VGG). The main benefits of our proposed model is the impact of using deep learning in localization and detection phase. Deep learning compared to classical object detection algorithms proves its efficiency as shown in Figures 5,6, and 7. The classic detection method that can be easy collision with different factor and may be limits on one actor per videos.

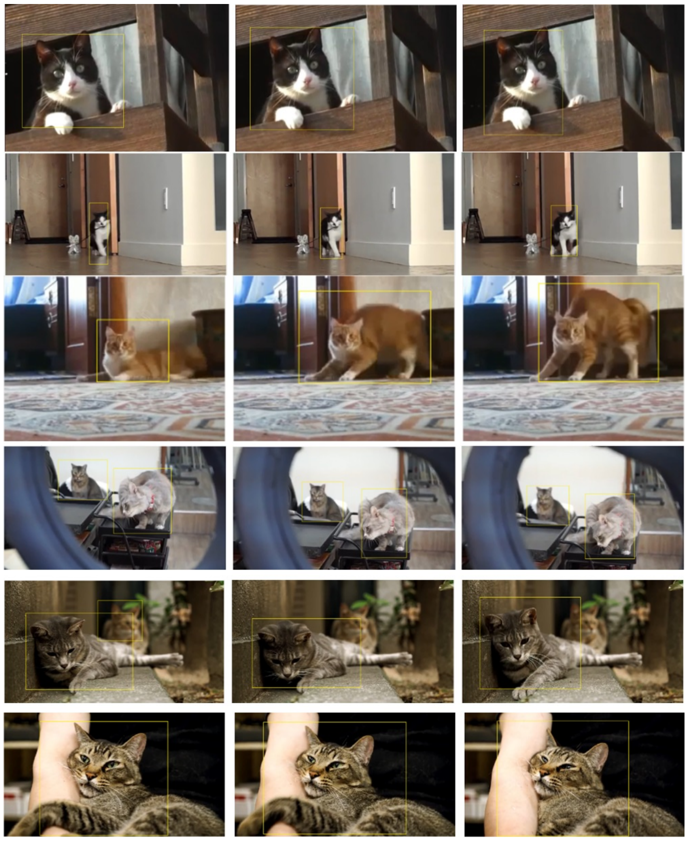

Fig. 3: Pet animal localization

Fig.5 compare the introduced model with model from [11, 12] as Clearly in (a) our proposed model can localize and classify more than one object at the same time which can't done using model appears in [11, 12] as shown in (b) and (c). This result is obtained because in other models they are depended on motion pixel which lead to the interested object in the whole frame are not localized as need. But, in the introduced model we search for the object of interest in the whole frame by region search and localized the interest class.

Fig. 6 shown one of strength of proposed model using Region to search for the object in image frame. Fig 6(a) is a result of introduced model in this paper and Clearly how the success that are achieving in localize and classify the pet animal but in (b) and (c) which are implementation according to the model in [11,12] the localize of the pet animal has fail. These results achieved because they are depended on moving pixel to localize the object in frame and the two cats in these shot are not moving for example.

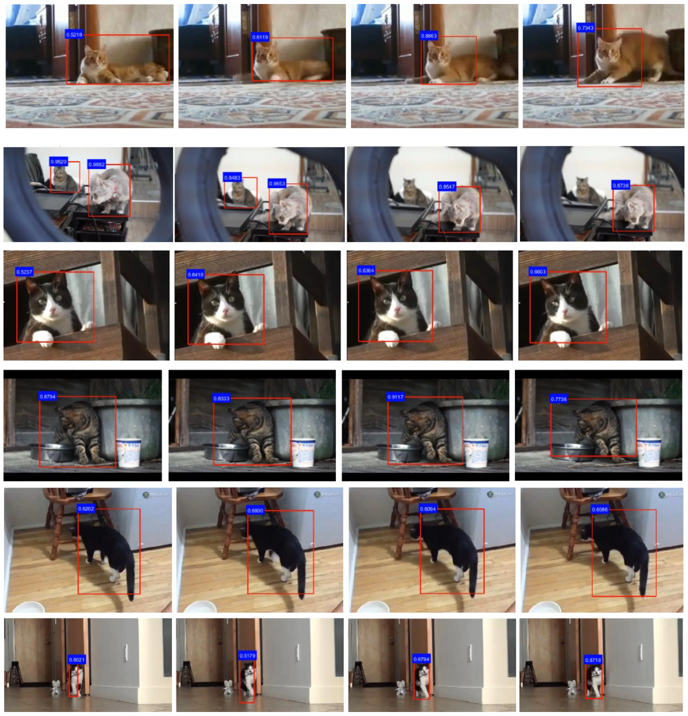

Fig. 4: Proposed model tracking and detection result against different action and shots for cats

