VISMAC 23

International Summer School on
Machine Vision

Deep Learning for visual object tracking

Prof. Christian Micheloni

University of Udine

Machine Learning and Perception (MLP) Lab.

Collaboration with Dr. Matteo Dunnhofer - University of Udine

Machine Learning and Perception (MLP) Lab.

Outline

1. Problem Definition
    1. Applications
    2. Definitions
    3. Challenges and Settings

2. Traditional Methods

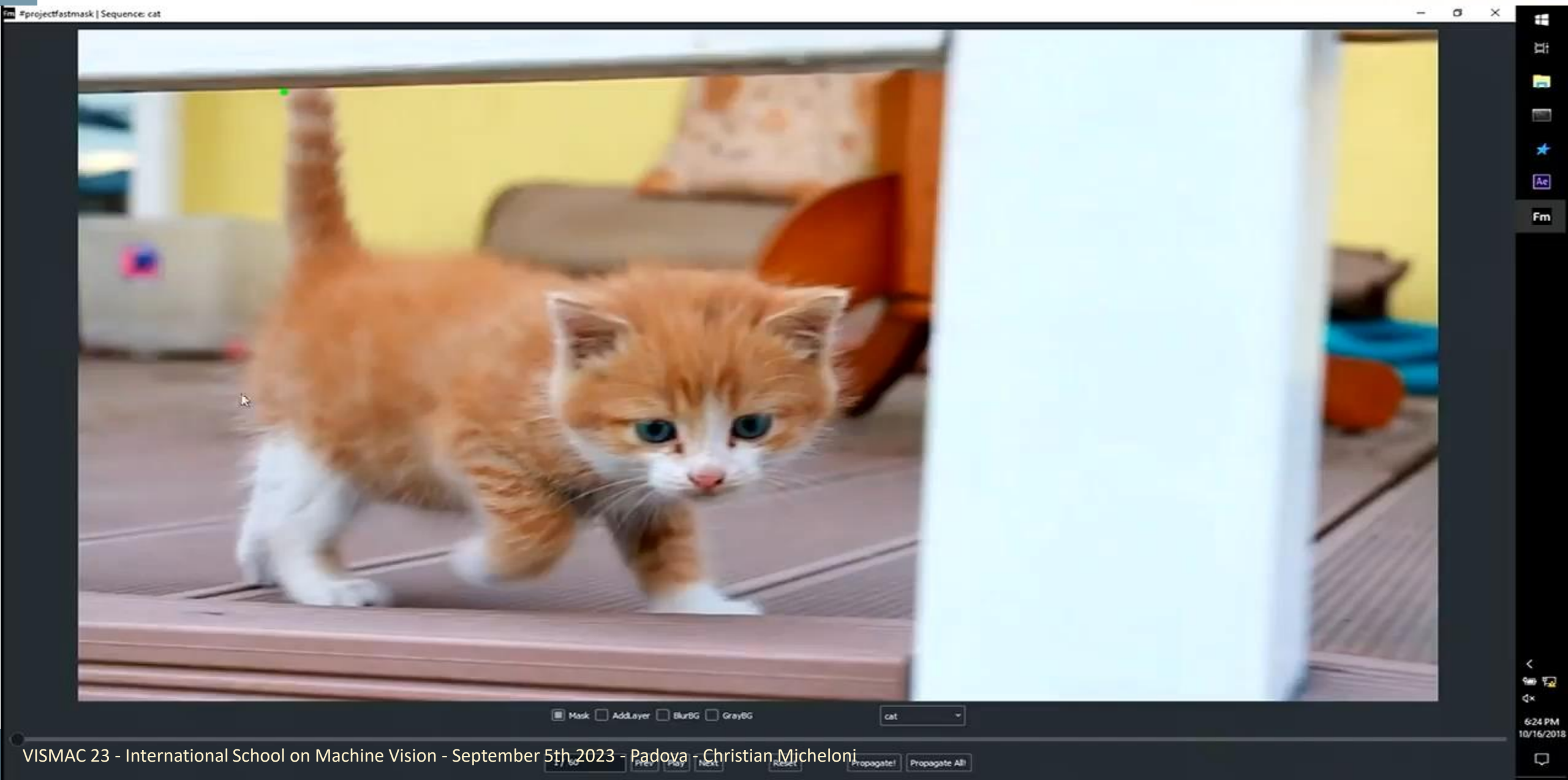
3. Deep-Learning Trackers
    1. Hybrid Methods
    2. Offline vs Online-Offline Learning
    3. Offline Trackers
    4. Online Trackers

Problem Definition

Visual Object Tracking

Applications

Applications



Applications

Applications

Definitions

"Tracking is the task of assigning consistent labels to the tracked objects in different frames of a video"

Object Tracking: A Survey, Yilmaz et al., 2006

"Online tracking is following the location of one target in a video starting from a selected region of interest in the first frame"

Visual Tracking: An Experimental Survey, Cucchiara et al., 2014

"The estimation of a time series representing the states of a target object in the consecutive frames"

Video Tracking: Theory and Applications, Maggio and Cavallaro, 2011

Definition

The **initial state** of a target object **is given**

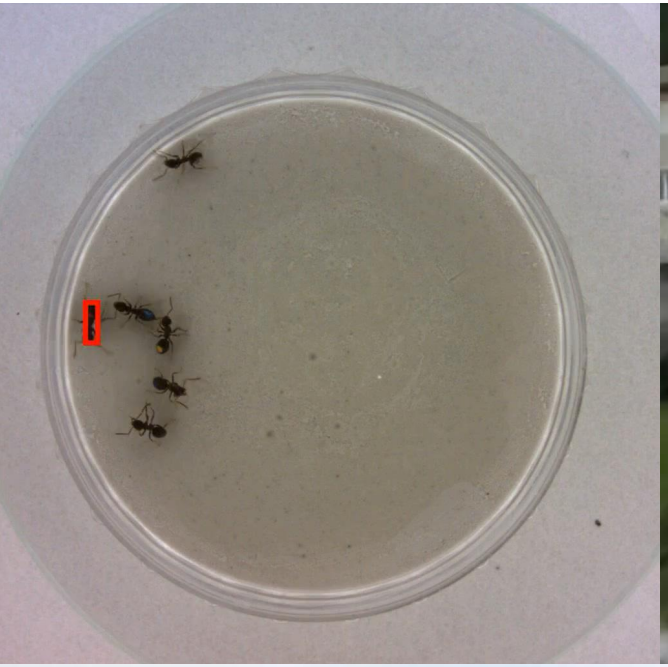
Deliver the new states of the target
in all the other frames of a video

Initialization

Tracking


1.00
CAVS PLAYOFF BASKETBALL PRESENTED BY DISCOUNT DRUG MART
DEFEND
CLE 0 14.5 0 IND
UBER
KeyBank

Generic Object Tracking



Target Representation - Barycenter

Target Representation – Bounding Box

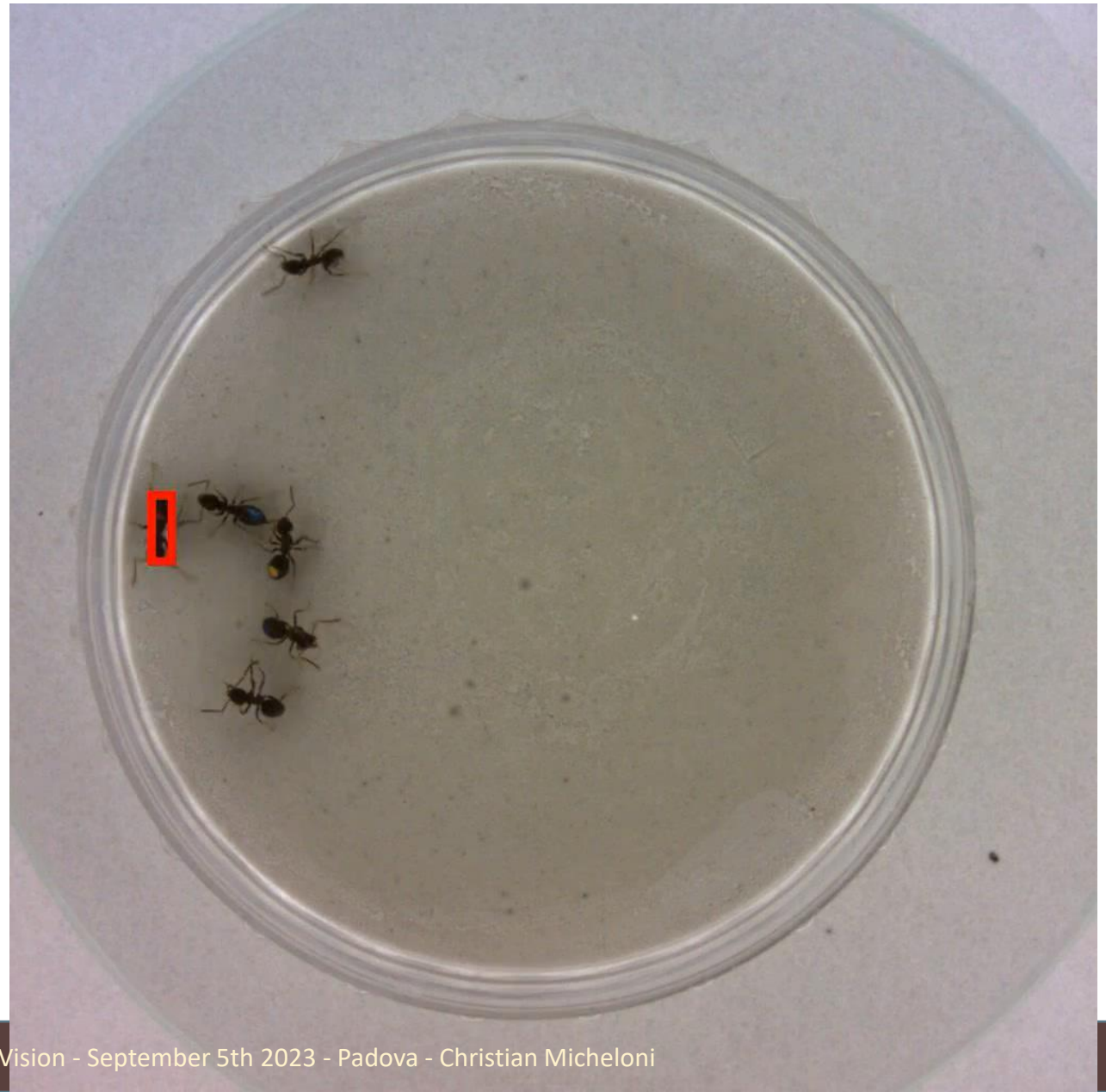
Representation most used at the state-of-the-art

