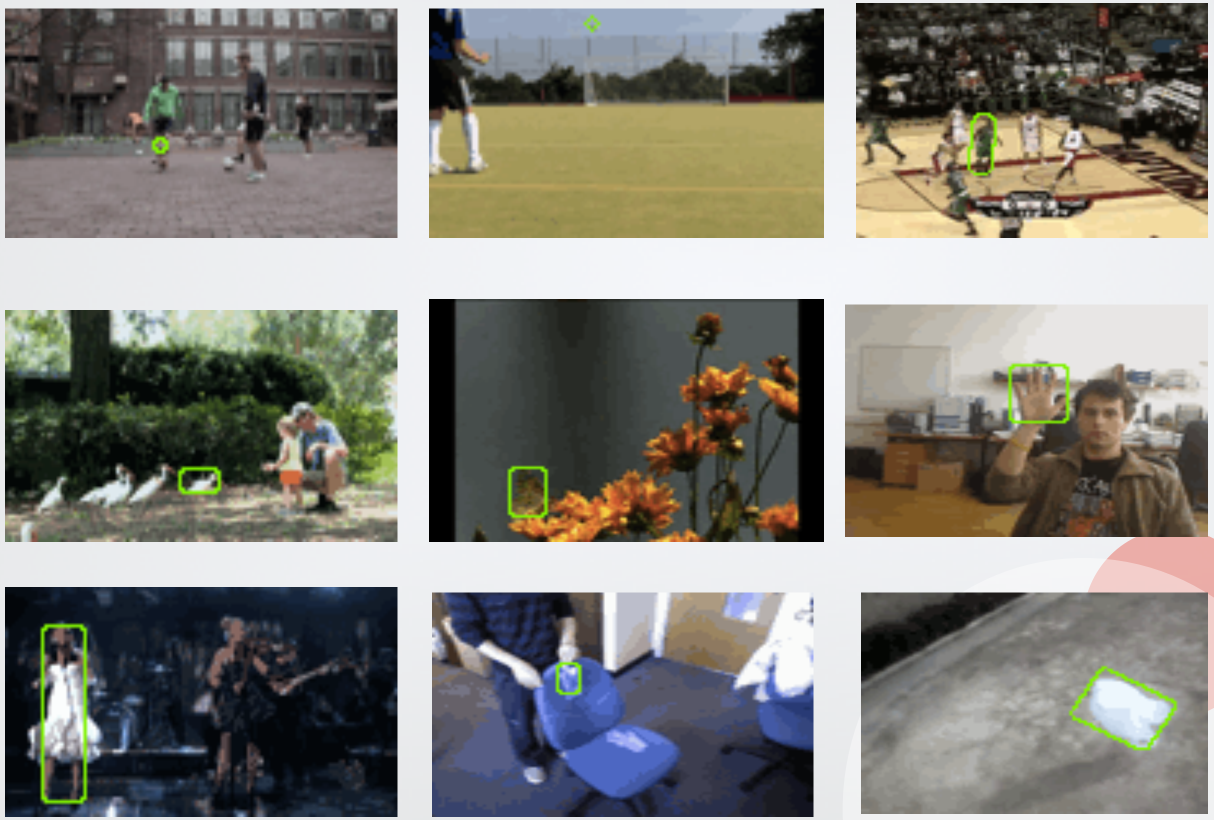
An Integrated Framework for Pedestrian Tracking

Taihong Xiao, Jinwen Ma

Department of Information Science, School of Mathematical Science and LMAM, Peking University

Why Pedestrian Tracking?

