Deep Tracking: Biologically Inspired Tracking with Deep Convolutional Networks

Afshin Dehghan, Meera Hahn, Si Chen
(adehghan@cs.ucf.edu, mhahn7@emory.edu, sichen@gwmail.gwu.edu)

CENTER FOR RESEARCH IN COMPUTER VISION

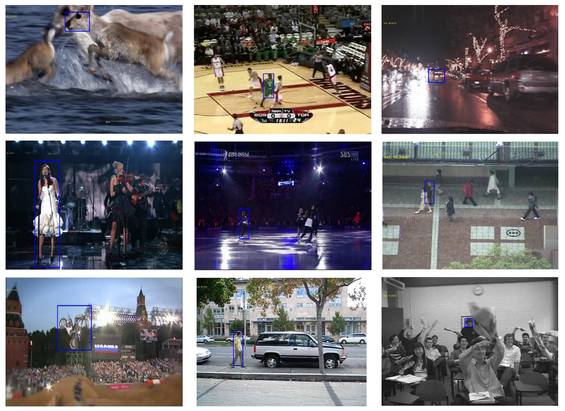
National Science Foundation
Research Experience for Undergraduates

- **Problem:**
  Online Visual Tracking in Videos

- **Challenges:**
  - Target Appearance
  - Occlusion
  - Abrupt Motion

- **Key Contribution:**
  - Using deep learning features instead of hand-crafted features
  - Learning discriminative features for target regions instead of conventional generative features
  - A new deep tracking framework which incorporate both spatial and temporal information
  - A new simple yet effective scale handling strategy

Tracking Examples:

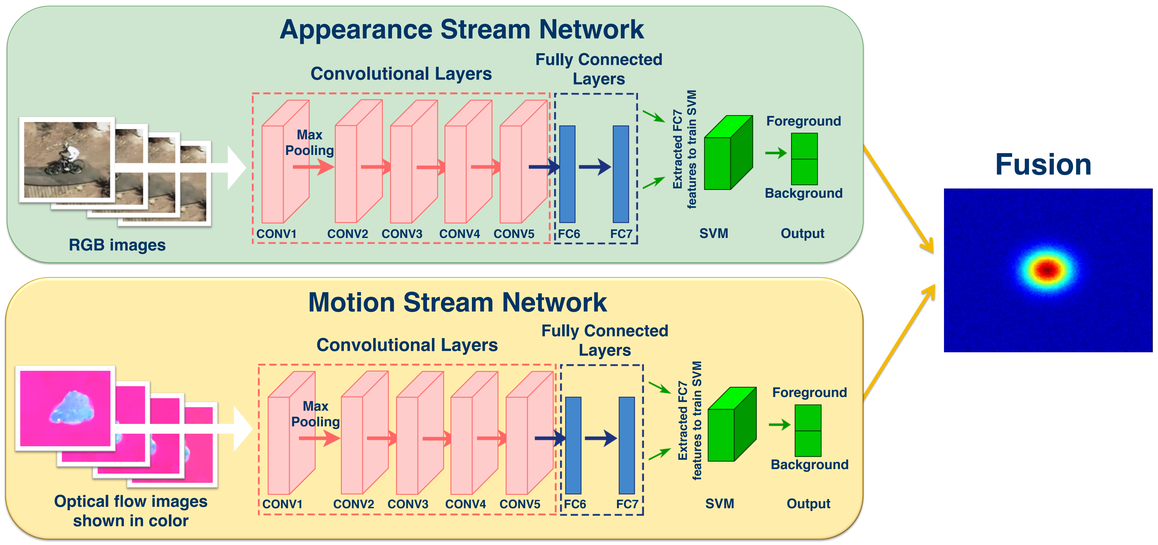


Network Structure:

- Uses the model from Krizhevsky, et al. [1]
- 5 convolutional layers, 3 fully connected layers

Deep Tracker Pipeline:

- Two stream network: appearance and motion
- Networks are pre-trained on ImageNET [2] and fine-tuned on each sequence
- RGB images and RGB optical flow images are used as input for the networks
- Features are extracted from the second fully connected network
- The scores from each stream are combined using late fusion