Target Representation – Segmentation

Challenges - Rotations

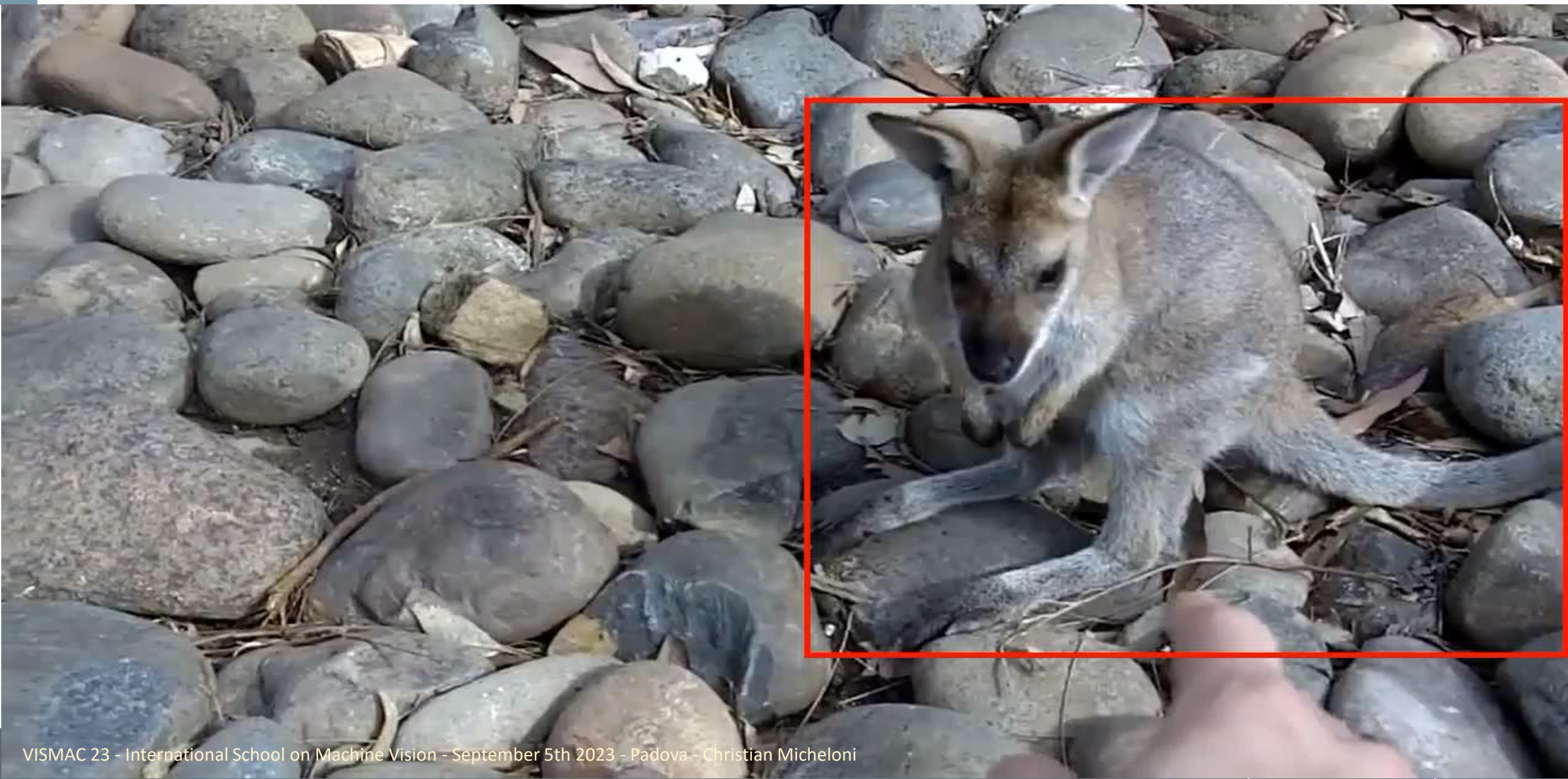
Challenges - Pose Variations

Challenges - Fast Motion



Challenges - Illumination Variations

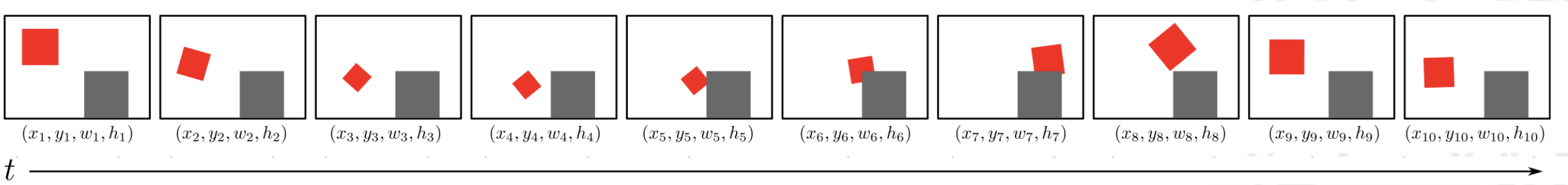
Challenges - Background Clutter



Challenges - Similar Objects

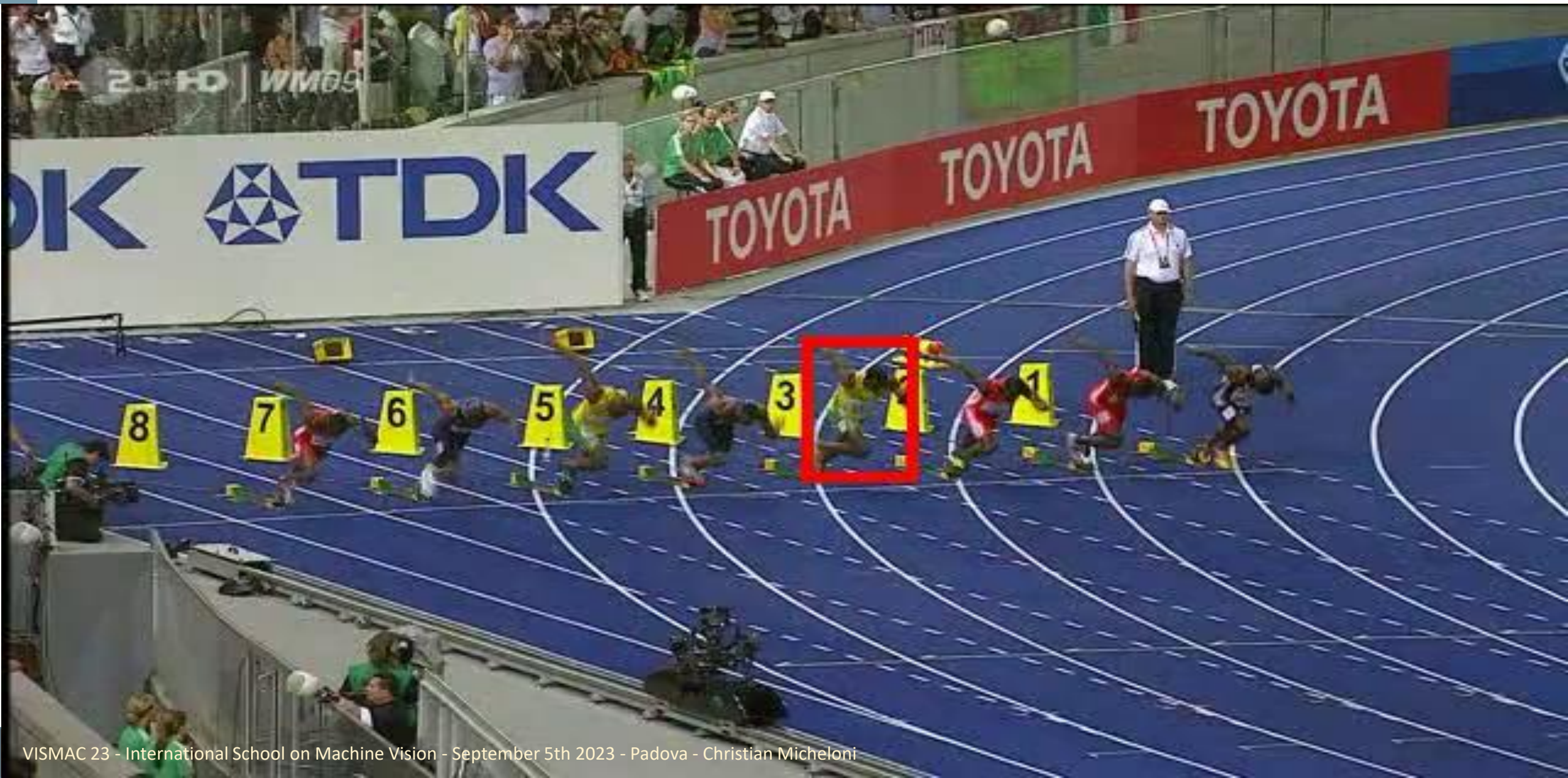
Challenges - Occlusions

Short-term Tracking

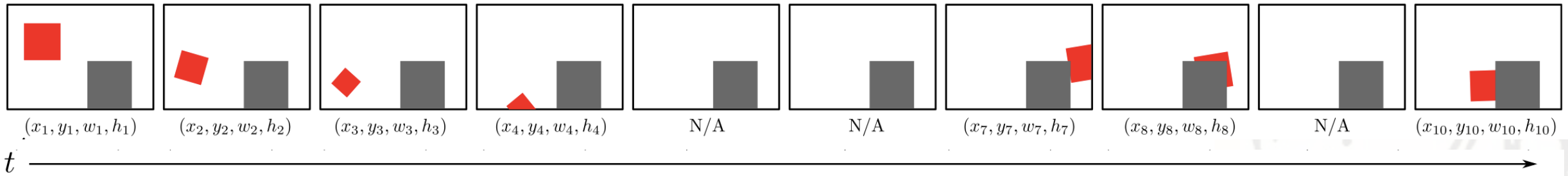


"Performance Evaluation Methodology for Long-Term Visual Object Tracking", Lukezič et al., TCyb, 2020

Short-term Tracking



Long-term Tracking

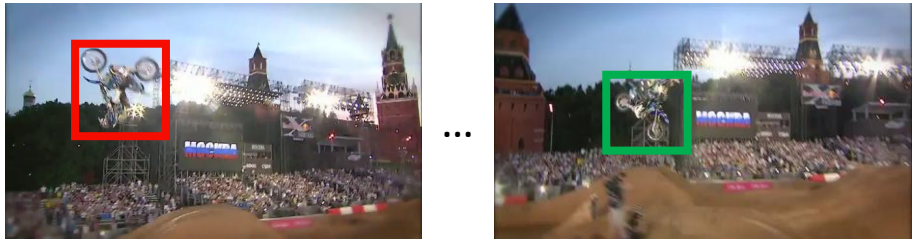


"Performance Evaluation Methodology for Long-Term Visual Object Tracking", Lukezič et al., TCyb, 2020

Long-term Tracking

Online Tracking

$t = 0$ t



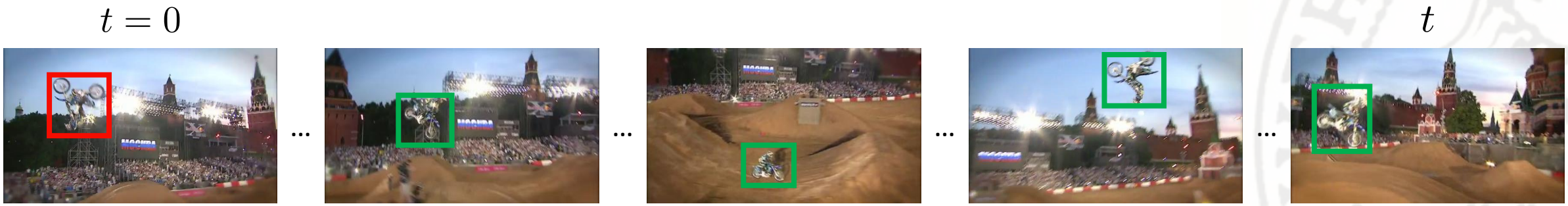
...

Online Tracking

$t = 0$ t

Online Tracking

Online Tracking



Setting used in real-time applications

Most studied in the literature

Offline Tracking

$t = 0$

Suitable for post analysis

Better handling of occlusions

Evaluation

Protocol

One-Pass Evaluation (OPE)

$t = 0$ $0 < t < T - 1$ $t = T - 1$

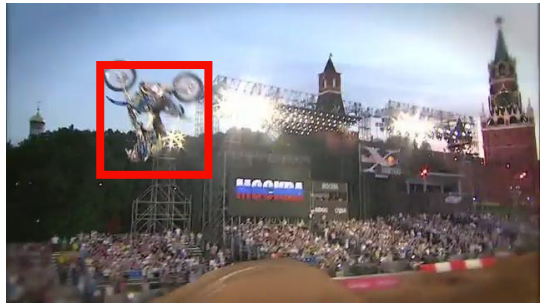
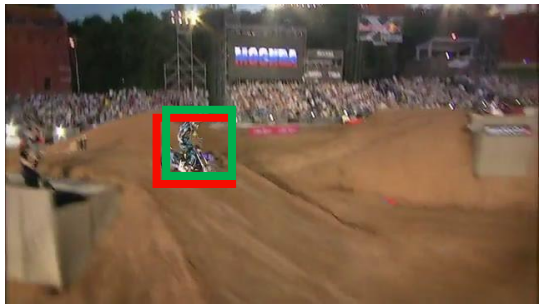
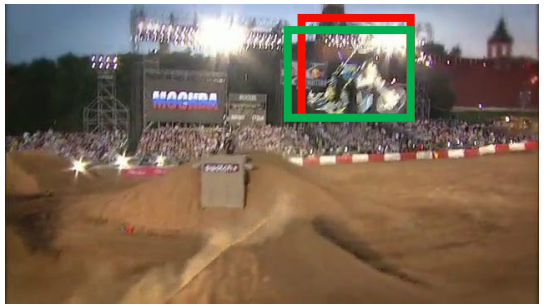
```
tracker.init(frame_0, init_state)
```

```
box_t = tracker.update(frame_t)   box_t = tracker.update(frame_t)   box_T1 = tracker.update(frame_T1)
```

"Object Tracking Benchmark", Wu et al., TPAMI 2015

Metrics

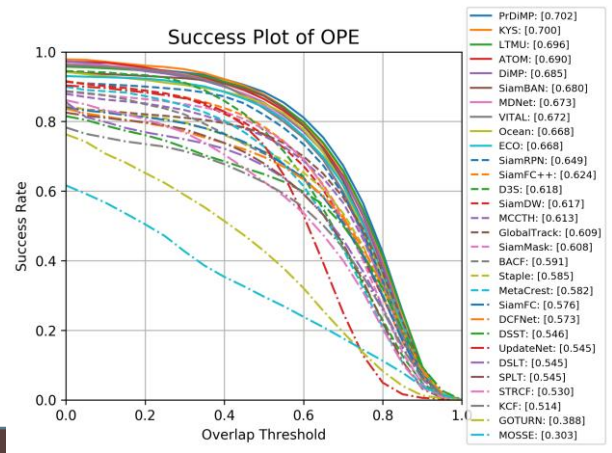
$t = 0$ $\quad 0 < t < T - 1$ $\quad t = T - 1$


$\text{IoU}(\square, \square)$ $\quad \text{IoU}(\square, \square)$ $\quad \text{IoU}(\square, \square)$

Average Overlap
Accuracy
Success Score
AUC

$$AO = \frac{1}{T} \sum_{t=0}^{T-1} \text{IoU}(\square_t, \square_t)$$



Evaluation Benchmarks

Visual Tracker Benchmark

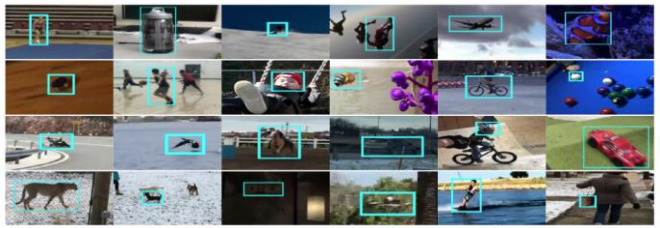
http://www.visual-tracking.net

OTB-50/100
100 videos
59K frames
10 object categories

`http://cvlab.hanyang.ac.kr/tracker_benchmark/`

VOT Challenge
> 200 videos
RGB short-term box/segmentation
RGB real-time short-term box/segmentation
RGB long-term box
RGBD short-term box

`https://www.votchallenge.net`

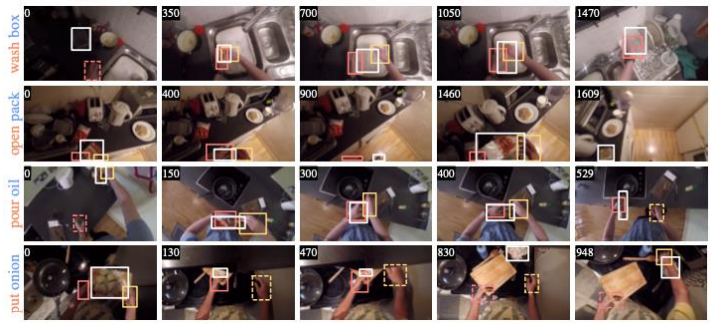


Need for Speed
100 videos
9 object categories
240 FPS videos

`http://ci2cv.net/nfs/index.html`

UAV123
123 videos
9 object categories
Object tracking from an UAV point-of-view

`https://cemse.kaust.edu.sa/ivul/uav123`



TREK-150
150 videos
34 object categories
Object tracking in FPV

`https://machinelearning.uniud.it/datasets/trek150/`

Traditional Methods

Template Matching

Similarity between the target patch and the frame

$$t = 0$$

"A Two-Stage Cross Correlation Approach to Template Matching", Goshtasby et al., TPAMI, 1984
"Fast Template Matching", Lewis, Vision Interface, 1995

Template Matching

Similarity between the target patch and the frame

$$t > 0$$

"A Two-Stage Cross Correlation Approach to Template Matching", Goshtasby et al., TPAMI, 1984
"Fast Template Matching", Lewis, Vision Interface, 1995

