

Siamese Tracking

Jesus Zarzar
October 16, 2018

Papers

- **Signature Verification**

“Signature Verification using a “Siamese” Time Delay Neural Network”

Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Sickinger and Roopak Shah
NIPS 1993

- **SiamFC**

“Fully-Convolutional Siamese Networks for Object Tracking”

Luca Bertinetto, Jack Valmadre, João F. Henriques, Andrea Vedaldi, Philip H. S. Torr
ECCV 2016

- **SiamRPN**

“High Performance Visual Tracking with Siamese Region Proposal Network”

Bo Li, Junjie Yan, Wei Wu, Zheng Zhu, Xiaolin Hu
CVPR 2018

Outline

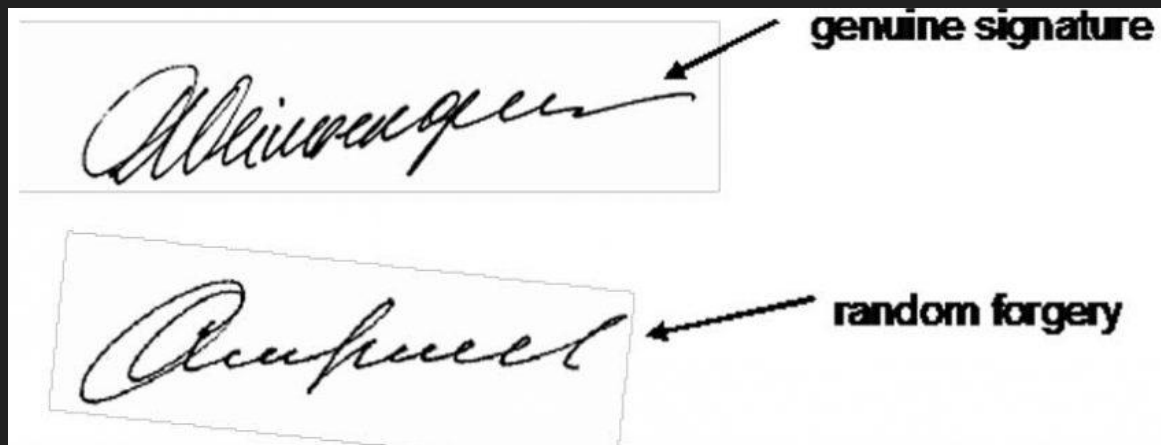
- Siamese Networks
 - Signature Verification
- Siamese Tracking
 - SiamFC
- State of the Art
 - SiamRPN

Siamese Networks - Signature Verification

- “Signature Verification using a “Siamese” Time Delay Neural Network”
Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Sickinger and Roopak Shah
NIPS 1993
- “Fully-Convolutional Siamese Networks for Object Tracking”
Luca Bertinetto, Jack Valmadre, João F. Henriques, Andrea Vedaldi, Philip H. S. Torr
ECCV 2016
- “High Performance Visual Tracking with Siamese Region Proposal Network”
Bo Li, Junjie Yan, Wei Wu, Zheng Zhu, Xiaolin Hu
VOT 2018

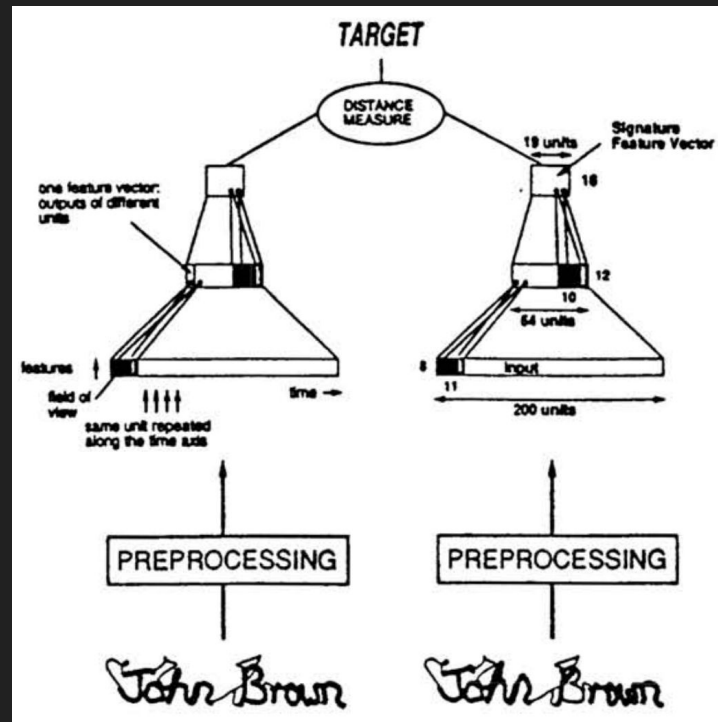
Siamese Networks - Signature Verification

