Target Tracking in Blind Range of Radars With Deep Learning

Chandrakanth.V[†1], A.V.S.N. Murthy[1] and Sumohana S. Channappayya[2]

[1]Defense Research and Development Laboratory (DRDL)
[2]Department of Electrical Engg., Indian Institute of Technology, Hyderabad
[†] chandrav@drdl.drdo.in

***Abstract*—Surveillance radars form the first line of defense in border areas. But due to highly uneven terrains, there are pockets of vulnerability for the enemy to move undetected till they are in the blind range of the radar. This class of targets are termed the 'pop up' targets. They pose a serious threat as they can inflict severe damage to life and property. Blind ranges occur by way of design in pulsed radars. To minimize the blind range problem, multistatic radar configuration or dual pulse transmission methods were proposed. Multistatic radar configuration is highly hardware intensive and dual pulse transmission could only reduce the blind range, not eliminate it. In this work we propose, elimination of blind range using deep learning based video tracking for mono static surveillance radars. Since radars operate in deploy and forget mode, visual system must also operate in a similar way for added advantage. Deep Learning paved way for automatic target detection and classification. However, a deep learning architecture is inherently not capable of tracking because of frame to frame independence in processing. To overcome this limitation, we use prior information from past detections to establish frame to frame correlation and predict future positions of target using a method inspired from CFAR in a parallel channel for target tracking.**

***Index Terms*—Deep Learning, CFAR , Radar, Video Tracking , Blind Range**

I. Introduction

Mono static radars are deployed in border areas with inaccessible terrains for territory defense. In certain locations the terrains are extremely uneven which camouflages the pop up targets. This class of targets are extremely dangerous as they move undetected until they are in blind range of radar and can inflict severe damage to life and property. In such hostile conditions one cannot afford blind ranges as they form pockets of vulnerability in the surveillance network. Presently to minimize blind range, dual pulse transmission technique is used as shown in Fig 1. This method uses two interleaved pulses of different durations to cater for short range and long range targets separately. This process reduces the blind range (DR1<DR2) but does not eliminate it completely. For such close range processing video tracking provides a feasible alternative. Before deep learning, centroid processing [1] and correlation processing [2] were the two prominent techniques used for video tracking. Centroid processing is a purely data driven approach and has no information regarding the features of the object like shape and size. It is also highly sensitive to background data. If the background data is stronger than the target data, the algorithm latches onto background clutter

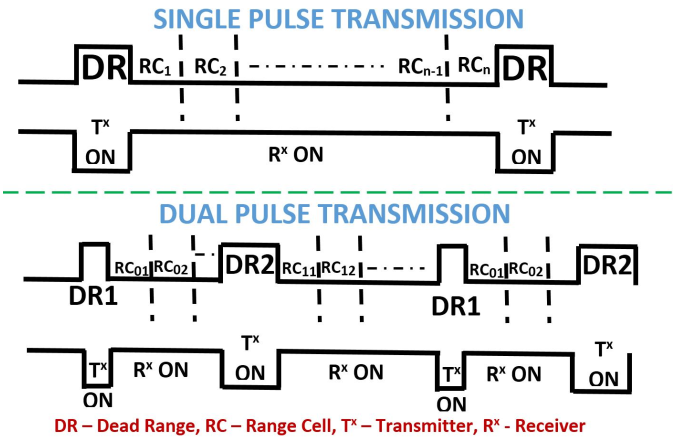


Fig. 1: Radar Blind Range (DR1 < DR2)

leading to track loss. The algorithm can not be applied to multi target scenario and the lost track can not be resuscitated without manual intervention.

Correlation processing requires a template to match the target in the incoming frames which is usually not available upfront. Both the methods are not suitable to work in deploy and forget mode as required for supplementing radar processing. With the advent of deep learning a new era of automated detection and feature extraction evolved. It completely removed manual intervention once the network is trained. Deep learning can be implemented in "deploy and forget" mode and hence can augment radar processing. However, there are two key issues that plagued deep learning for deploying in custom detection and tracking applications. First is the unavailability of labeled training data for custom targets and second is the frame to frame independence. This independent frame processing makes it diffcult for tracking using previous frame information. Most of the literature on CNNs published work for very generic data sets and the networks are not targeted for applications related to surveillance video processing. For our application we collected sample data set to mimic close range target movement as seen by radar on an Indian city street from a height of 25 meters with slant range of targets varying between 100-800 meters.

