Real Time Object Detection Deep Reinforcement Learning Model: RODRLM

[1] Mrs Sonal Tiwari ,[2] Dr. Shailja Sharma ,[3]Dr. Saurabh Tiwari

[1] Research Scholar, Department of Computer Science and Engineering RNTU Bhopal
[2] Associate Professor, Department of Computer Science and Engineering RNTU Bhopal
[3] Sr. Lecturer, Government S. V. Polytechnic College Bhopal

Abstract - Visual tracking is a challenging problem since it usually faces adverse factors, such as object deformation, fast motion, occlusion, and background clutter in practical applications. Reinforcement learning based Action-Decision Network (ADNet) has shown great potential for object tracking. However ,ADNet has some shortcomings in optimal action selection and action reward, and suffers from in efficient tracking. To this end, real time object detection deep reinforcement learning model come in tracking method , RODRLM improve efficiency and accuracy in real time object tracking . Reinforcement Learning based real time object detection framework using deep machine learning model is a unique and important technique using which user can get quality output and can use in many system for getting efficient results. In this work we propose object detection with deep reinforcement learning by which we train the agent to extract the features of sequence of the frame and with a trained agent we detect the object present in video. The proposed technique improve the feature extraction ability of its convolution layers. Then, in the reinforcement learning based training phase, both the selection criteria for optimal action and the reward function are redesigned separately to explore more appropriate action and eliminate useless action .Finally, an effective online adaptive update strategy is proposed to adapt to the appearance changes or deformation of the object during actual tracking. Specially, meta-learning is utilized to pursue the most appropriate parameters for the network so that the parameters are closer to the optimal ones in the subsequent tracking process.

Keywords: Visual tracking, reinforcement learning, meta-learning, multi-domain training.

I. INTRODUCTION

As it is known to us, letting the agent possess the ability to continuously learn and adapt from limited experience in non stationary environments is an important milestone on the path towards general intelligence .Visual tracking, as a prominent part of artificial intelligence, has received increasing attention in the past few decades. Similarly, with the development of machine learning, numerous learning networks have been widely used in visual tracking eld [10][11]. In 2017, for the first time, Yun et al.[12] proposed a novel tracker controlled by the , real time object detection deep reinforcement learning model . which is radically different from the existing trackers , Proposed tracker is developed to increase tracking accuracy and efficiency. The multi-domain training is soul of algorithm which makes the combination of training effective for proposed work. Selection criteria for optimal action and the reward function are redesigned separately to explore more appropriate action and eliminate useless action. an effective online adaptive update strategy is proposed to adapt to the appearance changes or deformation of the object during actual tracking. Experimental results demonstrate that the proposed tracker has advantages over ADNet and other techniques in terms of accuracy and efficiency .To improve the robustness and real-time performance of the tracker, research has been improved from three aspects. Firstly, the use of multi-domain training instead of supervised learning based training enables the tracker to learn the shared representation of different objects in the various training sequences, which allows the model to possess the ability to make single-step action decision. Secondly, the policy gradient based reinforcement learning is improved so that the tracker can capture the object by selecting more appropriate action and eliminating the useless action. Thirdly, the meta-learning based online adaptive update scheme is proposed to pursue the optimal parameters for the network so that the tracker can quickly adapt to new tracking tasks.

However, there are some following deficiencies in previous model. In the training phase based on deep reinforcement learning, previous model directly selects the action with the highest conditional probability as the optimal one. In fact, at the beginning of the training phase, we do not know which action is a good one. If a bad action is selected, it will easily lead the network to sample this action all the time, which is not conducive to training and exploring more appropriate action. Secondly, it's rewards for action are performed by dividing a video into several pieces. Specially, for a certain piece of training sequence, older model only focuses on whether the first frame and the last frame track successfully or not, and then assigns the calculated reward to all action corresponding to all frames. Such the reward method is not very reasonable. For some frames that fail to track, they also receive a positive reward if the tracker re-tracks the object in the next few frames. In addition, for a certain frame that tracks success, there may also be some useless action, which leads to a decrease in tracking efficiency. Moreover, there are insufficient tracking training sequences to fine-tune the network for specific tasks through pre-trained or transfer learning, like large-scale classified networks such as ImageNet [14], ResNet [1], and VGGNet [2].For the first issue raised above, the deep reinforcement learning based training method is improved to enhance the tracking accuracy and efficiency .Instead of selecting the action with the highest conditional probability, we directly sample an action according to the conditional probability distribution of all action to explore more appropriate action.

