Moving Object Tracking and Detection Based on Deep Learning

Mohammad Sanaul Haque, Beiji zou, Ishtiak Al Mamoon

Abstract— Object tracking is a key step in computer vision for video surveillance, public safety, and traffic analysis. Object detection and tracking are the two correlated components of Video Surveillance. Object detection in videos is the first step before performing complicated tasks such as tracking. Deep learning neural networks is a powerful programming paradigm which learns multiple levels of representation and abstraction of data such as images, sound, and text. In this paper Gaussian mixture model (GMM) based object detection, deep learning neural network-based recognition and tracking of objects using correlation filter is proposed, which can handle false detections, with improving the efficiency. The algorithm is designed to detect only cars and humans' while the performance is analyzed using True Positive Rate (TPR) and False Alarm Rate (FAR) as probabilistic metrics. The Experimental results of the proposed method are found to be better with an accuracy of 88%.

Index Terms— Object Detection, Object Tracking, Foreground Detection, Deep Learning, Computer Vision.

1 Introduction

THIS automatic detection, classification, and tracking of a number of objects is a challenging task for the wide range of applications, such as security, surveillance, traffic control and human-computer interaction. Real time tracking is still a problem and an active research topic because of object occlusions, random movements, complex background and varying illumination [1]. Moving object can be detected in two ways, (i) motion detection and (ii) motion estimation [2]. Motion detection is identifying varying regions from video frames using the fixed camera when objects are moving. Difficulties in object detection using the fixed camera are:

- Objects movement in a scene from frame to frame does not have constant movement.
- Objects in the scene may stop for some time and move further.
- Objects in the scene have different velocities.
- Moving objects in the scene may not cover a significant area of the frame this leads to recognition problem [3].

Substantial information about moving objects is required from frame to frame to track the objects properly. Deep Convolutional neural networks are best suited to do recognition and tracking of objects. Convolutional neural networks are the powerful visual model which has shown significant performance in many visual recognition problems.

A convolutional neural network is a combination of stacked convolutional layers and spatial pooling layers which are stacked alternately. The convolutional layer extracts feature maps using linear convolutional filters and nonlinear activation functions.

Mohammad Sanaul Haque, School of Information Science & Engineering Central South University, Hunan, China. Email: sana.sana@csu.edu.cn

Beiji Zou School of Information Science & Engineering Central South University, Hunan, China. Email: bjzou@csu.edu.cn

Ishtiak Al Mamoon, Eletrical and Computer Engineering Department Presidency University, Dhaka, Bangladesh, Email: ishtiakm@pu.edu.bd

Spatial pooling performs grouping of the local features using spatially adjacent pixels, this improves the robustness towards objects deformation [4].

In Convolutional Neural Networks (CNN) original image can be used as input without pre-processing of the image. CNN has been showing the best accuracy in large-scale image classification/recognition since it is combined with deep learning [5]. Researchers optimize the CNN model structure by using parameters to improve the accuracy. Most of the improved models use more time to train and test [6]. CNN models involve large data, which is to say the training images cannot be too few. Considering the advantages of background subtraction based object detection and deep learning neural network based tracking is employed in this work.

The organization of this manuscript is as follows. Section 2 introduces the related work and methods. In Section 3, the proposed method of combining foreground detector based object detection and deep learning neural network-correlation filter based object tracking is presented in detail. Experiments and result discussion are showed in Section 4 and finally a brief conclusion of the proposed work in Section 5.

2 Background study and Literature Review

There are many methods available for object detection which includes, frame differential method, background subtraction, and optical flow [7, 8]. Frame differential method calculates the absolute differences of the consecutive video frames [9, 10] followed by a threshold function to determine the changes. The goal of this method is to identify certain pixels in an image which is moving or static. If the threshold is not optimal, some of the frame differential methods, mentioned in [11], suffer from the problem of producing images which can corrupt by spot noises. Background subtraction method is used to detect moving object from the static background [12, 13]. This method uses the previous information of the image or some statistical information of the pixel in the video frames to build the background model. There are two main components in the

frame pixels. The First component with the largest variance belongs to background pixels; the second component contains pixels of the foreground. After matching the components background model needs to be updated. Optical flow method uses optical flow field and features of an object in an image [14, 15]. For a moving object, optical flow method uses the 3D object velocity mapped into 2D imaging surface, which is also known as image velocity. If there is no moving object in the imaging scene the optical flow field vectors remain smooth over the entire area. Moving objects carry different velocity vectors in the background scene [16].