- Electronic signature verification
 - Detect credit card frauds
- Comparison of two signatures
 - “Genuine” or “Forged”



Siamese Networks - Signature Verification

- Dataset
 - Digital-pen input signatures
 - 219 people, 10-20 signatures each
 - Bell Laboratories Cafeteria
- Two identical networks
- Handpicked features
- Encode two input feature vectors
 - Sample and Example
- Distance measure
 - Cosine similarity
 - 1 for Genuine
 - -1 for Forgerie



Siamese Networks - Signature Verification

- ~99% Genuine Accepted and ~80% Forgeries Rejected

Network	Input Features	Best Performance on:	
		Training Set	Validation Set
1, arch 1	pud acc-c acc-t spd $\cos\theta \sin\theta \cos\phi \sin\phi$	GA 97.0% FR 65.3%, 26 passes through set	GA 90.3% FR 74.8%, 6 passes through set
2, arch 1	same as 1, but pen up trajectory removed	GA 97.8% FR 60.0%, 11 passes through set	GA 93.2% FR 75.2%, 2 passes through set
3, arch 1	x y pud spd $\cos\theta \sin\theta$ $\cos\phi \sin\phi$	GA 99.8% FR 88.8%, 100 passes through set	GA 91.7% FR 74.2%, 32 passes through set
4, arch 1	same as network 3, but a larger training set	GA 98.2% FR 81.7%, 42 passes through set	GA 99.4% FR 80.5%, 42 passes through set
5, arch 2	same as 4, except ar- chitecture 2 was used	GA 98.6% FR 81.5%, 69 passes through set	GA 99.6% FR 80.1%, 44 passes through set

Siamese Tracking - SiamFC

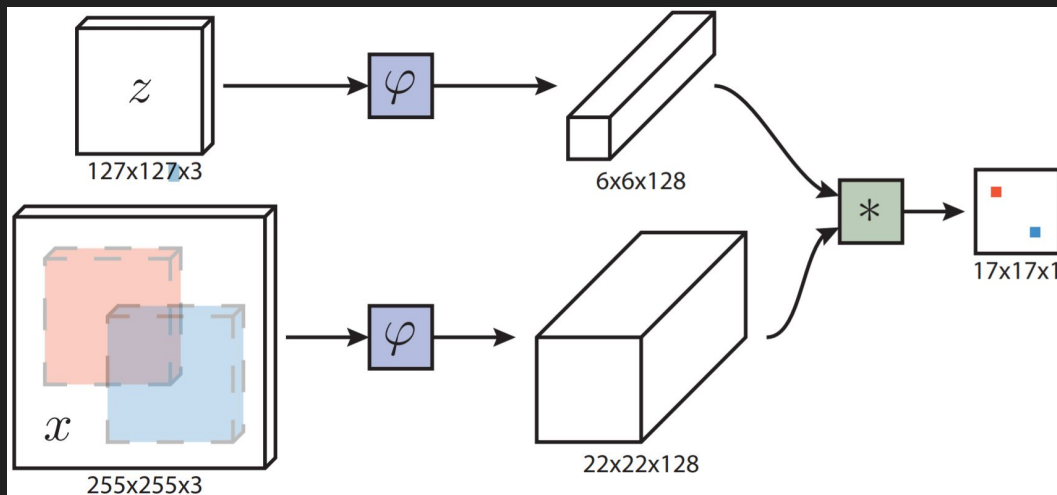
- “Signature Verification using a “Siamese” Time Delay Neural Network”
Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Sickinger and Roopak Shah
NIPS 1993
- “Fully-Convolutional Siamese Networks for Object Tracking”
Luca Bertinetto, Jack Valmadre, João F. Henriques, Andrea Vedaldi, Philip H. S. Torr
ECCV 2016
- “High Performance Visual Tracking with Siamese Region Proposal Network”
Bo Li, Junjie Yan, Wei Wu, Zheng Zhu, Xiaolin Hu
VOT 2018

Siamese Tracking - SiamFC



Siamese Tracking - SiamFC

- Exemplar z
- Search Image x
- Fully-convolutional encoder φ
 - Commutes with translation
- Output score map v
 - Cosine Window
 - Penalize large displacements



Siamese Tracking - SiamFC

- Encoding function φ
 - Based on AlexNet
 - Cut at conv5
 - No padding
 - Xavier Initialization

Layer	Support	Chan. map	Stride	Activation size		
				for exemplar	for search	chans.
				127×127	255×255	$\times 3$
conv1	11×11	96×3	2	59×59	123×123	$\times 96$
pool1	3×3		2	29×29	61×61	$\times 96$
conv2	5×5	256×48	1	25×25	57×57	$\times 256$
pool2	3×3		2	12×12	28×28	$\times 256$
conv3	3×3	384×256	1	10×10	26×26	$\times 192$
conv4	3×3	384×192	1	8×8	24×24	$\times 192$
conv5	3×3	256×192	1	6×6	22×22	$\times 128$

