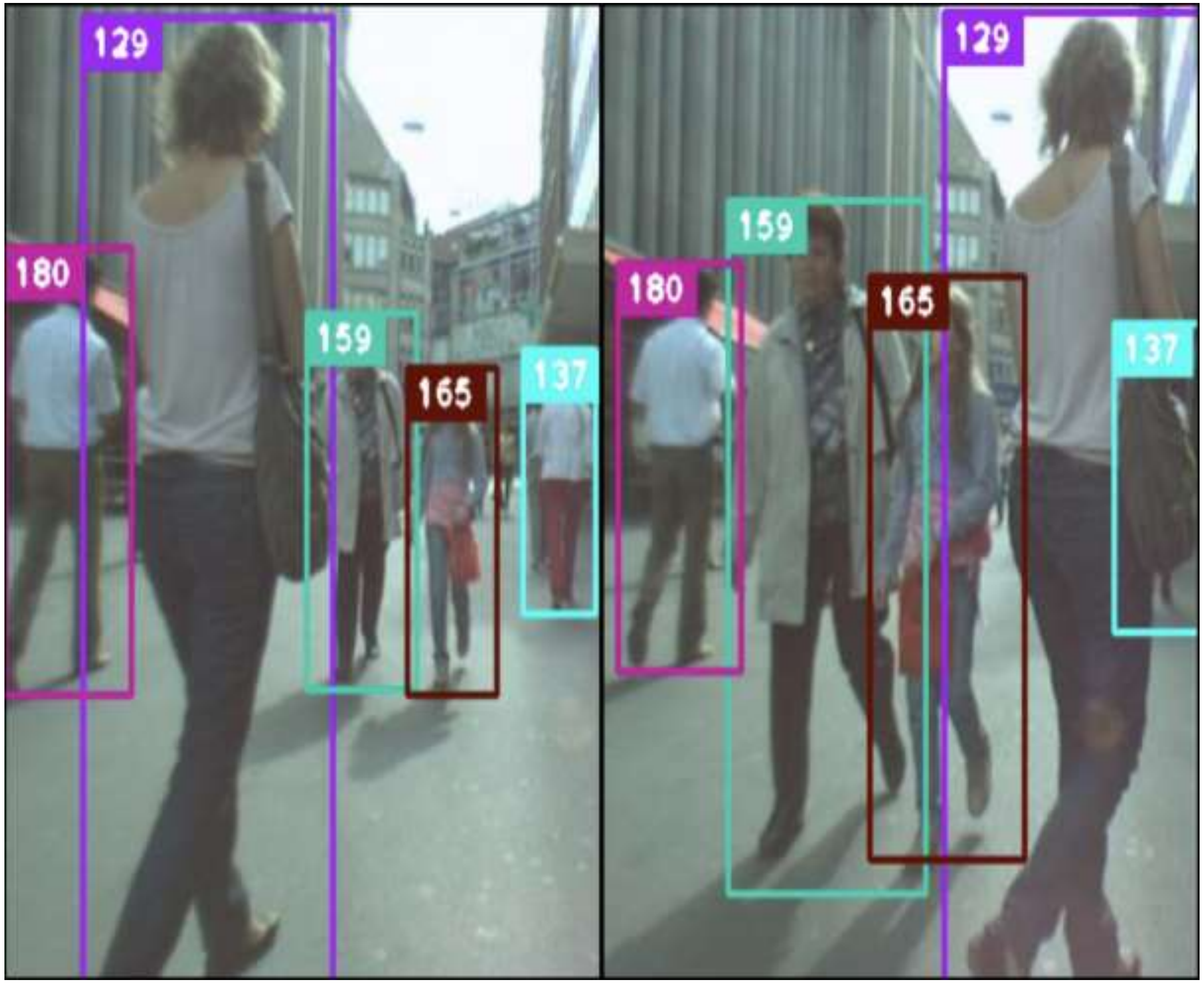
Deep learning for tracking and ultrasound images

Literature review

Lydia Abdemeziem



ENSEMBLE, PRENONS
LE CANCER DE VITESSE

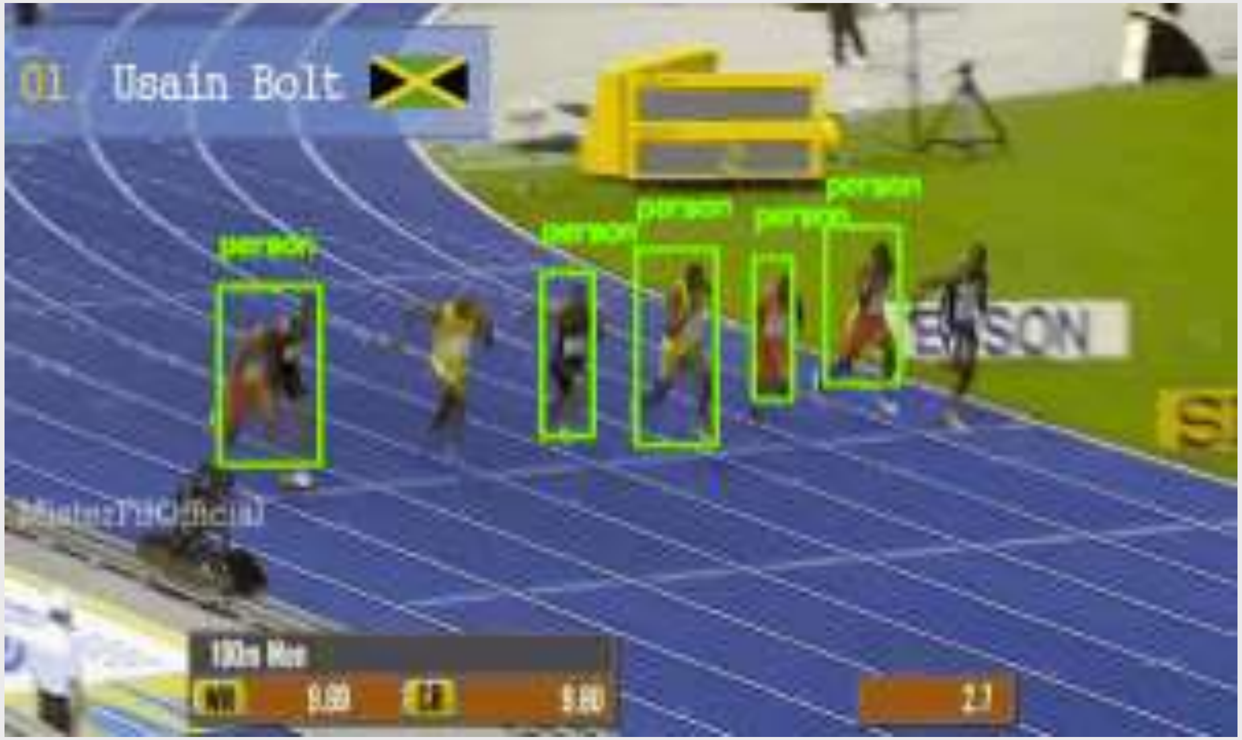
Summary

- Introduction to detection and tracking
- Popular detection method : YOLO
- Introduction to siamese networks : SiamFC
- Detection and tracking method : DeepSort
- Tracking and segmentation method : SiamMask
- US applications :
    - COSD-NN
    - Siam-U-Net
- Conclusion