Template Matching

Similarity between the target patch and the frame

$t > 0$

$\approx$

$$R(x,y) = \sum_{x',y'} (T(x',y') - I(x+x',y+y'))^2$$

Sum of Squared Differences

$$R(x,y) = \sum_{x',y'} (T(x',y') \cdot I(x+x',y+y'))$$

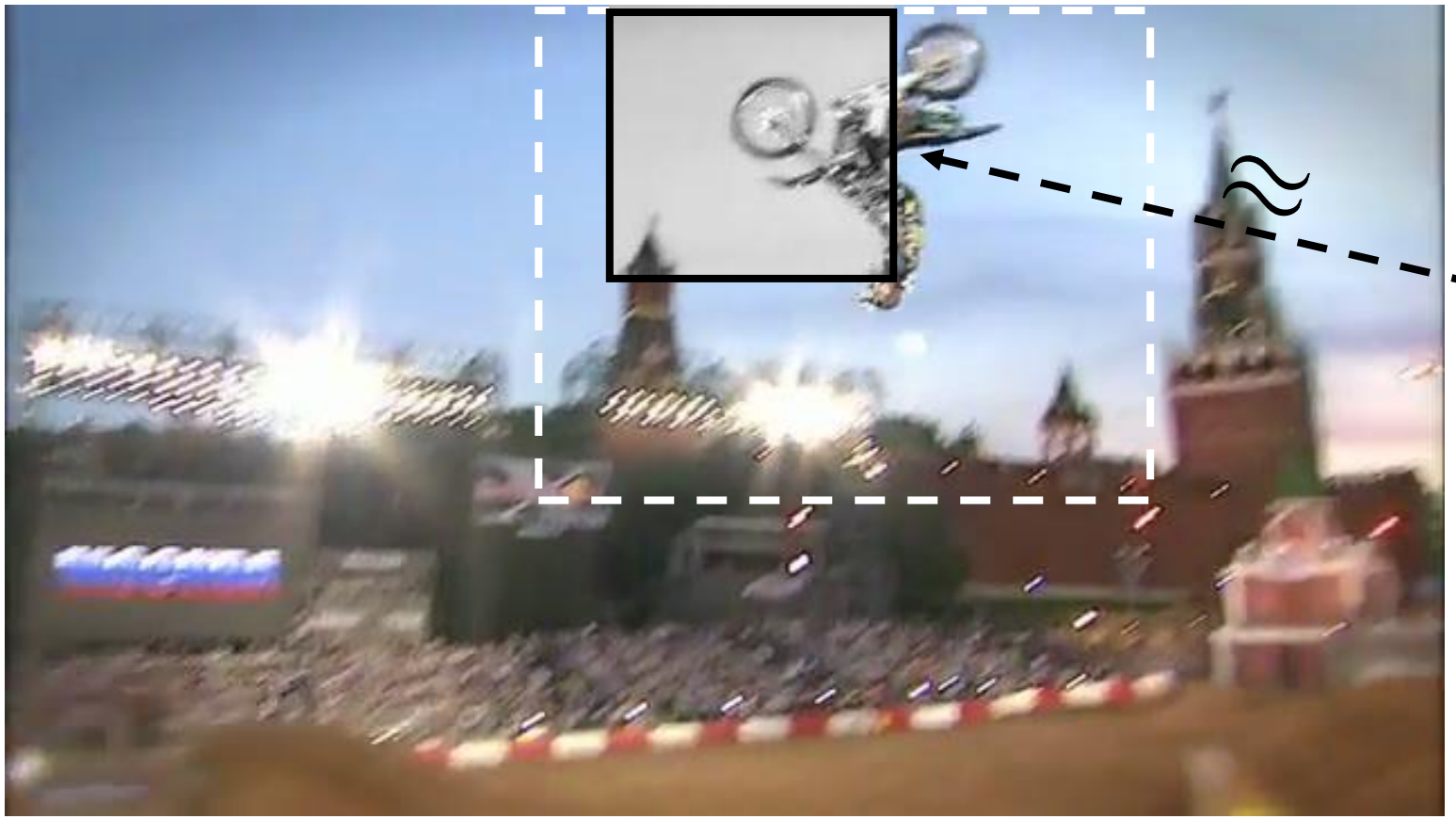
Correlation

"A Two-Stage Cross Correlation Approach to Template Matching", Goshtasby et al., TPAMI, 1984
"Fast Template Matching", Lewis, Vision Interface, 1995

Template Matching

Similarity between the target patch and the frame

$t > 0$



$\approx$

$$R(x,y) = \sum_{x',y'} (T(x',y') - I(x+x',y+y'))^2$$

Sum of Squared Differences

$$R(x,y) = \sum_{x',y'} (T(x',y') \cdot I(x+x',y+y'))$$

Correlation

"A Two-Stage Cross Correlation Approach to Template Matching", Goshtasby et al., TPAMI, 1984
"Fast Template Matching", Lewis, Vision Interface, 1995

Template Matching

Similarity between the target patch and the frame

$t > 0$

$\approx$

$$R(x,y) = \sum_{x',y'} (T(x',y') - I(x+x',y+y'))^2$$

Sum of Squared Differences

$$R(x,y) = \sum_{x',y'} (T(x',y') \cdot I(x+x',y+y'))$$

Correlation

"A Two-Stage Cross Correlation Approach to Template Matching", Goshtasby et al., TPAMI, 1984
"Fast Template Matching", Lewis, Vision Interface, 1995

Template Matching

Similarity between the target patch and the frame

$t > 0$

$\approx$

$$R(x,y) = \sum_{x',y'} (T(x',y') - I(x+x',y+y'))^2$$

Sum of Squared Differences

$$R(x,y) = \sum_{x',y'} (T(x',y') \cdot I(x+x',y+y'))$$

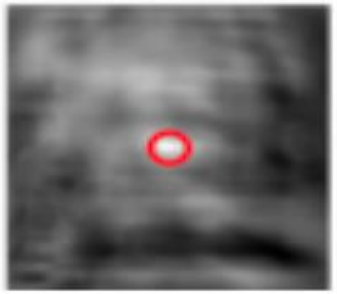
Correlation

"A Two-Stage Cross Correlation Approach to Template Matching", Goshtasby et al., TPAMI, 1984
"Fast Template Matching", Lewis, Vision Interface, 1995

Template Matching

Similarity between the target patch and the frame

$t > 0$



"A Two-Stage Cross Correlation Approach to Template Matching", Goshtasby et al., TPAMI, 1984
"Fast Template Matching", Lewis, Vision Interface, 1995

Tracking-by-Detection

Discriminate between the target and the surrounding background

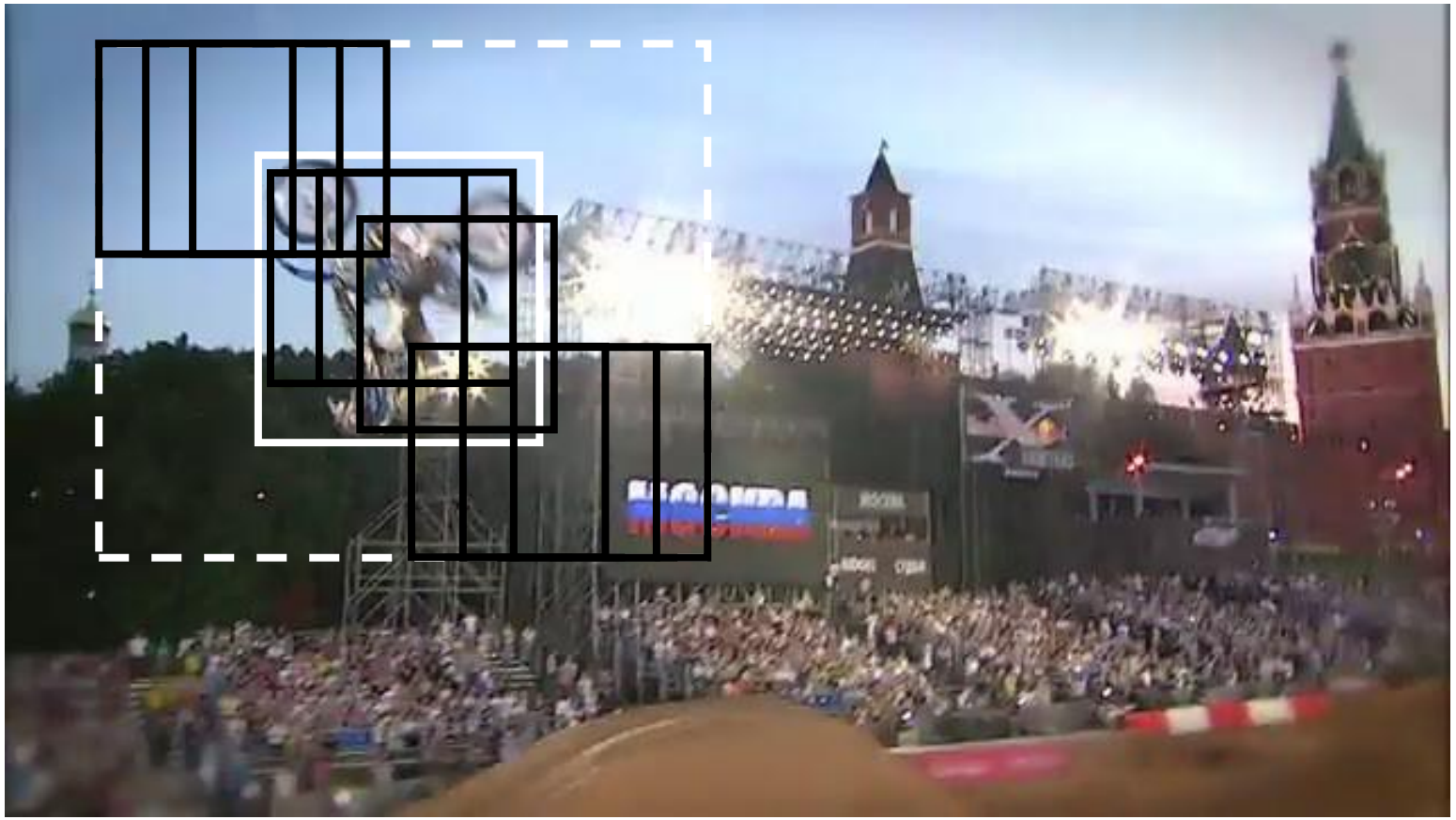
$t = 0$

"Real-time Tracking via Online Boosting", Grabner et al., BMVC 2006
"Robust Object Tracking with Online Multiple Instance Learning", Babenko et al., TPAMI 2011

Tracking-by-Detection

Discriminate between the target and the surrounding background

$$t = 0$$

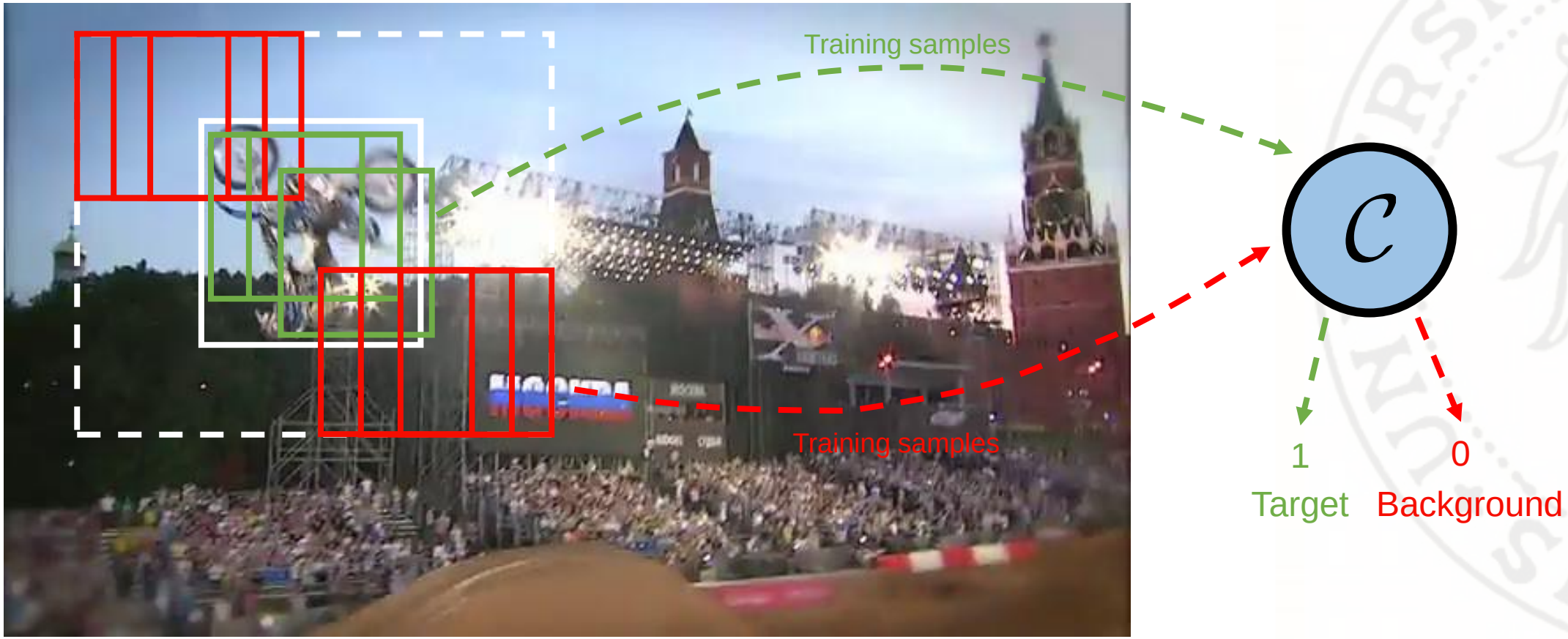


"Real-time Tracking via Online Boosting", Grabner et al., BMVC 2006
"Robust Object Tracking with Online Multiple Instance Learning", Babenko et al., TPAMI 2011

Tracking-by-Detection

Discriminate between the target and the surrounding background

$t = 0$

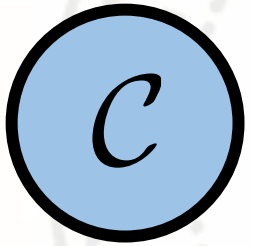


"Real-time Tracking via Online Boosting", Grabner et al., BMVC 2006
"Robust Object Tracking with Online Multiple Instance Learning", Babenko et al., TPAMI 2011

