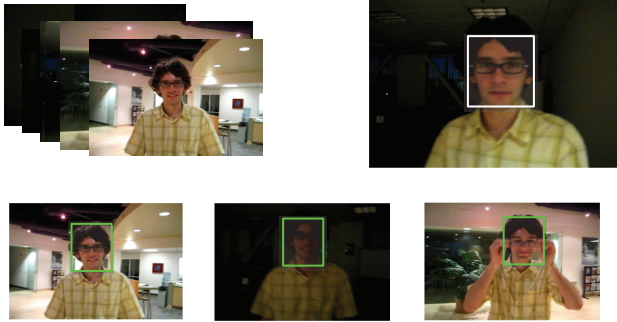


David Held, Sebastian Thrun, Silvio Savarese
Stanford University

Goal

Given an initial bounding box, track a single target through a video sequence

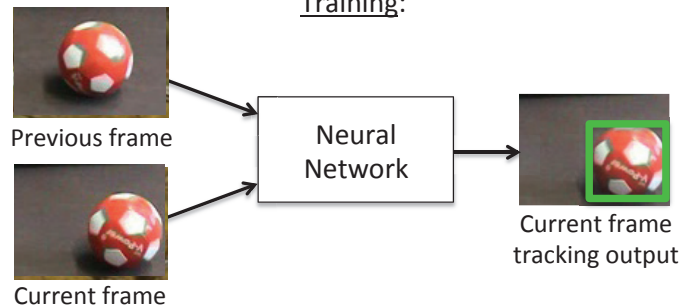


Desired properties

- Tracks generic objects (not class specific)
- Improves performance with data
- Real-time

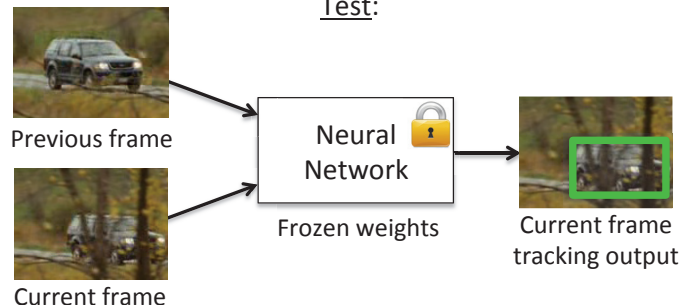
Learning Image Comparison

Training:



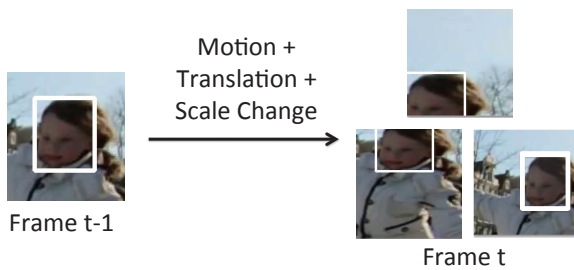
Network learns generic image comparison function

Test:

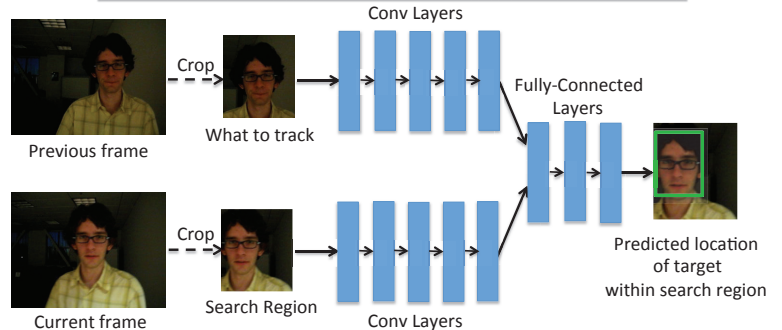


Network tracks novel objects (no finetuning)

Data Augmentation



Tracker Network



- Crop from previous frame defines "target object"
- Crop from current frame defines "search region"
- Both crops are input to neural network

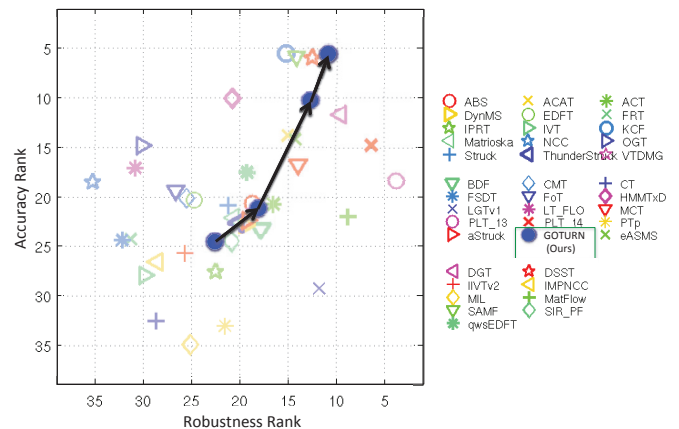
Results

Training data:

- 307 videos from ALOV300 [1] (removed 7 overlapping with test set)
- 239,283 images from ImageNet Detection Challenge [2]

Test set:

- 25 videos from VOT 2014 Tracking Challenge [3]



Dots represent training with 14, 37, 157, and 307 videos

Conclusion

- Improves performance with more training videos
- Runs at 100 FPS for a single video (not batch)
- No online training or fine-tuning required
 - Learns a generic image comparison function
- Tracker regresses directly to bounding box

Acknowledgments

We acknowledge the support of Toyota grant 1186781-31-UDARO and ONR grant 1165419-10-TDAUZ.

References

- [1] Smeulders, A.W., Chu, D.M., Cucchiara, R., Calderara, S., Dehghan, A., Shah, M.: Visual tracking: an experimental survey. Pattern Analysis and Machine Intelligence, (2014)
- [2] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. International Journal of Computer Vision (2014)
- [3] Kristan, M., Pflugfelder, R., Leonardis, A., Matas, J., Cehovin, L., Nebel, G., Vojić, T., Fernández, G., Lukežič, A., Dimitriev, A., et al.: The visual object tracking vot2014 challenge results. ECCV 2014 Workshops (2014)