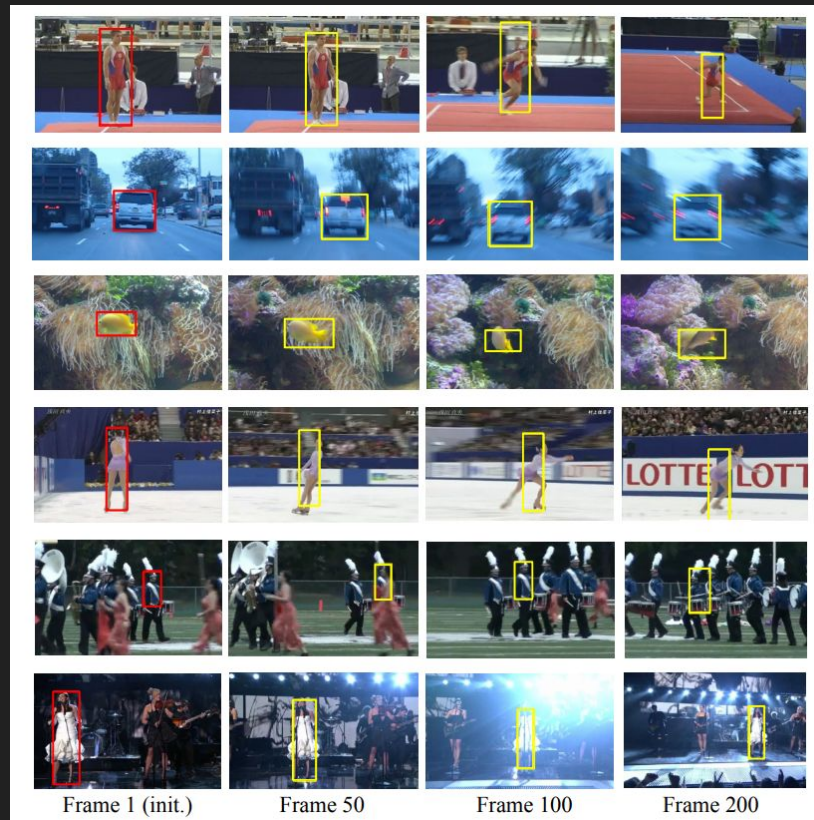
Siamese Tracking - SiamFC

- Trained on image pairs
 - ImageNet Video
 - 50 epochs, 50,000 pairs
 - Centered around GT
 - Padded with mean value
- Binary GT map y
- Score map v
- Loss map $l(y, v)$
- Averaged loss

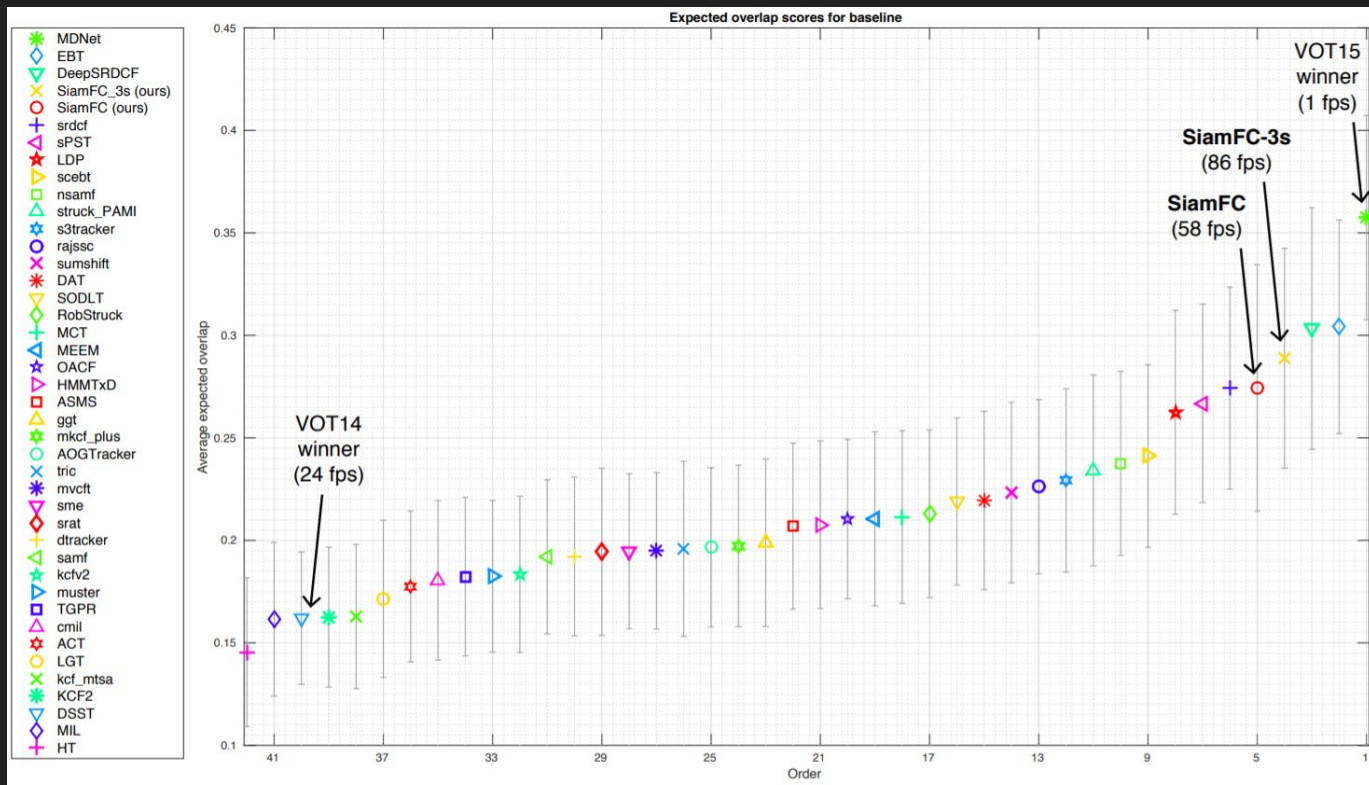
$$\ell(y, v) = \log(1 + \exp(-yv))$$
$$y[u] = \begin{cases} +1 & \text{if } k\|u - c\| \leq R \\ -1 & \text{otherwise} \end{cases} .$$