Tracking-by-Detection

Discriminate between the target and the surrounding background

$$t > 0$$

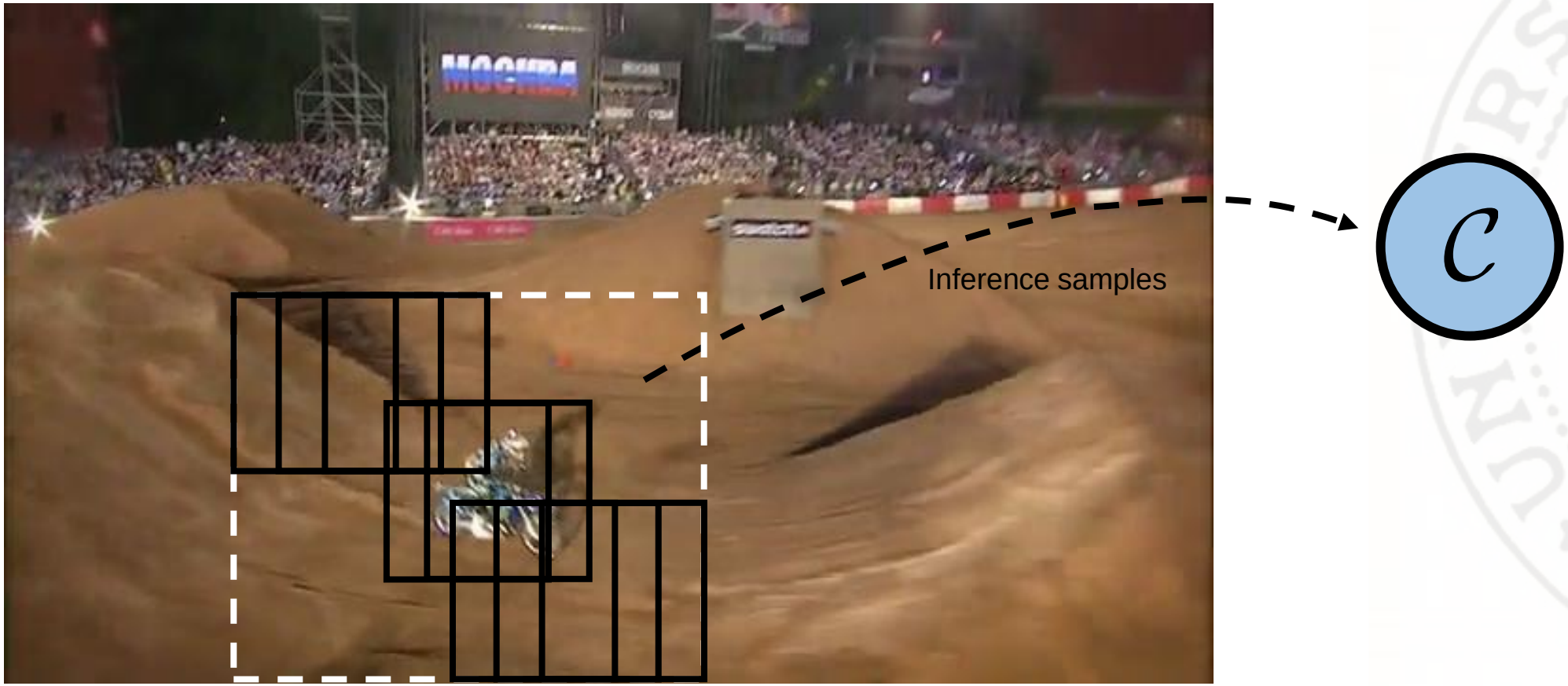


"Real-time Tracking via Online Boosting", Grabner et al., BMVC 2006
"Robust Object Tracking with Online Multiple Instance Learning", Babenko et al., TPAMI 2011

Tracking-by-Detection

Discriminate between the target and the surrounding background

$t > 0$



"Real-time Tracking via Online Boosting", Grabner et al., BMVC 2006
"Robust Object Tracking with Online Multiple Instance Learning", Babenko et al., TPAMI 2011

Tracking-by-Detection

Discriminate between the target and the surrounding background

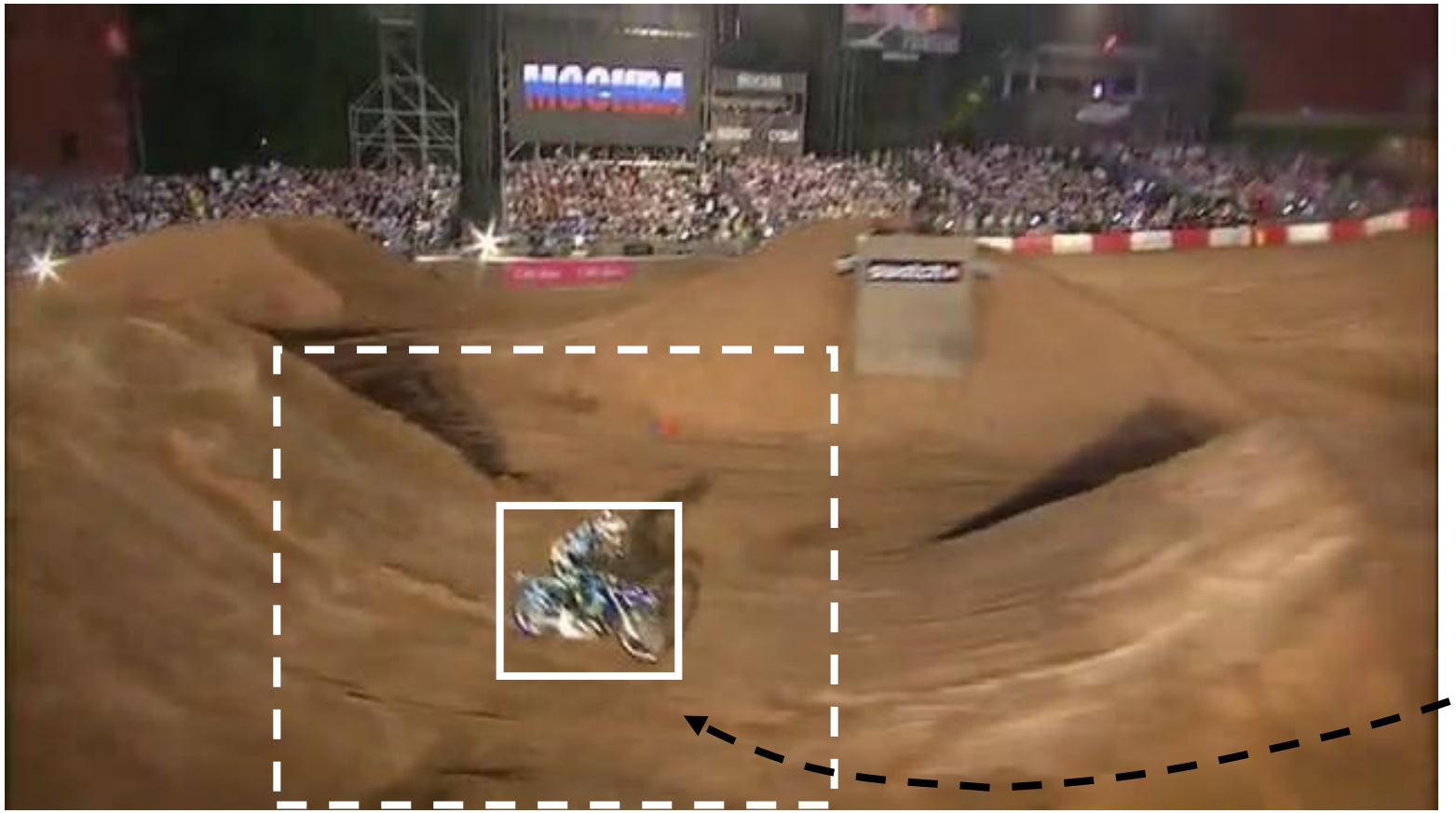
$$t > 0$$

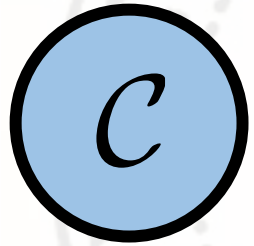
"Real-time Tracking via Online Boosting", Grabner et al., BMVC 2006
"Robust Object Tracking with Online Multiple Instance Learning", Babenko et al., TPAMI 2011

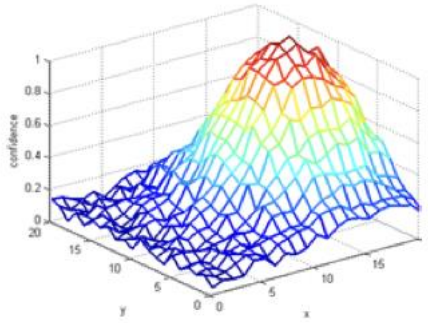
Tracking-by-Detection

Discriminate between the target and the surrounding background

$$t > 0$$





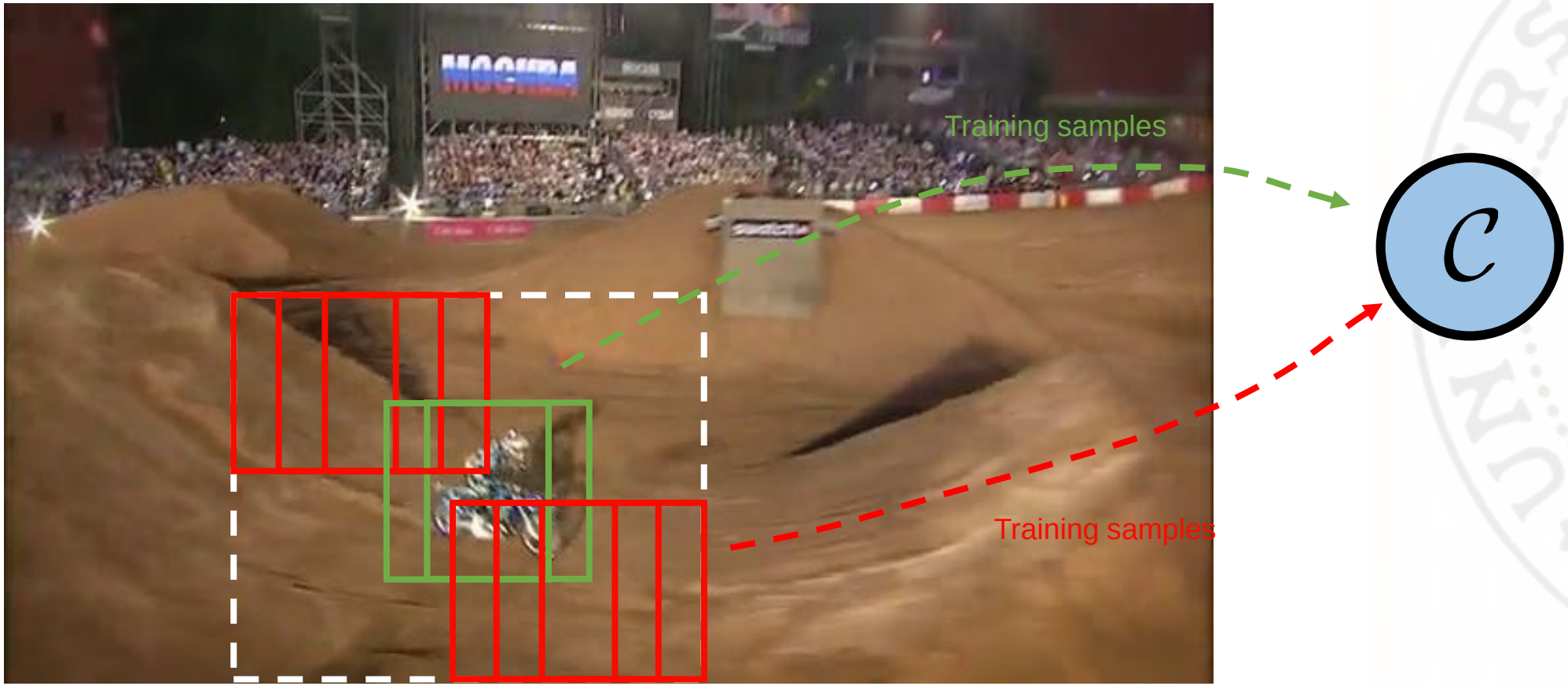


"Real-time Tracking via Online Boosting", Grabner et al., BMVC 2006
"Robust Object Tracking with Online Multiple Instance Learning", Babenko et al., TPAMI 2011

Tracking-by-Detection

Discriminate between the target and the surrounding background

$$t > 0$$



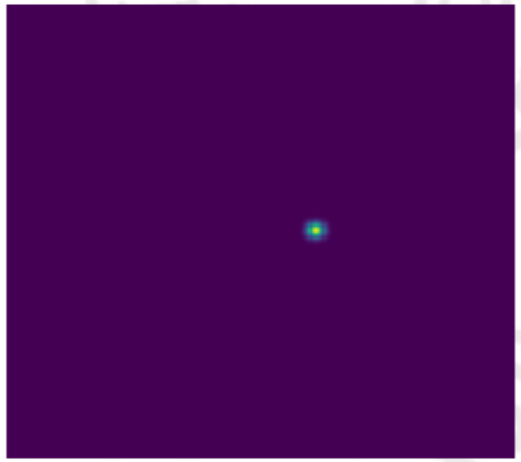
"Real-time Tracking via Online Boosting", Grabner et al., BMVC 2006
"Robust Object Tracking with Online Multiple Instance Learning", Babenko et al., TPAMI 2011

Discriminative Correlation Filters

Filtering frames with an online-learned target filter

$t > 0$

F



"Visual Object Tracking using Adaptive Correlation Filters", Bolme et al., CVPR 2010
"High-Speed Tracking with Kernelized Correlation Filters", Henriques et al., TPAMI 2015

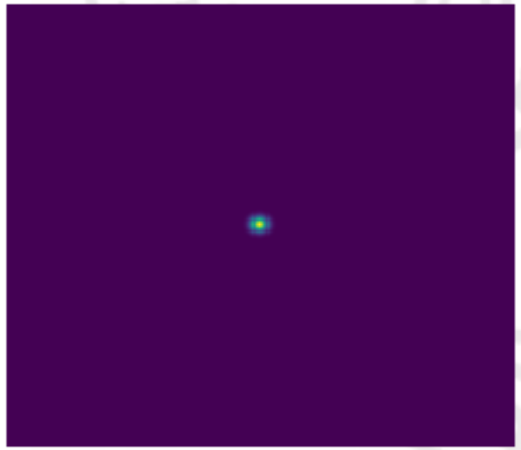
Discriminative Correlation Filters

Filtering frames with an online-learned target filter

$t = 0$

$$\arg\min_{\mathbf{F}} ||\mathbf{T} \otimes \mathbf{F} - \mathbf{G}||^2$$

$\mathbf{F}$



$\mathbf{G}$

"Visual Object Tracking using Adaptive Correlation Filters", Bolme et al., CVPR 2010
"High-Speed Tracking with Kernelized Correlation Filters", Henriques et al., TPAMI 2015