The primary contributions of this paper are summarized as follows. (1) The reinforcement learning based training method is improved by redesigning the decision criteria for optimal action to explore more appropriate action and redesigning the reward function to eliminate useless action. (2) A meta-learning based online adaptive update scheme for model parameters is proposed to adapt to the appearance changes or deformation of the object in the actual tracking.(3) The multi-domain training is incorporated into the ADNet model to further enhance the ability to learn the generic representation of different objects.

II. RELATED WORK

A video is a sequence of frames. Each frame can therefore be processed separately to reach an overarching goal. By keeping track of the spatial and temporal changes such as position, size, and shape between images, tracking can be accomplished. Object detection in videos verifies the presence of an object in an image. The object tracking occurs when that same object is followed through consecutive frames in a video feed. Visual object tracking can be generally divided into two categories, including a generative-based approach and a discriminate-based approach. Among them, the generative based approach usually extracts the features and learns them to generate a model representing the appearance of the object, and then the region that best matches the model is considered as the object in the image. While the discriminate based approach, from a mathematical point of view, poses the tracking problem as a binary classification problem, and its task is to distinguish the object from the surrounding background.

Convolutional neural network (CNN) has been proven with outstanding performance in a wide range of computer vision applications [2][7]. Despite this, the early CNN application [22] in tracking suffered from the data deficiency problem for training its network. To solve the data insufficiency, the transfer methods [9], [23] were proposed by using the pre-trained classification dataset (such as Image Net [18]). However, these methods still have limitations due to the gap between image classication and object tracking. The recently proposed methods [23], [24] tried to defeat this gap by training the networks with a large number of tracking training datasets [25]. Tao et al.[24] designed a Siamese deep neural network to learn a matching function, which is used to pursue the most similar candidates in a new frame through using only the original observation of the object from the rst frame. Cuiet al.[26] proposed a Recurrently Target-attending Tracker (RTT) by using multi-directional recurrent neural network to identify and exploit reliablepatches that facilitate the entire tracking process.

REINFORCEMENT LEARNING

Reinforcement Learning as described in [58], is a learning framework where an agent/learner interacts with the environment in a trial and error manner. In contrast with other machine learning methods, the agent is not told the proper actions to take. Instead, the agent explores the environment to achieve the maximum amount of future rewards (or statistically the highest sum of expected rewards), usually in search for a goal/objective (or a target space) represented numerically by a large reward .Therefore, this approach slightly differs from supervised learning, when, as knowledgeable designers, directly sample interactions and apply statistical pattern recognition. Since an agent is actuating and learning simultaneously, better methods need to be considered. Additionally, considering the hypothesis of a zero a priori knowledge about the environment a proper exploration/exploitation strategy is required. The final goal translates into progressively improve the sequence of actions given states of the environment, ultimately attaining the best policy/behavior for our agent. Every learning problem portrays similar modules: a learner (agent), a teacher (reward function), a performance measurement of how well the learning agent is behaving (usually tests represented numerically by rewards) and a couple more variables to be considered; several properties comprising a Reinforcement Learning problem are unraveled. In a Reinforcement Learning framework, an agent learns by reinforcement (as in psychology). A negative reward leads to an undesirable behavior. Conversely, a sequence of positive rewards lead towards a appreciable policy. As studied in [58], the aim of our agent is to maximize the accumulate sum of rewards:

III. ROPOSED TRACKER:

Proposed tracker is developed to increase tracking accuracy and efficiency. The multi-domain training is soul of algorithm which makes the combination of training effective for proposed work. Selection criteria for optimal action and the reward function are redesigned separately to explore more appropriate action and eliminate useless action. an effective online adaptive update strategy is proposed to adapt to the appearance changes or deformation of the object during actual tracking. Experimental results demonstrate that the proposed tracker has advantages over ADNet and other techniques in terms of accuracy and efficiency .To improve the robustness and real-time performance of the tracker, research has been improved from three aspects. Firstly, the use of multi-domain training instead of supervised learning based training enables the tracker to learn the shared representation of different objects in the various training sequences, which allows the model to possess the ability to make single-step action decision. Secondly, the policy gradient based reinforcement learning is improved so that the tracker can capture the object by selecting more appropriate action and eliminating the useless action. Thirdly, the meta-learning based online adaptive update scheme is proposed to pursue the optimal parameters for the network so that the tracker can quickly adapt to new tracking tasks.

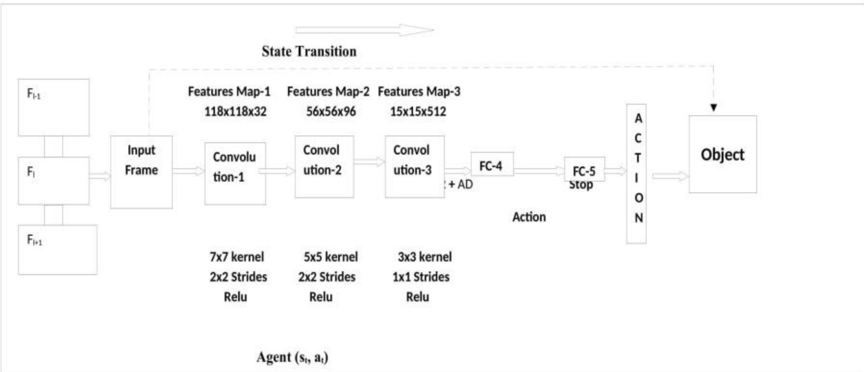

Figure 1 RODRLM Model

Proposed technique consists 11 kinds of action , these action can be divided into three categories: translation, scale change, and stop, where, the translation moves consist of four directional moves, {left, right, up, down}, and also include their two times larger moves; the scale changes are defined {scale up, scale down} [24].

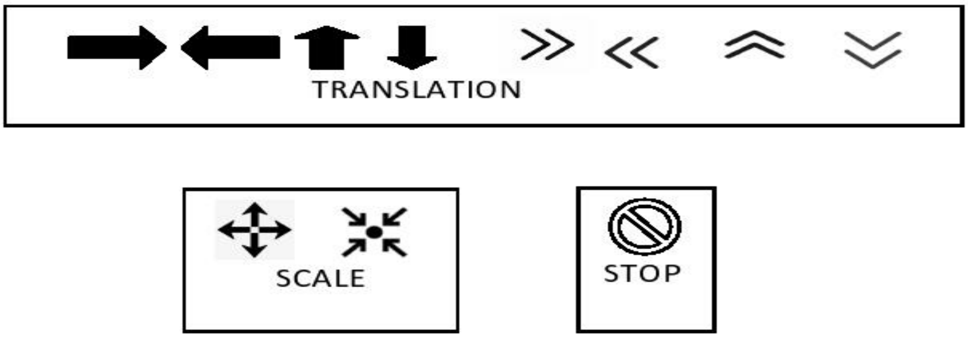

Figure 2 Action Dynamics

Take s as input of the network to calculate the conditional probability of 11 actions. To increase the degree of exploring more appropriate action, we randomly sampled an action from the action space $\{a_1, a_2, \ldots \ldots a_{11}\}$ directly according to the conditional probability distribution of all action,

$a = a_i , a_i$ is sampled from $p\ (a \in \{a_1 \ldots \ldots \ldots a_{11}\} | s;\ W_{rl})$,

the new ``box" and the new ``patch" are used as the input to the network to decide the next appropriate action until the stop action is selected or the number of action reaches a certain threshold.