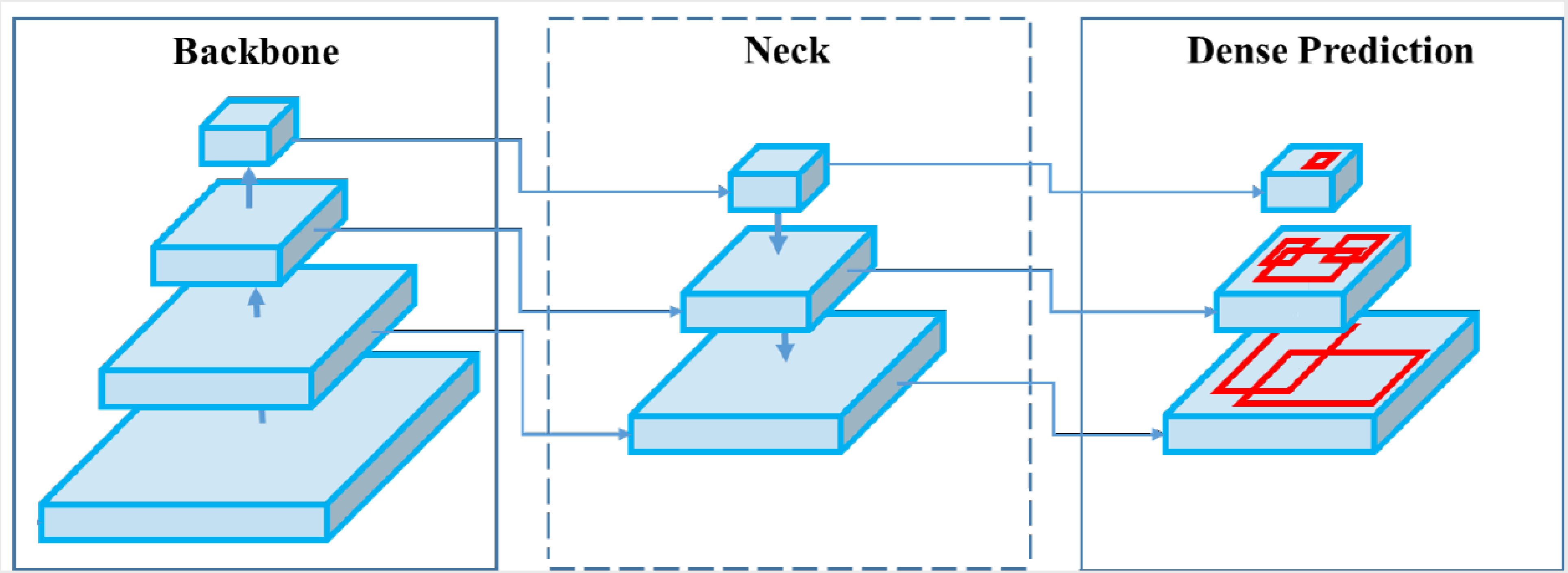
Deep learning and tracking

- Detection : detect an object in a frame
- Two levels of object tracking :
- Single Object Tracking(SOT)
- Multiple Object Tracking(MOT).

- Challenges :
  - Occlusion
  - Variation in viewpoints
  - Non stationary data



YOLOv4: Optimal Speed and Accuracy of Object Detection



- **We have 3 apparent blocks, after the input image:**
- Backbone
- Neck
- Head : Dense Prediction

Yolov4 - Architecture

Backbone

- A deep neural network composed mainly of convolution layers.
- Main objective is to extract the essential features
- The YoloV4 backbone architecture is composed of three parts:
  - Bag of freebies
  - Bag of specials
  - CSPDarknet53

Neck

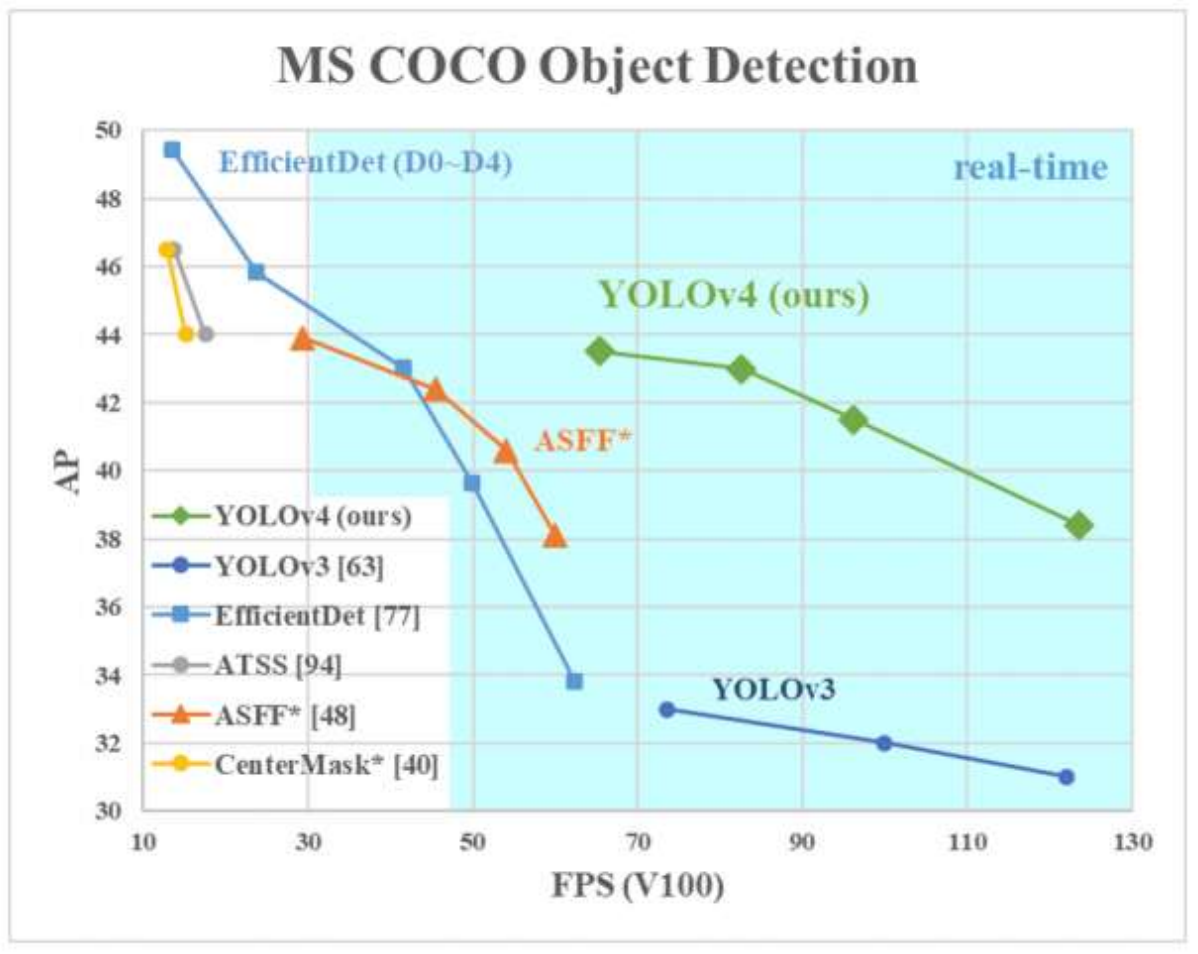
- Add extra layers between the backbone and the head (dense prediction block).
- Combine relevant features for detection

Head

- Locate bounding boxes
- Classify what's inside each box
- T same process as in YOLOv3. The network detects the bounding box coordinates (x,y,w,h) as well as the confidence score for a class.