In this work we addressed the following problems:

(a) Original Image 1 (b) Original Image 2 (c) Original Image 3

(d) YOLO Image 1 (e) YOLO Image 2 (f) YOLO Image 3

(g) Mask RCNN Image 1 (h) Mask RCNN Image 2 (i) Mask RCNN Image 3

Fig. 2: Performance of existing CNNs. Figures [a-c] show the original image to be processed for targets. Figures [d-e] show detection performance of YOLO network and figures [g-i] show the performance of MRCNN network on input images. It can be observed that the networks (using pre trained weights) are predicting false targets and require fine tuning of the weights.

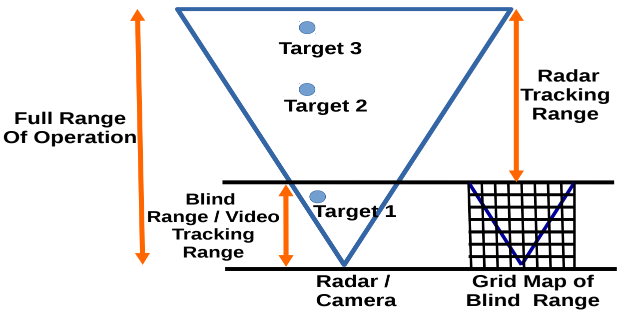


Fig. 3: Range mapping for proposed method. Blind range is grip mapped and stored in a look up table for range estimation.

- Multi target detection in blind range of radar using auxillary video channel with deep learning.
- Selective target detection using custom CNN.
- CFAR based target tracking in videos.

The rest of the paper is organized as follows. Section II briefly revises related work in this area. In section III we introduce the problem statement and discuss proposed solution with results. Section IV concludes our work with a brief on our plan for future research followed by references.

II. Related Work

A. *Classical Video Tracking*

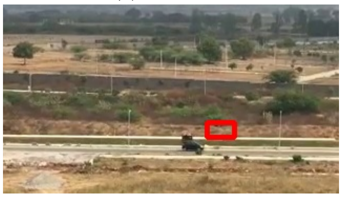
In a generic video tracking problem, we know the initial position of the target in the image. From that information we have to devise a technique to automatically detect and track the target in subsequent frames. This seemingly simple problem becomes extremely challenging because of real time issues such as orientation of the target, occlusions, movement of the target and relative scale variations. Earlier tracking algorithms used centroid of the image as a reference parameter to predict the position of target in next frame. This is termed as centroid tracking [1]. This method relies only on data and not on features of the target which made the algorithm unreliable. To include features of the target into account for detection, correlation processing [2] or template matching was proposed. This method used a reference template or apperance model of the target which was compared with the input data in sliding window fashion storing the correlation coefficient for each pixel shift in a vector. Several variations of this method were proposed in the literature [3] [4] [5] [6]. These methods are vulnerable to scaling, illumination changes and occlusions.

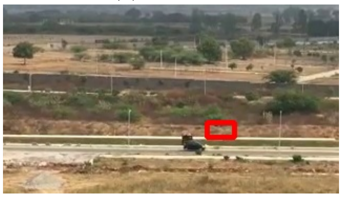

(a) Frame 5

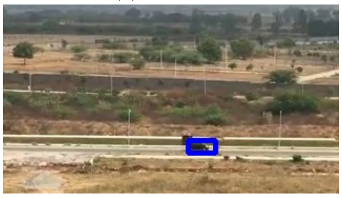

(b) Frame 8


(c) Frame 9


(d) Frame 10


(e) Frame 817

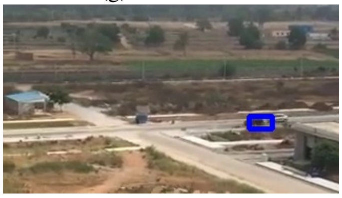

(f) Frame 818


(g) Frame 819


(h) Frame 820


(i) Frame 1076


(j) Frame 1077


(k) Frame 1078


(l) Frame 1079

Fig. 4: CFAR tracking:Blue Box -Target Detected and Acquired, Red Box- Coasted bounding box waiting for reacquiring the target