Many applications like visual recognition, speech recognition, and language processing require deep learning because of its good performance. Convolutional neural networks are widely used and studied among different types of deep learning neural network. During early days training data available was very less and computational capacity of systems was poor led to the problem of training of convolutional neural network without over fitting. The increased volume of the annotated data and the computational strengths of graphics processor units, convolutional neural networks training and validation became easy which increased the interest of researchers and they achieved state-of-the-art results on various tasks [17].

An on-line AdaBoost framework with Deep learning neural network architecture is proposed in [18], that having multi-level feature learning ability from a set of auxiliary images. An object proposal network with bounding box candidates to mitigate the object model refitting by decreasing hard negatives is proposed in [19]. A framework using CNN to coordinate object recognition system which learns and gains from images is proposed in [20]. This framework gathers images automatically in ordered classifications and learns images in high exactness, different On-Board PC can share proposed learning framework. A quick, completely parameterizable GPU execution of Convolutional neural network variations is proposed in [21]. These feature extractors are learned in a supervised way. These deep hierarchical designs accomplish the best-distributed results on benchmarks for object classification and handwritten digit recognition.

A variant of traditional convolution neural networks for multiple image recognition is proposed in [22]. This system uses binaries norm gradients (BING) method to recognize images followed by vectorization of deep convolutional neural networks which reduces network training and testing time. The advancement of the CPU, GPU, and the advancement of the parallel processing systems, led to the profundity extension of training data of convolutional neural networks. Drop-out method [23], rectified Linear Units [24], Stochastic Pool techniques [25] and other methods, increases the speed of the neural network training and reduces the over fitting problem. VGG model, Google Net models, Theano models, and Caffe models, enormously enhance the measure of image recognition. The existing methods may fail to specifically recognize/classify an object accurately because they are trained for a large number of object classes.

3 Proposed Methodology

The proposed method is depicted in Fig. 1. It consists of Gaussian Mixture Model (GMM) based foreground detector which detects all the moving objects in the input video frames. The detected moving object is cropped and given as an input to a pre-trained deep convolution neural network to recognize it as human or car. After recognition, if it belongs to the defined classes it is fed to the correlation filter to track till it disappears from the video frame.

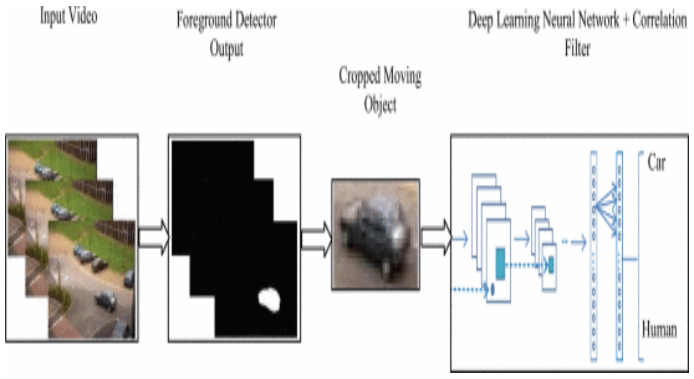

Fig. 1. Generic model of the proposed method

3.1 Foreground Detector Using Gaussian Mixture Model (GMM)

Define Gaussian mixture model in [26] is one of the background subtraction technique, which is mainly used for object detection and it is the base for many visual recognition application. The background in the video frames contains multidimensional variations and do not contain static components. The Gaussian mixture model is most suitable to understand the distribution of every pixel value in the background. The pixel value distribution can be described using (1).

$$V_1,V_2,V_3,\ldots\ldots V_t = I(x_i,y_i);\ 1 \le i \le t \qquad (1)$$

In (1), V_t denotes pixel value at time t in the video frame. x_i, y_i denotes the coordinate of the pixel in the frame. The probability of any pixel in a multidimensional Gaussian distribution is given as (2).

$$P(Vt) = \sum i=1n\omega i,\ t*f\ (Vt \mid \mu i,t,\sigma i,t) \qquad (2)$$

Where n is the number of Gaussian distribution which depends on background complexity. Generally, n is set between 3 and 5. ω_i,t is the weight function, μ_i,t is mean and $\sigma i,t$ is the covariance of ith Gaussian distribution at time t. $f(V_t \mid \mu_i,t,\sigma_i,t)$ is the normal distribution and the probability density function is given as in (3).