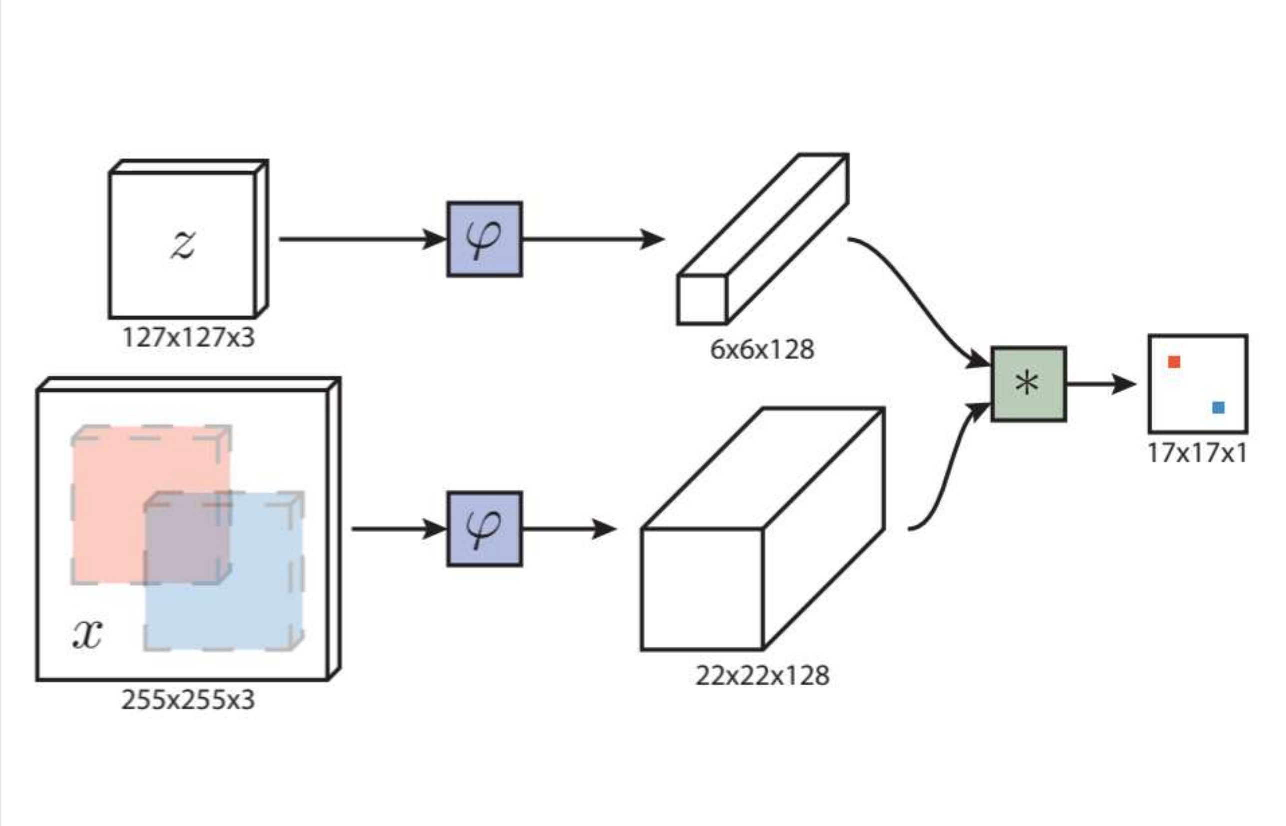
Results

- YOLOv4 runs twice faster than EfficientDet with comparable performance.
- Improves YOLOv3's AP and FPS by 10% and 12%, respectively.



Fully-convolutional siamese networks for object tracking

- **Two inputs different size :**
  - Smaller (z : target)
  - Bigger (x: search area)
- **Cross-correlation layer :** computes the similarity at all translated sub windows on a dense grid in a single evaluation
- **Output is a score map**



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

LITO Inserm institut Curie

SiamFC : training

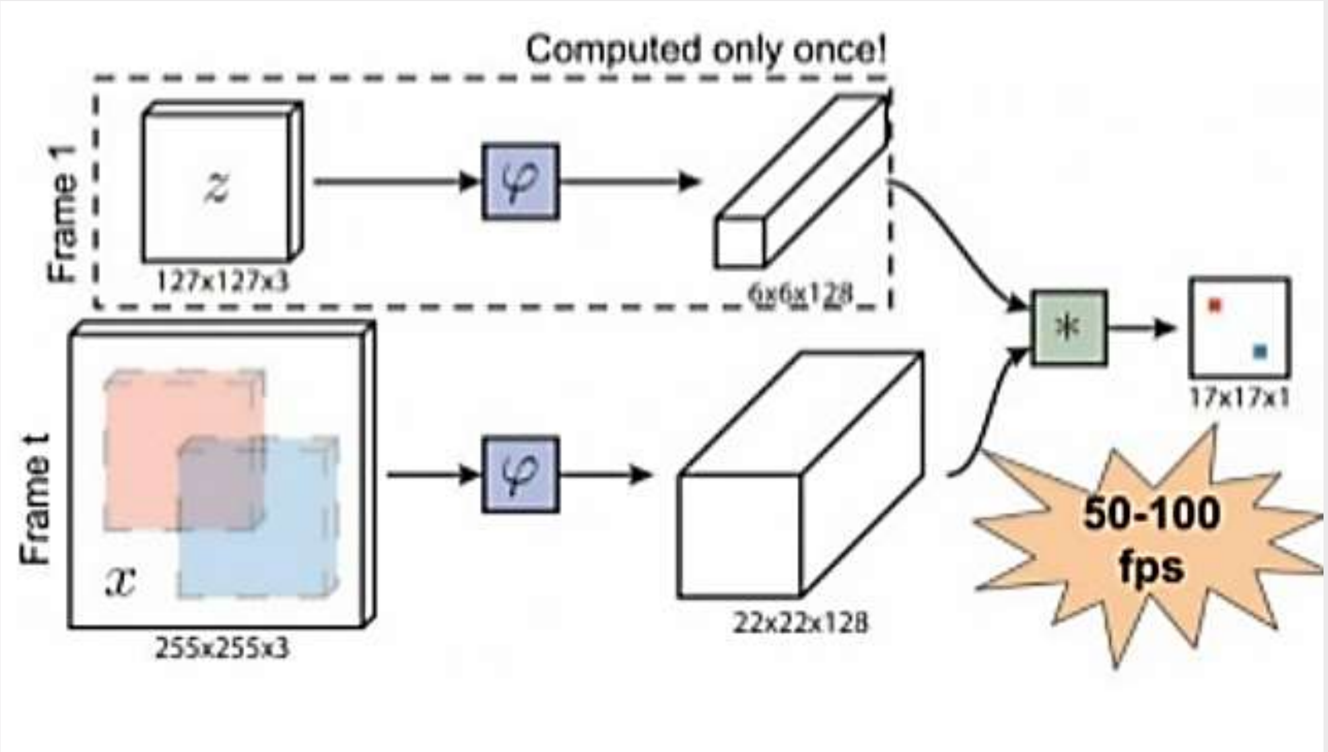
- Dataset build by extracting two patches with different amount of context for every labelled object. Then resized to 127x127 and 255x255.

- Pick random video and random pair of frames within the video
- Mean of logistic loss over all positions

