AUGMENTED STARTUPS
WHERE TECH ENTREPRENEURS COME TO LEARN
CREATE & INNOVATE

# Video Object Tracking and Segmentation

## SiamMask – A Unifying Approach

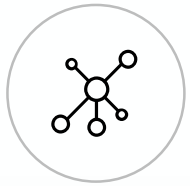
---

Paper Review

AUGMENTED STARTUPS
WHERE TECH ENTREPRENEURS COME TO LEARN
CREATE & INNOVATE

# How SiamMask Works?

SiamFC

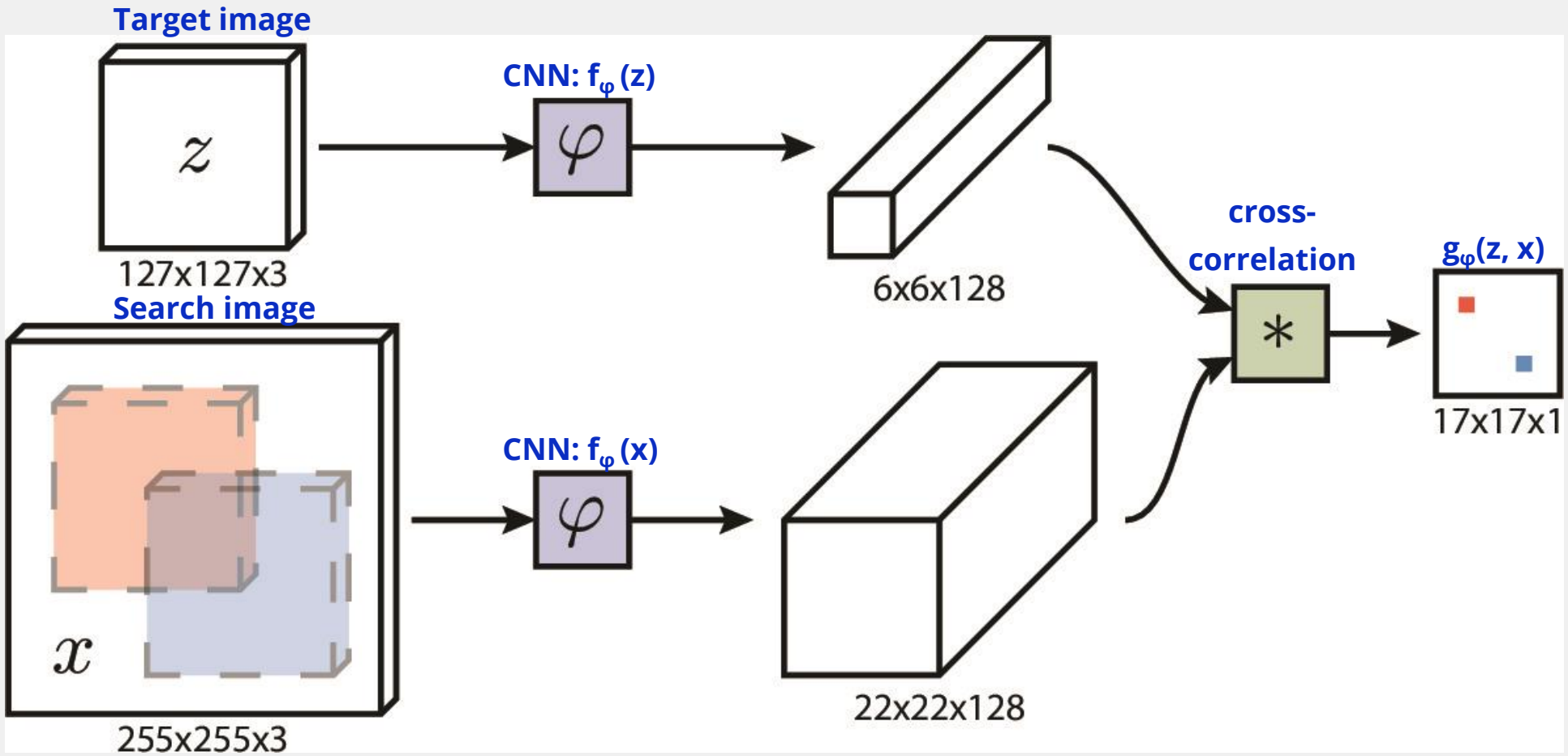


SiamRPN

SiamMask

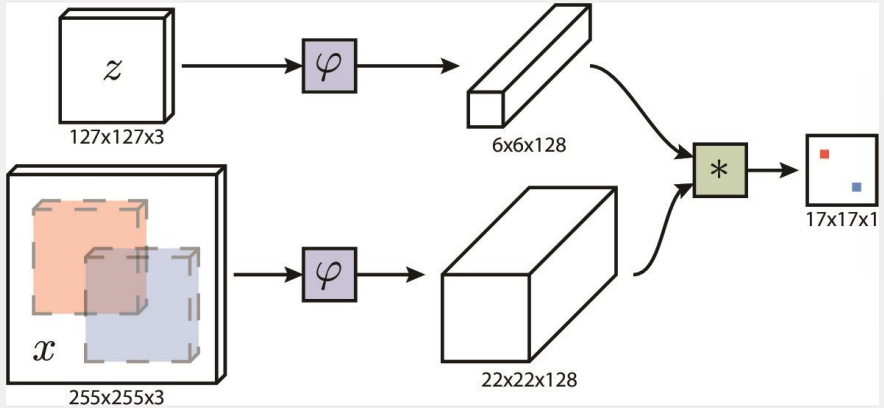
AUGMENTED STARTUPS
WHERE TECH ENTREPRENEURS COME TO LEARN
CREATE & INNOVATE

