IEEE Winter Conf. on Applications of Computer Vision 2018

Rotation Adaptive Visual Object Tracking with Motion Consistency

Litu Rout[1], Siddhartha[1], Deepak Mishra[1] and Rama Krishna Sai Subrahmanyam Gorthi[2]

[1]Department of Avionics, Indian Institute of Space Science and Technology, Thiruvananthapuram, Kerala, India

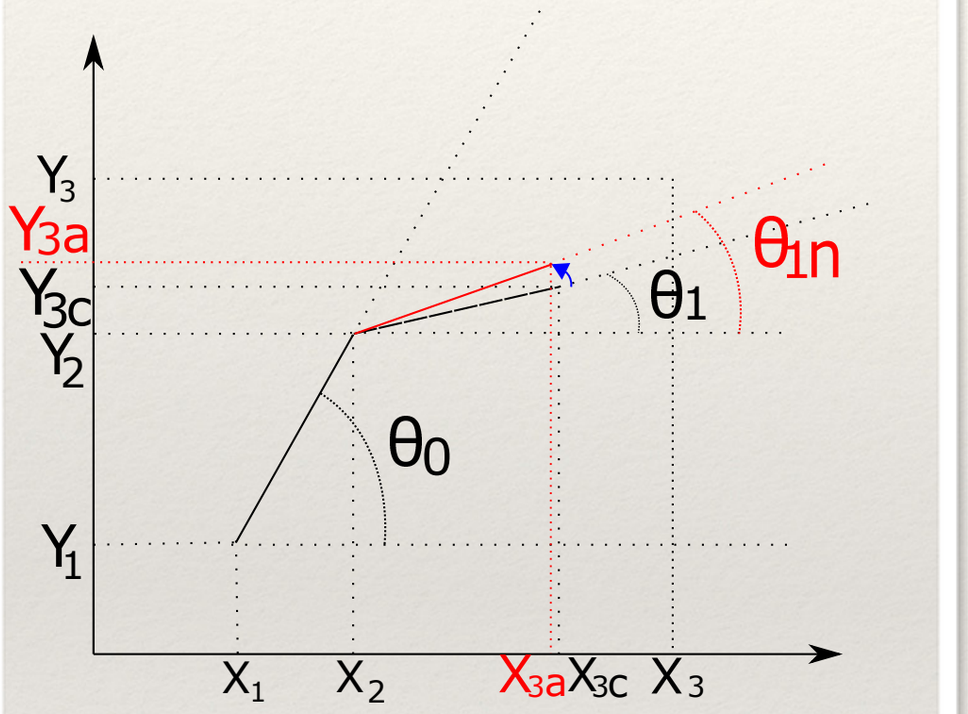
[2]Department of Electrical Engineering, Indian Institute of Technology Tirupati, Andhra Pradesh, India

liturout1997@gmail.com, deepak.mishra@iist.ac.in

Proposed methodology

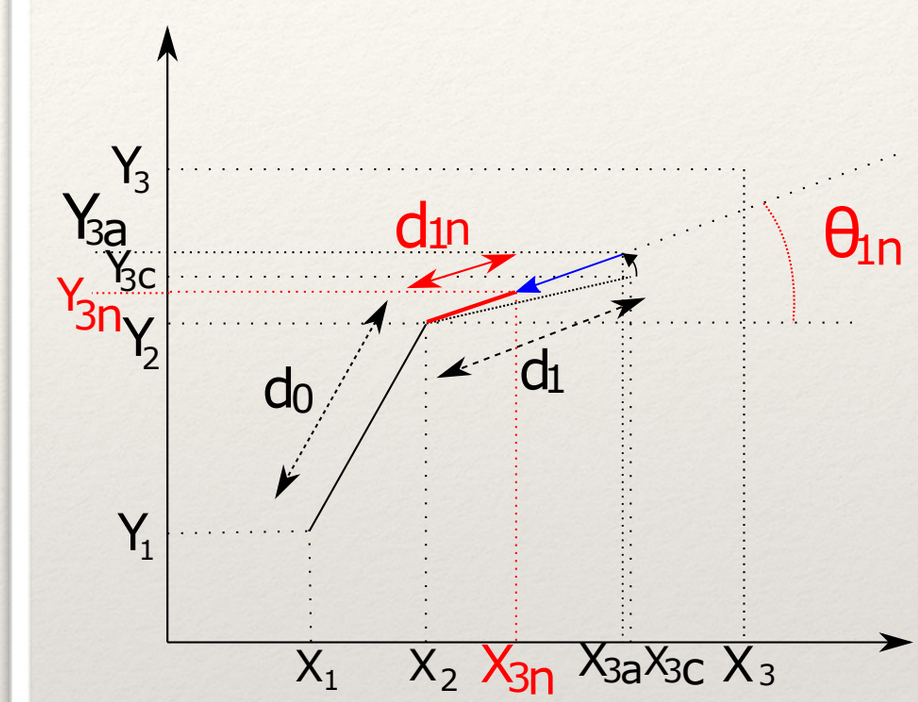
- A generic approach for incorporating
  - rotation invariance (RI) in object tracking
- Introduction of motion consistencies
  - Displacement consistency
  - Scale consistency