Training pairs extracted from the same video

SiamFC : traking pipeline

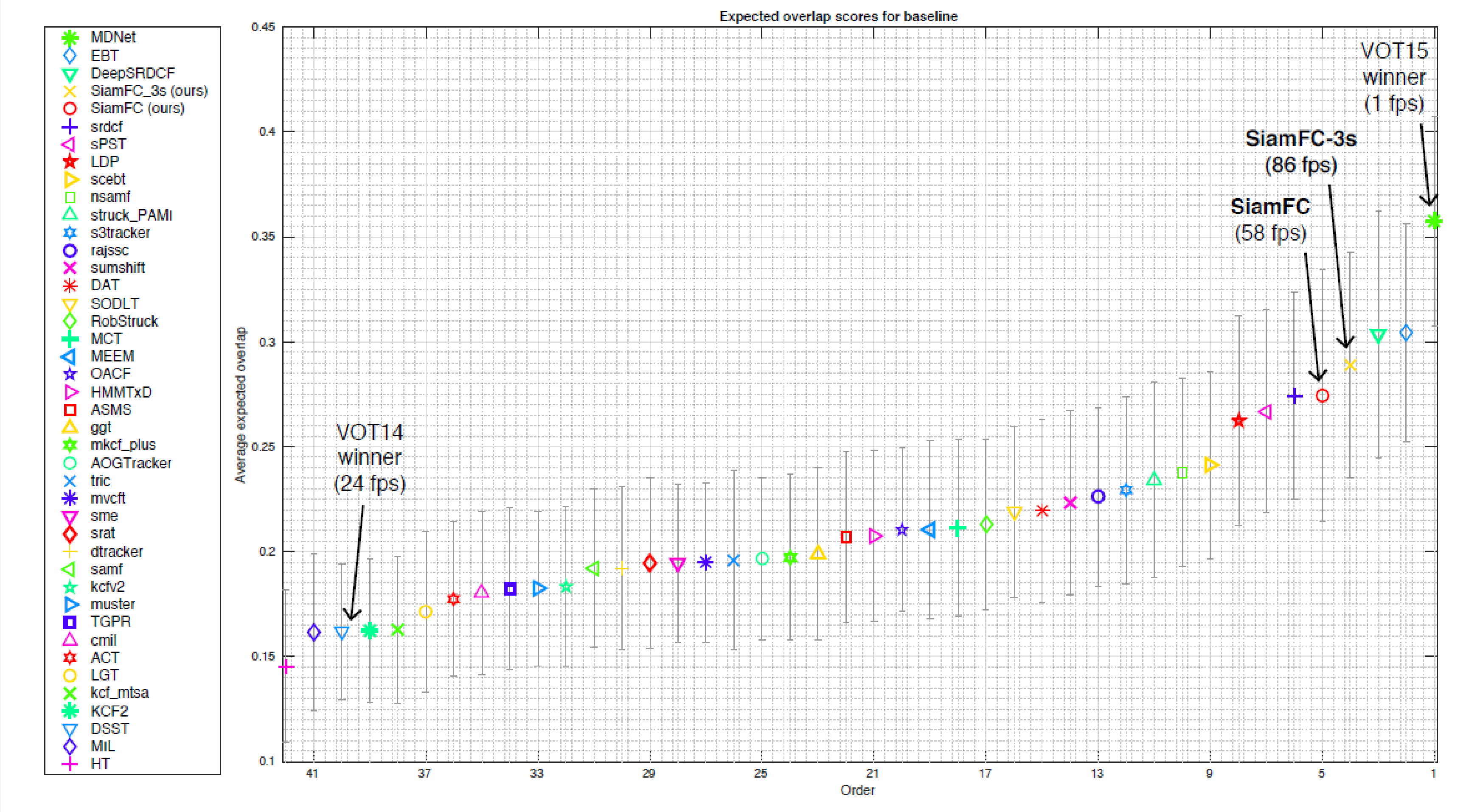
- Activation for the exemplar z only computed for the first frame
- Subwindow of x with max similarity sets the new location
- No update of target representation
- No bbox regression
- No fine-tuning > fast
- Tricks :
  - Pyramid of 3 scales
  - Cosine window to penalize large displacements



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

SiamFc : Evaluation

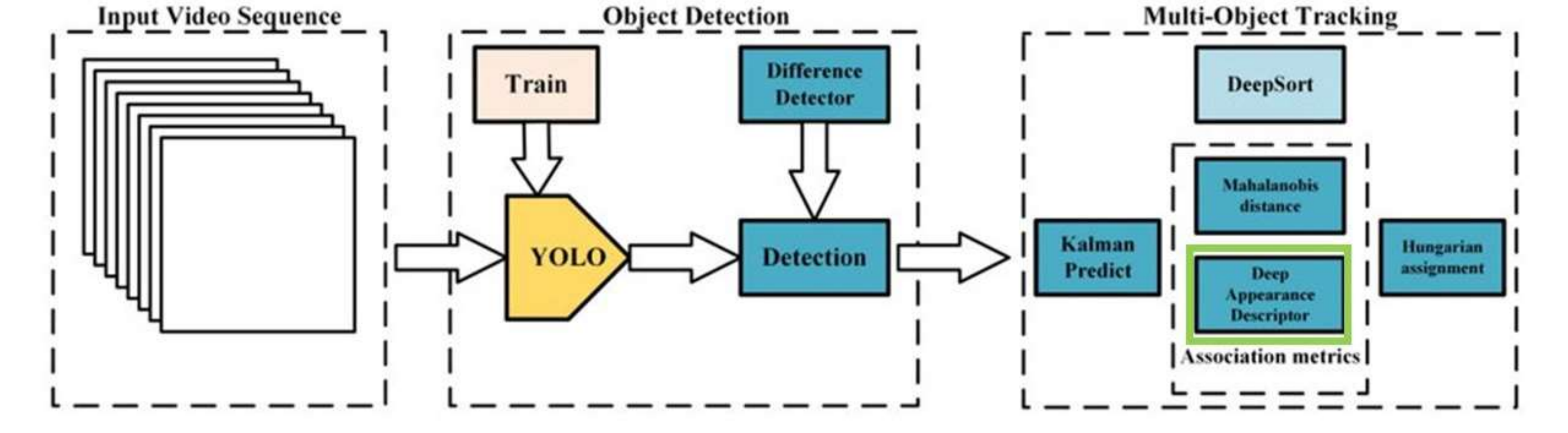
- At 1fps, the best tracker is almost 2 orders of magnitude slower than SiamFc with 86 fps
- state of the art for general trackets (VOT-15)
- None among the top 15 trackers operate above 20 fps



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

Simple online and realtime tracking with a deep association metric

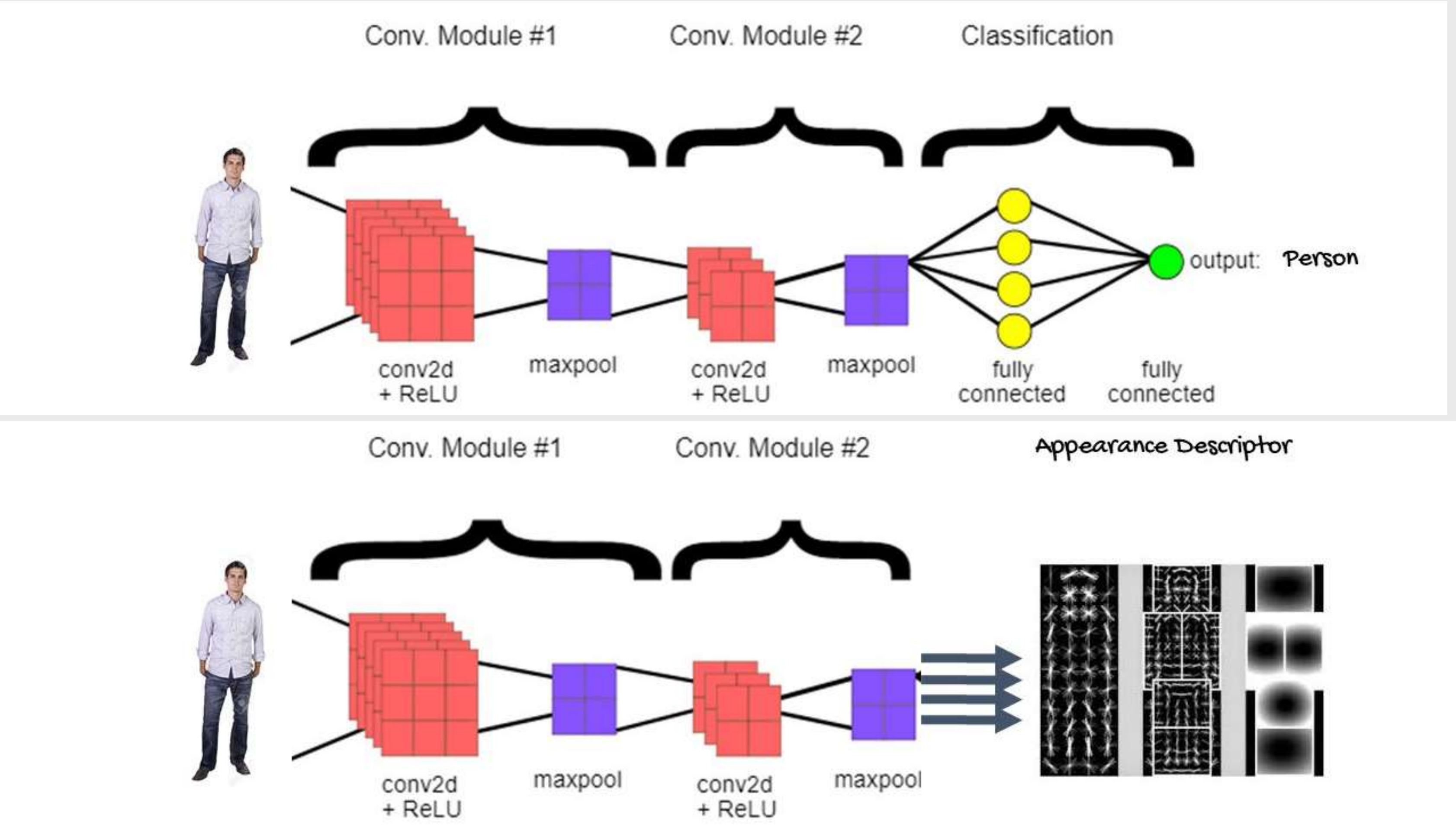
- Improvement of Simple Online Realtime Tracking (SORT)
- SORT comprises of 4 core components