# Fully Convolutional Siamese Network



AUGMENTED STARTUPS
WHERE TECH ENTREPRENEURS COME TO LEARN CREATE & INNOVATE

# SiamFC and SiamRPN




$$g_\varphi(z, x) = f_\varphi(z) * f_\varphi(x)$$

$L_{sim}$ **(Logistic Loss of SiamFC)**

$L_{box}$ $L_{score}$ **(Cross-entropy Losses of SiamRPN)**

AUGMENTED STARTUPS
WHERE TECH ENTREPRENEURS COME TO LEARN
CREATE & INNOVATE

# SiamMask

$$g_{\varphi}(z, x) = f_{\varphi}(z) * f_{\varphi}(x)$$

$$L_{sim}$$

$$L_{box}$$

$$L_{score}$$

**Predicted mask corresponding to the n-th RoW**

$$m_n = h_{\theta}(g_{\varphi}^{n}(z, x))$$

$$L_{mask}$$

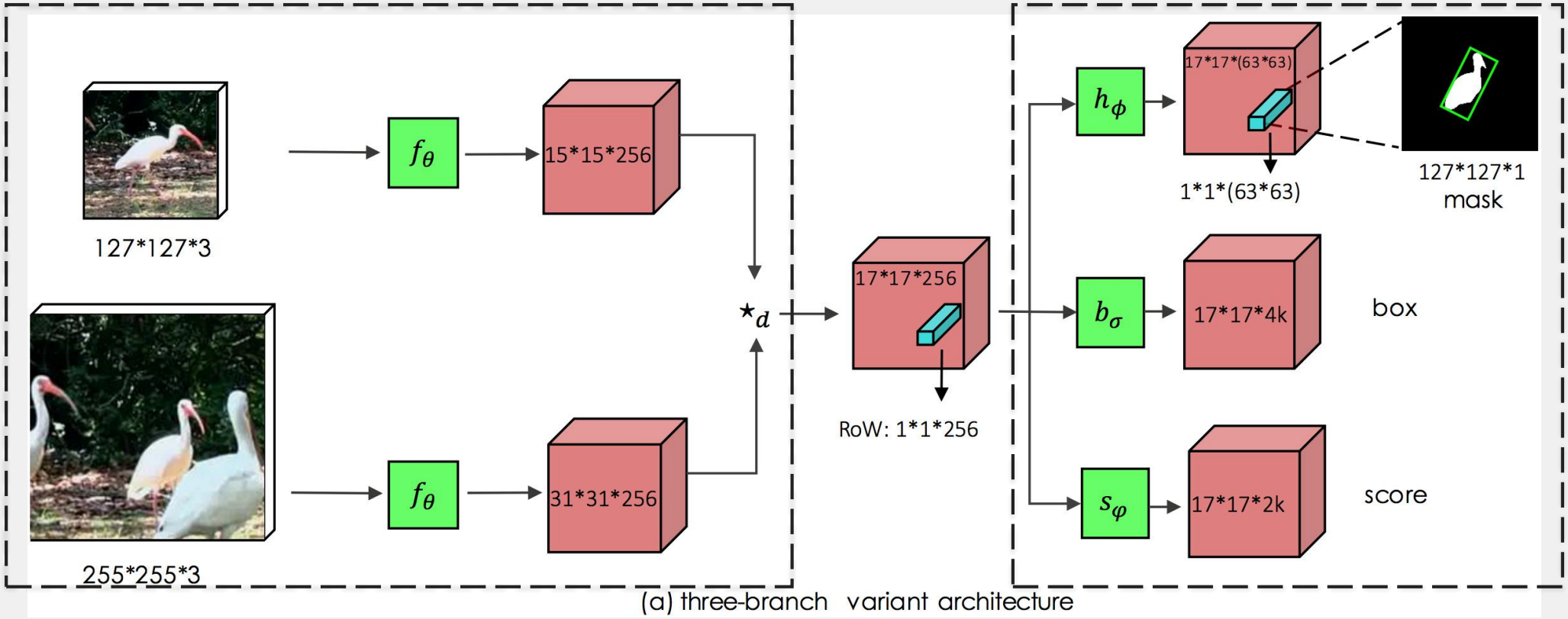
**Loss function for the mask prediction task**

AUGMENTED STARTUPS
WHERE TECH ENTREPRENEURS COME TO LEARN CREATE & INNOVATE

# SiamMask – variant1

## Fully convolutional network for RoW creation

## Neural Network for mask generation

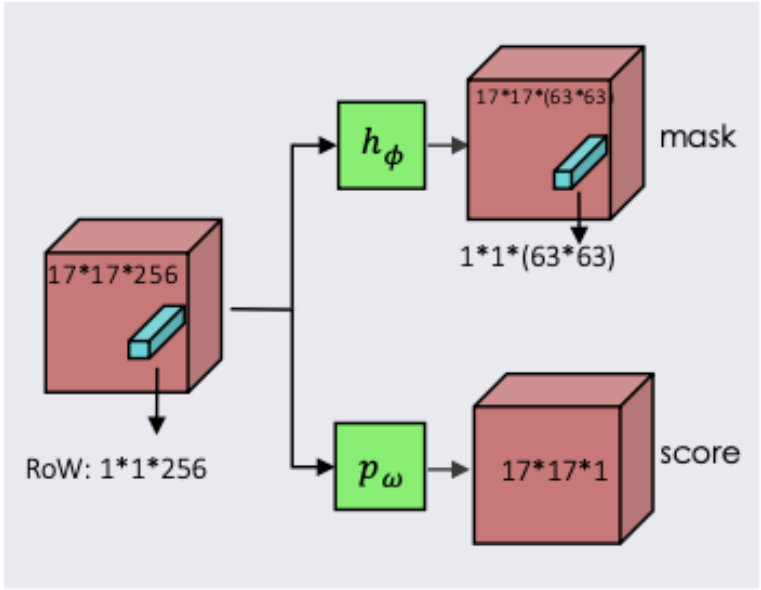



(a) three-branch variant architecture

$$\mathcal{L}_{3B} = \lambda_1 \cdot \mathcal{L}_{mask} + \lambda_2 \cdot \mathcal{L}_{score} + \lambda_3 \cdot \mathcal{L}_{box}$$