Motion Consistency: key idea



Angle consistency

$$[X_{3c}, Y_{3c}] = w \times [X_2, Y_2] + (1 - w) \times [X_3, Y_3]$$



Distance consistency

$$\theta_{1n} = w_\theta \times \theta_0 + (1 - w_\theta) \times \theta_1$$

$$d_{1n} = w_d \times d_0 + (1 - w_d) \times d_1$$

$$[X_{3n}, Y_{3n}] = [X_2, Y_2] + d_{1n} \angle \theta_{1n}$$

Scale consistency

- The conventional approach to estimate size of the target object is to form a scale pyramid and compute response map using each of these images.
- The corresponding scale of the response map having maximum response score among all these response maps determines the size of the target object and target centroid.
- if the position of the target centroid itself is corrupted due to the use of wining response map only, it will persist in subsequent frames. In this standard scenario, the response maps that correspond to different scales aren't used in determining the centroid.
- We propose to use Gaussian weighted average response map centred at the winning map and have variance as an additional hyper parameter. In this way we can incorporate the response maps that correspond to various scales in the scale pyramid.

Algorithm I : Scale Consistency using Gaussian weights

1. **Input parameters :**

   Let *responseMaps* represents the stack of response maps at each scale. μ represents the index of the winning response map. σ_{scale} represents the standard deviation of Gaussian weights. *scaleBins* numerically represents each scale i.e. *scaleBins(1)* represents the first scale, *scaleBins(2)* represents the second scale and so on. Let N represents the total number of scales used in the scale pyramid.

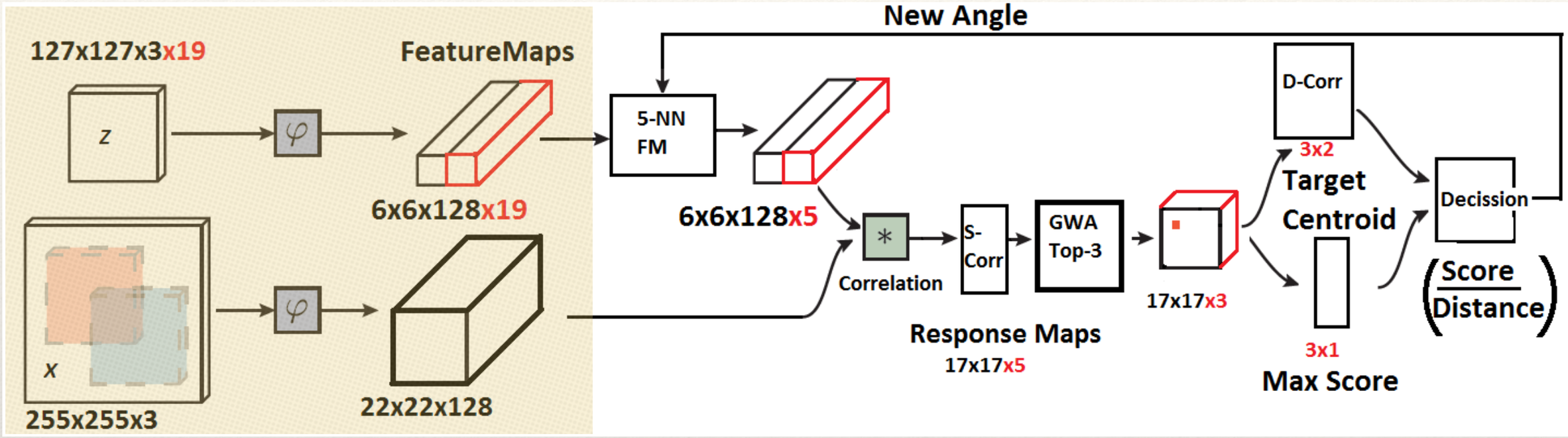
2. **Computation of scale weights and updation of *responseMap*:**