Wojke, Nicolai Bewley, Alex Paulus, Dietrich

DeepSort

- DeepSort adds a pre-trained neural net to generate features for objects



DeepSort : Results

- In comparison to SORT, ID switches reduce from 1423 to 781. This is a decrease of approximately 45%.

		MOTA ↑	MOTP ↑	MT ↑	ML ↓	ID ↓	FM ↓	FP ↓	FN ↓	Runtime ↑
KDNT [16]*	BATCH	68.2	79.4	41.0%	19.0%	933	1093	11479	45605	0.7 Hz
LMP_p [17]*	BATCH	**71.0**	**80.2**	**46.9%**	21.9%	434	**587**	7880	**44564**	0.5 Hz
MCMOT_HDM [18]	BATCH	62.4	78.3	31.5%	24.2%	1394	1318	9855	57257	35 Hz
NOMTwSDP16 [19]	BATCH	62.2	79.6	32.5%	31.1%	**406**	642	**5119**	63352	3 Hz
EAMTT [20]	**ONLINE**	52.5	78.8	19.0%	34.9%	910	**1321**	**4407**	81223	12 Hz
POI [16]*	**ONLINE**	**66.1**	79.5	**34.0%**	20.8%	805	3093	5061	**55914**	10 Hz
SORT [12]*	**ONLINE**	59.8	**79.6**	25.4%	22.7%	1423	1835	8698	63245	**60 Hz**
Deep SORT (Ours)*	**ONLINE**	61.4	79.1	32.8%	**18.2%**	**781**	2008	12852	56668	40 Hz

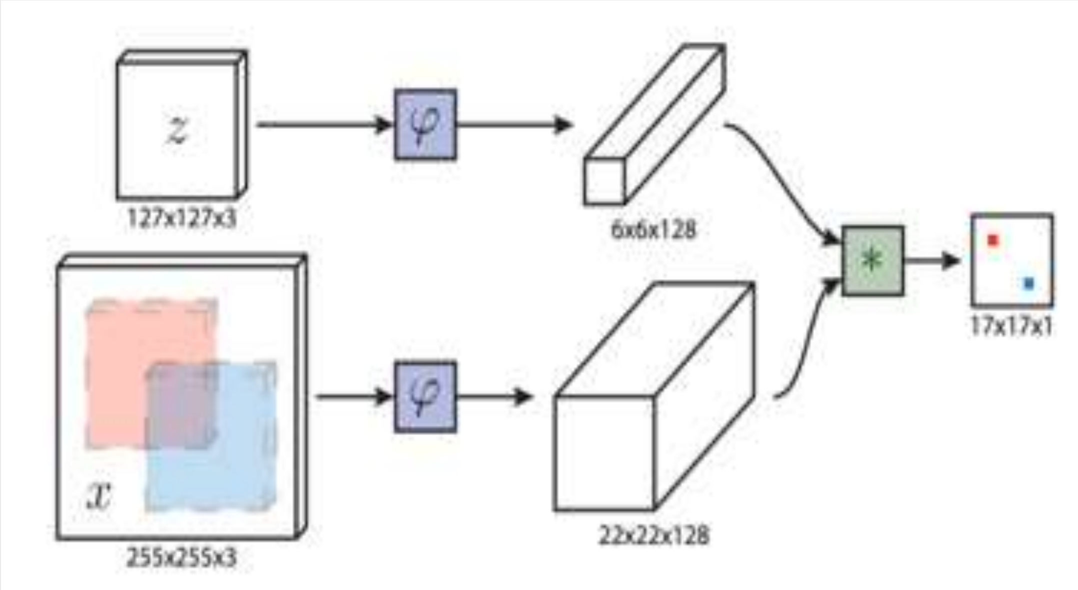
Fast online object tracking and segmentation: A unifying approach

- Visual object tracking
- Semi supervised object segmentation
- In real time

Wang, Qiang Zhang, Li Bertinetto, Luca Hu, Weiming Torr, Philip H.S.

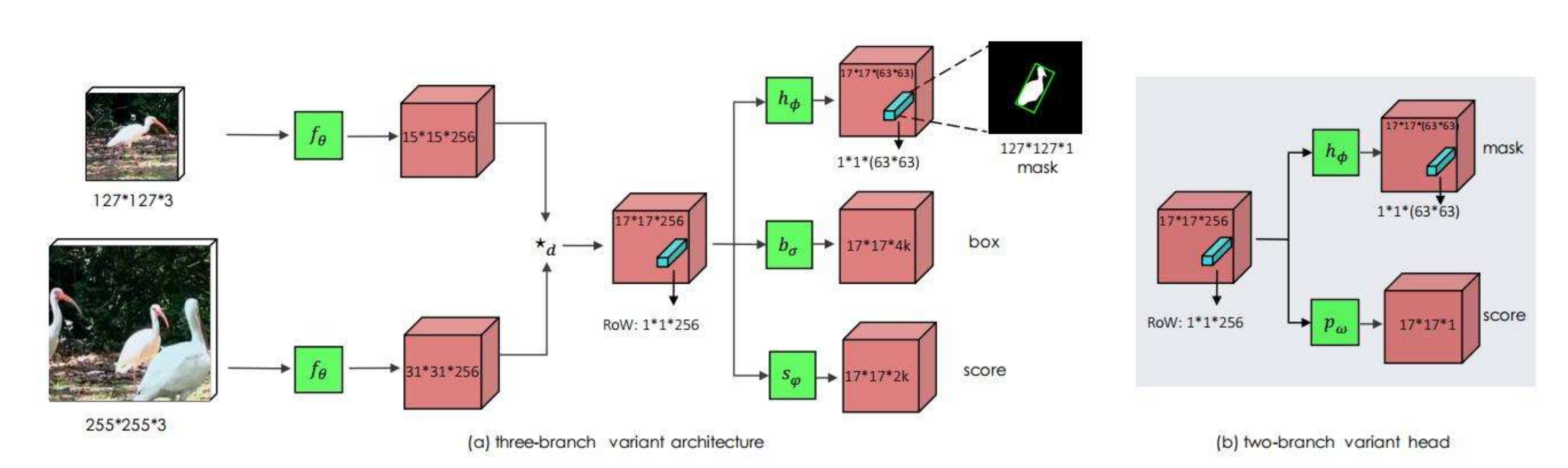
• Basic Idea

- **Fully-convolutional Siamese architecture**
- **Agnostic to the siamese fully convolutional method**
- **3 tasks :**
  1. Learn measure of similarity between the target object and multiple candidates in a sliding window
  2. Bounding box regression using a Region Proposal Network
  3. Class agnostic binary segmentation