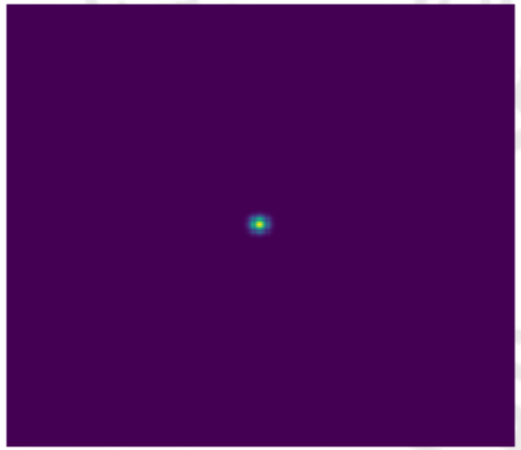
Discriminative Correlation Filters

Filtering frames with an online-learned target filter

$t = 0$

$$\arg\min_{\mathbf{F}} ||\mathbf{T} \otimes \mathbf{F} - \mathbf{G}||^2$$

$\mathbf{F}$



$\mathbf{G}$

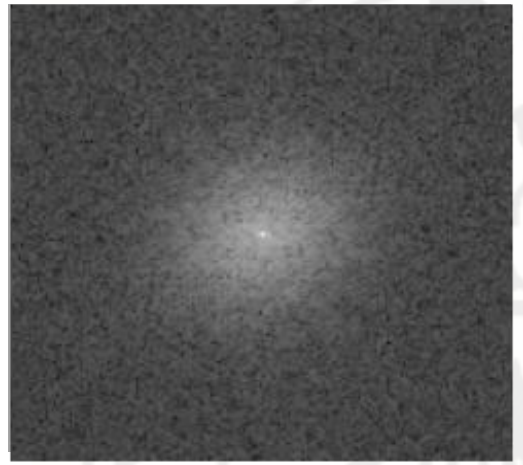
"Visual Object Tracking using Adaptive Correlation Filters", Bolme et al., CVPR 2010
"High-Speed Tracking with Kernelized Correlation Filters", Henriques et al., TPAMI 2015

Discriminative Correlation Filters

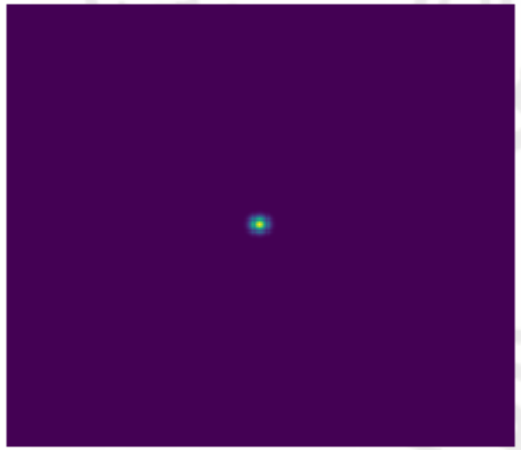
Filtering frames with an online-learned target filter

$t = 0$

$$\arg\min_{\mathbf{F}} ||\mathbf{T} \otimes \mathbf{F} - \mathbf{G}||^2$$



$\mathbf{F}$



$\mathbf{G}$

"Visual Object Tracking using Adaptive Correlation Filters", Bolme et al., CVPR 2010
"High-Speed Tracking with Kernelized Correlation Filters", Henriques et al., TPAMI 2015

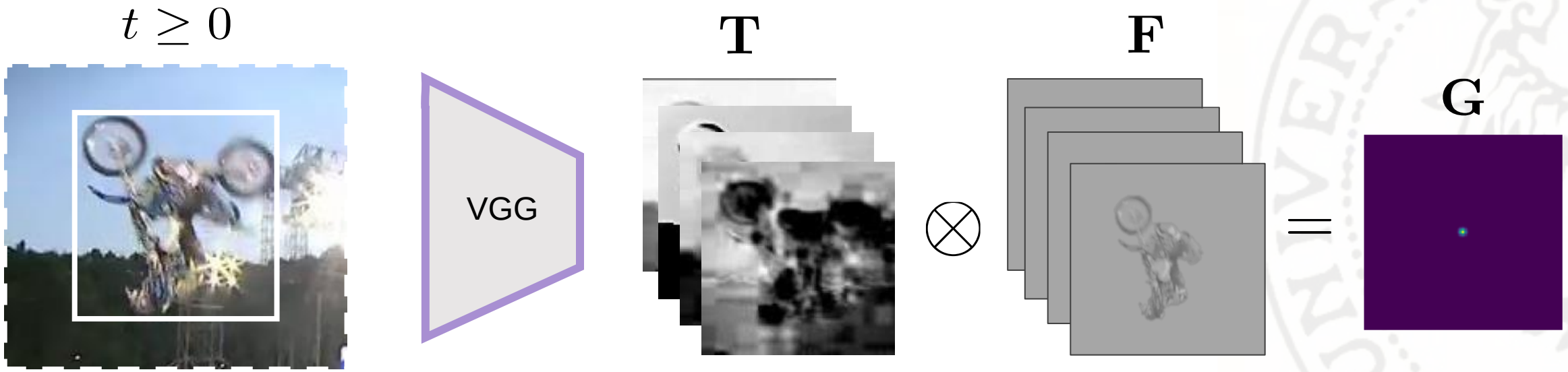
Discriminative Correlation Filters

Deep Learning Methods

Hybrid Methods

DeepSRDCF

Exploit CNN features in the DCF domain



"Convolutional Features for Correlation Filter Based Visual Tracking", Danelljan et al., ICCVW 2015

DeepSRDCF

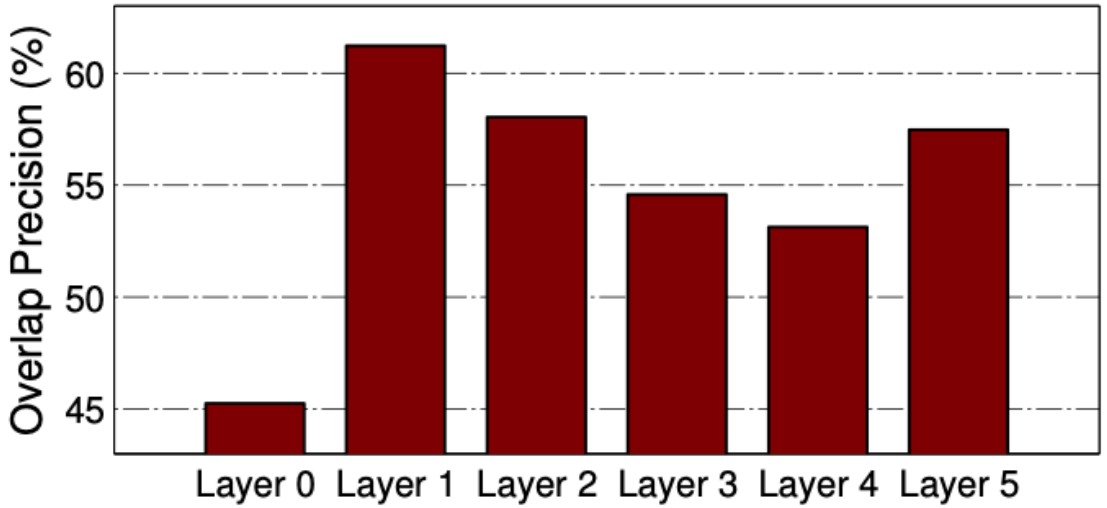


Figure 2. Comparison of tracking performance when using different convolutional layers in the network. The mean overlap precision over all color videos in the OTB dataset is displayed. The input RGB image (layer 0) provides inferior performance compared to the convolutional layers. The best results are obtained using the first convolutional layer. The performance then degrades for each deeper layer in the network, until the final layer.

"Convolutional Features for Correlation Filter Based Visual Tracking", Danelljan et al., ICCVW 2015

DeepSRDCF

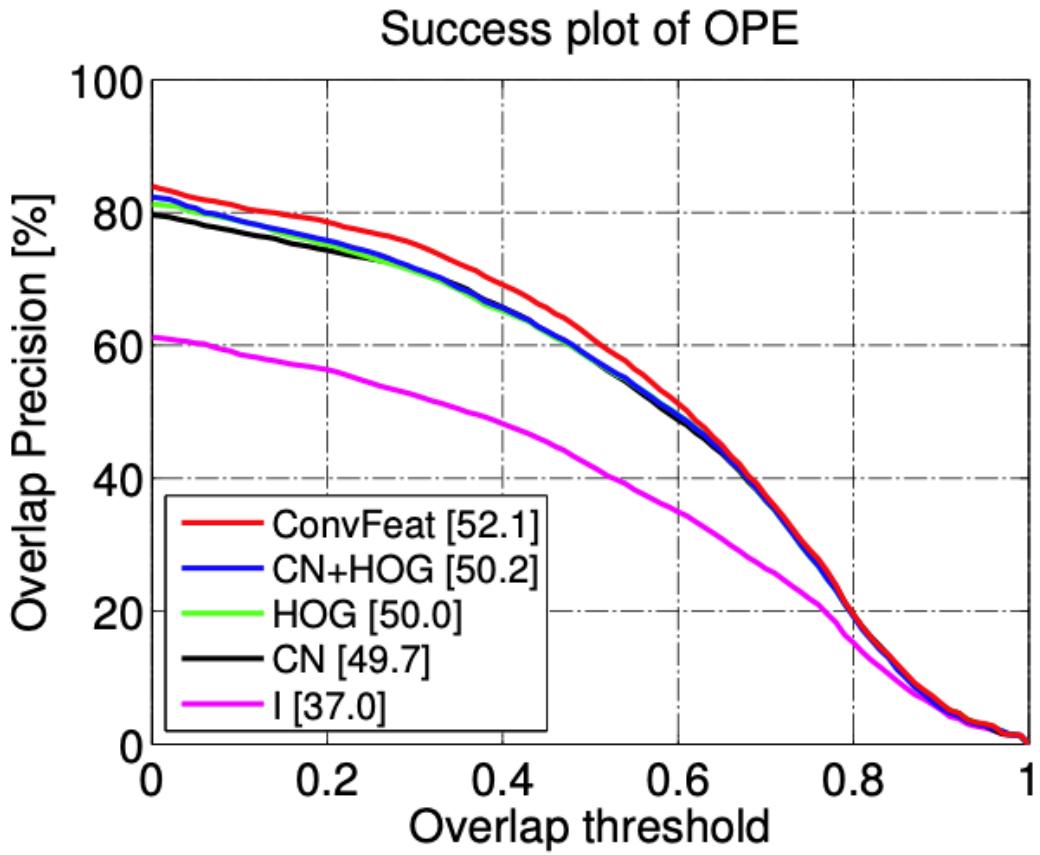
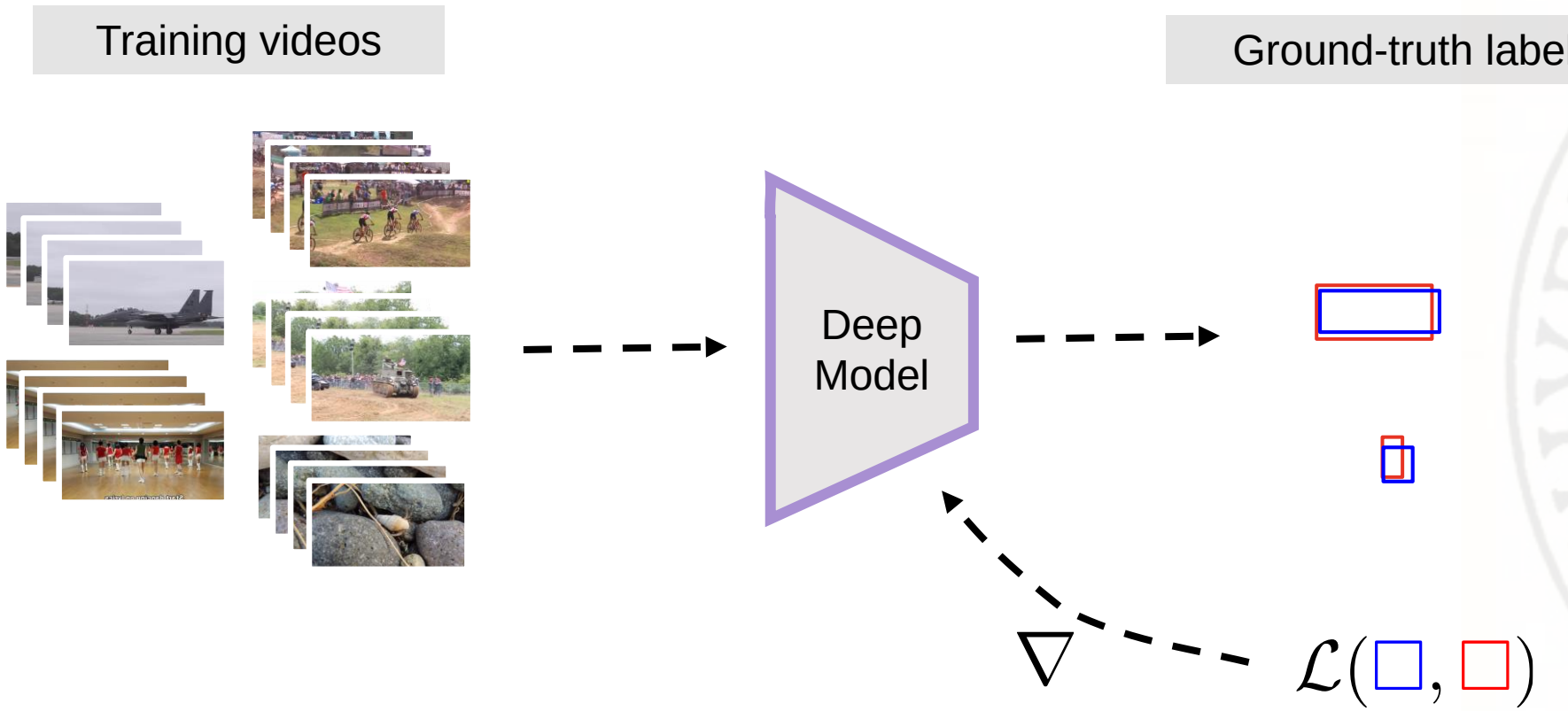


Figure 3. Comparison of the first layer convolutional features with different handcrafted features: HOG, CN and I (image intensity).