General Object Tracking

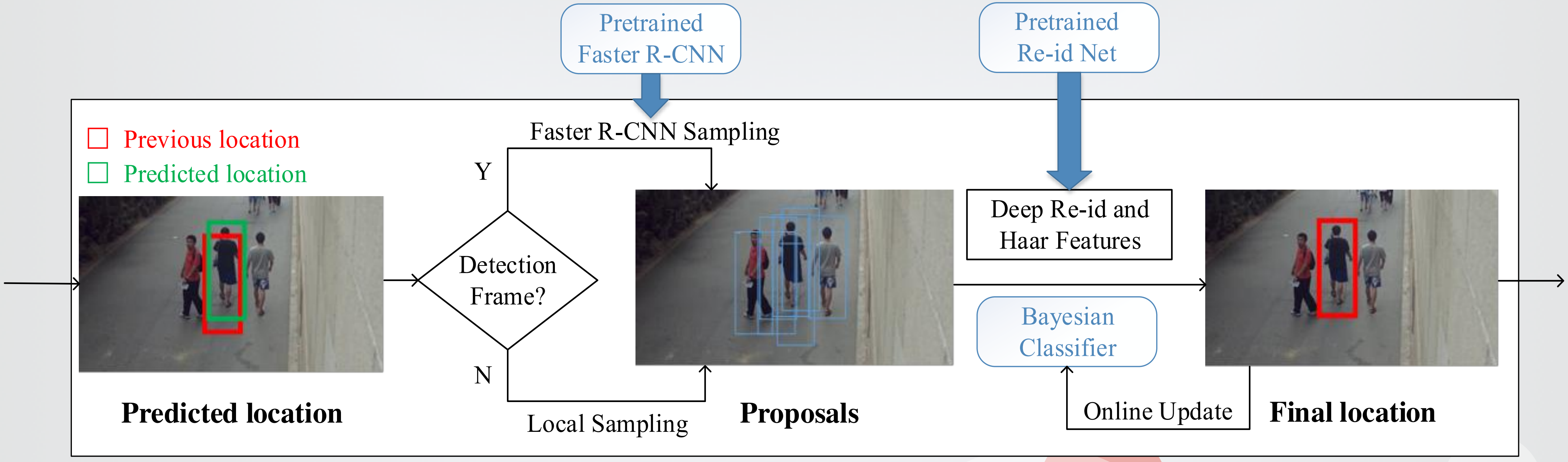


VOT Challenges

Pedestrian Tracking

1. Various gestures, appearances and poses
2. Distraction from similar person
3. Complete occlusion

The ILFPT Model Overview



Frame *t*

Target Prediction

Assume that we are tracking on frame t, the speed of the target in the current frame can be estimated by

$$v_x^t = \rho v_x^t + (1-\rho)\Delta_x^{t-1},$$
$$v_y^t = \rho v_y^{t-1} + (1 - \rho)\Delta_y^{t-1}$$

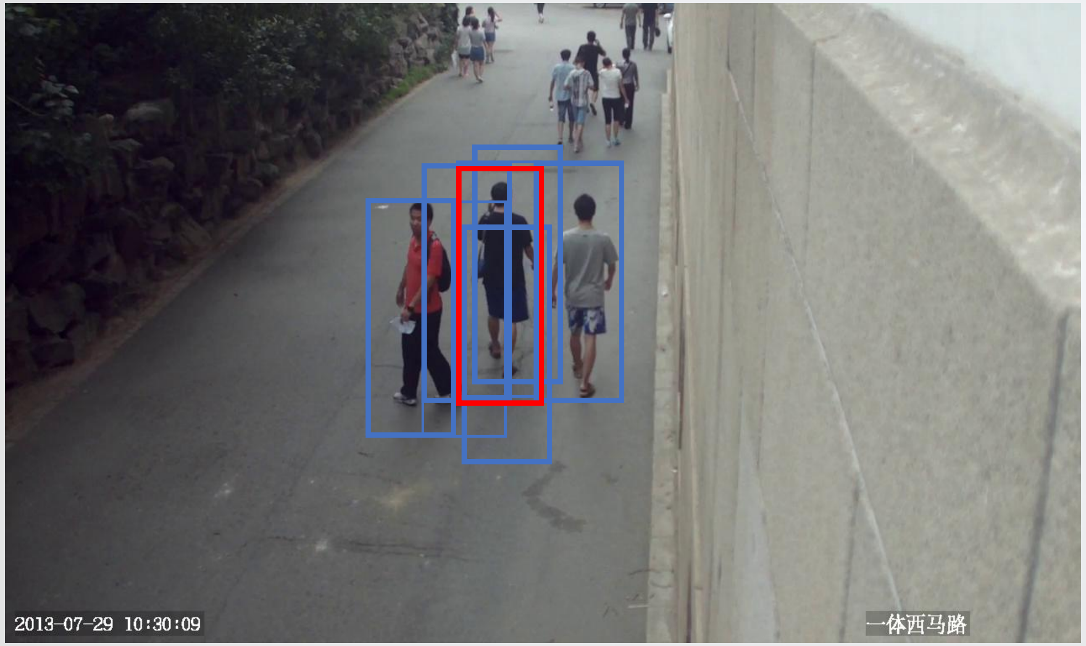
where v_x^{t-1} and v_y^{t-1} are respectively the horizontal and vertical speed in frame $t-1$, $\rho \in [0,1]$ is the momentum factor that controls the weights of the previous speed and $\Delta_x^{t-1} = x^{t-1} - x^{t-2}, \Delta_y^{t-1} = y^{t-1} - y^{t-2}$