Fully-convolutional Siamese architecture

Method – Network setup



- **2 variants**
- **SiamFC (2 branch)**

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

- **SiamRPN (3 branch)**

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

Evaluation

	SiamMask-*Opt*	SiamMask	SiamMask-2B	DaSiamRPN [71]	SiamRPN [31]	SA_Siam_R [17]	CSRDCF [37]	STRCF [32]
EAO ↑	**0.387**	**0.380**	0.334	0.326	0.244	0.337	0.263	0.345
Accuracy ↑	**0.642**	**0.609**	0.575	0.569	0.490	0.566	0.466	0.523
Robustness ↓	0.295	0.276	0.304	0.337	0.460	0.258	0.318	**0.215**
Speed (fps) ↑	5	55	60	160	**200**	32.4	48.9	2.9

Comparison with the state-of-the-art under EAO, Accuracy, and Robustness on the VOT-2018 benchmark.

	EAO ↑	A ↑	R ↓	EAO ↑	A ↑	R ↓	Speed ↑
SiamMask-box	0.363	0.584	0.300	0.412	0.623	0.233	**76**
SiamMask	0.380	0.609	**0.276**	0.433	0.639	**0.214**	55
SiamMask-*Opt*	**0.387**	**0.642**	0.295	**0.442**	**0.670**	0.233	5

Results on VOT-2018 benchmark (left) and VOT-2016 benchmark (right). Speed is measured in frames per second (fps).

	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 [60]	✔	✔	**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
MSK_b [45]	✔	✘	69.6	-	-	-	-	-	0.1
SFL [11]	✔	✔	76.1	90.6	12.1	76.0	85.5	10.4	0.1
FAVOS [10]	✘	✔	82.4	96.5	4.5	79.5	89.4	5.5	0.8
RGMP [63]	✘	✔	81.5	91.7	10.9	82.0	90.8	10.1	8
PML [9]	✘	✔	75.5	89.6	8.5	79.3	93.4	7.8	3.6
OSMN [66]	✘	✔	74.0	87.6	9.0	72.9	84.0	10.6	8.0
PLM [70]	✘	✔	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**	**55**

Results on DAVIS 2016 (validation set). FT and M respectively denote if the method requires fine-tuning and whether it is initialised with a mask (✔) or a bounding box (✘). Speed is measured in frames per second (fps).

Datasets :

Benchmarks for the evaluation of the object tracking task: VOT2016 & VOT-2018

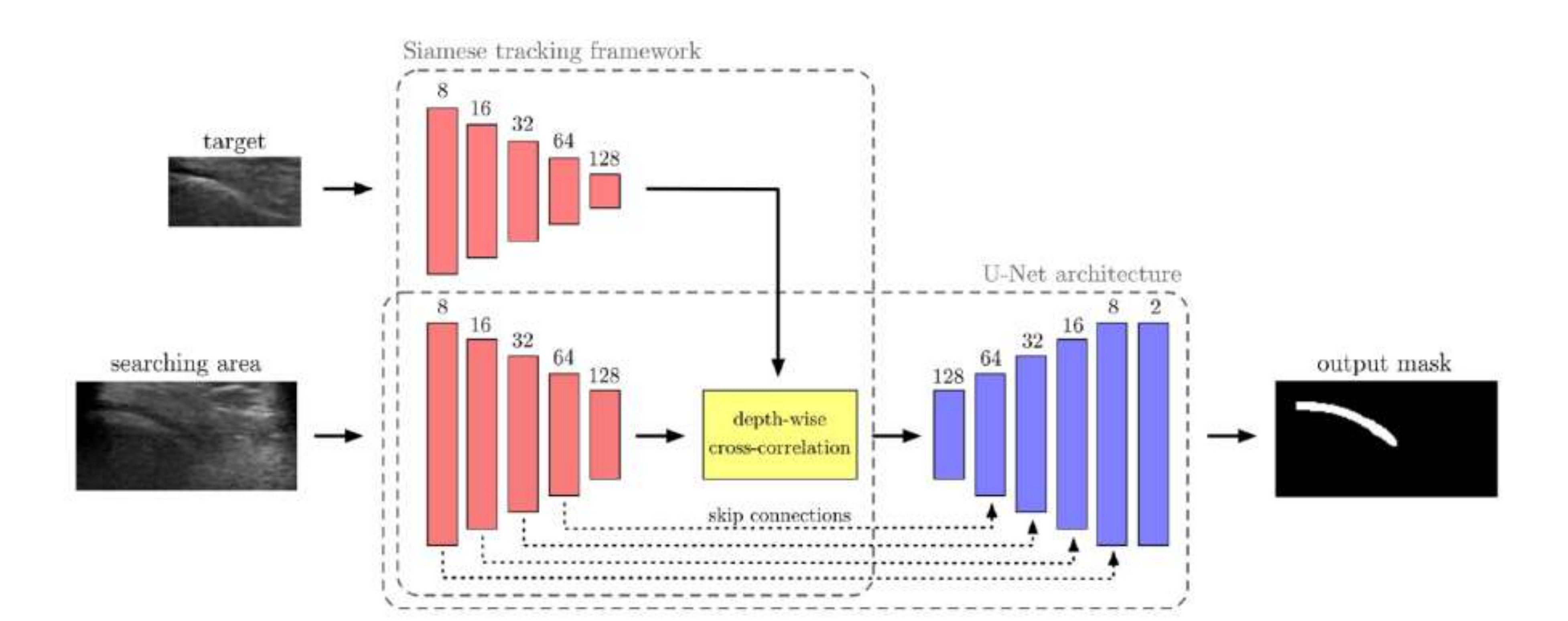
Benchmarks for Video Object Segmentation: DAVIS-2016, DAVIS-2017 and YouTube-VOS

• Results

Qualitative results of SiamMask for sequences belonging to both object tracking and video object segmentation benchmarks.

Failure cases: motion blur and "non-object" instance.

- **Siam-U-Net: encoder-decoder siamese network for knee cartilage tracking in ultrasound images**