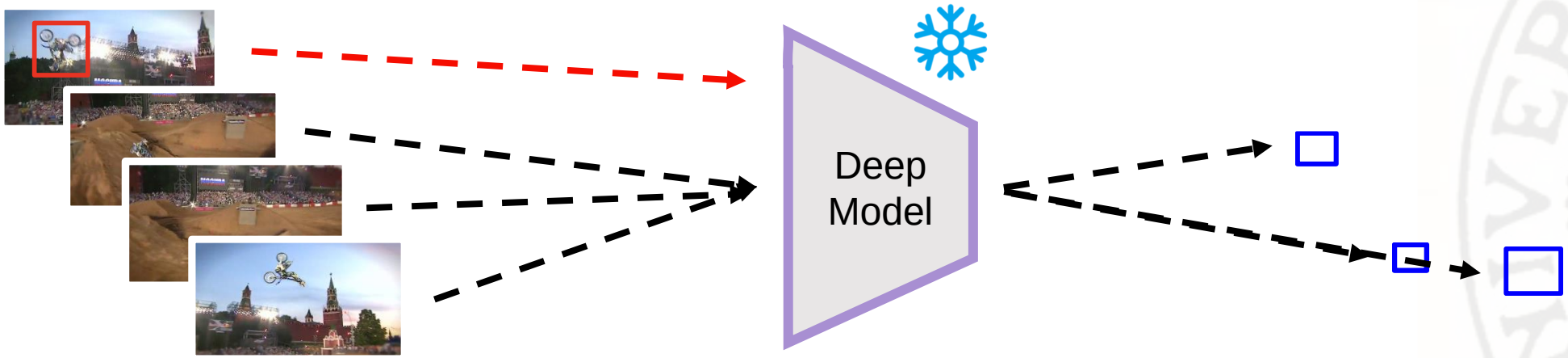
"Convolutional Features for Correlation Filter Based Visual Tracking", Danelljan et al., ICCVW 2015

Offline Learning
vs
Offline/Online Learning

Offline Trackers - Training



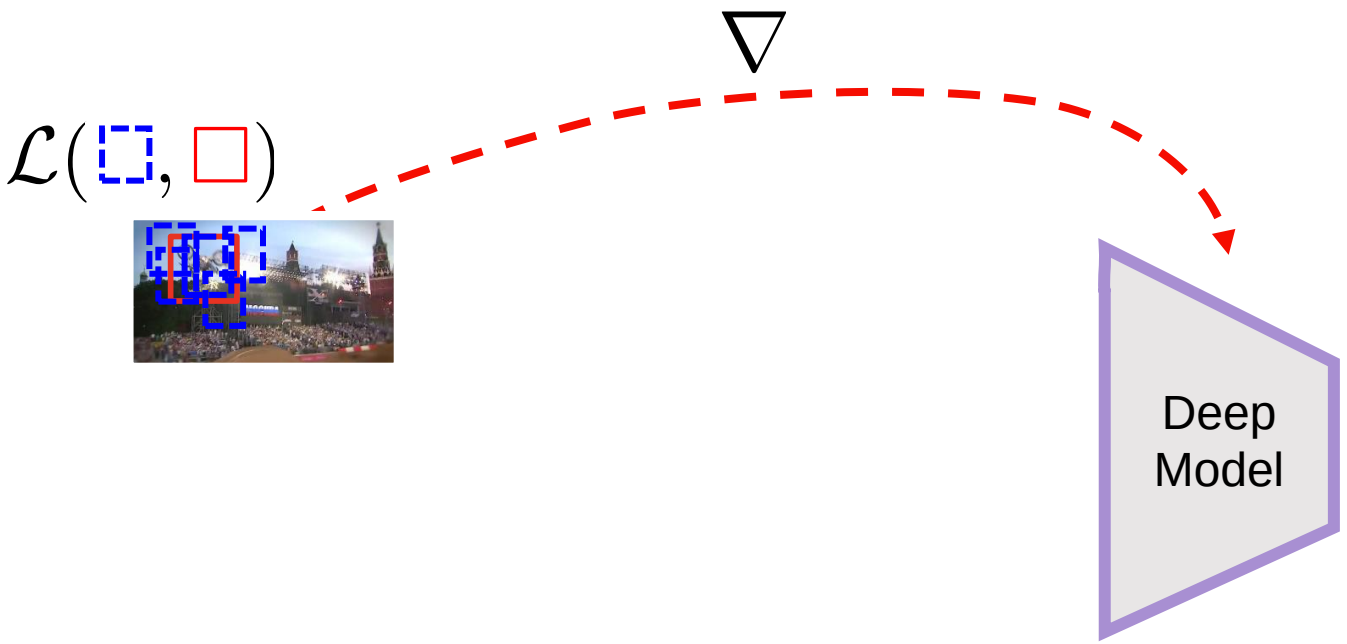
Offline Trackers - Tracking



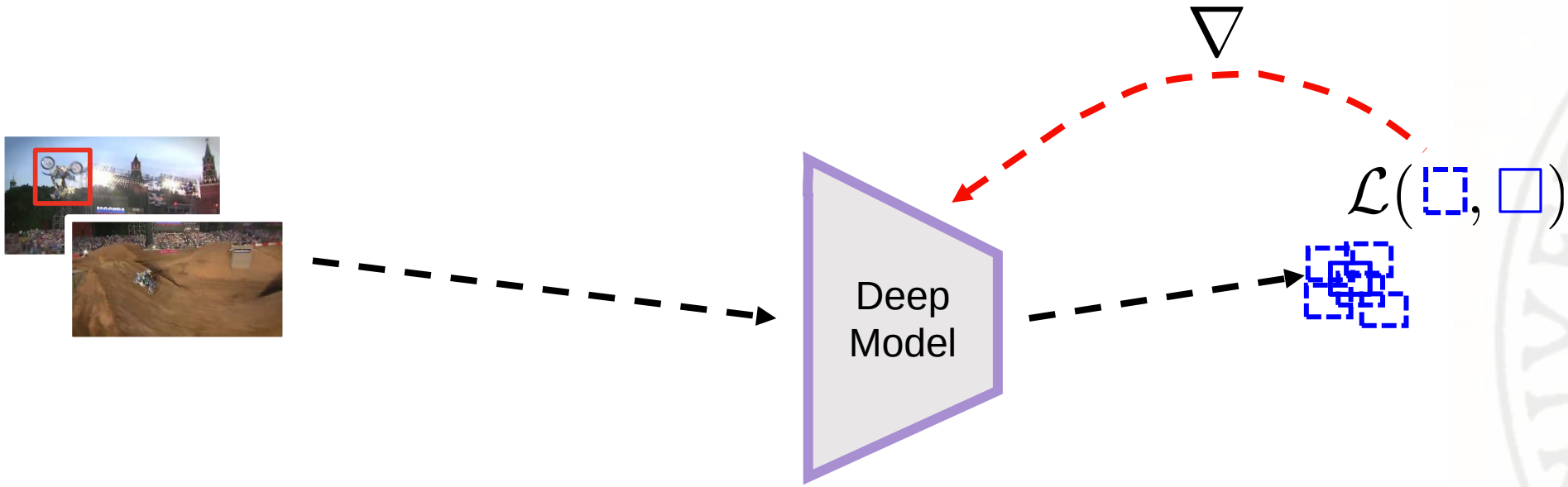
Fast processing times

Not robust to continuously changing scenes

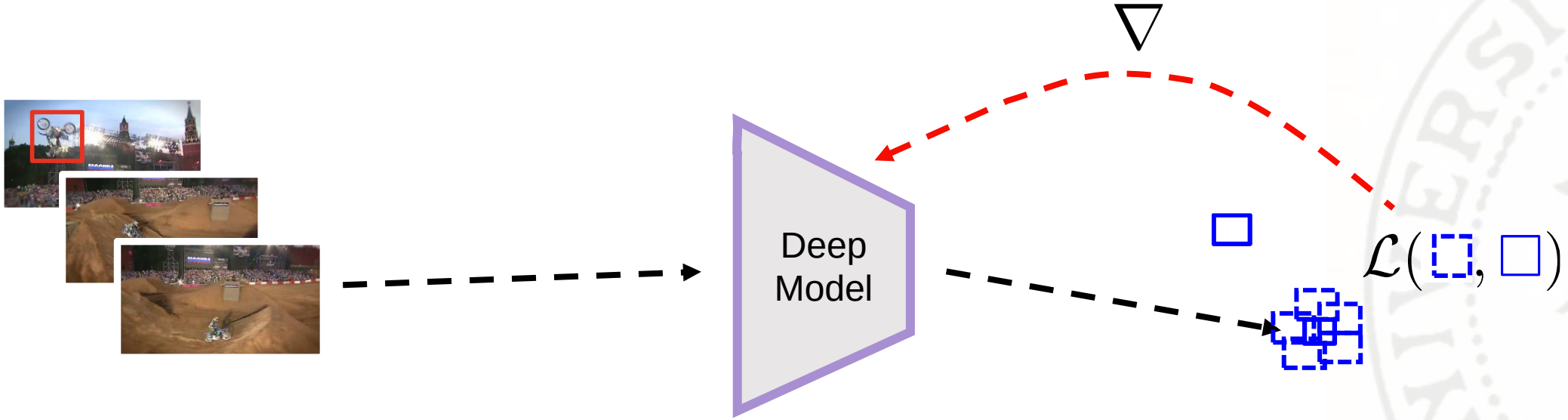
Offline/Online Trackers - Tracking



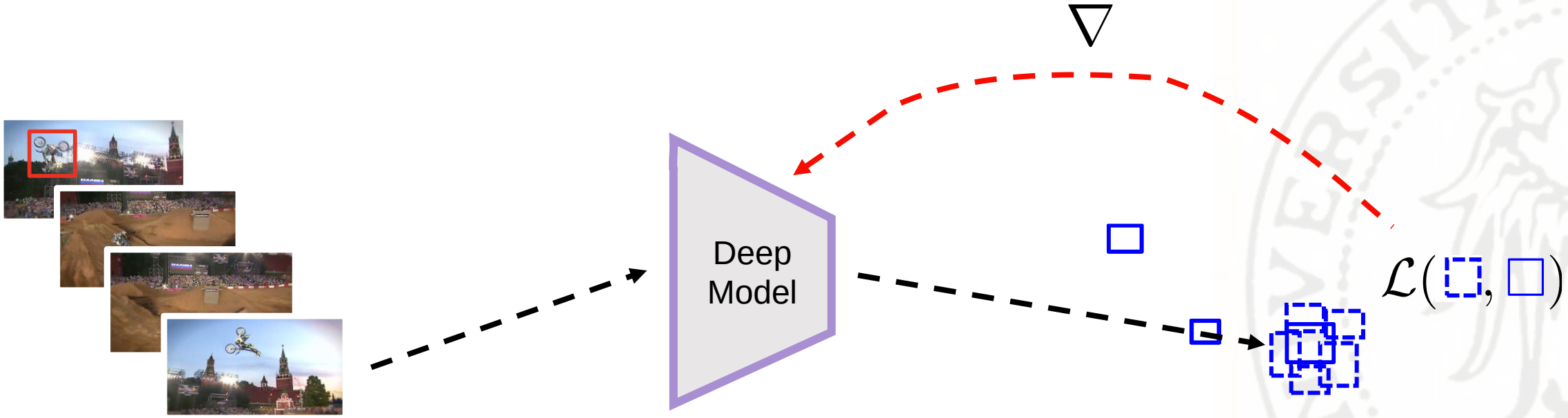
Offline/Online Trackers - Tracking



Offline/Online Trackers - Tracking



Offline/Online Trackers - Tracking



Robust to continously changing scenes

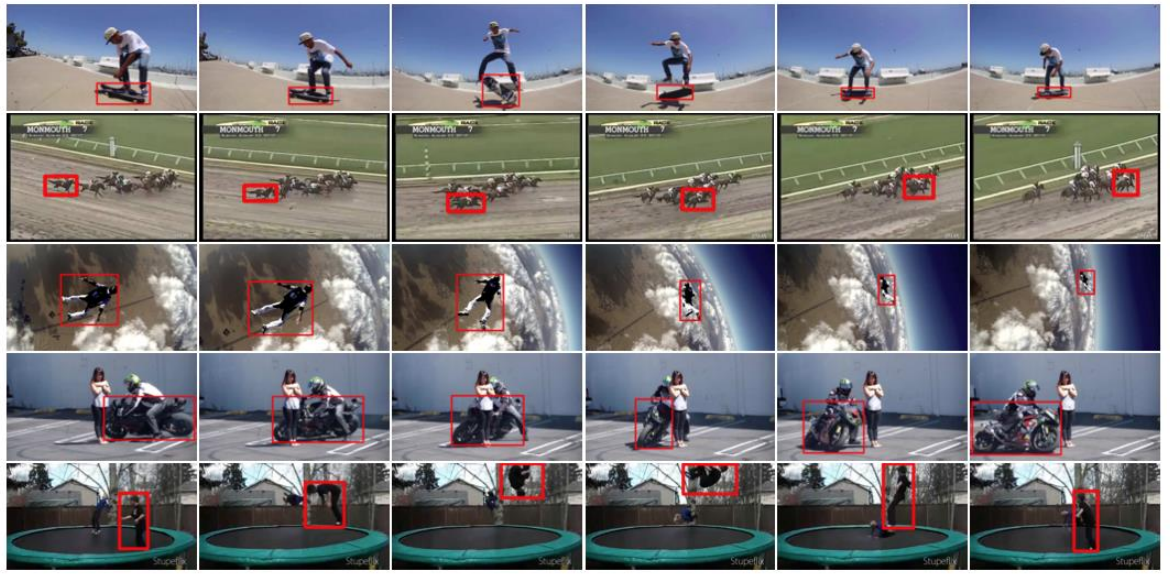
Slow processing speed

Datasets - TrackingNet

~30K videos
- 30132 training videos
- 511 testing videos

Dense bounding-boxes (> 14M)
- Hand-labeled at 1 FPS
- Interpolated at the other frames

27 object categories



https://tracking-net.org

"TrackingNet: A Large-Scale Dataset and Benchmark for Object Tracking in the Wild", Mueller et al., ECCV 2018

Datasets - LaSOT

1400 videos at 30 FPS
- 1220 videos for training
- 280 for testing

70 object categories
- 20 videos for each one

1m23s avg length
- Long-term settings

Language descriptions

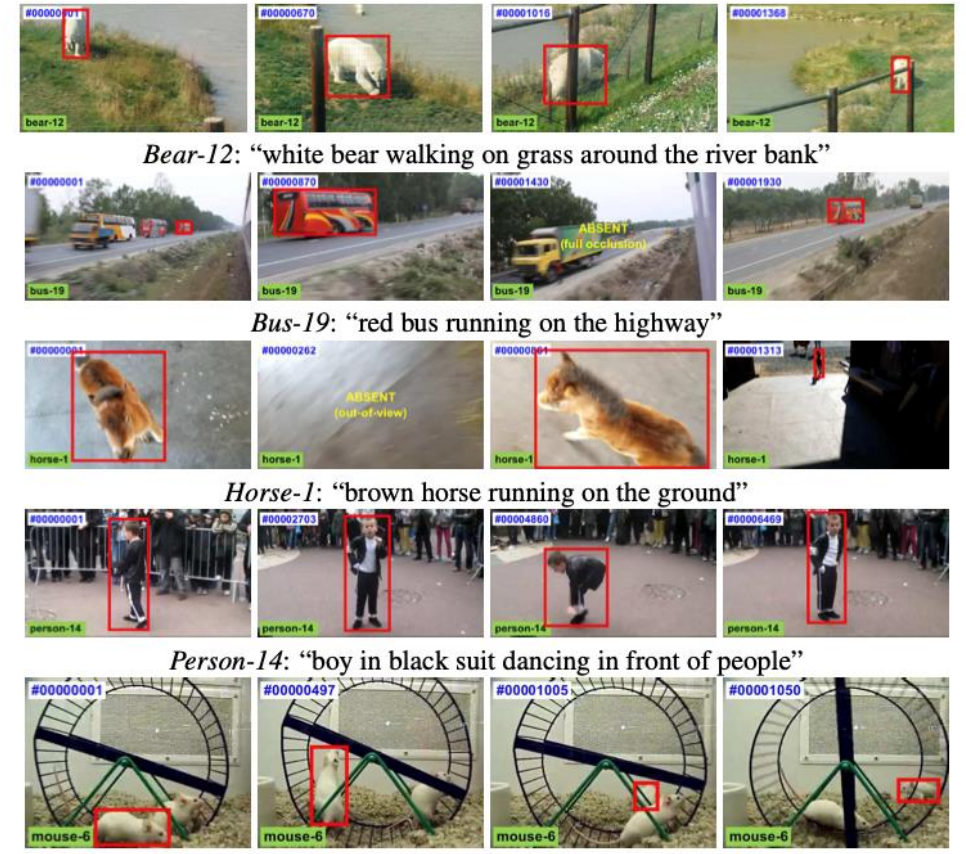


Figure 2. Example sequences and annotations of our LaSOT. We focus on long-term videos in which target objects may disappear, and then re-enter the view again. In addition, we provide natural language specification for each sequence. Best viewed in color.