Siamese Tracking - SiamFC

- Tracking
 - Search image 4 times previous box
 - 5 Scales $1.025^{\{-2,-1,0,1,2\}}$
 - No model update



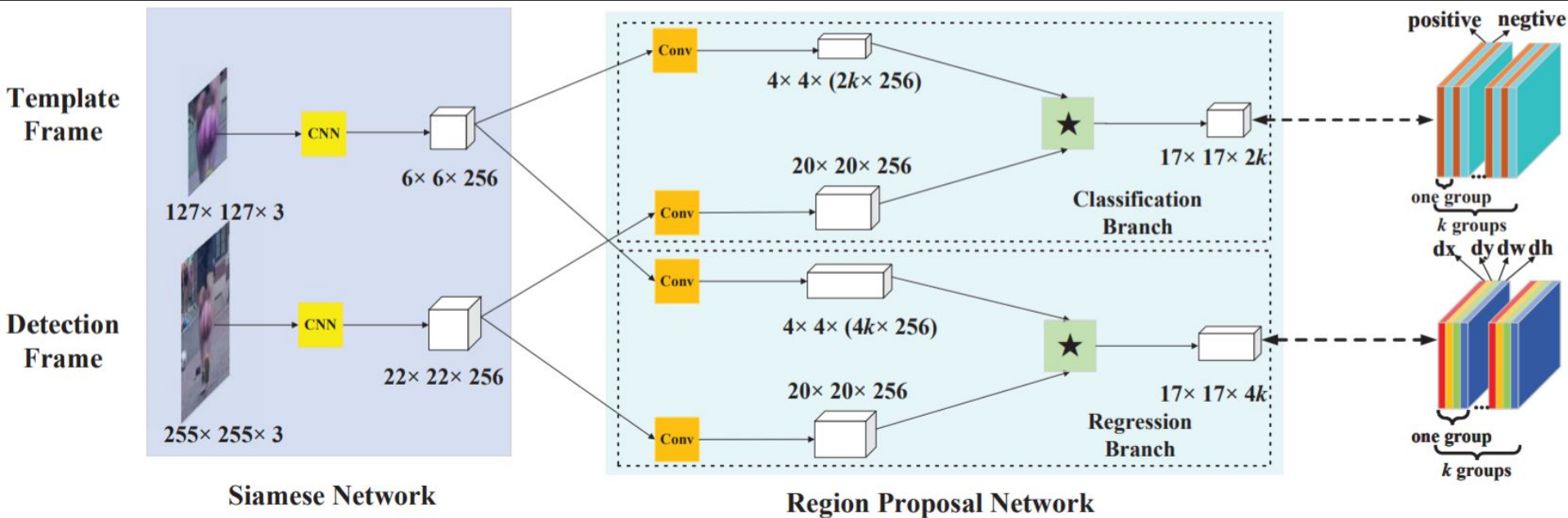
Siamese Tracking - SiamFC



State of the Art - SiamRPN

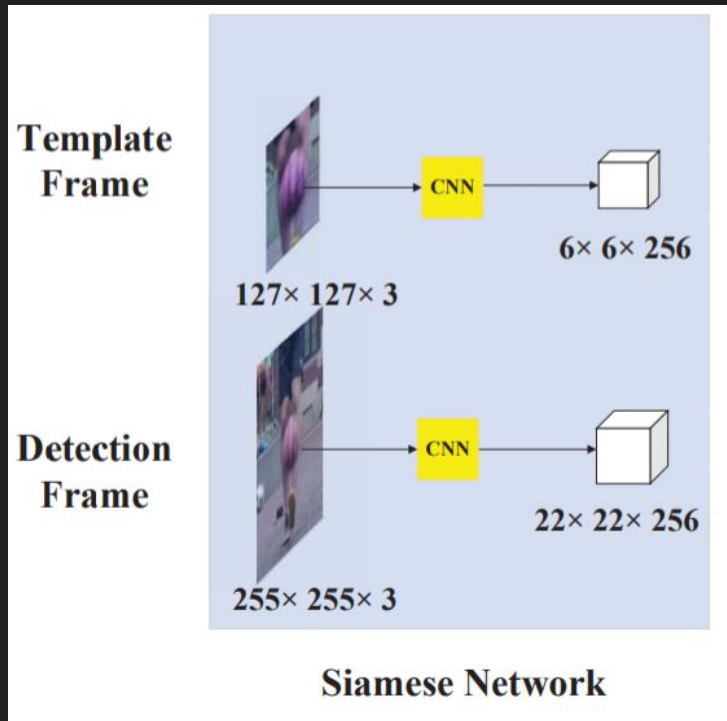
- “Signature Verification using a “Siamese” Time Delay Neural Network”
Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Sickinger and Roopak Shah
NIPS 1993
- “Fully-Convolutional Siamese Networks for Object Tracking”
Luca Bertinetto, Jack Valmadre, João F. Henriques, Andrea Vedaldi, Philip H. S. Torr