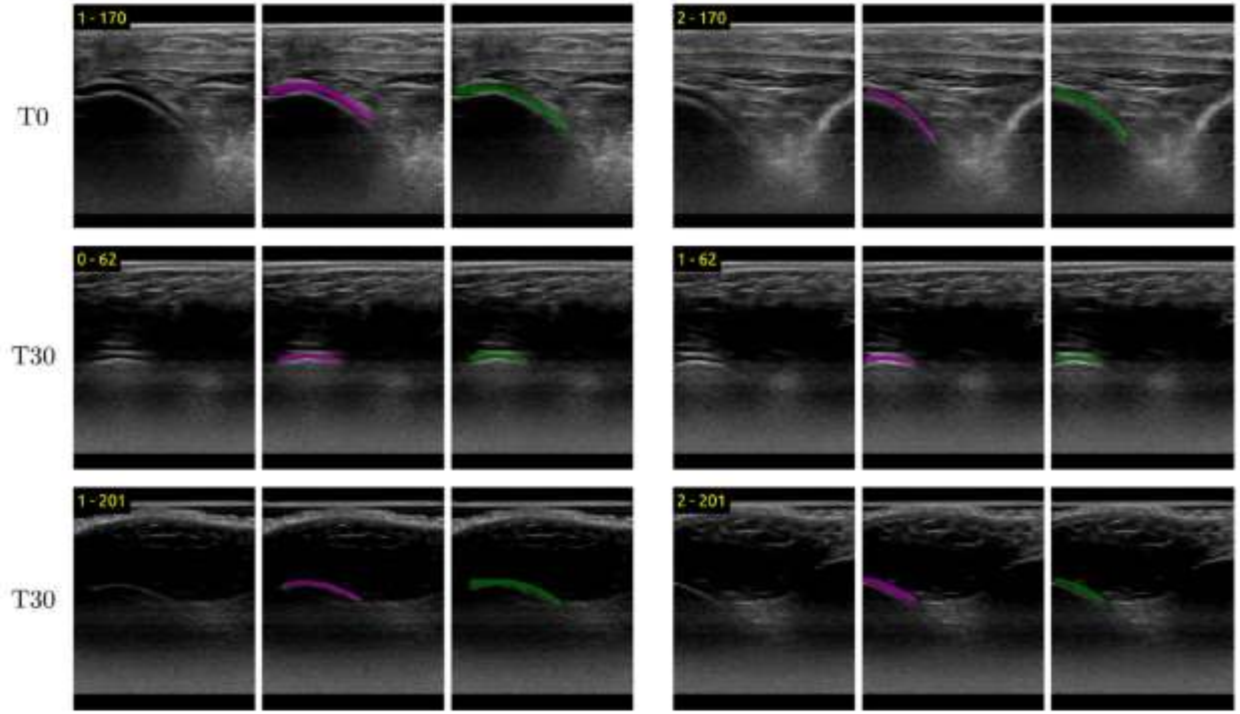


Matteo Dunnhofer *, Maria Antico , Fumio Sasazawa , Yu Takeda , Saskia Camps , Niki Martinel , Christian Micheloni , Gustavo Carneiro , Davide Fontanarosa

Evaluation

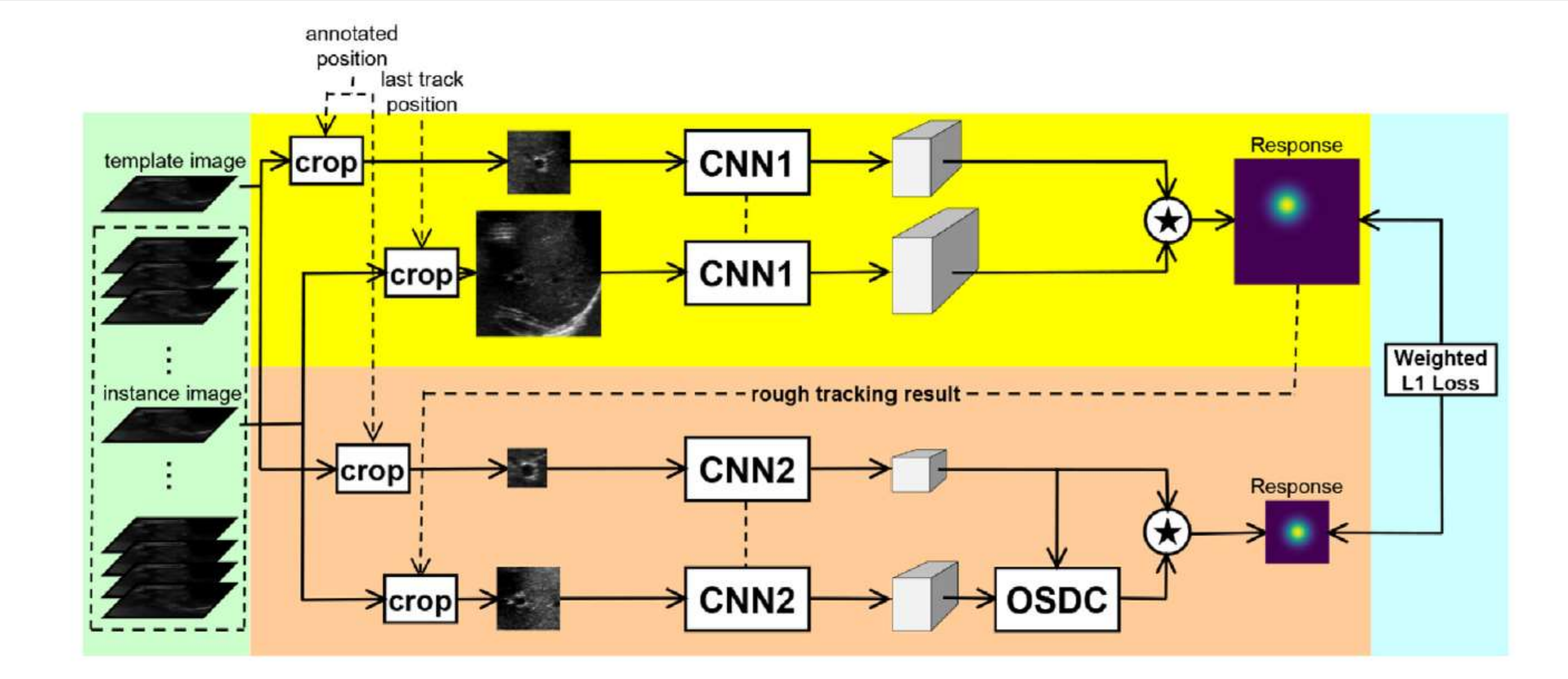
Results of Siam-U-Net obtained, on Evaluation 1, for the temporal (left column of results) and for the spatio-temporal (right column) tracking settings.

Split	Temporal tracking average DSC	Spatio-temporal tracking average DSC
1	0.74 ± 0.15	0.73 ± 0.16
2	0.69 ± 0.20	0.71 ± 0.16
3	0.69 ± 0.16	0.70 ± 0.15
4	0.69 ± 0.17	0.68 ± 0.18
5	0.73 ± 0.14	0.73 ± 0.14
6	0.69 ± 0.15	0.68 ± 0.16
Total	0.70 ± 0.16	0.71 ± 0.16



Matteo Dunnhofer *, Maria Antico , Fumio Sasazawa , Yu Takeda , Saskia Camps , Niki Martinel , Christian Micheloni , Gustavo Carneiro , Davide Fontanarosa

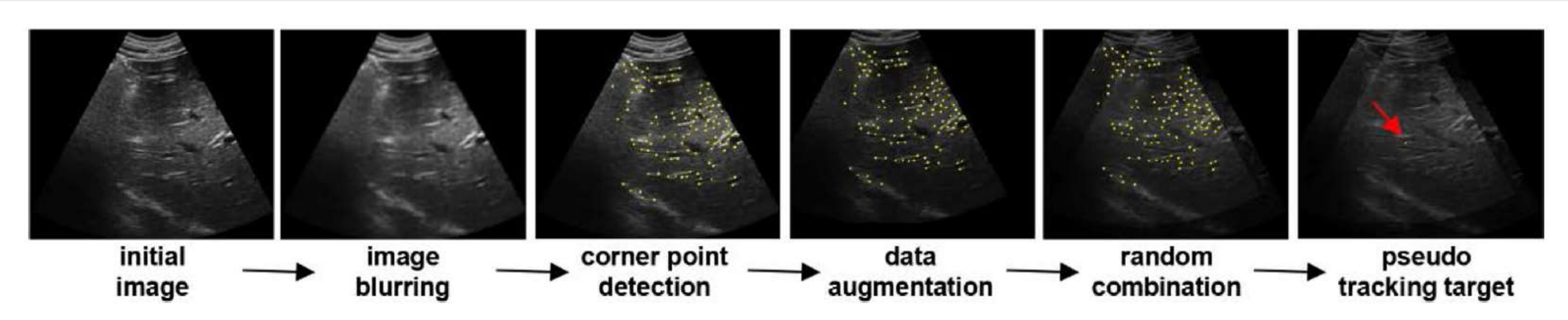
Cascaded one-shot deformable convolutional neural networks: Developing a deep learning model for respiratory motion estimation in ultrasound sequences



Fei Liu , Dan Liu , Jie Tian , Xiaoyan Xie , Xin Yang , Kun Wang

Method – Network setup

- Procedure of the unsupervised training strategy:
- For an image randomly selected from an ultrasound sequence
- blurring is firstly applied to suppress noise for corner point detection
- Data augmentation and combination with an image from a random timestamp simulate the perturbation of respiratory motion
- one point is selected randomly as pseudo tracking target for network training
- The initial image is regarded as template image and the processed image as instance image.