http://vision.cs.stonybrook.edu/~lasot/

"LaSOT: A High-quality Benchmark for Large-scale Single Object Tracking", Fan, Lin et al., CVPR 2019

Datasets - GOT-10k

10000 10FPS videos

- 9660 videos for training
- 180 for validation
- 180 for testing (sequestered)

536 object categories

16s avg length

Per-frame occlusion-level annotation

http://got-10k.aitestunion.com

"GOT-10k: A Large High-Diversity Benchmark for Generic Object Tracking in the Wild", Huang et al., TPAMI 2019

Offline Trackers

GOTURN - Tracking

$\vdots$

b_t

t

$t+1$

$\vdots$

"Learning to Track at 100 FPS with Deep Regression Networks", Held et al., ECCV 2016

GOTURN - Tracking

$\vdots$

b_t

t

$t+1$

$\vdots$

"Learning to Track at 100 FPS with Deep Regression Networks", Held et al., ECCV 2016

GOTURN - Tracking

$\vdots$

b_t

t

227 x 227 x 3

$t+1$

227 x 227 x 3

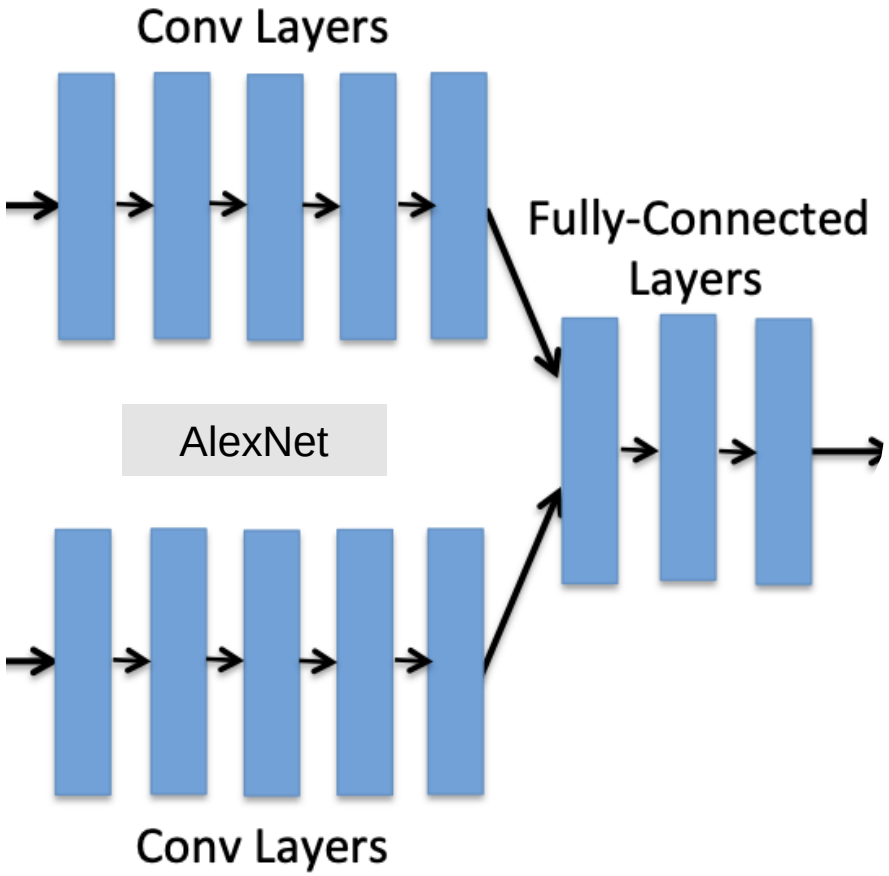
$\vdots$

"Learning to Track at 100 FPS with Deep Regression Networks", Held et al., ECCV 2016

GOTURN - Tracking

227 x 227 x 3

227 x 227 x 3



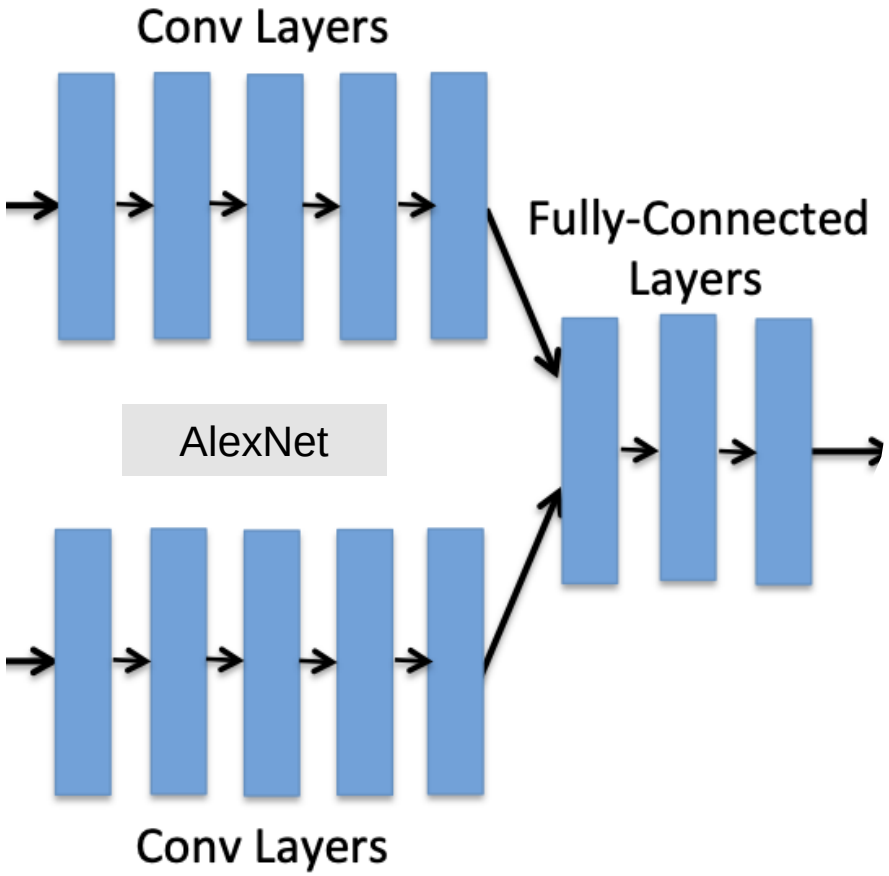
for training

"Learning to Track at 100 FPS with Deep Regression Networks", Held et al., ECCV 2016

GOTURN - Tracking

227 x 227 x 3

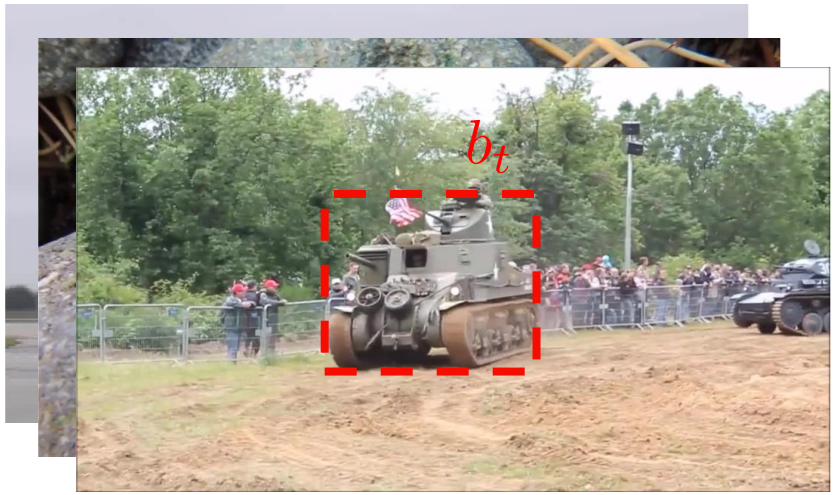
227 x 227 x 3



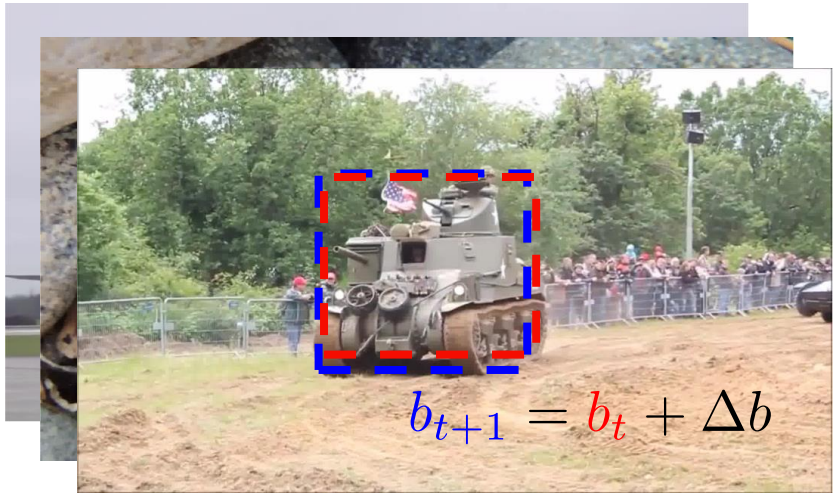
for training

"Learning to Track at 100 FPS with Deep Regression Networks", Held et al., ECCV 2016

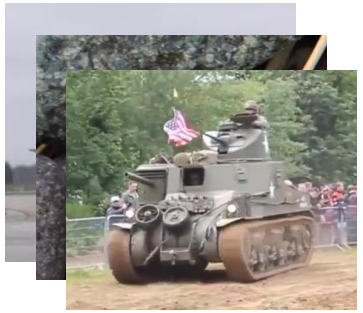
GOTURN - Training



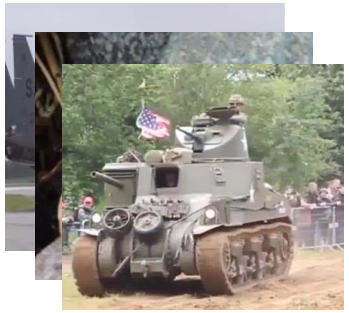
t



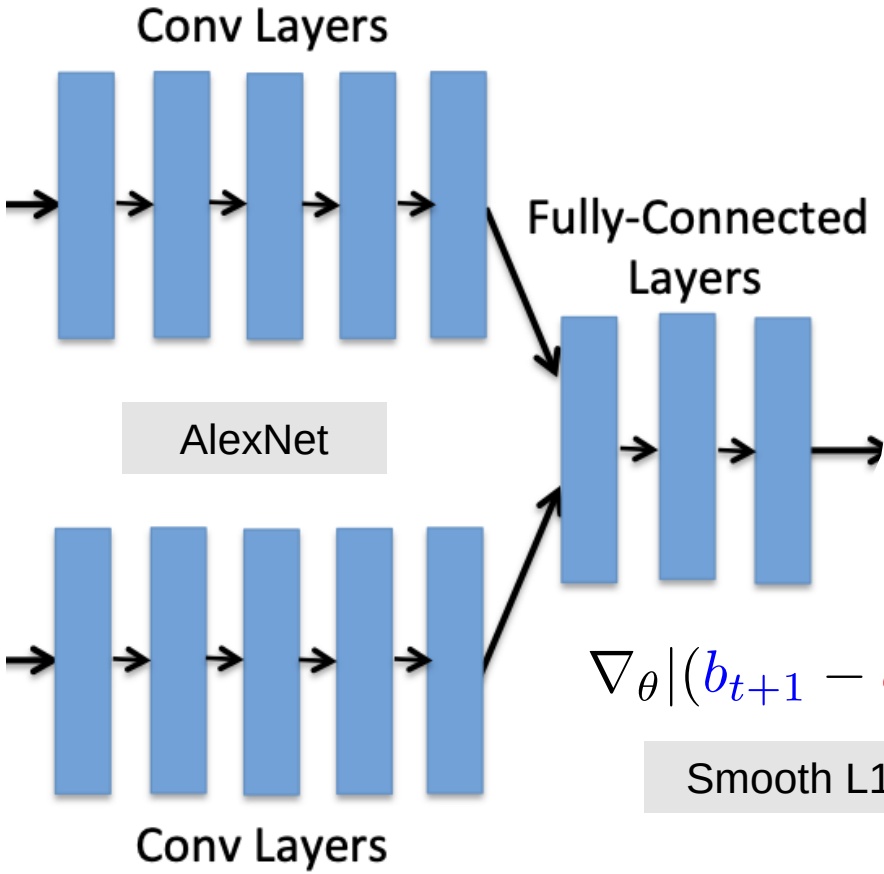
$t+1$



227 x 227 x 3



227 x 227 x 3



"Learning to Track at 100 FPS with Deep Regression Networks", Held et al., ECCV 2016

SiamFC - Tracking

$t = 0$

$t > 0$

"Fully-Convolutional Siamese Networks for Object Tracking", Bertinetto et al., ECCVW 2016

SiamFC - Tracking

$t = 0$

$t > 0$

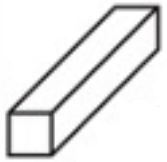
"Fully-Convolutional Siamese Networks for Object Tracking", Bertinetto et al., ECCVW 2016

SiamFC - Tracking

$t = 0$

CNN

T



$t > 0$

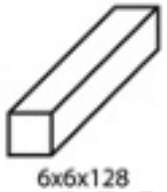
"Fully-Convolutional Siamese Networks for Object Tracking", Bertinetto et al., ECCVW 2016