   (a) Define weights for each scale as
   $scaleWeights = \frac{1}{\sqrt{2\times\pi}\times\sigma_{scale}} \exp^{-(\frac{scaleBins-\mu}{\sigma_{scale}})^2}$

   (b) $responseMap = \sum_{i=1}^{N}[responseMaps(i) \times scaleWeights(i)]$

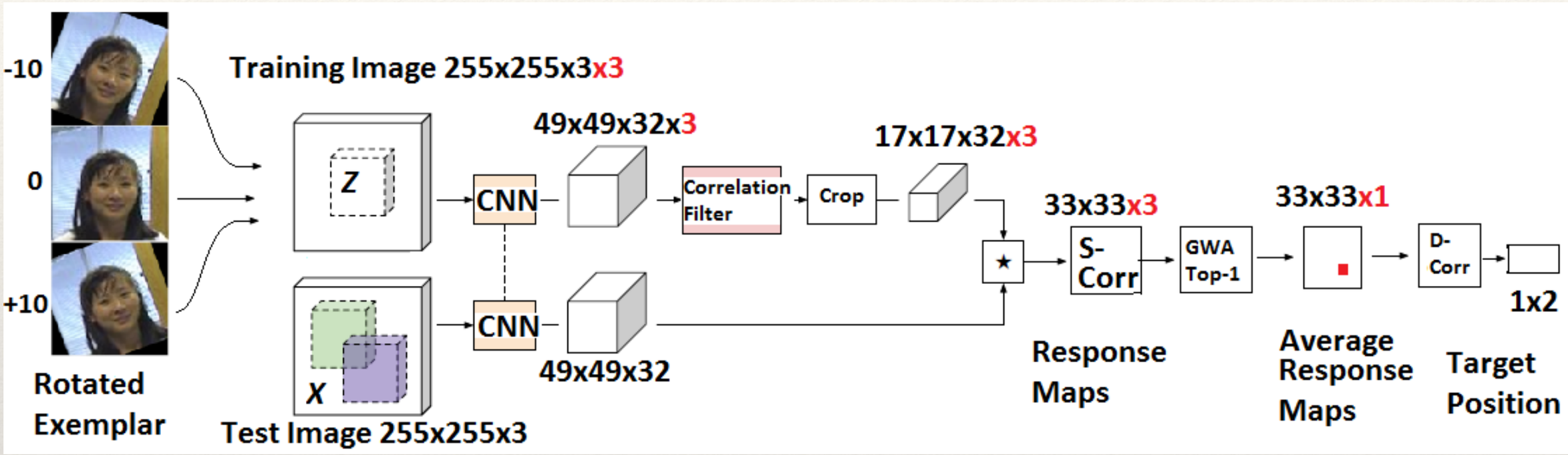
3. **Output response map :** The output of this algorithm is the Gaussian weighted average *responseMap*.

Rotation Invariant Siamese Fully Convolutional Network (SiameseFC)



The conventional SiameseFC extracts features for cropped exemplar, instead exemplar image can be rotated uniformly from -180° to 180° at an interval of θ and corresponding features can be extracted. Here, $\theta = 20^\circ$. So there would be 19 feature maps instead of only one. Assume initial newAngle to be 0°. 5-NN FM block passes 5 nearest neighbour feature maps based on the new angle of rotation. Let S-Corr and D-Corr blocks represent scale and displacement corrections respectively. GWA block computes Gaussian weighted average response map centred at top 3 maximum score response maps. Decision block computes the ratio of maximum score i.e. probability of detection and the corresponding distance from the previous location. The maximum ratio determines the final target centroid and the new angle of rotation which is determined as the angle corresponding to maximum ratio.