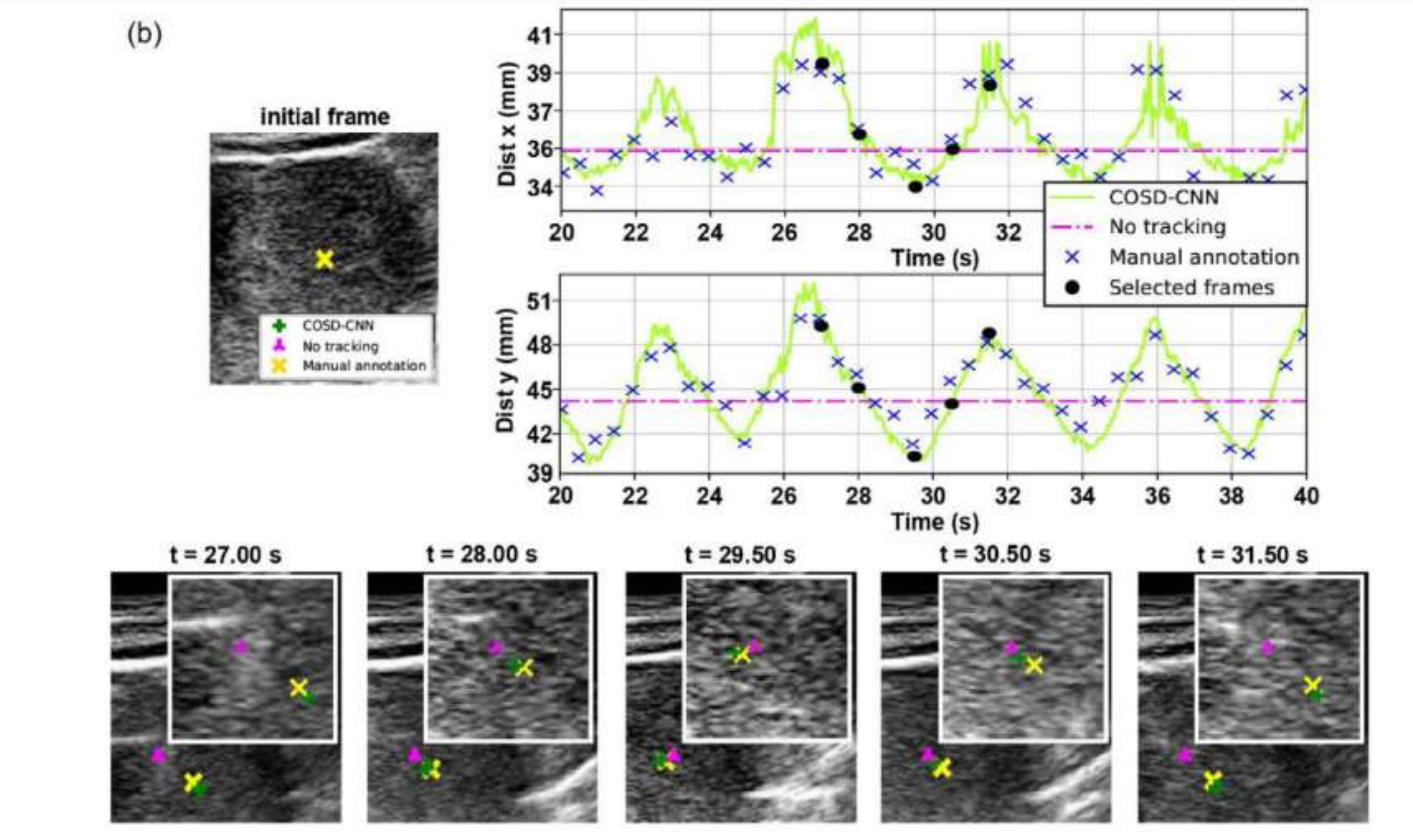
Fei Liu , Dan Liu , Jie Tian , Xiaoyan Xie , Xin Yang , Kun Wang

Evaluation

Tracking performance comparisons for COSD-CNN with other state-of-the-art methods and human observers on FSYSU and CLUST 2D test set. Best performance among the automatic tracking methods are highlighted in **bold**.

CLUST set	Mean (mm)	Std (mm)	TE95th (mm)
No tracking	6.45	5.11	16.48
Nouri and Rothberg (2015)	3.35	5.21	14.19
Kondo (2015)	2.91	10.52	5.18
Makhinya and Goksel (2015)	1.44	2.80	3.62
Gomariz et al. (2019)(baseline)	1.34	2.57	2.95
Hallack et al. (2015)	1.21	3.17	2.82
Jeungyoon et al. (2019)	0.85	0.80	2.32
Williamson et al. (2018b)	0.74	1.03	1.85
Shepard et al. (2017)	0.72	1.25	1.71
COSD-CNN	**0.69**	**0.67**	**1.57**
Observer 1	0.46	0.36	1.13
Observer 2	0.47	0.34	1.08
Observer 3	0.44	0.32	1.03
FSYSU set	Mean (mm)	Std (mm)	TE95th (mm)
No tracking	5.77	2.31	9.67
COSD-CNN	**1.52**	**0.77**	**3.10**
Observer 1	1.05	0.44	1.57
Observer 2	1.19	0.41	1.66

Results



Fei Liu , Dan Liu , Jie Tian , Xiaoyan Xie , Xin Yang , Kun Wang

Conclusion

- ✓ Fast evolving field
- ✓ Tracking in US images is tricky
- ✓ Need for data
- ✓ Test Yolo and Siamese NN

Bibliography

- Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. *ArXiv*. *https://arxiv.org/abs/2004.10934*

- Bertinetto, L., Valmadre, J., Henriques, J. F., Vedaldi, A., & Torr, P. H. S. (2016). Fully-convolutional siamese networks for object tracking. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9914 LNCS*(April 2019), 850–865. https://doi.org/10.1007/978-3-319-48881-3_56

- Dunnhofer, M., Antico, M., Sasazawa, F., Takeda, Y., Camps, S., Martinel, N., Micheloni, C., Carneiro, G., & Fontanarosa, D. (2020). Siam-U-Net: encoder-decoder siamese network for knee cartilage tracking in ultrasound images. *Medical Image Analysis*, *60*, 101631. https://doi.org/10.1016/j.media.2019.101631

- Liu, F., Liu, D., Tian, J., Xie, X., Yang, X., & Wang, K. (2020). Cascaded one-shot deformable convolutional neural networks: Developing a deep learning model for respiratory motion estimation in ultrasound sequences. *Medical Image Analysis*, *65*. https://doi.org/10.1016/j.media.2020.101793

- Wojke, N., Bewley, A., & Paulus, D. (2018). Simple online and realtime tracking with a deep association metric. *Proceedings - International Conference on Image Processing, ICIP*, *2017-Septe*, 3645–3649. https://doi.org/10.1109/ICIP.2017.8296962

- Wang, Q., Zhang, L., Bertinetto, L., Hu, W., & Torr, P. H. S. (2019). Fast online object tracking and segmentation: A unifying approach. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2019-June*, 1328–1338. https://doi.org/10.1109/CVPR.2019.00142