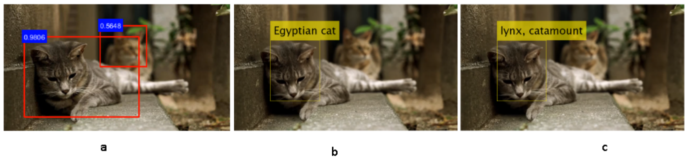

Fig. 5: The proposed tracking PET animal approach comparing with other methods: (a) proposed approach, (b)GoogleNet model [11], (c) VGG model [12]

In the localization phase the more the detect of the object is Cleary the more of the classification are success. In Fig. 7 we try to show how the model introduced in this paper are success to localize the full object as shown in Fig.7 (a). In Fig. 7 (b) and (c) which are the results of model in [11, 12] can Cleary observe only moving parts are localize and the results of classify is depended on this part these may lead to incorrect results in classification phase.

As show in table (1) we can see that depending on Deep learning network model as an object detection method can lead to classification result more accurate than depending on method like optical flow or background subtraction proposed by [11] and [12]. Our proposed model gives better detection rate other than two other models (e.g. GoogleNet, VGG). The proposed model localize and detect the object of interest which is cat during different cat actions (e.g. Pounce,

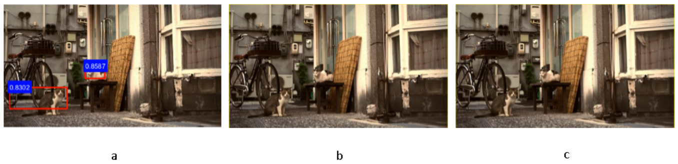

Fig. 6: (a)localization result of proposed mode, (b) localization result using GoogleNet [11], (c) localization result using VGG [12]

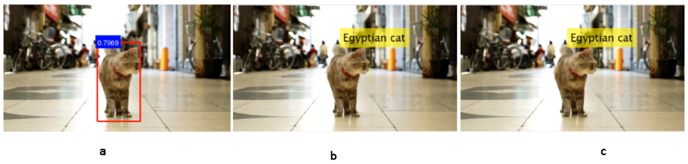

Fig. 7: (a) Full body of cat are localize by proposed method (b) Localization of moving part only by GoogleNet [11] (c) Localization of moving part only by VGG [12]

Grooming, Leap, Step, and Eat). Fig. 8 shows the accuracy achieved by proposed model using Fast R-CNN compared with other GoogleNet and VGG models. The accuracy is estimated against different videos, with the result of detection and classification is calculated over the whole video shots. The proposed model gives the highest accuracy compared with other models for all videos except video 2.

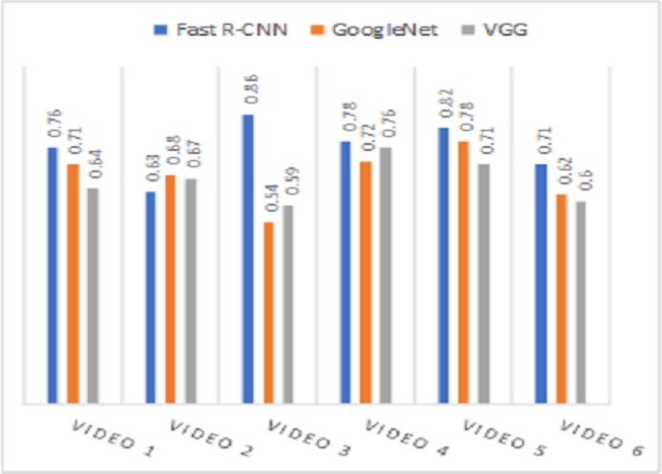

Fig. 8: Classification and tracking per frame accuracy on the different videos

V. CONCUSSIONS AND FUTURE WORK

We proposed in this paper a model for localizing tracking pet animal. This paper showed how reliable can be obtained from using deep learning solution. This paper used Fast R CNN Model as conventional neural network (CNN) network. It's used to detect and classify the pet animal. The Pet animal that consider our object of interest is "cat". We tested this model on different video sequence with different cat actions. This model is considering one of the first models interested in this area of research. We hope in future to upgrade this system to identification and tracking the pet animal in multi-camera streaming and provide a real-time model for that case.