Therefore, the center of tracking window in frame t can be predicted as

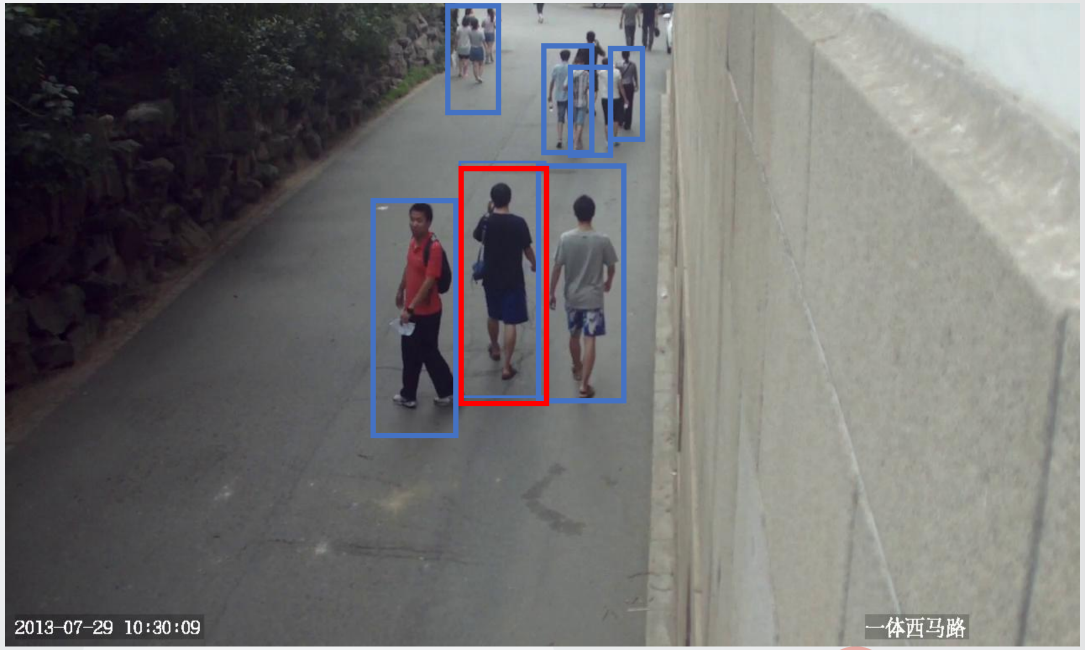
$$x^t = x^{t-1} + v_x^t$$
$$y^t = y^{t-1} + v_y^t$$

Two Sampling Techniques

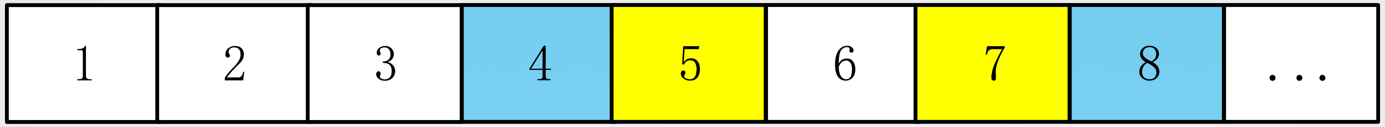
Local sampling



Faster RCNN sampling



Detection Frames & Non-detection Frames



Detection Frames

- Periodical Frames
- Feed-back Frames

Non-detection Frames

1. Faster RCNN sampling is activated in detection frames.
2. Faster RCNN sampling helps adapt the bounding box size.
3. Switching scheme improves speed.

