Fully-Convolutional Siamese Networks for Object Tracking

Luca et. al
University of Oxford: Reference

– Review by Kapil Wanaskar, SJSU

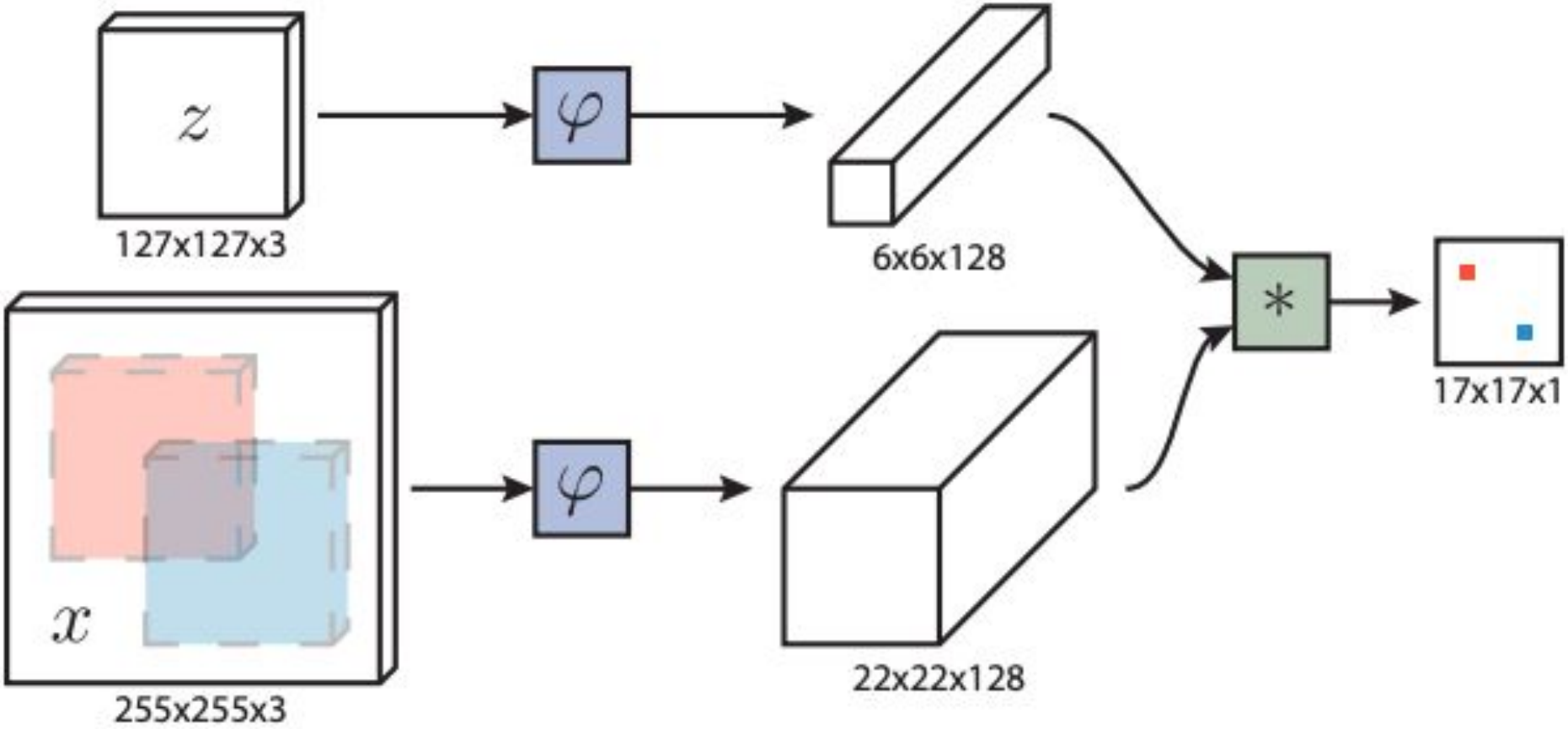

Fig. 1: Fully-convolutional Siamese architecture