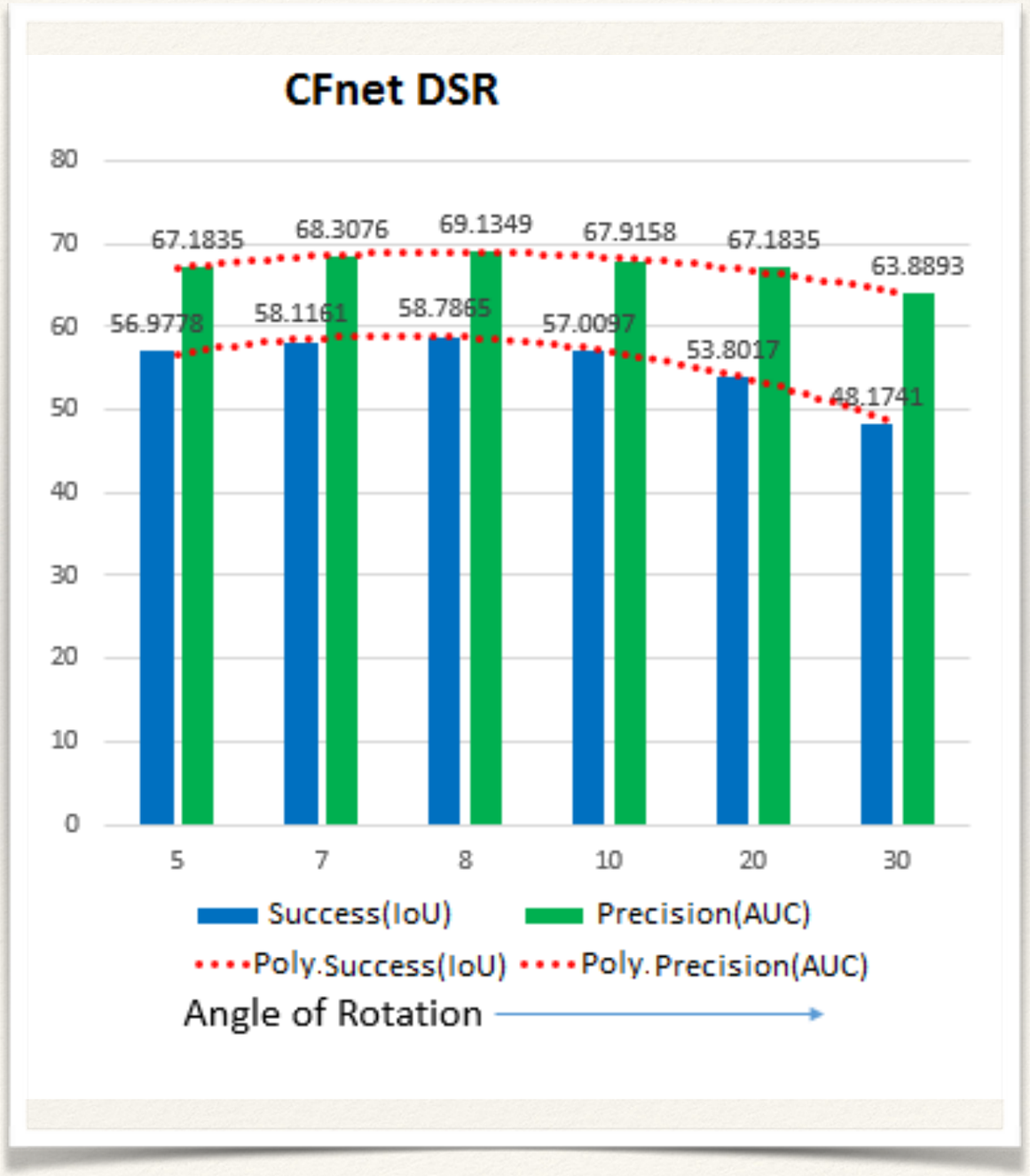
Rotation invariant correlation filter network



Rotation Invariant Correlation filter network. The input exemplar image is rotated by $\theta = [-\zeta, 0^{\circ}, +\zeta]$. Here, $\zeta = 10^{\circ}$ represents the angle of rotation of the exemplar. The angle of rotation 0° represents the actual cropped exemplar image obtained after each iteration. Thus, the three feature maps of rotated exemplar are correlated with the feature map of instance image which produce three most probable response maps. Let S-Corr and D-Corr blocks represent scale and displacement corrections respectively. The S-Corr block performs scale correction on these three response maps. The GWA block computes Gaussian weighted average response map centred at the winning response map. The D-Corr block performs displacement correction and computes the final target centroid.