Online Learning Model

A positive sample set $\mathcal{S}^+$ and a negative sample set $\mathcal{S}^-$ are initialized for storing new pedestrian patterns and updating online model. For each candidate sample c, we obtain a compressed low-dimensional feature $\mathbf{v} = (v_1, v_2, \dots, v_n)^T \in \mathbb{R}^n$.

$$R(\boldsymbol{v}) = \log \frac{p(y=1|\mathbf{v})\}}{p(y=0|\mathbf{v})\}} = \sum_{i=1}^{n} \log \frac{p(v_i|y=1)}{p(v_i|y=0)}$$

The conditional probabilities are assumed to be Gaussian distributed with four parameters $(\mu_i^1, \sigma_i^1, \mu_i^0, \sigma_i^0)$

$$p(v_i|y=1) = \mathcal{N}(\mu_i^1, \sigma_i^1), \quad p(v_i|y=0) = \mathcal{N}(\mu_i^0, \sigma_i^0)$$

Online Learning Model

Update rules for parameters

$$\sigma_i^j \leftarrow \sqrt{\lambda\left(\sigma_i^j\right)^2 + (1-\lambda)\left(\tilde{\sigma}_i^j\right)^2 + \lambda(1-\lambda)\left(\mu_i^j - \tilde{\mu}_i^j\right)^2}$$

$$\mu_i^j \leftarrow \lambda\mu_i^j + (1-\lambda)\tilde{\mu}_i^j$$

$$(i = 1,2,\dots,n;\ j = 0,1)$$

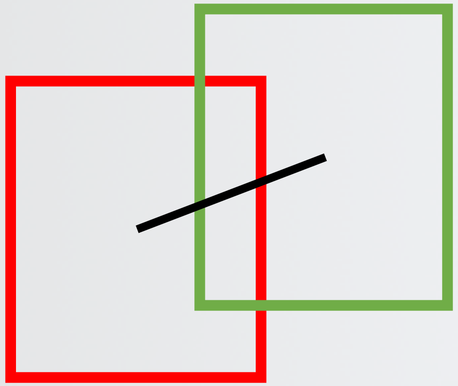
where $\lambda \in (0,1)$ is the inertial factor that controls the updating speed and

$$\tilde{\mu}_i^j = \frac{1}{|S_j|}\sum_{k\in S_j} v_i(k)$$

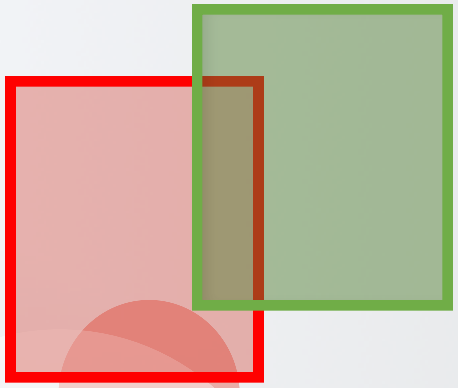
$$\tilde{\sigma}_i^j = \sqrt{\frac{1}{|S_j|}\sum_{k\in S_j}\left(v_i(k) - \tilde{\mu}_i^j\right)^2}$$