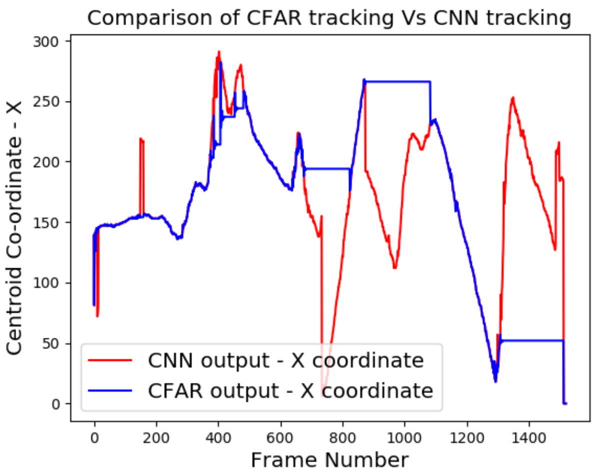


Fig. 5: CFAR tracking - X coordinate

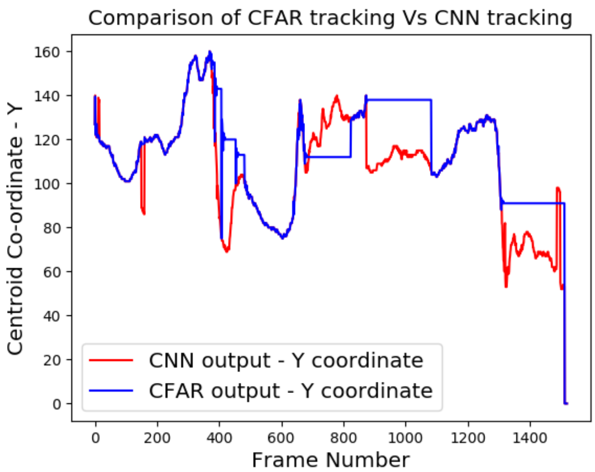


Fig. 6: CFAR tracking - Y coordinate

B. Deep Learning based Video Tracking

With deep learning a new era of automatic target detection and classification evolved. However, most of the work is concentrated on very short range automotive radars [7] [8] and Synthetic Apertuere Radars [9] [10]. CNNs have significantly improved the state of the art in the applications of object detection and classification [11] [12] [13]. These networks are designed in layers with each layer capturing the information from input frame at different level of abstraction. The breakthrough in CNNs was the absence of hand crafted feature description. The network automatically learns the information from the input image and in the final layer converts the information into a feature vector which is then processed by fully connected neural network (FCN). By virtue of design CNNs were capable only for detection and localization. Tracking an object required previous state information of the target which is not available in CNN processing.

C. Radar Tracking and Data Fusion

Target tracking based on Kalman filters [14] [24] is well established area. The works in [15] [16] covers the topic in great detail for the interested reader. Some of the recent works are mentioned in [17] [18] [19]. Hou [20] presented a similar method of fusing range information from radar data with orientation information (azimuth, elevation) from image for target tracking. Golrokh et.al [21] presents the IR data fusion with radar data with similar processing as [20] towards the application of avian monitoring system. They use basic image processing techniques like background subtraction, thresholding, noise suppression, tracking and feature extraction for avian tracking. Wang [22] proposed active contour tracking fused with MMW radar for vehicle detection and tracking.

All the above works used optical sensors to augment radar processing and improve target detection and tracking. The range considered is common for both the sensors and none of the works addressed the problem of blind range. The proposed work differs from the above works, as the range considered by the sensors in our case is mutually exclusive. We used radar output for open range processing and optical sensor output for blind range processing in monostatic radars using deep learning.