AUGMENTED STARTUPS
WHERE TECH ENTREPRENEURS COME TO LEARN
CREATE & INNOVATE

# SiamMask – Variant 2

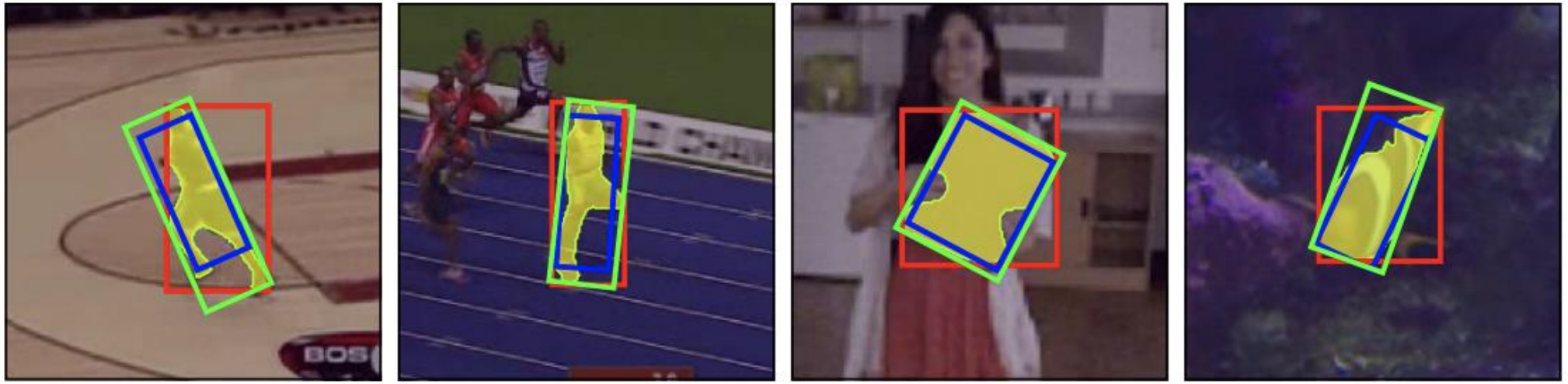



(b) two-branch variant head

$$\mathcal{L}_{2B} = \lambda_1 \cdot \mathcal{L}_{mask} + \lambda_2 \cdot \mathcal{L}_{sim}$$

AUGMENTED STARTUPS
WHERE TECH ENTREPRENEURS COME TO LEARN CREATE & INNOVATE

# SiamMask – Box Generation



**Minmax: the axis-aligned rectangle containing the object**
**MBR: the minimum bounding rectangle**
**Opt: the rectangle obtained via the optimisation strategy proposed in VOT-2016**