Finally stop action is the only action which stops the process and indicates successful identification of single object and terminates process and place the mask on initial position for multiple objects present in the environment it will restart process . With the stop action mask will overlap ground truth and intersection over union (IOU) computed which is used in many visual attention models to calculate the effective identification

REWARD FUNCTION

A reward in RL is part of the feedback from the environment. When an agent interacts with the environment, it can observe the changes in the state and reward signal through its actions, if there is change. It can then use this reward signal (can be positive for a good action or negative for a bad action) to draw conclusions about how to behave in a state. The goal, in general, is to solve a given task with the maximum reward possible.

$$W_{rl} \leftarrow W_{rl} + \sum_{l}^{L} \sum_{l}^{T} \Delta_{W_{rl}} \log p\ (a_t , l | s_t , l ; W_{rl}) * r_{t,l,}$$

Where $r_{t,l}$ denotes the reward for each action, $\nabla_{W_{rl}}$ denotes the gradient of W_{rl} , t and l and denote the time steps t= 1,.......T and the frame indices l= 1,..........L

, respectively .If the reward is positive, the probability of the corresponding action update is relatively large; otherwise, the probability of the corresponding action update is relatively small. The proposed reward function allocates each action with a relatively accurate reward, which can greatly enhance the performance of actual tracking.

POLICY FOR REAL TIME OBJECT DETECTION

The intention of the agent is to find a policy to convert bounding box in a manner that it find suitable action which enhance the reward when it interact with the environments. Policy will guide the agent to take decision at any state s it needs to select action a . Proposed method solve this problem by formulate problem as a deep reinforcement learning problem using multi domain -Learning .

We used algorithm proposed by R. J. Williams, et al[100] for policy gradient algorithm. Approach works on calculating action value function using neural network and work better than traditional ADN-net learning approach. The frame is crop from the input video and only process once to estimate all possible action value , also gather various features with reply memory which improve data efficiency .

The policy gradient based reinforcement learning is improved so that the tracker can capture the object by selecting more appropriate action and eliminating the useless action. The meta-learning based online adaptive update scheme is proposed to pursue the optimal parameters for the network so that the tracker can quickly adapt to new tracking tasks ,After the multi-domain training is incorporated into ADNet to replace the supervised learning-based training, the model has the ability to make single-step action decision. But in actual tracking, the model must possess the ability to make multistep sequential action decision to finalize the location of the object in each frame .All above problem guide us to the deep reinforcement- learning which is a model -free algorithm. Moreover the huge state and action space, the input data is high dimensional images by which it is aiming to find out concepts such as object representation openly from the raw input data. Thus, to integrate the two ideas of Q-learning and learning from raw input data the technique called MDL is applied.We are using multi -domain training network, which is used to predict a single -step action for object tracking . It receive 112×112 three channel patches as the input ,and has six hidden layers , including three convolutional layers and three fully connected layers and produce output as 11 kinds of actions for object detection.Customized OTB dataset is used for implementing proposed scheme ,For this we have inserted many other video which used for training and testing purposes for proposed model.

We implemented this work with Anaconda python 3.5 combined with tensor flow. The experiments are carried out in TensorFlow [28] software installed on a 64-bit Windows OS, which runs on an Intel(R) Core(TM) i7-7700k CPU @3.60GHz, 32GB RAM.

IV RESULTS AND DISCUSSION

The performance of the proposed approach is evaluated based on the challenging customized online object tracking benchmark (OTB) [28] in the experiment.VOT2013 [29], VOT2014 [30], VOT2015 [31] datasets were selected as training samples, and VOT-100 dataset [28] with 100 video sequences (includingOTB-50) was selected as the test samples. To further validate the effectiveness of our tracker, 9 state of- the-art trackers, including ADNet [24], CCOT [32], MDNet [23], MCPF [23], ECO-HC [45], DCFNet [26], MEEM [31], DSST [46] and KCF [21], are selected to compare with our tracked object detection model. Table 1 show the precision and success performance based on center location error and overlap ratio, respectively.