III. Proposed Methodology

A. Problem Statement

Our objective is to design an automated target tracker (ATT) to detect and track targets in blind range of radar. Figure 3 shows the proposed method of operation. Target 1 is in the blind range of radar and is processed in optical channel. The area covered by optical channel is grid mapped. Each pixel location is translated to range and stored in a look up table. Targets 2 and 3 are detected by radar and complete data is generated every scan by fusing data from both the channels. To the best of our knowledge this problem has not been addressed in the literature, the proposed CNN based solution with CFAR processing is the first if its kind. It has two fold advantages.

(a) For radar surveillance applications, we can train the CNN extensively to track the targets for most of the possible scenarios.

(b) For tracking applications we can deploy custom trained CNNs to track specific object of interest using CFAR.

B. Proposed solution and Results

As mentioned earlier, in the deep learning method every frame is a new input and the network only detects the object in the current frame. The next frame is an entirely new input with no information available from the previous frame. But to track a target, past observations have to be taken into account. So we used a parallel channel to store this information for prediction. We used this information to create a track vector which stores the target co-ordinates in every frame and predicts the next probable position of target in subsequent frame using the CFAR technique [23]. In this work we used CA CFAR technique with only prior window for tracking. With the physical limitations on target manouverability (excluding passing through occlusions), we can model the movement of the bounding box co-ordinates as uniform distribution. In our case we used the target co-ordinates as range cell information rather than their strengths and predicted the next target location by uniformly averaging the past 'n' measurements. There is a bound set on the measured co-ordinates based on target parameters which is used to classify the observation as valid or invalid.

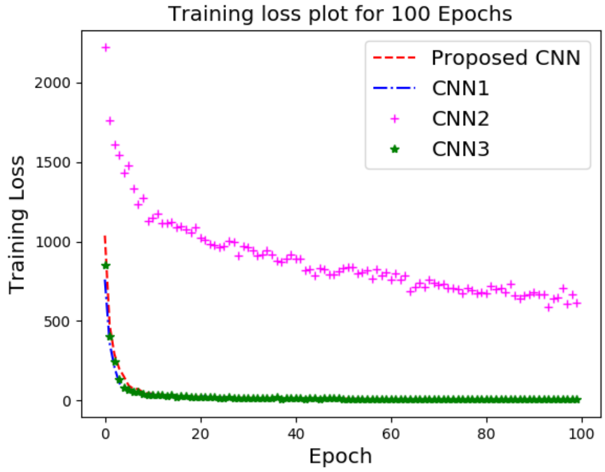


Fig. 7: Training Loss

Figure 2 shows the performance of YOLO and Mask RCNN. It can be observed that the networks are classifying detected targets wrongly and in some frames predicting false targets. This is because the weights we used are trained for generic datasets where the targets have been viewed from a different perspective. The present dataset mimics radar surveillance as seen from a height of 25 meters above ground. To work with this data we have designed custom CNNs as mentioned in Table 1. For specific target tracking we only used regression head for bounding box prediction. Bounding box regression attempts to learn a function to map proposal box (P) to ground truth (G). $\hat{G}$ is the predicted bounding box. CFAR tries to predict the center of the predicted bounding box $(\hat{G})$. Each box is specified by four coordinates $(x, y, w, h) where (x, y)$ are the top left coordinates of $\hat{G}$ and (w, h) are width and height of $\hat{G}$ respectively. We use mean square error (MSE,eq-9) cost function with Adam optimization [25]for bounding box regression. Each CNN differs in a few layers and the performance of the CNNs against training and validation losses are shown in Figs 7-8. After designing the CNN we added a parallel CFAR channel for tracking the targets. The tracking performance of CFAR is shown in Fig 4. To test the performance of the proposed CFAR tracking we have introduced intentional frame jumps in the input video. It can be observed that the predicted track window maintains its course in subsequent images. In figure 4(e-h) the algorithm re-aquires the target when the co-ordinates are within the predefined bounds of CFAR. Figs 4(i-l) show that the bound is user configurable and it can be observed that the window acquires the target at a significant distance if properly designed. The performance of CFAR on centroid co-ordinates is shown in Figs 5-6. It can observed that abrupt target jumps are ignored and the data is smoothed to maintain the target track. The target is successfully reaquired after coasting thus improving tracking efficiency.