TABLE I: Shows per-action detection accuracy, as shown fast R-CNN depended on itself to localize The object which lead to more accuracy on classification result

Model/Action	Pounce	Grooming	Leap	Step	Eat
Fast R-CNN	.82	.94	.96	.94	.92
GoogleNet	.71	.68	.76	.91	.82
VGG	.72	.69	.74	.92	.84

We hope to provide datasets for a different kind of pet animal to make the learning phase more general.

REFERENCES

[1] A. Ojo, E.Curry E, T. Janowski, Designing Next Generation Smart City Initiatives - harbessing Findings and Lessons from a Study of ten Smart City programs, The 22nd European Conference on Information Systems (ECIS 2014), vol. 2050, pp.1-14, 2014.

[2] B. Tilo, J. Calic, and B.T. Thomas, Tracking Animals in Wildlife Videos Using Face Detection. Proc. of European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology, November 2004.

[3] K. Santosh, and S. K. Singh, Biometric recognition for pet animal, Journal of Software Engineering and Applications, vol. 7, no. 5, pp. 470-482, 2014.

[4] K. Santosh, and S. K. Singh, Monitoring of pet animal in smart cities using animal biometrics, Future Generation Computer Systems,Volume 83, pp.553-56, June 2018.

[5] G. Ross, et al. Rich feature hierarchies for accurate object detection and semantic segmentation, Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580-587, 2014.

[6] H. Bappy, Jawadul, and K. Amit Roy-Chowdhury: CNN based region proposals for efficient object detection, The IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25-28 Sept. 2016

[7] R. Prajit, Object Detection in Video using Faster R-CNN, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.

[8] Everingham, M.; Eslami, S.M.A.; van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A: The pascal visual object classes challenge, A retrospective. Int. J. Comput. Vis., pp: 98-136, 2015.

[9] Ren, S.; He, K.; Girshick, R.; Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks, In Proceedings of the Advances in Neural Information Processing Systems, Montréal,QC, Canada, 7–12 December, pp. 91-99, 2015

[10] Sa, Inkyu, et al., Deepfruits: A fruit detection system using deep neural networks, Sensors, 16(8), 1222, August 2016.

[11] https://www.mathworks.com/company/newsletters/articles/deep-learning-for-computer-vision-with-matlab.html.

[12] https://devblogs.nvidia.com/parallelforall/deep-learning-for-computer-vision-with-matlab-and-cudnn/.

[13] S. Talari , M. Shafie-khah, P. Siano , V. Loia, A. Tommasetti and J. P. S. Catalão : A Review of Smart Cities Based on the Internet of Things Concept, Energies 2017, vol.10, no.(4), 421 doi: 10.3390 /en10040421

[14] A. Z.K.Simonyan,: Very deep convolutional networks for large-scale image recognition, ICLR, Hilton San Diego Resort and Spa, May 7-9, 2015.

[15] Y. LeCun, Y. Bengio and G. Hinton: Deep learning , Nature 521, pp. 436-444, 28 May 2015 , doi:10.1038/nature14539

[16] V. Sze, Yu. Yang, and J. S. Emer, Efficient Processing of Deep Neural Networks: A Tutorial and Survey, CoRR jornal, 2017

[17] S. Wang , Abdel-rahman Mohamed , R. Caruana , J. Bilmes , M. Plilipose , M. Richardson, K. Geras, G. Urban, and O. Aslan: Analysis of Deep Neural Networks with the Extended Data Jacobian Matrix, Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 2016. JMLR: W and CP volume 48, 2016

[18] A. Krizhevsky, S. Ilya and H.Geoffrey: ImageNet Classification with Deep Convolutional Neural Networks, Advances in Neural Information Processing Systems 25, pp. 1097-1105, 2012