ECCV 2016
- “High Performance Visual Tracking with Siamese Region Proposal Network”
Bo Li, Junjie Yan, Wei Wu, Zheng Zhu, Xiaolin Hu
VOT 2018

State of the Art - SiamRPN



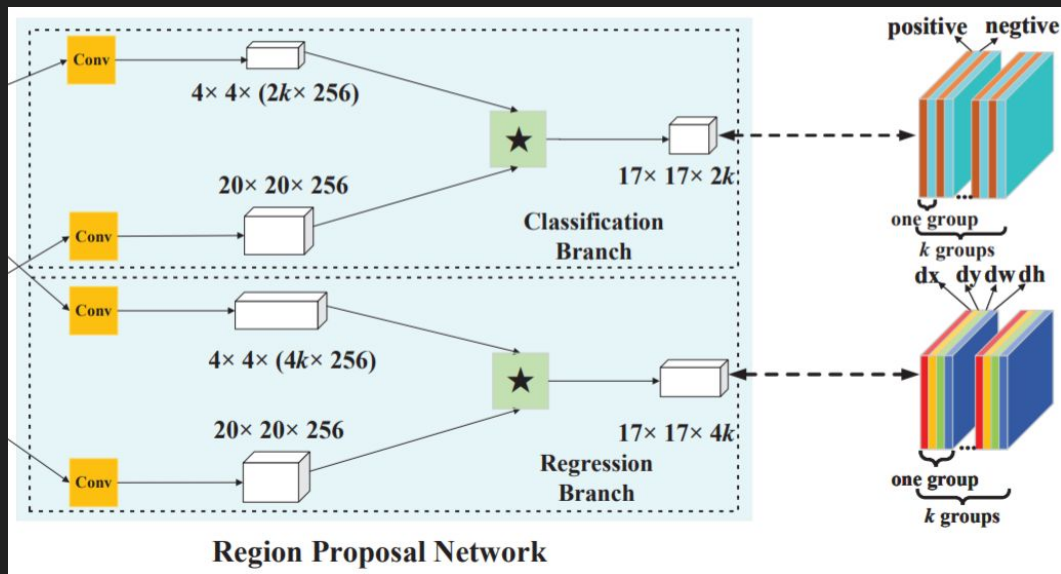
State of the Art - SiamRPN

- φ from SiamFC
- First 3 conv layers fixed



State of the Art - SiamRPN

- Similar to Faster R-CNN
 - k Anchors
 - positive/negative anchor map
 - $dx/dy/dw/dh$ anchor map
- Anchors
 - Pre-defined boxes
 - Fixed width, height ratios
 - Ratios [0.33, 0.5, 1, 2, 3]
 - Single scale
 - Centered anywhere in image