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(G_t - \hat{G}_t)^2 \quad (1)$$

TABLE I: Some of the configurations considered

Architecture
CNN 1 - proposed CNN
Conv(32,3,3)->ReLU->Conv(64,5,5)->ReLU ->Max pool-> Dropout -> Conv(32,5,5)->ReLU->Conv(32,3,3)-> ReLU->Dense(64)->ReLU->Dense(4)
CNN 2
Conv(32,3,3)->ReLU->Conv(64,5,5)->ReLU -> Max pool-> Dropout -> Conv(32,5,5)->ReLU->**Batch Norm** -> **Conv(64,3,3)**->ReLU->Dense(64)->ReLU->Dense(4)
CNN 3
Conv(32,3,3)->ReLU->Conv(64,5,5)->ReLU -> Max pool-> Dropout -> Conv(32,5,5)->ReLU-> Batch Norm -> **Conv(32,3,3)**->ReLU-> Conv(64,5,5)-> ReLU->**Max pool(2,2)**-> **Dropout**->Dense(64)->ReLU->Dense(4)
CNN 4
Conv(32,3,3)->ReLU->Conv(64,5,5)->ReLU ->Max pool-> Dropout-> Conv(32,5,5)->ReLU-> Batch Norm -> Conv(32,3,3)->ReLU->Conv(64,5,5)->ReLU-> Max pool-> Dropout-> Dense(64)->ReLU-> ->**Dense(32)**-> ReLU->Dense(4)

[a]Max Pool = maxpool(2,2), Dropout = dropout(0.5)
[b]Bold Font indicates change w.r.t above layer configuration

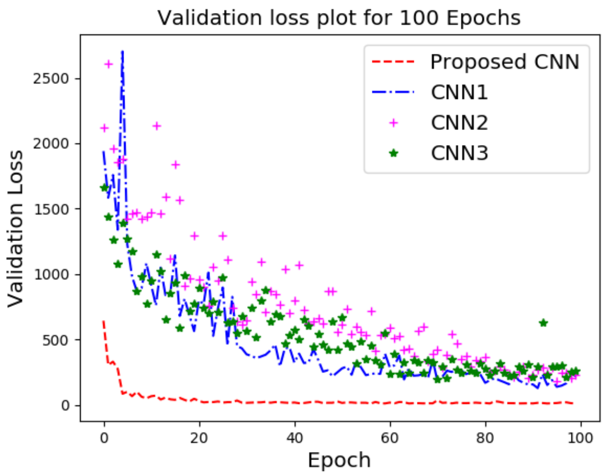


Fig. 8: Validation Loss

IV. Conclusion

In this work we have presented a CNN based architecture to detect and track targets of interest in blind range of radar. The algorithm can be used in addition to radar processing to cater for complete range processing. We have used CFAR processing for target tracking with custom threshold setting to cater for occlusions in the field. Presently the algorithm is proved with RGB data and we are presently acquiring IR data for testing the algorithm.

References

[1] Venkateswarlu R., Sujata K. V., Rao B. V., Centroid Tracker and Aim Point Selection, SPIE Vol. 1697 Acquisition, Tracking and Pointing VI, pp.520-529, 1992.

[2] Ellis JG, kramer, K.A and stubberud SC. "Image correlation based video tracking". in 2011, "21st International Conference on systems Engineering " (pp132-136).IEEE.

[3] D.S.Bolme, J.R.Beveridge, B.A.Draper and Y.M.Lui,"Visual object tracking using adaptive correlation filters",in Proc. IEEE conference Comput. Vis Pattern Recognition., jun 2019 pp2544-2550

[4] M.Danelljan, G Hager, F.S.Khan and M.Felsberg,"learning Spatially regularized correlation filtes for visual tracking",in Proc Int. Conf Comput. Vis Dec 2015,pp. 4310-4318

[5] Danelljan M, Hager G, Shahbaz Khan F, Felsberg M. Convolutional features for correlation filter based visual tracking. In proceedings of the IEEE International Conference on Computer Vision Workshops 2015 (pp. 58-66).