AUGMENTED STARTUPS
WHERE TECH ENTREPRENEURS COME TO LEARN
CREATE & INNOVATE

# Implementation Details

**Network Architecture**

ResNet-50 until the final convolutional layer of the 4-th stage as the backbone

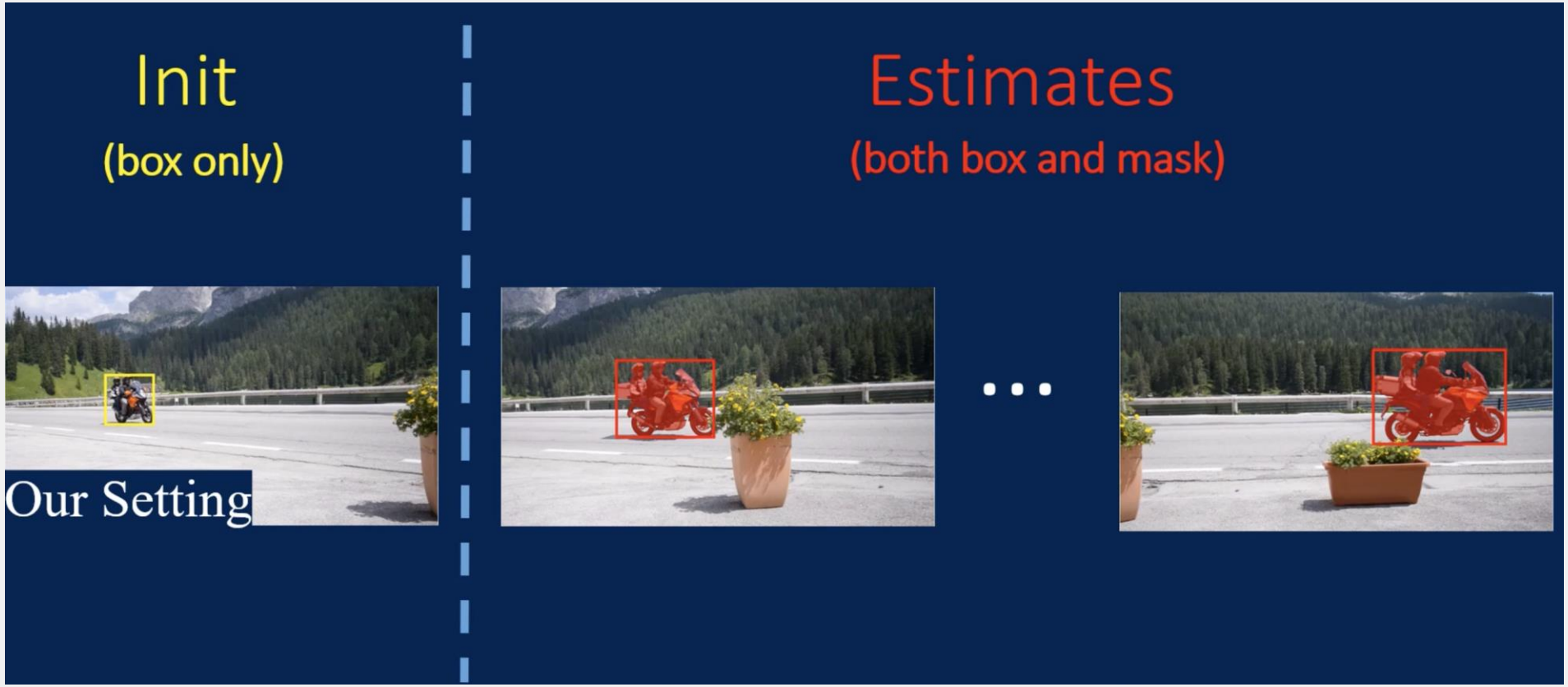
**Training**

examplar and search image patches of 127×127 and 255×255 pixels respectively

**Inference**

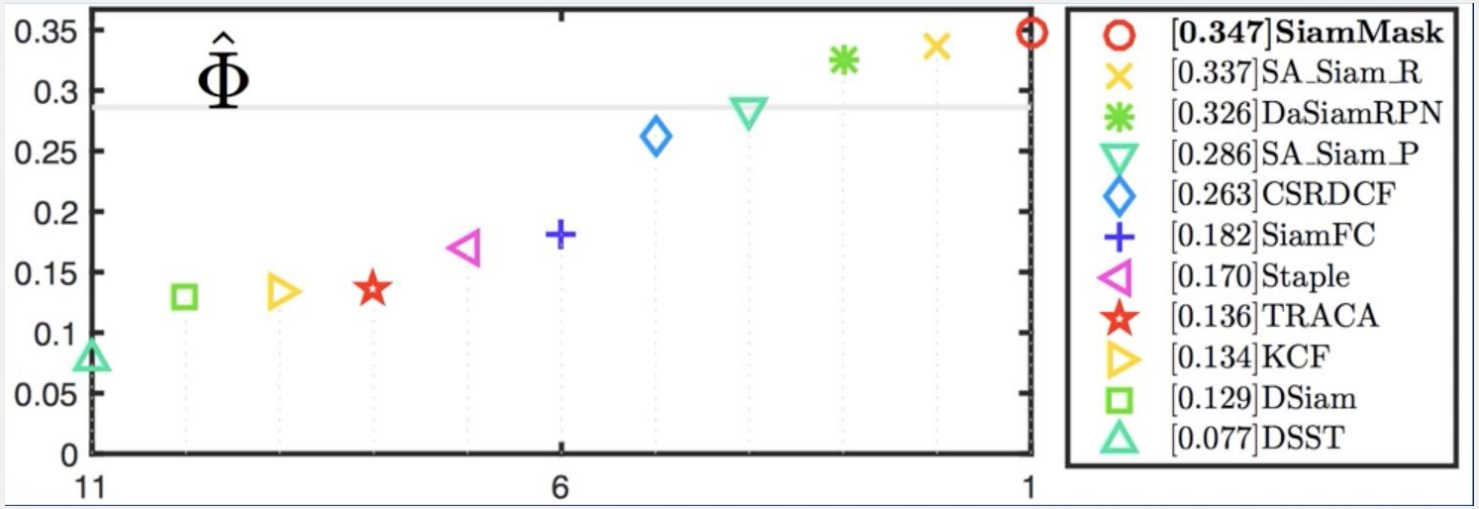
evaluated once per frame, without any adaptation

# SiamMask Performance



AUGMENTED STARTUPS
WHERE TECH ENTREPRENEURS COME TO LEARN CREATE & INNOVATE