State of the Art - SiamRPN

- Classification Branch

- Positive - IoU > 0.6
- Negative - IoU < 0.3

- Regression Branch

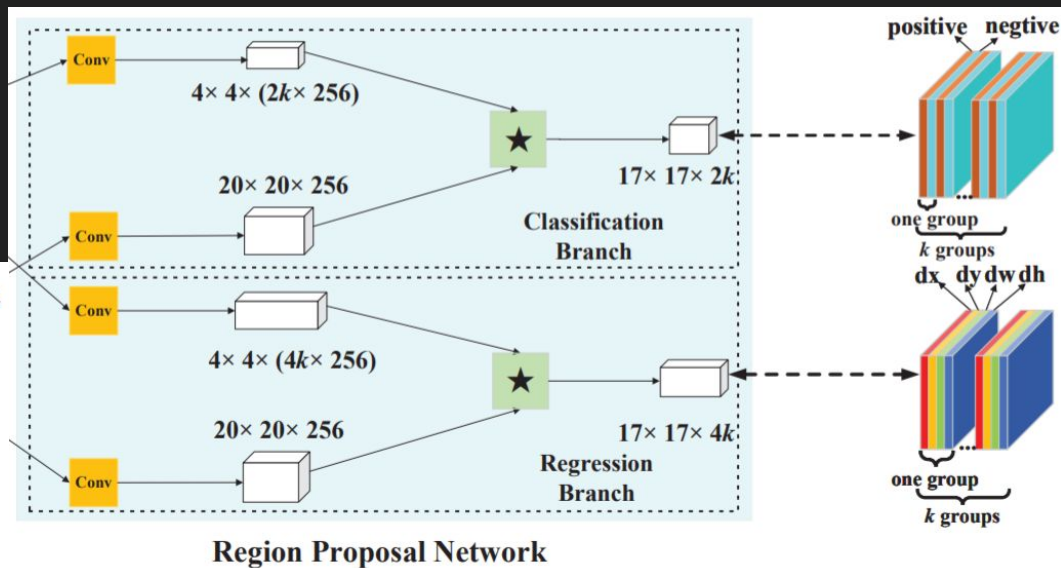
- dx/dy/dw/dh w.r.t. GT

$$x_i^{pro} = x_i^{an} + dx_l^{reg} * w_l^{an}$$

$$y_j^{pro} = y_j^{an} + dy_l^{reg} * h_l^{an}$$

$$w_l^{pro} = w_l^{an} * e^{dw_l}$$

$$h_l^{pro} = h_l^{an} * e^{dh_l}$$



State of the Art - SiamRPN

- Trained on pair of frames
 - ILSVRC - Random frame difference
 - Up to 100 frame difference
 - Youtube-BB - Continuously
 - Template - Detection pairs
 - 64 anchors per pair
 - At most 16 positive anchors
 - 50 epochs
- Classification Branch
 - Cross-Entropy loss L_{cls}
- Regression Branch
 - Smooth L_1 loss L_{reg}

$$\begin{aligned} loss &= L_{cls} + \lambda L_{reg} \\ L_{reg} &= \sum_{i=0}^3 smooth_{L1}(\delta[i], \sigma) \\ smooth_{L1}(x, \sigma) &= \begin{cases} 0.5\sigma^2 x^2, & |x| < \frac{1}{\sigma^2} \\ |x| - \frac{1}{2\sigma^2}, & |x| \geq \frac{1}{\sigma^2} \end{cases} \\ \delta[0] &= \frac{T_x - A_x}{A_w}, \quad \delta[1] = \frac{T_y - A_y}{A_h} \\ \delta[2] &= \ln \frac{T_w}{A_w}, \quad \delta[3] = \ln \frac{T_h}{A_h} \end{aligned}$$

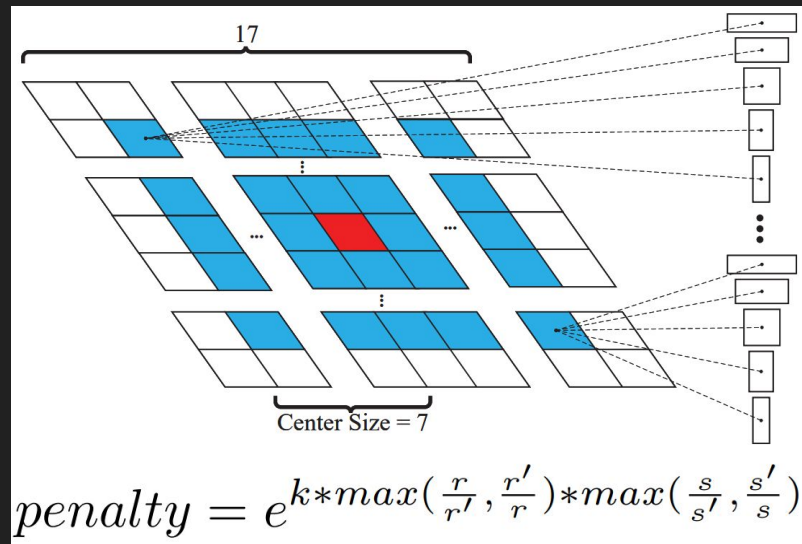
σ - hyperparameter

T - ground truth box

A - anchor box

State of the Art - SiamRPN

- Proposal Selection
 - Top K from Classification Maps
 - Discard anchors far from center
 - Apply cosine window
 - Size and Ratio change penalty
 - Re-ranked
 - Non-Maximum-Suppression
- Target box linearly interpolated