Optimal angle of rotation in CFnet-DSR

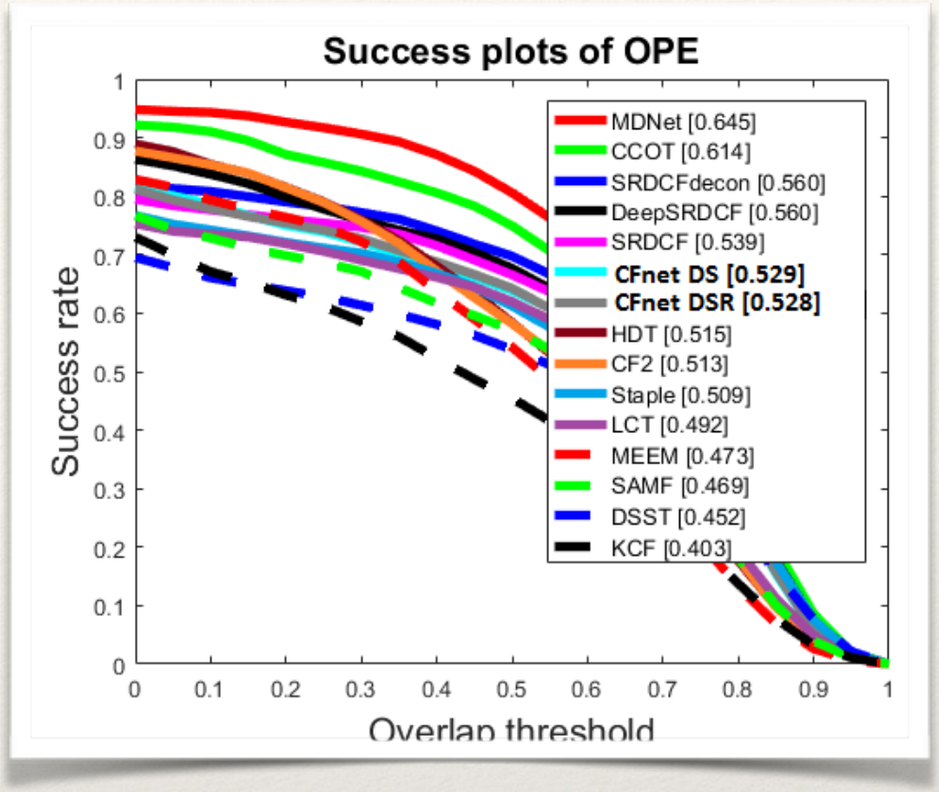
- Success (AUC) and Precision (Threshold) versus various angle of rotation.
- A rough estimation to obtain optimal angle of rotation ζ