Precision describes as accuracy of measurement and intersection of union tells how much object actually detected in particular action. Calculation of both gives the accuracy of object detection .following table compares proposed technique with other existing research and gives the results and with these results we can say that proposed research leads object detection in smarter way.

The progress of each part enhances the actual overall performance of the real time tracker to some extent. Among them, the improvement of meta-learning based random update is the most obvious, which also indirectly indicates that the appropriate parameters are very helpful for accommodating the appearance changes or deformation of the object. At the first frame, the ground truth is added with the Gaussian noise and sampled to generate positive and negative sample sets. Last 10 frames are taken with IoU >0.6 to improve confidence for tracker. Precision and success rates are shown in following graph for calculating precision vs. center location error one pass for frames considered. Compared with different research and results shows the enhancement in factor compared to other method.

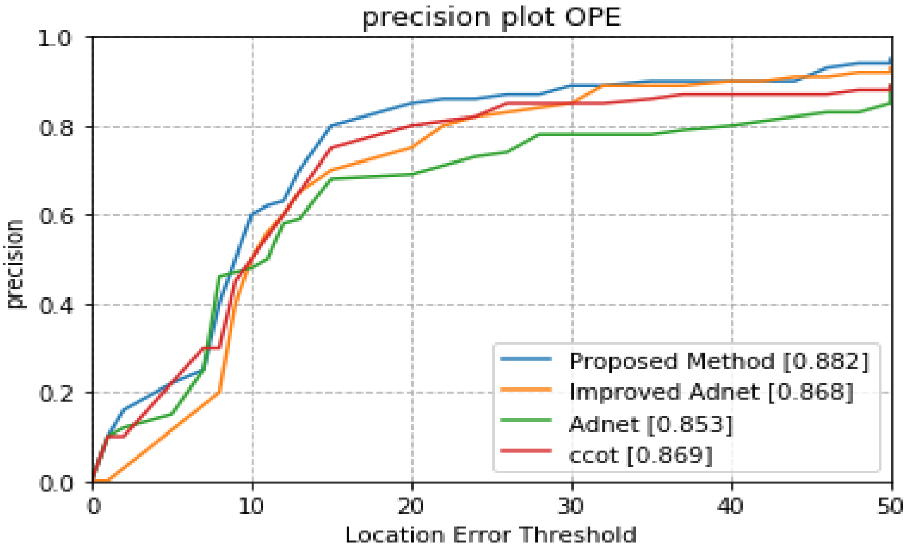

Figure 3 Precision Comparison chart for Proposed method with others

As shown in plot proposed technique gives 0.882 precision compared to improved ADNET [0.868], ADNET [0.853], and CCOT [0.869]. Reinforcement training helped to calculate action dynamics in such a way that minimum action is used to track correct object in video sequence Softmax calculates the value of action which is used to calculate precision.