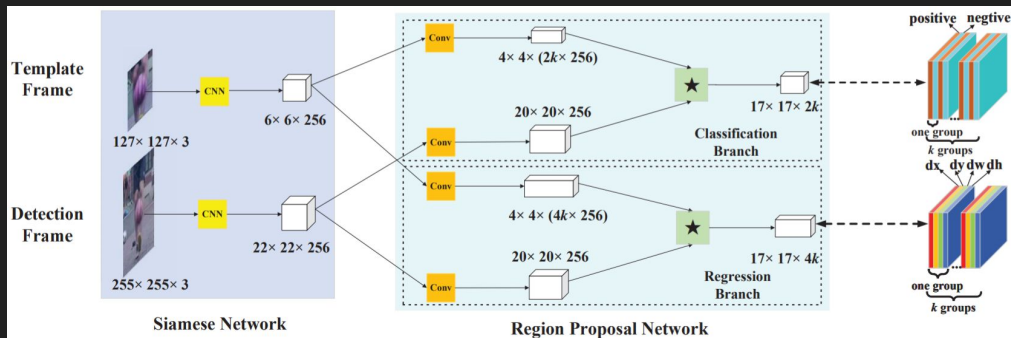
k - hyperparameter

r, r' - ratio of previous and current frame

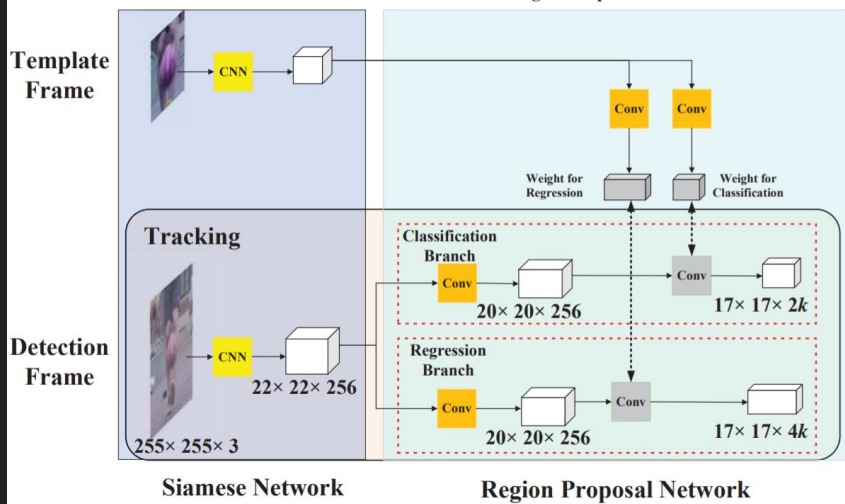
s, s' - scale of previous and current frame