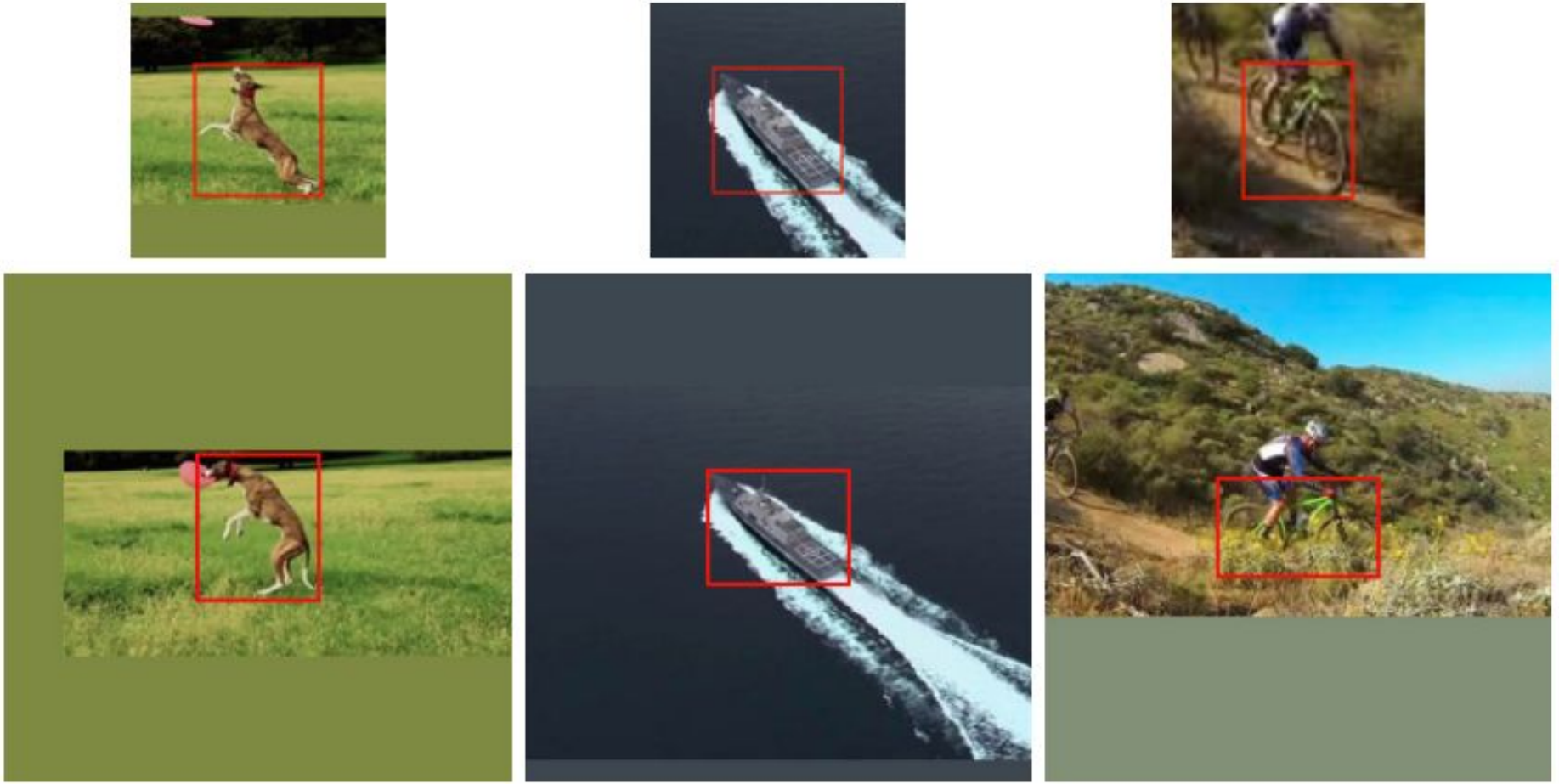

Fig. 2: Training pairs extracted from the same video

Layer	Support	Chan. map	Stride	Activation size for exemplar	Activation size for search	Activation size 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	$\times 128$

