RSINet: Rotation-Scale Invariant Network for Online Visual Tracking

Yang Fang (fangyang968@gmail.com), Geun-Sik Jo* and Chang-Hee Lee
Inha University, KAIST

Introduction

Siamese network trackers

- **No model update** and cannot learn target-specific variation adaptively
- Axis-aligned model contains **extra noise**
- Weak at **rotation** and **scale** estimation

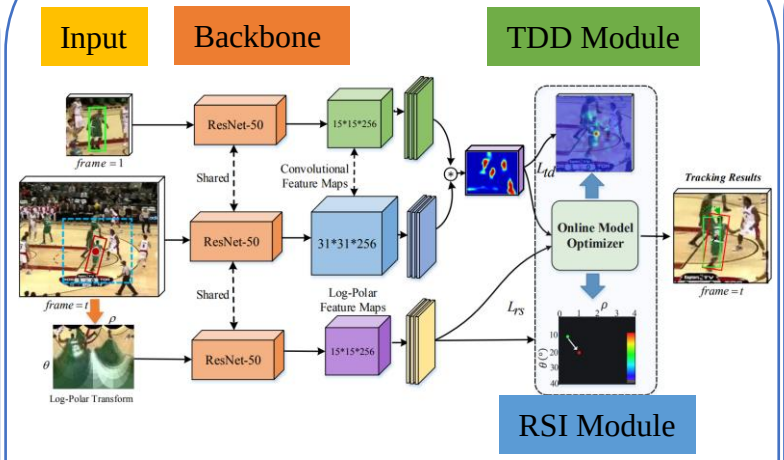
Proposed RSINet tracker

- **Model update** adaptively and dynamically
- Object-aligned model without **extra noise**
- Tailored for **rotation** and **scale** estimation

Algorithm 1: Proposed RSINet Tracker.

Input: Pre-trained Network model $\mathbf{M}$ and Initial frame I_0 with annotation.
Output: Estimated target state $\mathcal{O}_t^* = (x_t, y_t, s_t, r_t)$; Updated model filters $\boldsymbol{h_t}$.
while *frame* $t \leq$ *length(video sequence)* **do**
  Feed new frame into Siamese network to predict new target state (x_t, y_t, s_t, r_t).
  if $(t \mid 10)$ **then**
    Calculate spatio-temporal energy ε, in [9]
    if $\varepsilon \geq \kappa\varepsilon_0$ **then**
      Derive steepest descend update rate α_s, [10]
      $\alpha \leftarrow \min(\frac{1}{\varepsilon}, \alpha_s)$, [11]
    end
    Update tracking model filter $h^{t+1} = h^t + \alpha\nabla L(h^t)$.
  end
  $t = t + 1$
end

Rotation-Scale Invariant Network



- Target-Distractor Discrimination (TDD) module:

Score map: $s(x, w) = m \cdot (x * w) + (1 - m) \cdot \max(0, x * w)$

TDD Loss: $L_{td}(\mathbf{w}) = \frac{1}{N} \sum_{(x,y)\in S} \|s(x, w) - y\|^2 + \|\gamma * w\|^2$

- Rotation-Scale Invariance Module (RSI) module:

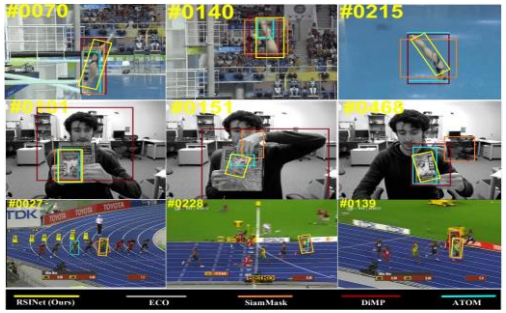
$$f(I^{lp}, \mathbf{h}) = \psi_3(h_3 * \psi_2(h_2 * \psi_1(h_1 * I^{lp})))$$

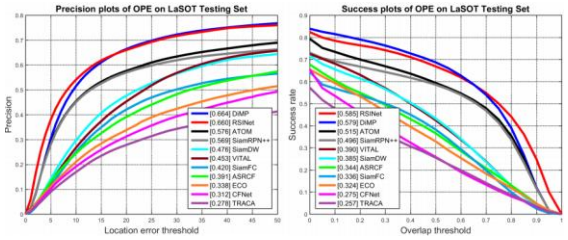
Rotation-Scale formulation in log-polar:

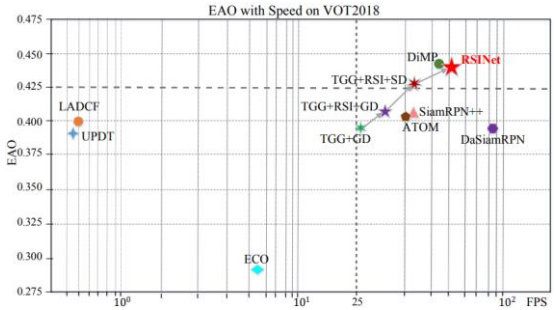
$$I_{t+1}^{lp}(\rho, \theta) = I_t^{lp}(\rho - \Delta\rho, \theta - \Delta\theta)$$

RSI Loss: $L_{rs}(\mathbf{h}) = \sum_{i=1}^{N} \|\mathcal{R}(f(I_i^{lp}, h), g_i)\|^2 + \sum_j \lambda_j \|h_i\|^2$

Experimental evaluation







Conclusion

Proposed RSINet enables target-distractor model and rotation-scale model learning simultaneously. It keeps a good balance between tracking accuracy (0.604 on vot18) and running efficiency (45 FPS)