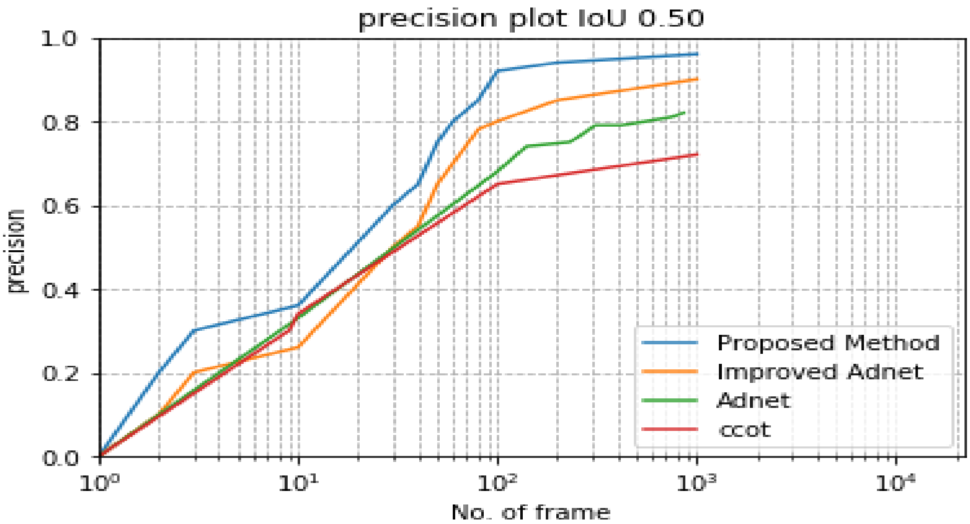

Figure 3 Comparison chart for Proposed method with others

In the experiment, the ratio of the frames using re-detection to the whole frames was around 9%, and the ratio of the frames requiring more than five actions to capture the target to the whole frames was only around 4%, that is, most of the frames require fewer than five actions to pursue the target in each frame and this is the strength of this research.

COMPARISON OF EVALUATION INDICATORS OF DIFFERENT RESEARCH WITH PROPOSED MODEL

To further validate the effectiveness of our tracker, 9 state-of- the-art trackers, including ADNet [24], CCOT [32], MDNet [23], MCPF [23], ECO-HC [45], DCFNet [26], MEEM [31], DSST [46] and KCF [21], are selected to compare with our tracker. Table 1 show the precision and success performance based on center location error and overlap ratio, respectively

table 1.

	PROPOSED	ADNET	CCOT	MDNET	ADNET	MCPF	ECO-HC	DCFNET	MEEM	KCF	DSST
Precision	87.1%	86.8%	86.9%	86.3%	84.4%	84.5 %	82.7 %	77.8%	75.4 %	66.8 %	66.5 %
IoU	.655	.651	.653	.648	.628	.613	.625	.598	.515	.462	.502

Precision and IoU performance

	PROPOSED	ADNET	CCOT	MDNET	ADNET	MCPF	ECO-HC	DCFNET	MEEM	KCF	DSST
FPS	**4.1**	**3.9**	**<1**	**<1**	**3.0**	**<1**	**16.0**	**29.0**	**20.1**	**124.6**	**16.1**
GPU	**X**	**O**	**O**	**O**	**O**	**O**	**X**	**O**	**X**	**X**	**X**

Table 2 Comparison of frame per second (FPS) of RODRLM vs. others.

COMPUTATION OF TRAINING TIME FOR RODRLM

RODRLM executed all instructions in 10 th generation Intel core I7 CPU. Because it's way too easy to train the real time detector with minimum feature and eliminate complex computation as well. Following table will give you the view of training time comparison with other algorithm and different model.

Table 3 Training Time Comparisons

Method	Training Time
Atari 2600 Games Gorilla[45]	*4 days*
A3C, FF [45]	*1 day on CPU*
Faster RCNN on image data[46]	*18 hour*
SSD [46]	*14*
YoloV3[46]	*10*
Proposed RODRLM-NET	*9 hr 54minutes*

V RESULTS DISCUSSION

The use of deep reinforcement learning (DRL) enables even partially labelled data to be effectively utilized for multi domain learning. Through evaluation of the customize OTB/VOT dataset, the proposed tracker is validated to accomplish a competitive performance that is three times quicker than state-of-the-art, deep network–based trackers. The plots of precision and success rate for one-pass evaluation (OPE) including the area-under-the curve (AUC) score over all 100 video sequences are presented. It can be seen from Figure that our improvements are effective.

Precision and success rates are shown in graph 6.1 for calculating precision vs. center location error one pass for frames considered. Compared with different research and results shows the enhancement in factor compared to other method. In Fig no 6.2 -- Overlap Ratio and precision Shown and compared with existing research and better performance presented. detectors claims that having more than seventy percent intersection area is covered between predicted and ground truth box many researchers claims the accuracy is better than other detection shown. Table 6.1, proposed research succeeded to enhance the quality of real time object detection ADNET having 86.8% precision which also gives the better quality of detection we found that multi domain training is technique which takes the proposed work in better way.shows comparison of computational speed of different trackers. In this FPS

(Frame per Second) defines how rapidly your object detection model processes your video and generates the required output. ADNet, the proposed tracker has significant advantages in terms of precision and success rate, and its tracking speed is also faster by 1 frame per second. Than training time computation and comparison shown in table 6.3 and that training time of RODRLM-net is very comparative and better.