$$f\ (V \mid \mu,\sigma)=12\pi m2 \mid \sigma \mid 12e-12(Vt\mu)T\sigma-1(Vt-\mu) \qquad (3)$$

In the above expression, m is the dimension of pixel value V_t; σ is the covariance of the grey pixel if V_t is a greyscale pixel. Suppose V_t is RGB value then σ is the covariance matrix of

many dimensions. The components of RGB are independent of each other with the same variance which reduces computational complexity [16]. Multidimensional covariance matrix can be expressed as in (4).

$$\sigma=\begin{pmatrix} \sigma R2000\sigma G2000\sigma B2 \end{pmatrix} =\sigma 2I \quad (4)$$

σ2R , σ2G and σ2B are the variances of R, G, and B components respectively. σ2R=σ2G=σ2B=σ, where σ is the standard deviation of R, G, and B pixels. According to (2) GMM individual pixel distribution can be formulated using the parameters n, ωi,t, μi,t and σi,t based on the background complexity. Once the GMM is initialized, the variables of Gaussian distribution can be updated using (5).

$$\omega\hat{}i,t=(1-\alpha)\omega i,t+\alpha f(Vt\,|\,\mu i,t,\sigma i,t)\mu\hat{}i,t = (1-\rho)\omega i,t+\rho Vt\sigma\hat{}2i,t$$
$$\omega\hat{}i,t =(1-\rho)\sigma 2i,t+\rho(Vt-\mu i,t)T(Vt-\mu i,t) \quad (5)$$

Where ωˆi,t,μ^i,t, and σˆ2i,t are the parameters used for the evaluation of ωi,t ,μi,t and σi,tat time t. To define the update speed the learning rate parameter α is used, which is expressed as below in (6)

$$\rho=\alpha 12\pi m2\,|\,\sigma\,|\,12e-12(V_{t\mu})T_{\sigma}-1(V_{t-\mu}) \quad (6)$$

V_t and the k_{th} Gaussian distribution is used in order to check whether Vt and μi,t, match according to (7)

$$|\,V_{t,\mu n,t}\,|<D\sigma_{i,t} \quad (7)$$

In (7) if both the terms match then $P(k\,|\,V_{t,\mu i,t,\ \sigma i,t})=1$, otherwise $P(k\,|\,V_{t,\mu i,t,\ \sigma i,t})=0$; D usually is equal to 2.5. The mixture of Gaussian distribution has to be updated and finally, B Gaussian distributions are used as representation of the background based on the following equation (8):

$$B=\text{argminb}\,(\sum i=1b>T) \quad (8)$$

Here T is the minimum weights proportion that represents the background. The value of mean and variance do not change if the model does not match with the current pixel value, only weight is updated. This algorithm consumes time but very effective in background removal [16].

3.2 Deep Learning Neural Network

Deep learning neural network uses larger data sets in order to perform efficient recognition/classification. The Realistic scenario shows considerable variation, in order to recognize objects it is essential to use larger training data sets. Recently it has become possible to use labeled data with millions of images. Deep neural networks are hard to train but the deeper training provides best results when compared to shallow networks.

Some of the data sets include LableMe [27] which has thousands of segmented images and ImageNet [28] which has more than fifteen million labeled images. Deep CNN's can be treated as modern multilayer perceptrons. These algorithms have become state of the art in computer vision challenges [29]. Recent advances provide affordable GPU's to use larger datasets for training & validation and also a platform for researchers to have further insight into more complex (deeper) network models [30]. Traditional NN (neural network) add one or two hidden layers but CNN can have more hidden layer [31, 32].

Designing a Deep learning CNN's with some numbers of hidden layers is an application or designer dependent. CNN includes convolutional layers which are called feature extraction layer, fully connected which is an intermediate layer, pooling layer which is used for dimensionality reduction and nonlinear function like sigmoid, hyperbolic and rectified units.

The architecture of ImageNet-Caffe-Alexnet which is used in this work is shown in Fig. 2. AlexNet has 8 weight layers of which 5 are convolutional layers and last 3 are fully-connected layers, and it consists of 3 max-pooling layers following the first, second and fifth convolutional layers. 96 filters are there in first convolutional layer each of size 11×11 with a stride of 4 pixels and padding with 2 pixels. Other layers have a stride and padding of 1 pixel. 256 filters are there in second convolutional layer each of size 5×5. The third convolutional layer has 384 filters, fourth has 384 and the fifth layer has 256 filters each of these with a size of 3×3. Alexnet has rectified linear unit (ReLu) for handling nonlinearity which is given by

$$f(x)=\max(0, x)$$

The advantage of the ReLu is it provides faster training when compared to traditional sigmoid and Tanh functions.