Evaluation Methodology

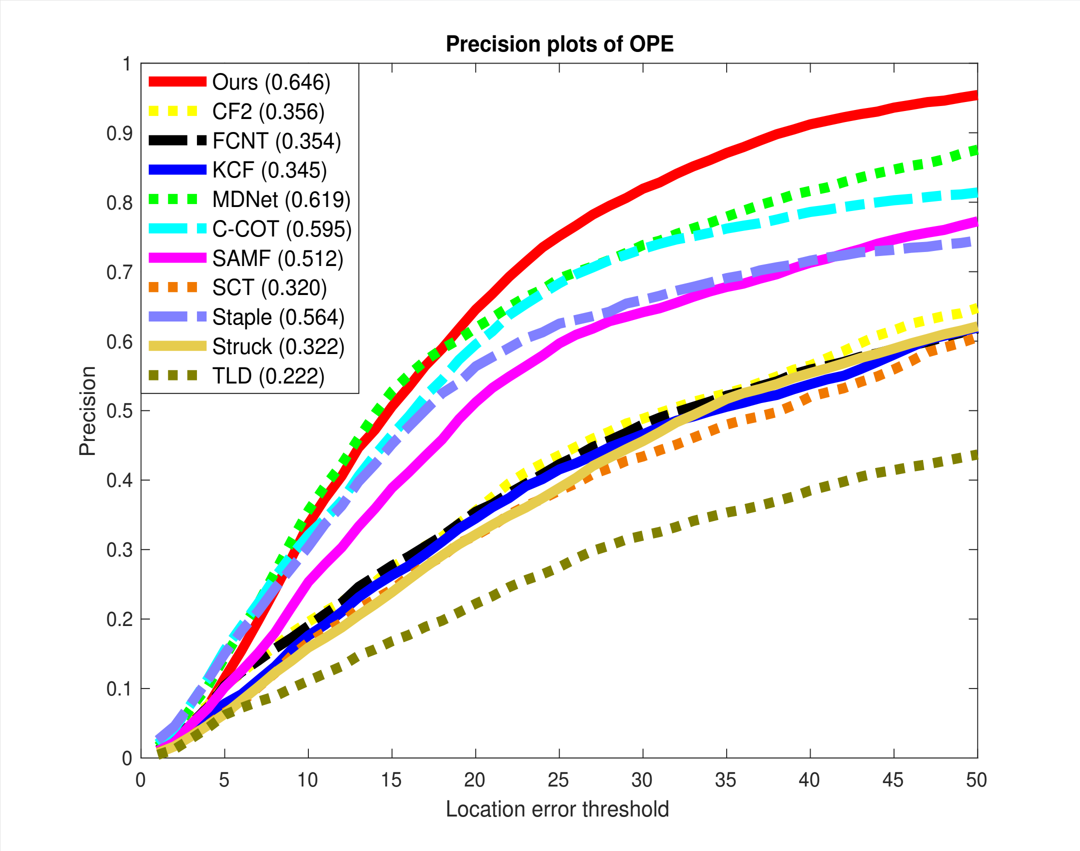
Center Location Error (CLE)