# SiamMask Performance – VOT2018



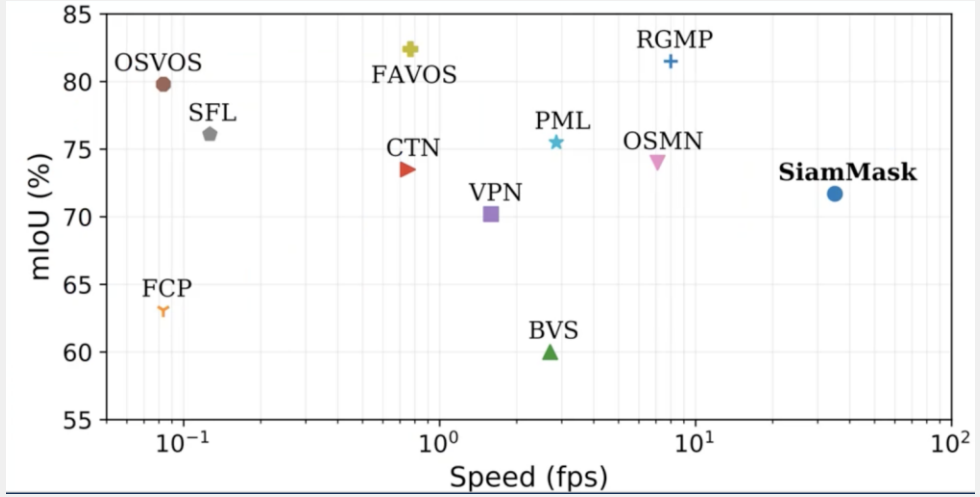

| | SiamMask | SiamMask-2B | DaSiamRPN [72] | SiamRPN [31] | SA_Siam_R [17] | CSRDCF [37] | STRCF [32] | LSART [56] | ECO [15] |
|---|---|---|---|---|---|---|---|---|---|
| EAO ↑ | **0.347** | 0.334 | 0.326 | 0.244 | 0.337 | 0.263 | 0.345 | 0.323 | 0.280 |
| Accuracy ↑ | **0.602** | 0.575 | 0.569 | 0.490 | 0.566 | 0.466 | 0.523 | 0.495 | 0.484 |
| Robustness ↓ | 0.288 | 0.306 | 0.337 | 0.460 | 0.258 | 0.318 | **0.215** | 0.218 | 0.276 |
| Speed (fps) ↑ | 35 | 40 | 160 | **200** | 32.4 | 48.9 | 2.9 | 1.7 | 3.7 |