Local receptive fields and shared weight are the two components of CNN [33]. Local information in the small region of the image is termed as local receptive fields. Shared weight and biases for neurons in the hidden layers of CNN has many advantages. In [34]it has been shown that first convolutional layers act as Gabor filter. Using many convolutional layers provides a broad feature matrix. The advantage of shared weights is that the number of parameters used in the network decreases rapidly, which reduces the training times and also, helps us to construct deeper networks.

CNN's are susceptible to over fitting because of large parameterization and representation. Over fitting can be eliminated by using many techniques, which guarantees the generality of the learned parameters of the specific target problem. In the network, convolutional layers are followed by pooling layers to down sample the current representation of the image, which reduces the number of parameters, carried to the next layer and also increases the computational efficiency. Further, the use of drop out layer randomly reduces the hidden neurons in the training processes, which avoids the over fitting. The initial CNN training to perform generalized object recognition and classification can be optimized using a technique called transfer learning.

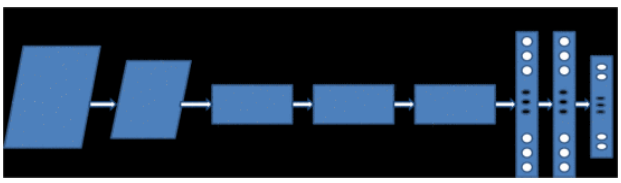

Fig. 2.Architecture of ImageNet-Caffe-Alexnet

3.3 Transfer Learning

In the proposed method, the pretrained ImageNet-Caffe-Alexnet network is used and optimized to perform object tracking. The hidden layers exhibit distinct feature representation characteristics in which lower layers are for features extraction and higher layers hold information that is specific to the recognition/classification task.

The Alexnet has 60 million parameters, 650,000 neurons, and trained over the ImageNet dataset on an image classification problem. Instead of designing a new CNN with random parameter initialization, which is time consuming it is better to use a pretrained CNN model and fine-tune its parameters to perform our specific recognition/classification domain. The last three fully connected layers of the pre-trained Alexnet can classify 1000 classes. So these three layers should be fine-tuned for our classification task. A fully connected layer, a softmax layer, and a classification output layer should transfer last three layers to the new classification task. The output of Deep learning neural network is the detected object with predefined class. This filters the information, what to be tracked. The selected objects from deep learning neural networks are used as the target to track using correlation filter.

3.4 Correlation Filter Used for Tracking

The recognized objects from deep learning neural network are treated as target initially in the first frame. The target is tracked by correlating the filter in next frame. The objects recognized by deep learning neural network in subsequent frames are used for tracking and the maximum correlation output value indicates the target and its new position. The coordinates of the object location is then updated based on that new location.

During tracking, changes in appearance of objects by changing the rotation, scale, pose and lighting variations are often. Therefore, the filters are required to adapt to these dynamic changes for tracking the object efficiently purpose [35]. The filter learned from frame i is computed as as (9):

$$H*i=\eta G_{io}F_{io} + (1-\eta)\, H*i-1 \qquad (9)$$

Where H and F are the 2D Fourier transform of the input image and of the filter H respectively and G is the correlation as given in (10).

$$G=F_o \qquad (10)$$

4 Expected Results and Discussion

The proposed method is examined to check the effectiveness and tracking efficiency. The proposed method is implemented in MATLAB 2018 installed on i5 processor system with Nvidia GEFORCE GT 710 GPU. The proposed method execution time is 0.4s per frame and it is evaluated using different video sequences from Performance Evaluation of Tracking and Surveillance (PETS) database [36] and also videos in [37], which contain many challenging parameters like random movement, scale variation, pose variation and shadows. The experimental evaluation of the proposed method for various videos is listed in Table I.

The original video frame, tracking results and detected object label is shown in Fig. 3.