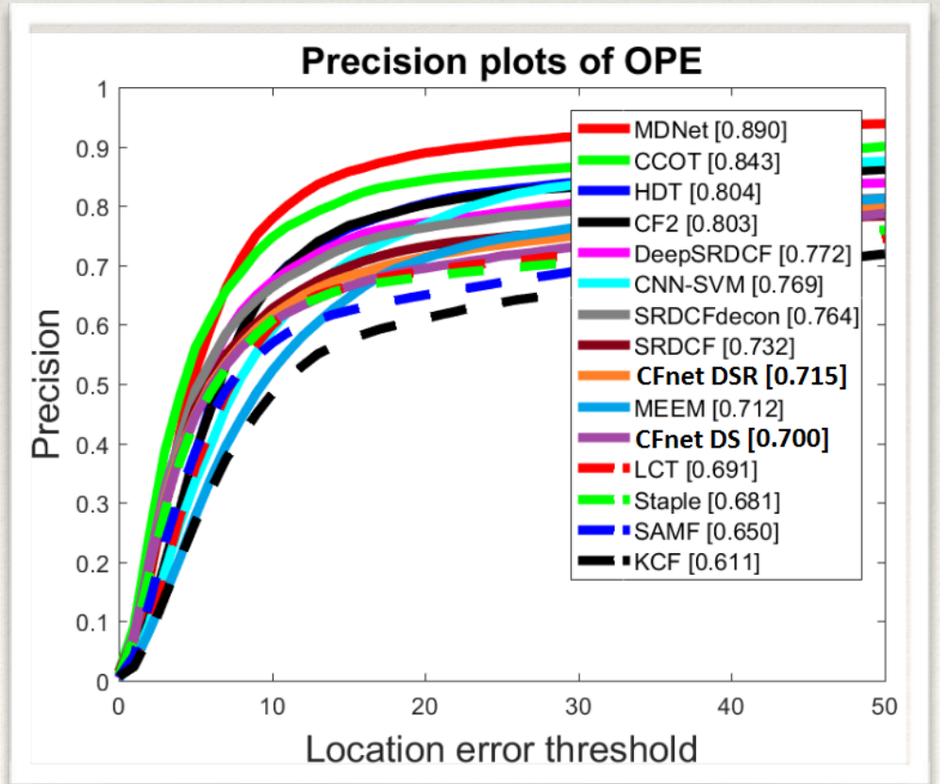


Experimental Results

CFnet

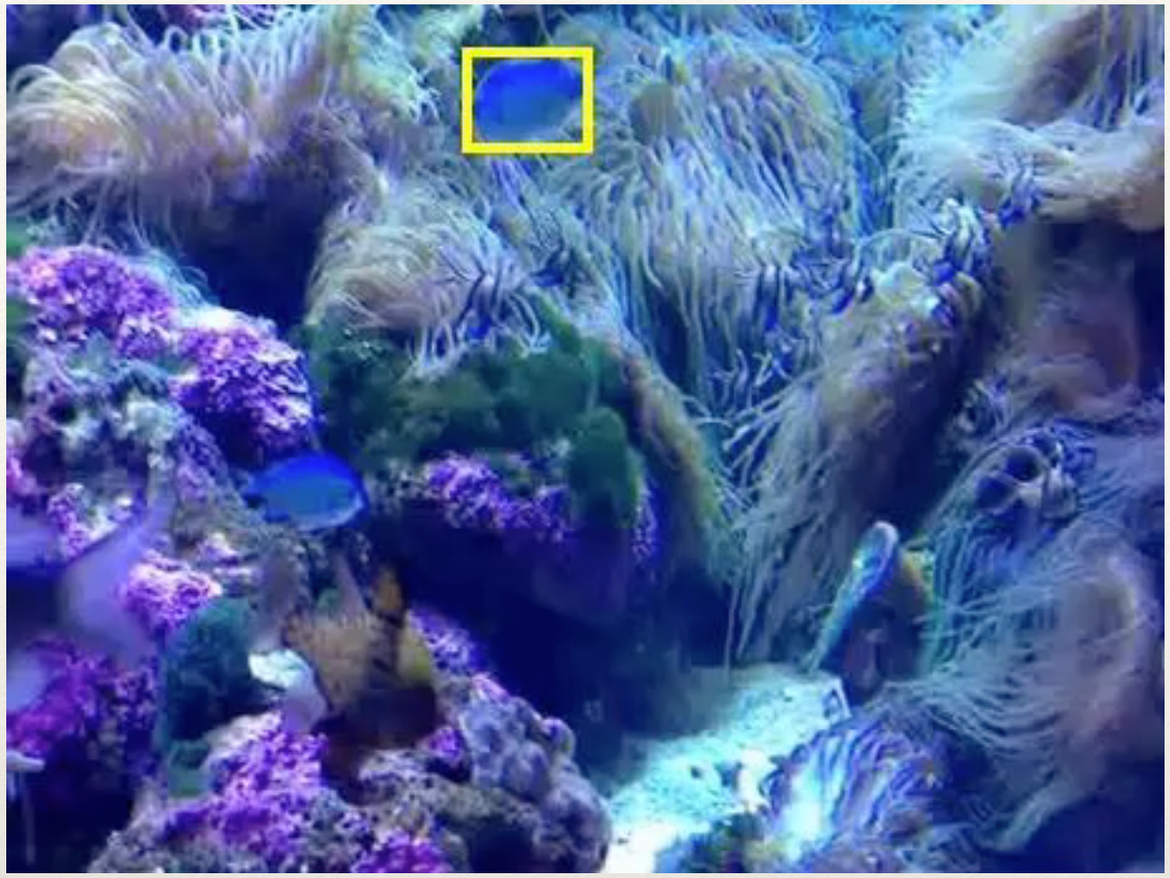
CFnet-DSR





Conclusion

- The work demonstrated a way to incorporate Rotation Invariance (RI) in generic object tracking.
- The introduction of scale and displacement consistency enhanced the degree of smoothness on physical movement variables such as speed and angles.
- The success rate improved by 4.6% whereas precision, by 6.75% relative to baseline approach on OTB dataset.
- The Proposed Siamese DSR gave a drastic improvement in robustness rank by 15.7% and accuracy rank by 14.3% on VOT 2016 database.
- Our future research may include replacing the simple CNN present in both Siamese and CFnet architectures with a very deep CNN.



SiameseFC

SiameseFC-DSR

Thank you for your attention

Acknowledgement

- WACV organization committee and anonymous reviewers
- IIST Trivandrum and DST (Govt. of India for financial support)
- CVVR lab members