[6] Danelljan M, Hager G, Shahbaz Khan F, Felsberg M. Learning spatially regularized correlation filters for visual tracking. In proceedings of the IEEE international conference on computer vision 2015 (pp. 4310-4318).

[7] Lombacher J, Hahn M, Dickmann J, Whler C. Potential of radar for static object classification using deep learning methods. In2016 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM) 2016 May 19 (pp. 1-4). IEEE.

[8] Wheeler TA, Holder M, Winner H, Kochenderfer MJ. Deep stochastic radar models. In2017 IEEE Intelligent Vehicles Symposium (IV) 2017 Jun 11 (pp. 47-53). IEEE.

[9] Chen S, Wang H. SAR target recognition based on deep learning. In2014 International Conference on Data Science and Advanced Analytics (DSAA) 2014 Oct 30 (pp. 541-547). IEEE.

[10] Wilmanski M, Kreucher C, Lauer J. Modern approaches in deep learning for SAR ATR. InAlgorithms for synthetic aperture radar imagery XXIII 2016 May 14 (Vol. 9843, p. 98430N). International Society for Optics and Photonics.

[11] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. InProceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 779-788).

[12] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich fea-ture hierarchies for accurate object detection and semanticsegmentation. In CVPR, 2014.

[13] He K, Gkioxari G, Dollr P, Girshick R. Mask r-cnn. In Proceedings of the IEEE international conference on computer vision 2017 (pp. 2961-2969).

[14] Genovese, Anthony F. The interacting multiple model algorithm for accurate state estimation of maneuvering targets. Johns Hopkins APL technical digest 22.4 (2001): 614-623.

[15] Y. Bar-Shalom ve T. E. Fortmann. Tracking and Data Association, New York, Academic Press, 1988.
[16] Gilbert, Anton L.,et al."A Real Time Video Tracking System". IEEE Transactions on PAM, VolPAMI-2, No.1, Jan 1980.
[17] Hoffmann F, Ritchie M, Fioranelli F, Charlish A, Griffiths H. Micro-Doppler based detection and tracking of UAVs with multistatic radar. In2016 IEEE Radar Conference (RadarConf) 2016 May 2 (pp. 1-6). IEEE.
[18] Liu G, Wang L, Zou S. A radar-based blind spot detection and warning system for driver assistance. In2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC) 2017 Mar 25 (pp. 2204-2208). IEEE.
[19] Segal S, Logvinenko A, Slapak A. Occlusion Handling in Radar for Detection of Obstacles Based on Tracking Model. In2018 19th International Radar Symposium (IRS) 2018 Jun 20 (pp. 1-10). IEEE.
[20] Hou Z, Han C. A target tracking system based on radar and image fusion. InSixth International Conference of Information Fusion, 2003. Proceedings of the 2003 Jul 8 (Vol. 2, pp. 1426-1432). IEEE.
[21] Mirzaei G, Jamali MM, Gorsevski PV, Frizado J, Bingman VP. Data fusion of ir and marine radar data. In2013 Asilomar Conference on Signals, Systems and Computers 2013 Nov 3 (pp. 565-568). IEEE.
[22] Wang X, Xu L, Sun H, Xin J, Zheng N. On-road vehicle detection and tracking using MMW radar and monovision fusion. IEEE Transactions on Intelligent Transportation Systems. 2016 Apr 29;17(7):2075-84.
[23] Rohling H(1983). Radar CFAR Thresholding in clutter and multiple target situations. IEEE transactions on Aerospace and Electronic Systems. AES -19, 608-621.
[24] Skolnik M.I., "Introduction to Radar Systems", New York, McGraw Hill Book Co.1980.
[25] Zhang Z. Improved Adam optimizer for deep neural networks. In2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS) 2018 Jun 4 (pp. 1-2).