Fig. 3. Output of the proposed method for different videos, first column is original video frames and second column is tracking results

(a) Frame #55 (b) Frame #163 (c) Frame # 374 (d) Frame #38 (e) Frame #735 (f) Frame #429

TABLE 1
PROPOSED METHOD IN DIFFERENT SENARIO

Videos	Proposed Method					
	Frame Number	TP	FP	FN	FAR	TPR
HighwayII_raw	55	8	0	0	0	1
	72	5	0	0	0	1
	108	4	1	0	0.2	1
Highway	163	3	0	0	0	1
	172	3	0	0	0	1
	190	2	2	1	0.5	0.66
Laboratory_rawII	374	1	0	0	0	1
	963	2	1	0	0.33	1
	972	1	1	0	0.5	1
View_001	38	4	0	0	0	1
	196	4	0	1	0	0.8
	246	5	0	1	0	0.83
View_003	735	7	0	1	0	0.87
	1025	8	1	2	0.11	0.8
	1884	8	2	1	0.2	0.88
View_007	429	3	1	0	0.25	1
	669	4	1	1	0.2	0.8
	1089	3	2	1	0.4	0.75

False Alarm Rate (FAR) is the ratio between total false positives to the sum of the total number of true positives and false positives in a frame.

$$FAR=FPTP+FP \qquad (11)$$

FAR is a metric used to find the miss detection percentage in a video.

True Positive Rate (TPR) is the ratio between the total true positives to the total number of objects in a frame.

$$TPR=TPTP+FN \quad (12)$$

TPR is a metric used to find the true detection percentage of the identified object in a video, where True Positive (TP) is the correct detection of the objects in the frame. True Negative (TN) is correctly detecting an object into negative. False Positive (FP) is detecting incorrect object into positive. False Negative (FN) is detecting an incorrect object into negative.

During the tracking of moving objects from the first frame till the last frame many objects enters the frame and leaves the frame. Some objects reappear in the frame at different time instants. So the object information should be suitably maintained by the deep learning neural network to track objects in an efficient way. In the proposed method label is assigned for every object recognized by the deep learning neural network. If the object leaves the frame the label of that object is deleted.

4 Conclusion

The moving object Detection and tracking using deep learning neural network is proposed in this paper. Transfer learning using deep learning neural network is effective because of its recognition accuracy. The proposed method is examined to track the objects effectively using the popular datasets and the results are analyzed using probabilistic approaches. The accuracy of the proposed method is 88%. The results prove that the proposed method is effective in tracking the moving objects.

References

[1] Mengjie Hu, Zhen Liu, Jingyu Zhang, Guangjun Zhang, "Robust object tracking via multi-cue fusion", Signal Processing, vol. 139, pp. 86-95, Oct. 2017.

[2] AM Tekalp, Digital video processing, New Jersey: Prentice Hall, 1995.

[3] B.N. Subudhi, S Ghosh, P.K. Nanda, A. Ghosh, "Moving object detection using spatio-temporal multilayer compound Markov Random Field and histogram thresholding based change detection", Multimedia Tools and Applications, vol. 76, no. 11, pp. 13511-13543, June 2017.

[4] Shiqi Yu, Sen Jia, Chunyan Xu, "Convolutional neural networks for hyperspectral image classification", Neurocomputing, vol. 219, pp. 88-98, Jan. 2017.

[5] Tianming Liang, Xinzheng Xu, Pengcheng Xiao, "A new image classification method based on modified condensed nearest neighbor and convolutional neural networks", Pattern Recognition Letters, vol. 94, pp. 105-111, July 2017.

[6] X.X. Niu, C.Y. Suen, "A novel hybrid CNN-SVM classifier for recognizing handwritten digits", Pattern Recognition, vol. 45, pp. 1318-1325, 2012.

[7] M Huang, S Yen, "A real-time and color-based computer vision for traffic monitoring system", IEEE International Conference on Multimedia and Expo (ICME 2004), vol. 3, pp. 2119-2122, 2004.

[8] S Peng, "Flow detection based on traffic video image processing", Journal of Multimedia, vol. 8, no. 5, pp. 519-526, Oct 2013.

[9] A Lipton, H Fuiyoshi, R Patil, "Moving target classification and tracking from real-time video", Proceedings of IEEE Workshop on Applications of Computer Vision, pp. 8-14, 1998.

[10] R. Zhang, L Yang, K Liu, X Liu, "Moving objective detection and its contours extraction using level set method", International Conference on Control Engineering and Communication Technology, pp. 778-781, 2012.

[11] P Spagnolo, T Dorazio, M Leo, A Distante, "Moving object segmentation by background subtraction and temporal analysis", Image and Vision Computing, vol. 24, no. 5, pp. 411-423, 2006.