AUGMENTED STARTUPS
WHERE TECH ENTREPRENEURS COME TO LEARN CREATE & INNOVATE

# SiamMask Performance – DAVIS 2016/2017

| | FT | M | $\mathcal{J}_{\mathcal{M}}\uparrow$ | $\mathcal{J}_{\mathcal{O}}\uparrow$ | $\mathcal{J}_{\mathcal{D}}\downarrow$ | $\mathcal{F}_{\mathcal{M}}\uparrow$ | $\mathcal{F}_{\mathcal{O}}\uparrow$ | $\mathcal{F}_{\mathcal{D}}\downarrow$ | Speed |
|---|---|---|---|---|---|---|---|---|---|
| OnAVOS [61] | ✔ | ✔ | **86.1** | **96.1** | 5.2 | **84.9** | **89.7** | 5.8 | 0.08 |
| MSK [45] | ✔ | ✔ | 79.7 | 93.1 | 8.9 | 75.4 | 87.1 | 9.0 | 0.1 |
| $\text{MSK}_b$ [45] | ✔ | ✘ | 69.6 | - | - | - | - | - | 0.1 |
| SFL [12] | ✔ | ✔ | 76.1 | 90.6 | 12.1 | 76.0 | 85.5 | 10.4 | 0.1 |
| FAVOS [11] | ✘ | ✔ | 82.4 | 96.5 | 4.5 | 79.5 | 89.4 | 5.5 | 0.8 |
| RGMP [64] | ✘ | ✔ | 81.5 | 91.7 | 10.9 | 82.0 | 90.8 | 10.1 | 8 |
| PML [10] | ✘ | ✔ | 75.5 | 89.6 | 8.5 | 79.3 | 93.4 | 7.8 | 3.6 |
| OSMN [67] | ✘ | ✔ | 74.0 | 87.6 | 9.0 | 72.9 | 84.0 | 10.6 | 8.0 |
| PLM [71] | ✘ | ✔ | 70.2 | 86.3 | 11.2 | 62.5 | 73.2 | 14.7 | 6.7 |
| VPN [24] | ✘ | ✔ | 70.2 | 82.3 | 12.4 | 65.5 | 69.0 | 14.4 | 1.6 |
| **SiamMask** | ✘ | ✘ | 71.7 | 86.8 | **3.0** | 67.8 | 79.8 | **2.1** | **35** |

| | FT | M | $\mathcal{J}_{\mathcal{M}}\uparrow$ | $\mathcal{J}_{\mathcal{O}}\uparrow$ | $\mathcal{J}_{\mathcal{D}}\downarrow$ | $\mathcal{F}_{\mathcal{M}}\uparrow$ | $\mathcal{F}_{\mathcal{O}}\uparrow$ | $\mathcal{F}_{\mathcal{D}}\downarrow$ | Speed |
|---|---|---|---|---|---|---|---|---|---|
| OnAVOS [61] | ✔ | ✔ | **61.6** | **67.4** | 27.9 | **69.1** | **75.4** | 26.6 | 0.1 |
| OSVOS [7] | ✔ | ✔ | 56.6 | 63.8 | 26.1 | 63.9 | 73.8 | 27.0 | 0.1 |
| FAVOS [11] | ✘ | ✔ | 54.6 | 61.1 | 14.1 | 61.8 | 72.3 | 18.0 | 0.8 |
| OSMN [67] | ✘ | ✔ | 52.5 | 60.9 | 21.5 | 57.1 | 66.1 | 24.3 | 8.0 |
| **SiamMask** | ✘ | ✘ | 51.1 | 60.5 | **-1.1** | 55.0 | 64.3 | **1.9** | **35** |