Table 1: Architecture of convolutional embedding function

Frame 1 (init.) Frame 50 Frame 100 Frame 200

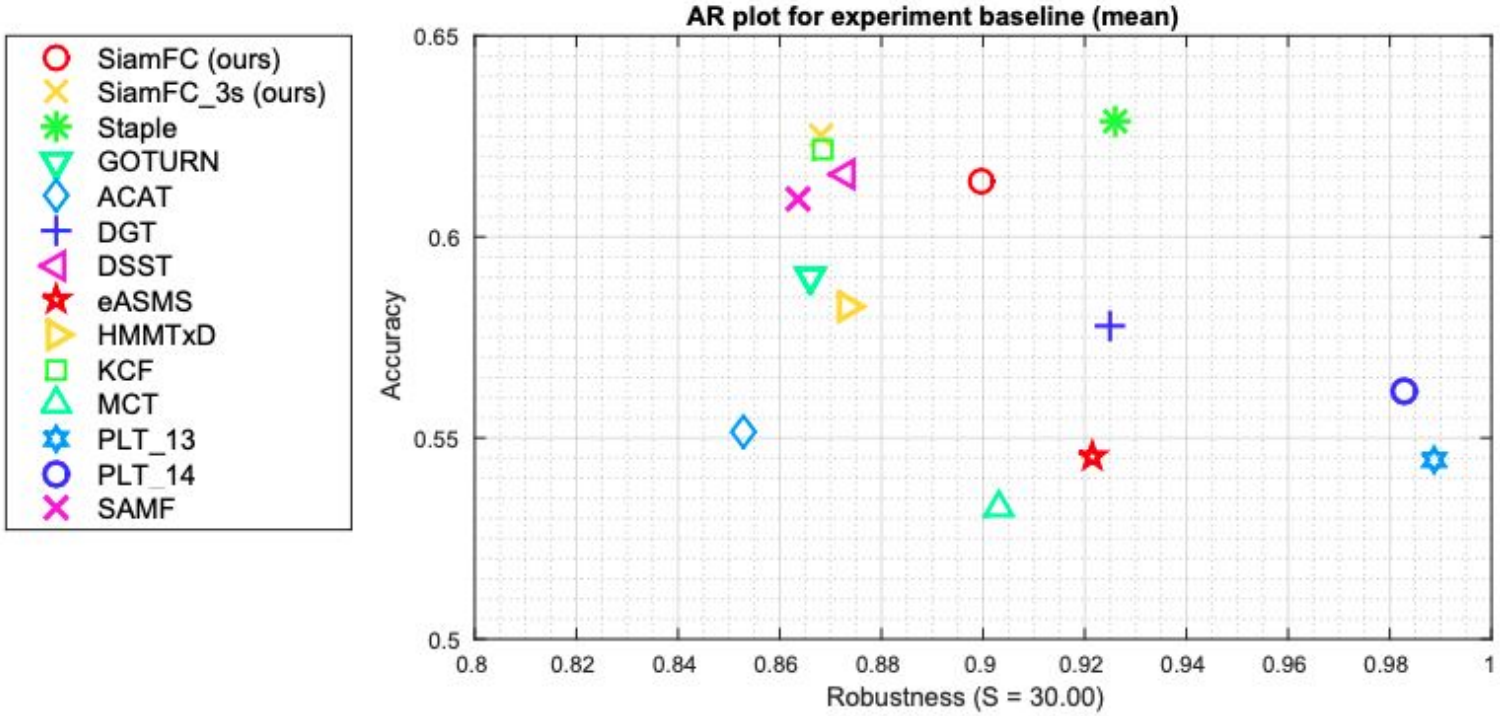

Fig. 4: VOT-14 Accuracy-robustness plot. Best trackers are closer to the top-right corner.