[12] I Haritaoglu, D Harwood, L Davis, "W: Real-time surveillance of people and their activities", IEEE Trans. Pattern Anal. Mach. Intell, vol. 22, no. 8, pp. 809-830, 2000.

[13] S Mao et al., "Rapid vehicle logo region detection based on information theory", Computers and Electrical Engineering, vol. 39, no. 3, pp. 863-872, 2013.

[14] S Mao, M Ye, X Li, F Pang, J Zhou, "Performance of optical flow techniques", International Journal of Computer Vision, vol. 12, no. 1, pp. 42-77, 1994.

[15] J Barron, D Fleet, S Beauchemin, "Optical flow-motion history image (OF-MHI) for action recognition", Signal Image Video Processing, vol. 9, no. 8, pp. 1897-1906, 2015.

[16] D.M Tsai, W.Y. Chiu, M.H Lee, A fast method for moving object detection in video surveillance image, 2016, [online] Available: 10.1007/s11760-016-1030-2.

[17] J Gu et al., "Recent Advances in Convolutional Neural Networks", CoRR, vol. abs/1512.07108, 2015.

[18] Xiangzeng Zhou, Lei Xie, Peng Zhang, Yanning Zhang, "An Ensemble Of Deep Neural Networks For Object Tracking", Proceedings of IEEE International Conference on Image Processing (ICIP), pp. 843-847, 2014.

[19] Gao Zhu, Fatih Porikli, Hongdong Li, "Robust Visual Tracking with Deep Convolutional Neural Network based Object Proposals on PETS", IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1265-1272, 2016.

[20] J Hyeok, L Myung-jae, Young-guk Ha, "Integrated Learning System for Object Recognition from images based on Convolutional Neural Network", Proceedings of International Conference on Computational Science and Computational Intelligence, pp. 824-828, 2016.

[21] C.C Dan, Ueli Meier, Jonathan Masci, M.G Luca, S Jurgen, "Flexible High Performance Convolutional Neural Networks for Image Classification", Proceedings of the Twenty-Second international joint conference on Artificial Intelligence, vol. 2, pp. 1237-1242, 2011.

[22] Huiwei Shi, Xiaodong Mu, Shuyang Wang, "BVCNN: a multi-object image recognition method based on the convolutional neural networks", IEEE Proceedings of International Conference on Virtual Reality and Visualization, pp. 81-84, 2015.

[23] P Baldi, P J Sadowski, "Understanding dropout", Neural Information Processing Systems, pp. 2814-2822, 2013.

[24] V. Nair, G.E. Hinton, "Rectified linear units improve restricted boltzman machines", International Conference on Machine Learning, pp. 807-814, 2010.

[25] D.Z Mattew, F. Rob, "Stochastic pooling for regularization of deep convolutional neural networks", International Conference on Learning Representations, 2013.

[26] C Stauffer, W.E.L Grimson, "Adaptive background mixture models for real-time tracking", Proceedings of IEEE Computer Society Com-

puter Vision and Pattern Recognition, vol. 2, pp. 246-252, 1999.
[27] R. Girshick, J. Donahue, T. Darrell, J. Malik, "Labelme: a database and web-based tool for image annotation", International journal of computer vision, vol. 77, no. 1, pp. 157-173, 2008.
[28] Jia Deng et al., "ImageNet: A Large-Scale Hierarchical Image Database", CVPR09, 2009.
[29] R. Girshick, J. Donahue, T. Darrell, J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", Computer Vision and Pattern Recognition IEEE, pp. 580-587, 2014.
[30] Krizhevsk Alex, Sutskever Ilya, Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks", Advances in Neural Information Processing Systems, pp. 1097-1105, 2012.
[31] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, "Gradient based learning applied to document recognition", Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov 1998.
[32] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition", CoRR, vol. absI1409.1556, 2014.
[33] Goodfellow Ian, Y. Bengio, A. Courville, Deep learning, MIT Press, 2016.
[34] J. Yosinski, J. Clune, Y. Bengio, H. Lipson, "How transferable are features in deep neural networks?", Advances in Neural Information Processing Systems, pp. 3320-3328, 2014.
[35] David S. Bolme, J. Ross Beveridge, Bruce A. Draper, Yui Man Lui, "Visual Object Tracking using Adaptive Correlation Filters", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2544-2550, 2010.
[36] http://www.cvg.reading.ac.uk/PETS2009/data/website/a.html.
[37] [online] Available: http://cvrr.ucsd.edu/aton/testbed