Finally compare and validated Performance of RODRLM-net vs. model in literature and shown the competitive performance.

VI Conclusion and Future work

Proposed work uses model named RODRLM-net which contains three convolutions and two fully connected layers. Convolution layer used for feature extraction and identification of object is done with deep reinforcement learning .Softmax classifier is used for defining the label of class. The main contribution of this project is to defining model and combining the multi domain training with Meta learning in this way executed novel deep learning framework, using Tensor flow. This implementation can simply be customized by future researchers in sort to conduct additional experiments with different datasets, neural network architecture, or the agent architecture.

limitation of existing models is overcome through the multi-domain training ,improve the criteria of the most favourable action selection and three convolution network layer ,with increase patches and reinforcement learning based training and utilizing meta -learning for the parameters of the network during object tracking to improve the efficiency .For video 128 training patches and 8 frames are randomly selected for each patch size of frame .Proposed tracker consist fewer layers in network and its computational complexity is lower as compared to existing deep networks for tracking. According to the experimental results, the speed of the proposed tracker improves 4.2 frames per second, which is 38.5 % faster than the ADNet. Proposed tracker is compared with many existing network tracker, and we found much better result as previous tracker in precision and success rate.

The research also can find the distance between object by making suitable changes in architecture and algorithm. Today we all facing covid-19 pandemic for making social distancing in professional way we need to calculate distance between objects and may organize the behaviour and many other factor related to object. For this type of computation we need more complex algorithm and architecture.

Finally, we need to believe that we require object detection systems for nano-robots or for robots that will discover areas that have not been seen by humans, such as depth parts of the ocean or other planets, and the detection systems will have to discover to new object classes as they are encountered.

References

1. K. He, X. Zhang, S. Ren, and J. Sun, ``Deep residual learning for image recognition," in Proc. CVPR, Las Vegas, NV, USA, Jun. 2016,pp. 770-778.
2. K. Simonyan and A. Zisserman, ``Very deep convolutional networks for large-scale image recognition," in Proc. ICLR, 2015,pp.1-14
3. Caselitz T. Tipaldi G.D. Dewan, A. and W. Burgard. Motion-based detection and tracking in 3d lidar scans. In 2016 IEEE International Conference on Robotics and Automation (ICRA), pages 4508–4513, 2016
4. Ross Girshick, Jeff Donahue, Student Member, IEEE, Trevor Darrell, Member ,IEEE, and Jitendra Malik, Fellow, IEEE. “Region-Based Convolutional Networks for Accurate Object Detection and segmentation” Copyright © 2016, IEEE
5. Bouvrie J (2006) 1 Introduction Notes on Convolutional Neural Networks. do http://dx.doi.org/10.1016/j.protcy.2014.09.007
6. Bozorg et al. Bahraini. SLAM in dynamic environments via ML-RANSAC.Mecha- tronics,49:105–118, 2018
7. K. He, G. Gkioxari, P. Dollar, and R. Girshick, ``Mask R-CNN,"IEEE Trans. Pattern Anal. Mach. Intell., to be published. doi:10.1109/ TPAMI.2018.2844175
8. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proceedings of the IEEE conference on computer vision and pattern recognition , 2016, pp. 779–788
9. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille``DeepLab: Semantic image segmentation with deep Convolutional nets,atrous convolution, and fully connected CRFs,"IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 4, pp. 834-848, Apr. 2018
10. H. Nam and B. Han, ``Learning multi-domain convolutional neural networks for visual tracking," in Proc. CVPR, Las Vegas, NV, USA, Jun. 2016,pp. 4293-4302
11. Q. Wang, J. Gao, J. Xing, M. Zhang, and W. Hu, ``DCFNet: Discriminant correlation lters network for visual tracking," 2017,arXiv:1704.04057.[Online]. Available: https://arxiv.org/abs/1704.04057
12. S. Yun, J. Choi, Y. Yoo, K. Yun, and J. Y. Choi, ``Action-decision networks for visual tracking with deep reinforcement learning," in Proc. CVPR ,Honolulu, HI, USA, Jul. 2017, pp. 13491358.
13. E. Park and A. C. Berg, ``Meta-tracker: Fast and robust online adaptation for visual object trackers," in Proc. ECCV, Munich, Germany, 2018,pp. 587-604
14. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang,A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, ``ImageNet large scale visual recognition challenge,"Int. J. Comput. Vis.,vol. 115, no. 3, pp. 211-252, Dec. 2015
15. S. Ray Essentials of machine learning algorithms (with python and codes), 2017.[Online]. Available: http://www.analyticsvidhya.com/blog/2015/08/common machine learning-algorithms
16. Ross Girshick, Jeff Donahue, Student Member, IEEE, Trevor Darrell, Member, IEEE, and Jitendra Malik, Fellow, IEEE. “Region-Based Convolutional Networks for Accurate Object Detection and segmentation ”Copyright © 2016, IEEE
17. Ross Girshick “Fast R-CNN”arXiv:1504.08083v2[cs.CV].

18. Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun "Faster R CNN: Towards Real-Time Object Detection with Region Proposal Networks". arXiv:1506.01497.
19. H. Nam and B. Han. Learning multi-domain convolution neural networks for visual tracking. arXiv preprint arXiv:1510.07945, 2015.
20. Y. Wu, J. Lim, and M.-H. Yang. Object tracking benchmark. IEEE Transaction on Pattern Analysis and Machine Intelligence, 37(9):1834–1848, 2015.
21. M. Kristan, R. P flug felder, A. Leonardis, J. Matas, L. Cˇ ehovin, G. Nebehay, T. Vojir, G. Fernandez, and A. Lukeˇziˇc. The visual object tracking vot2014 challenge results. In Visual Object Tracking Workshop 2014 at ECCV2014, 2014.
22. H. Li, Y. Li, and F. Porikli, ``Robust online visual tracking with a single Convolutional neural network,'' in Proc. ACCV . Cham, Switzerland: Springer, 2014, pp. 194-209
23. C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang, ``Robust visual tracking via hierarchical Convolutional features,''IEEE Trans. Pattern Anal. Mach.Intell., to be published. doi:10.1109/TPAMI.2018.2865311
24. R. Tao, E. Gavves, and A. W. M. Smeulders, ``Siamese instance search for tracking,'' in Proc. CVPR, Las Vegas, NV, USA, Jun. 2016, pp. 1420-1429.
25. Y. Wu, J. Lim, and M. H. Yang, ``Object tracking benchmark,'' IEEE Trans .Pattern Anal. Mach. Intell., vol. 37, no. 9, pp. 1834-1848, Sep. 2015
26. Z. Cui, S. Xiao, J. Feng, and S. Yan, ``Recurrently target-attending tracking,'' in Proc. CVPR, Las Vegas, NV, USA, Jun. 2016, pp. 1449-14
27. D. Hubel and T. Wiesel. Receptive fields and functional architecture of monkey striate cortex .Journal of Physiology (London), 195, pages 215–243, 1980.
28. K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics, 36,pages 193–202, 1980.
29. Bottou L. Bengio Y. LeCun, Y. and P. Haffner. Gradient-based learning applied to document recognition .Proceedings of the IEEE, 86(11) , pages 2278–2324, 1980.
30. Wolf L. Bileschi S. Serre, T. and M. Riesenhuber. Robust object recognition with cortex-like mechanisms .IEEE Trans. Pattern Anal. Mach. Intell., 29(3), pages 411– 426,1980.
31. Samer L. Hijazi, Rishi Kumar, and Chris Rowen. Using convolutional neural networks for image recognition. 2015.