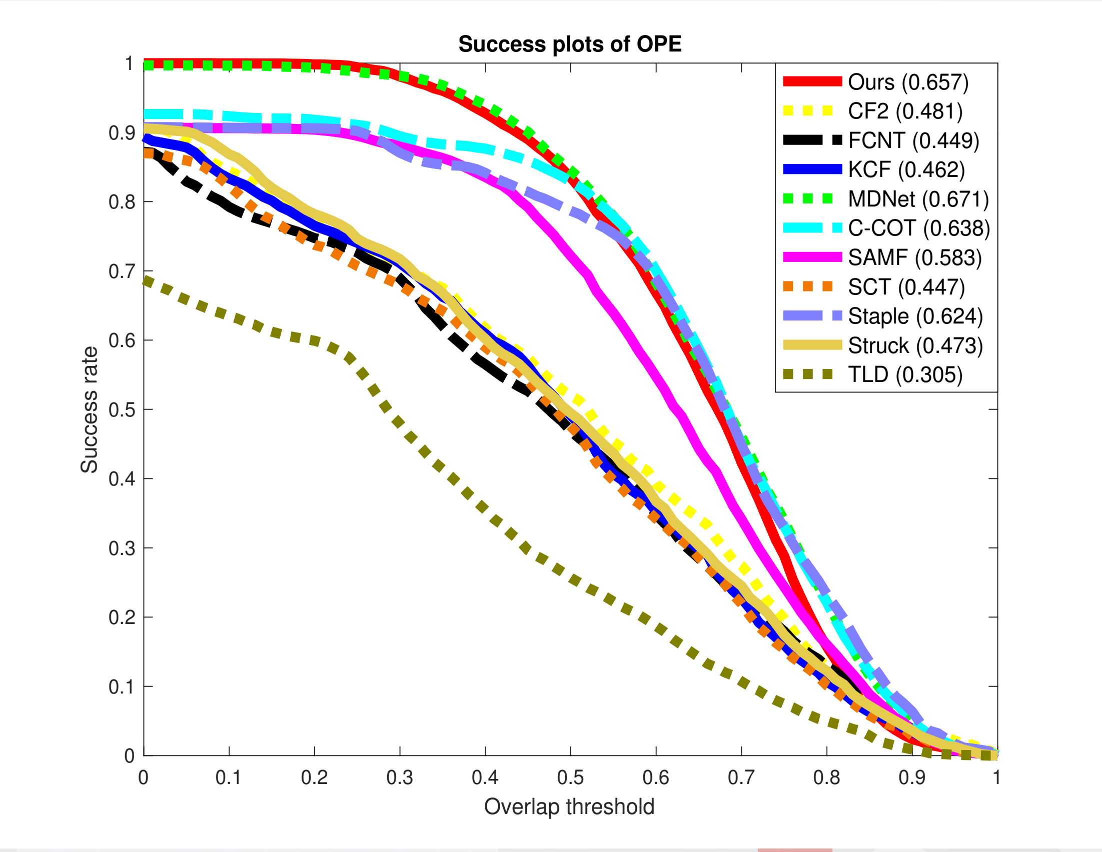
Pascal VOC overlap ratio (VOR)