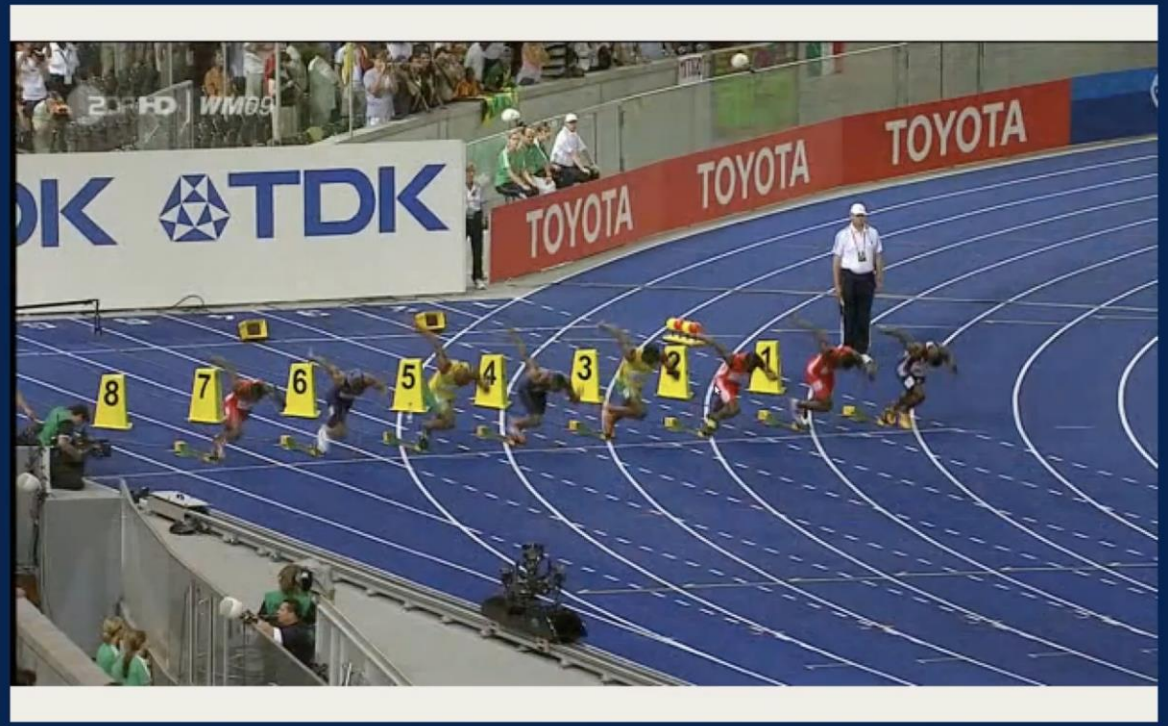


# SiamMask Performance – Experimental Results

UNIVERSITY OF OXFORD

SiamMask runs at 35 fps (on GPU)

(video slowed down for better visualization)




VOT sequences

AUGMENTED STARTUPS
WHERE TECH ENTREPRENEURS COME TO LEARN CREATE & INNOVATE

# Important Links

**Blog**

http://www.robots.ox.ac.uk/~qwang/SiamMask/

**Paper**

https://arxiv.org/pdf/1812.05050.pdf

**Code**

https://github.com/augmentedstartups/SiamMask

AUGMENTED STARTUPS
WHERE TECH ENTREPRENEURS COME TO LEARN
CREATE & INNOVATE

# Video Object Tracking and Segmentation

## SiamMask – A Unifying Approach

---

Paper Review