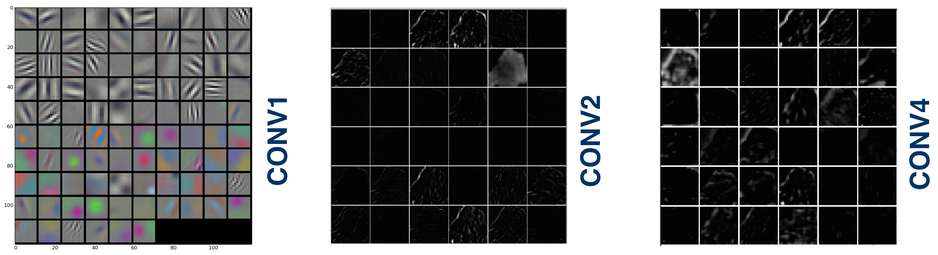
Update:

- Temporal network: updates every 3 frames
- Spatial network: updates every 50 frames or if confidence values drop below defined threshold
  - Handles occlusion and illumination

Handling Scale Change:

- Tracker choses the position of the target for the current frame based on testing
- Then 20 different scales are tested at that chosen location
  - Tested scale sizes are chosen based on scale size of the last frame

Feature Visualization:



Convolutional Neural Network:

- **Collecting Training samples:**
  - Collect labeled positive and negative training samples
  - Data Augmentation: perturbation and rotation



Blue: Annotated frame
Yellow: Positive & negative training samples

Experiments:

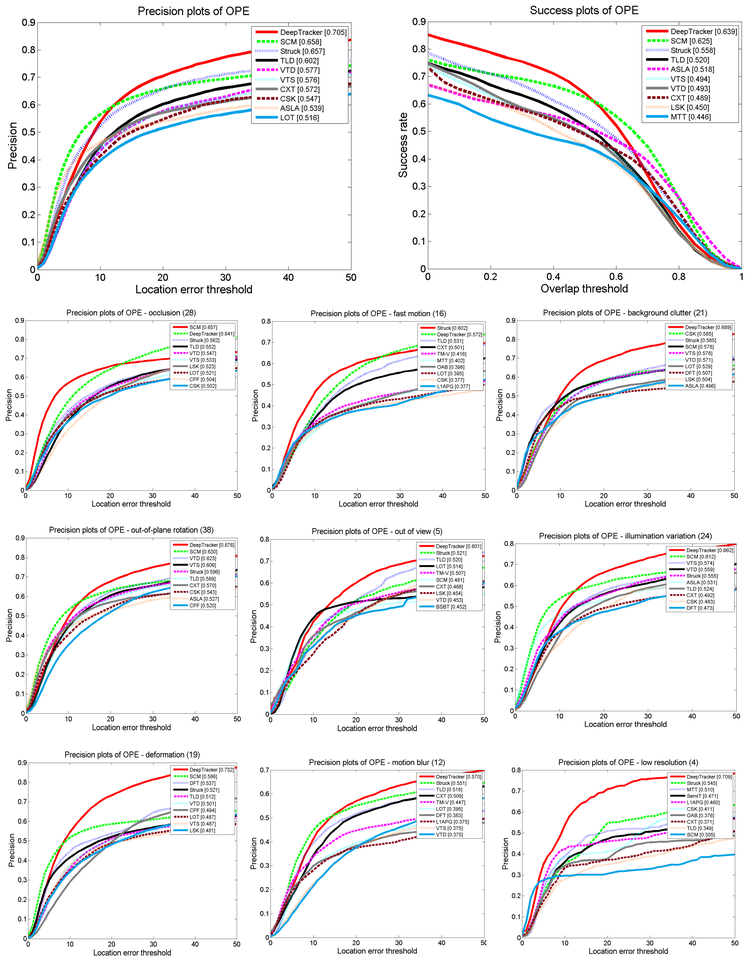
- Tracking Benchmark [3]
- 50 sequences

Results: Deep Tracking vs. State of the Art Trackers

F Score Comparison $F = 2 \cdot \frac{precision * recall}{precision + recall}$

	Deep Tracking	Struck	SCM
Average	65.92	55.19	61.64

Success and Precision Plots



References

[1] Krizhevsky, Alex, Ilya Sutskever, Geoffrey E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks." *NIPS 2012: Neural Information Processing Systems.*

[2] Jia, Yangqing. "Caffe: An Open Source Convolutional Architecture for Fast Feature Embedding." 2013 <http://caffe.berkeleyvision.org>.

[3] Wu, Yi, Jongwoo Lim, Ming-Hsuan Yang. "Online Object Tracking: A Benchmark." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 2013.

[4] Zhong, Wei, Huchuan Lu, Ming-Hsuan Yang. "Robust Object Tracking via Sparsity-based Collaborative Model." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2012).*