Experimental Results

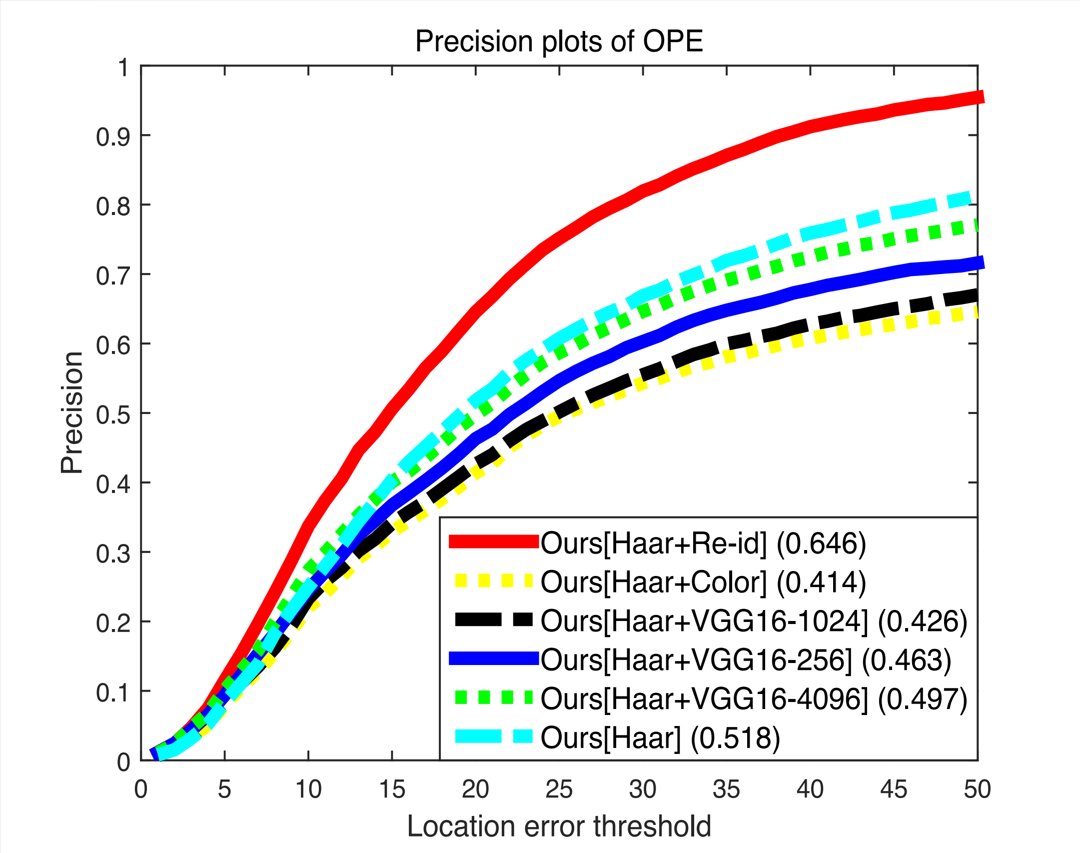


Precision Curve

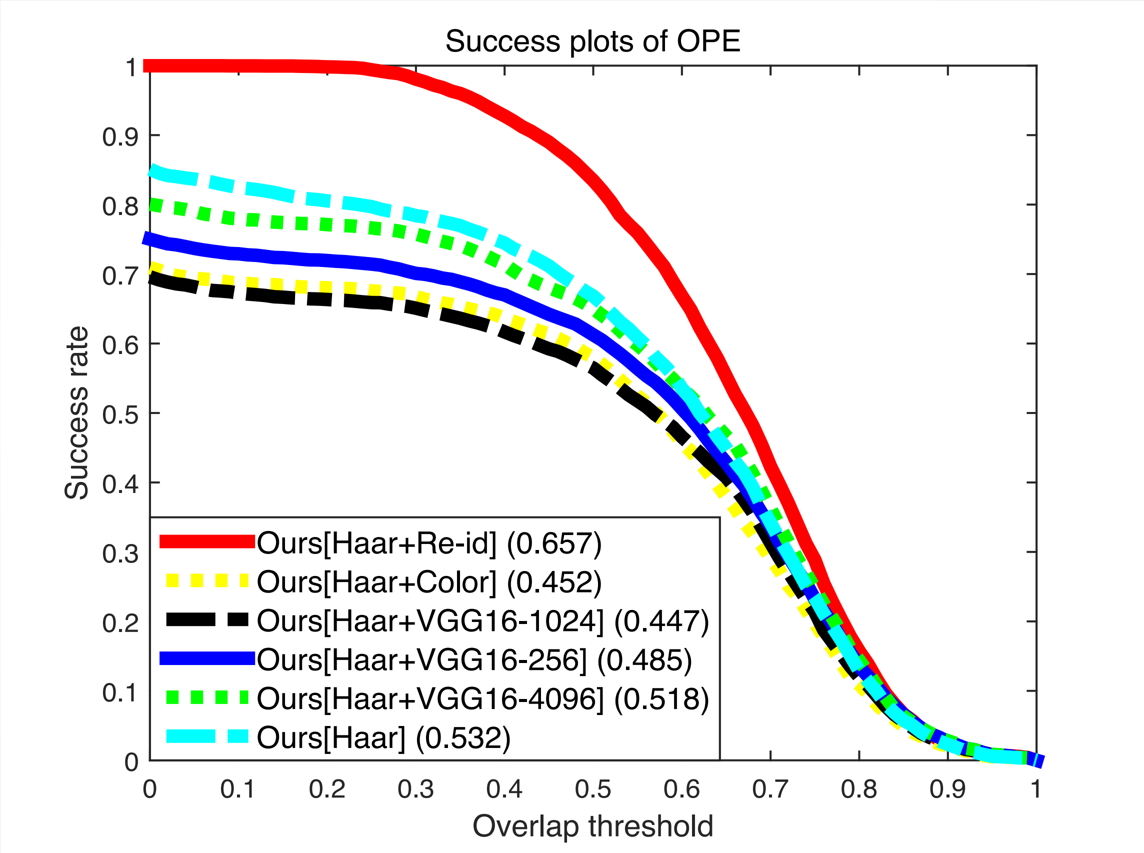


Success Curve

Effectiveness of Deep Re-id Feature



Precision Curve



Success Curve

Video Demo

https://www.youtube.com/watch?v=HQIi0Z9b4Pw

Thank you!