State of the Art - SiamRPN

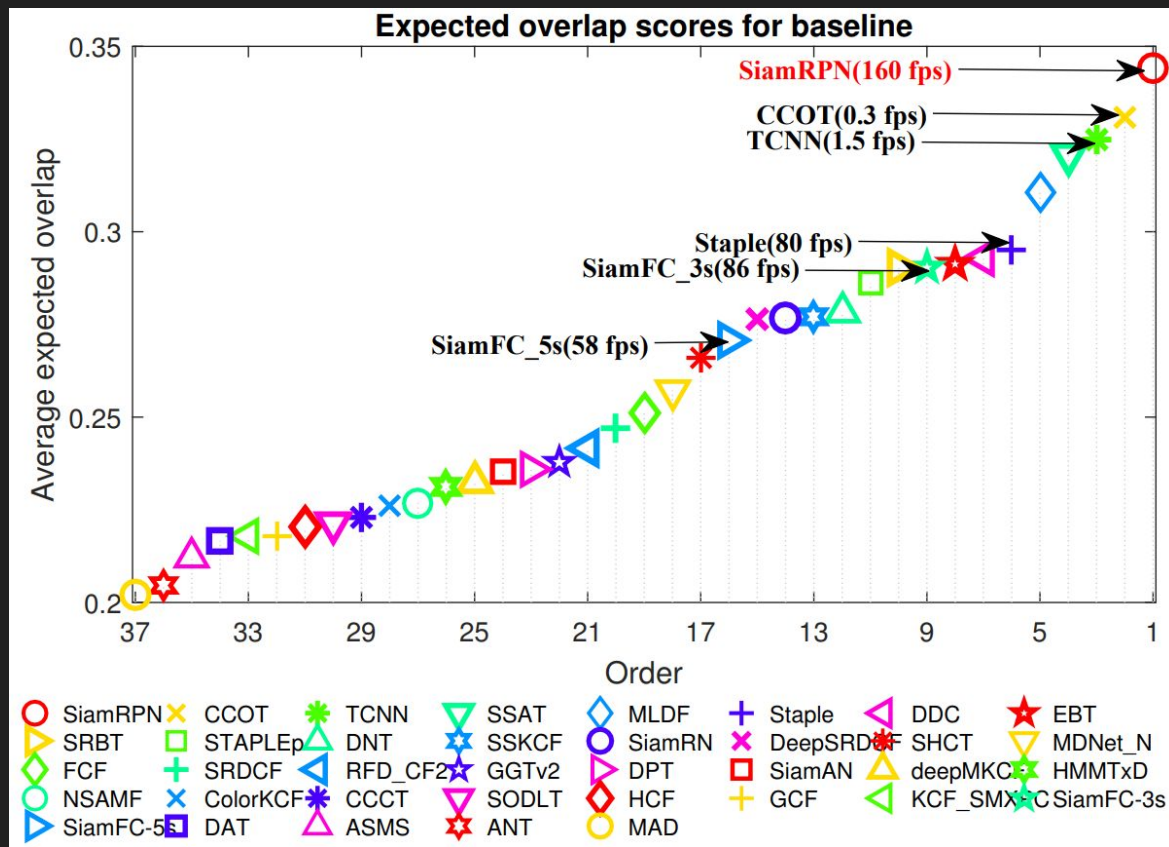
- Dynamic Model Tracking
 - Template = Previous Frame



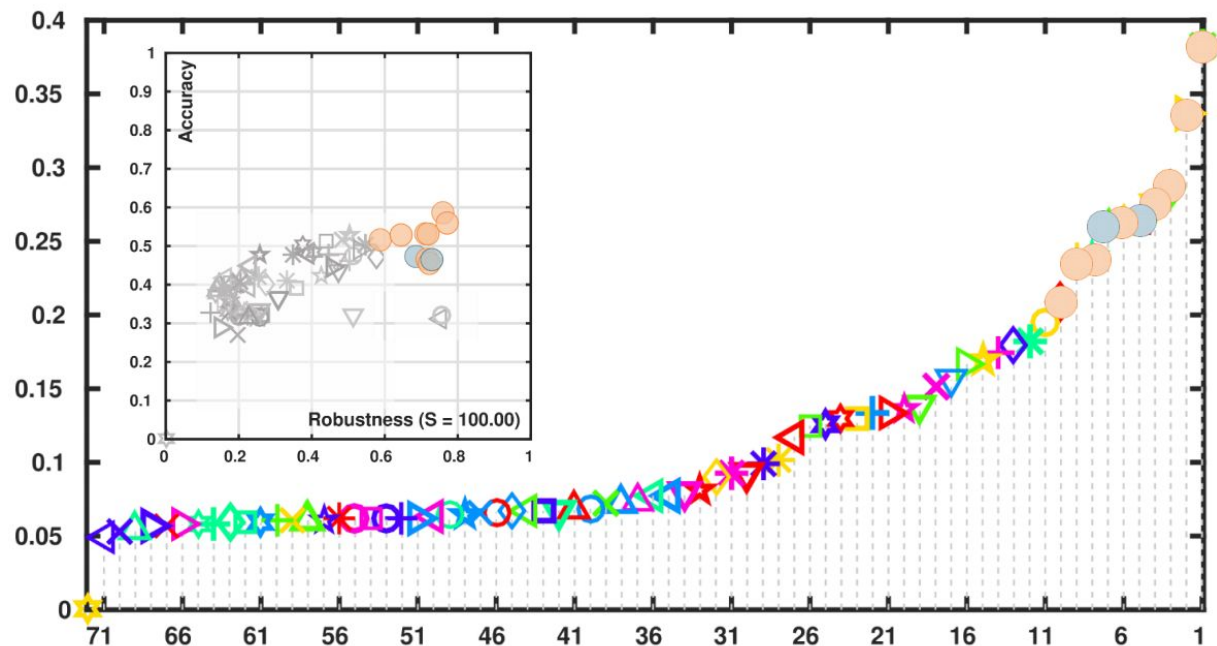
- One-Shot Detection
 - Template = Frame Zero
 - Online Tracking



State of the Art - SiamRPN



State of the Art - SiamRPN



Two classes:



Approach: [Siamese](#)
(e.g., SiamFc¹)

Features: CNN

Hardware: GPU



Approach: [CSRDCF²](#)

Features: HOG+CN

Hardware: CPU

Thank You