SiamFC - Tracking

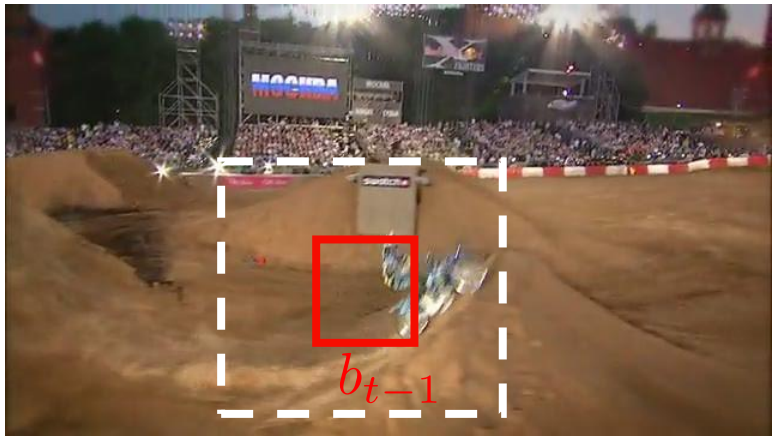
$t = 0$

CNN

T

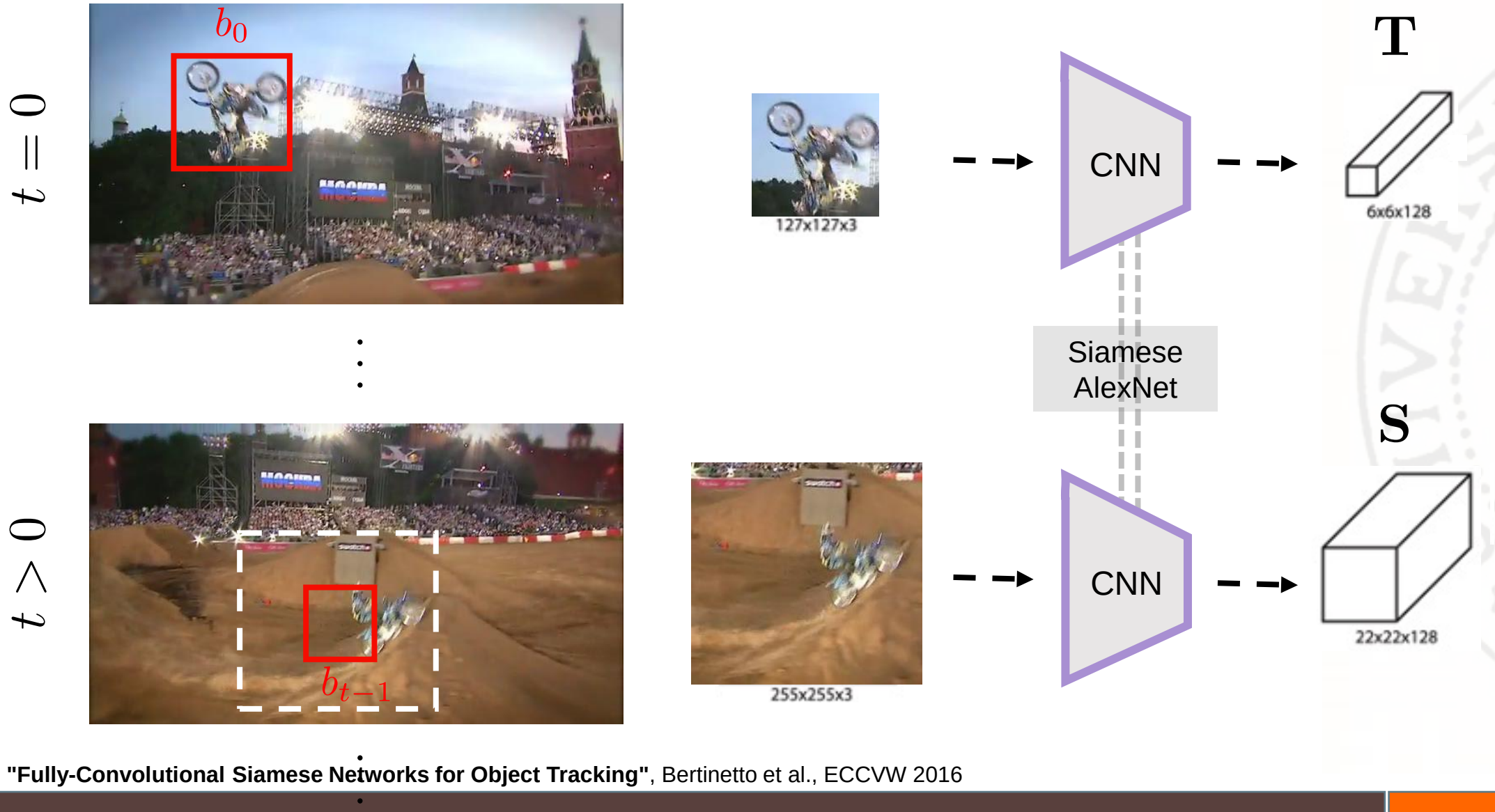


$t > 0$



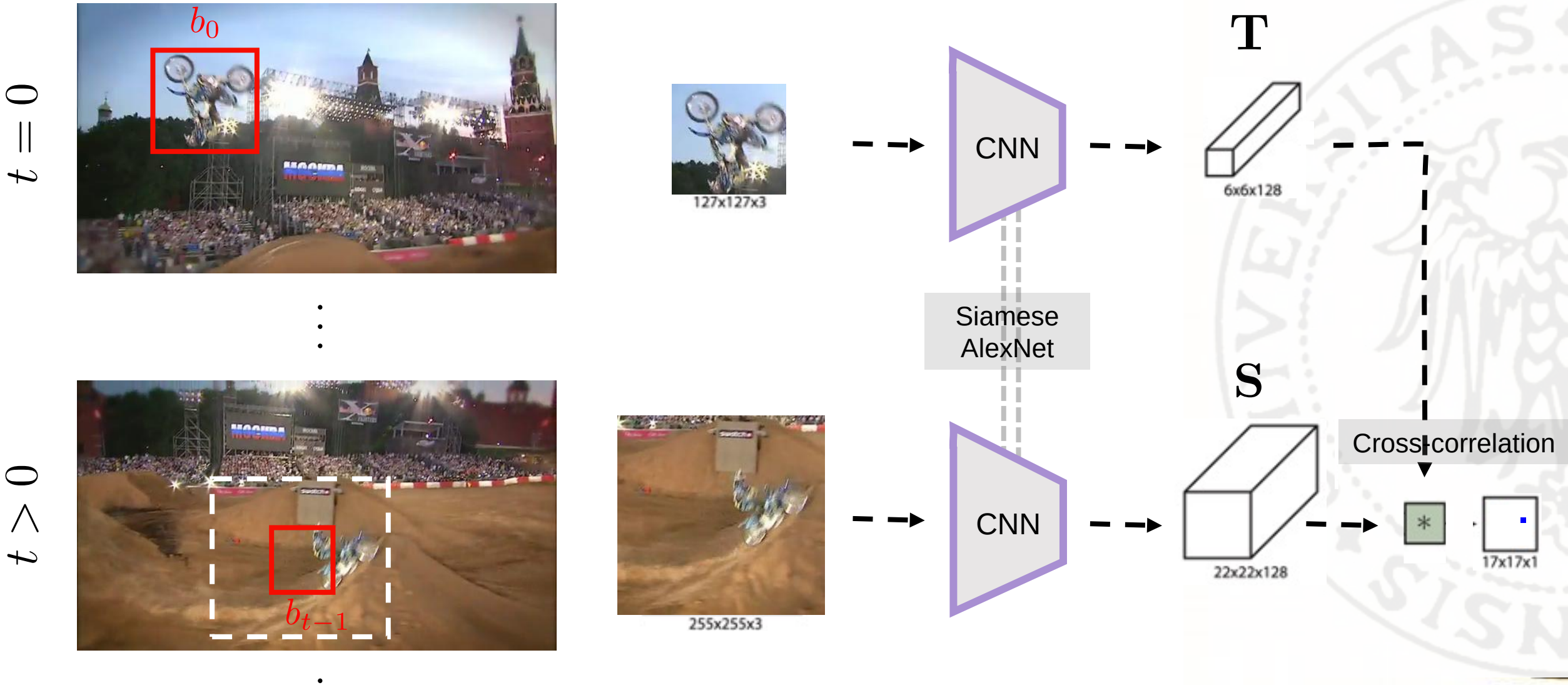
"Fully-Convolutional Siamese Networks for Object Tracking", Bertinetto et al., ECCVW 2016

SiamFC - Tracking



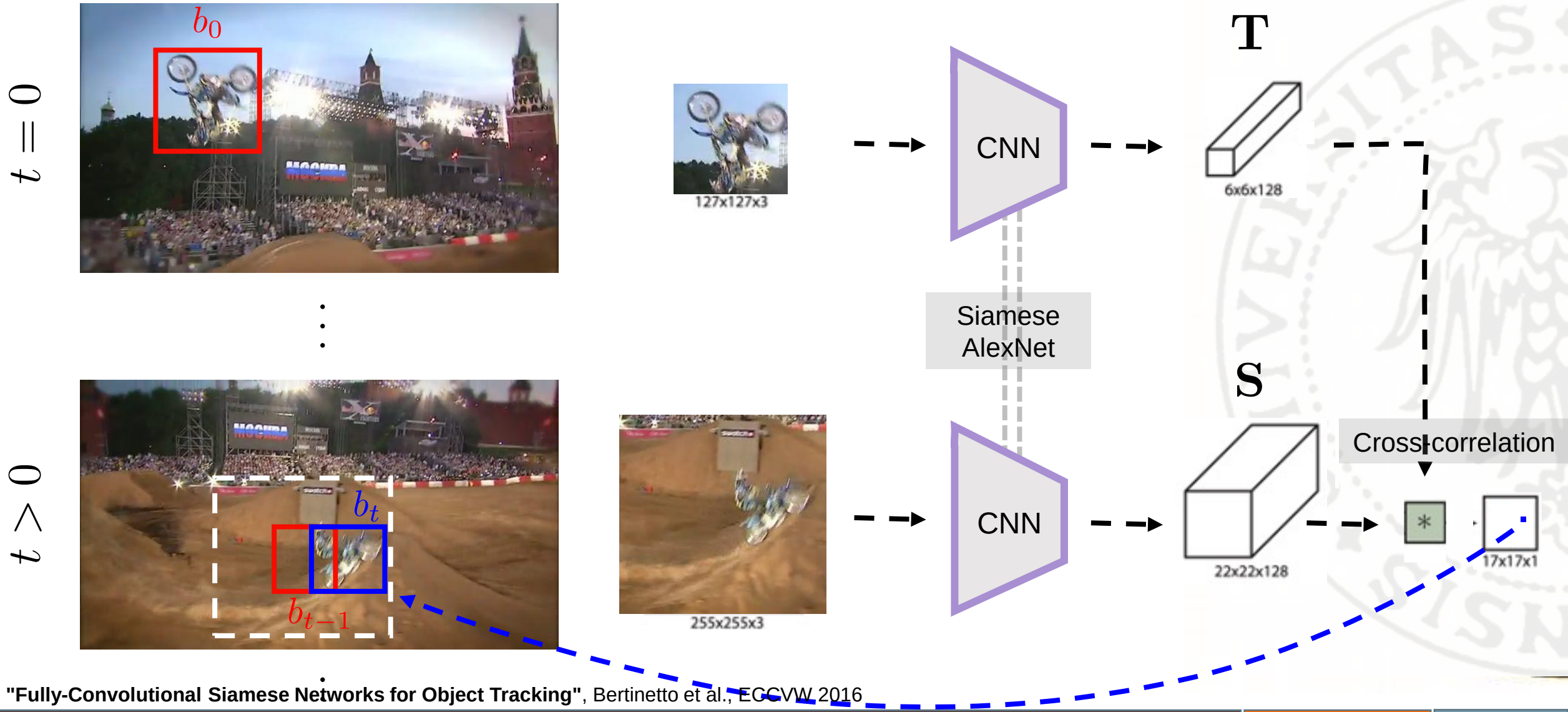
"Fully-Convolutional Siamese Networks for Object Tracking", Bertinetto et al., ECCVW 2016

SiamFC - Tracking



"Fully-Convolutional Siamese Networks for Object Tracking", Bertinetto et al., ECCVW 2016

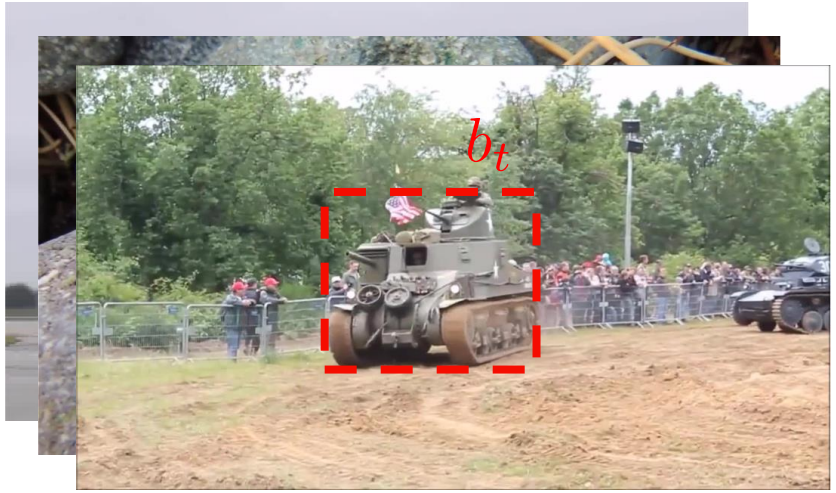
SiamFC - Tracking



"Fully-Convolutional Siamese Networks for Object Tracking", Bertinetto et al., ECCVW 2016

SiamFC - Training

t

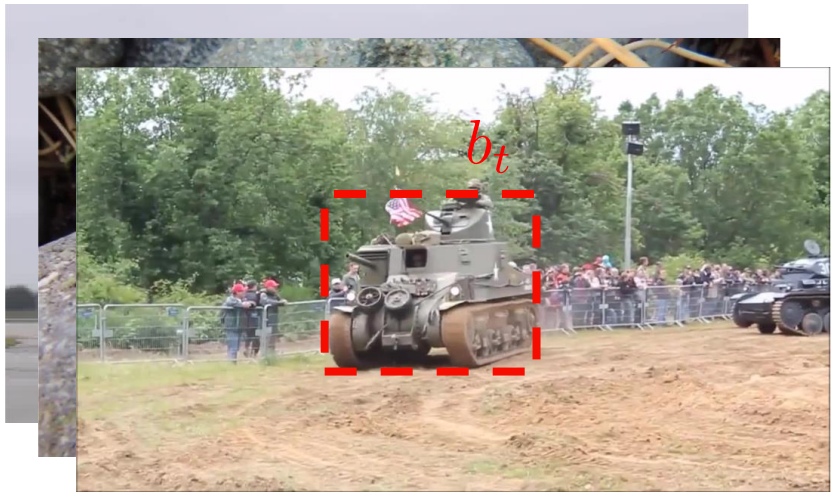


$t + k$

"Fully-Convolutional Siamese Networks for Object Tracking", Bertinetto et al., ECCVW 2016

SiamFC - Training

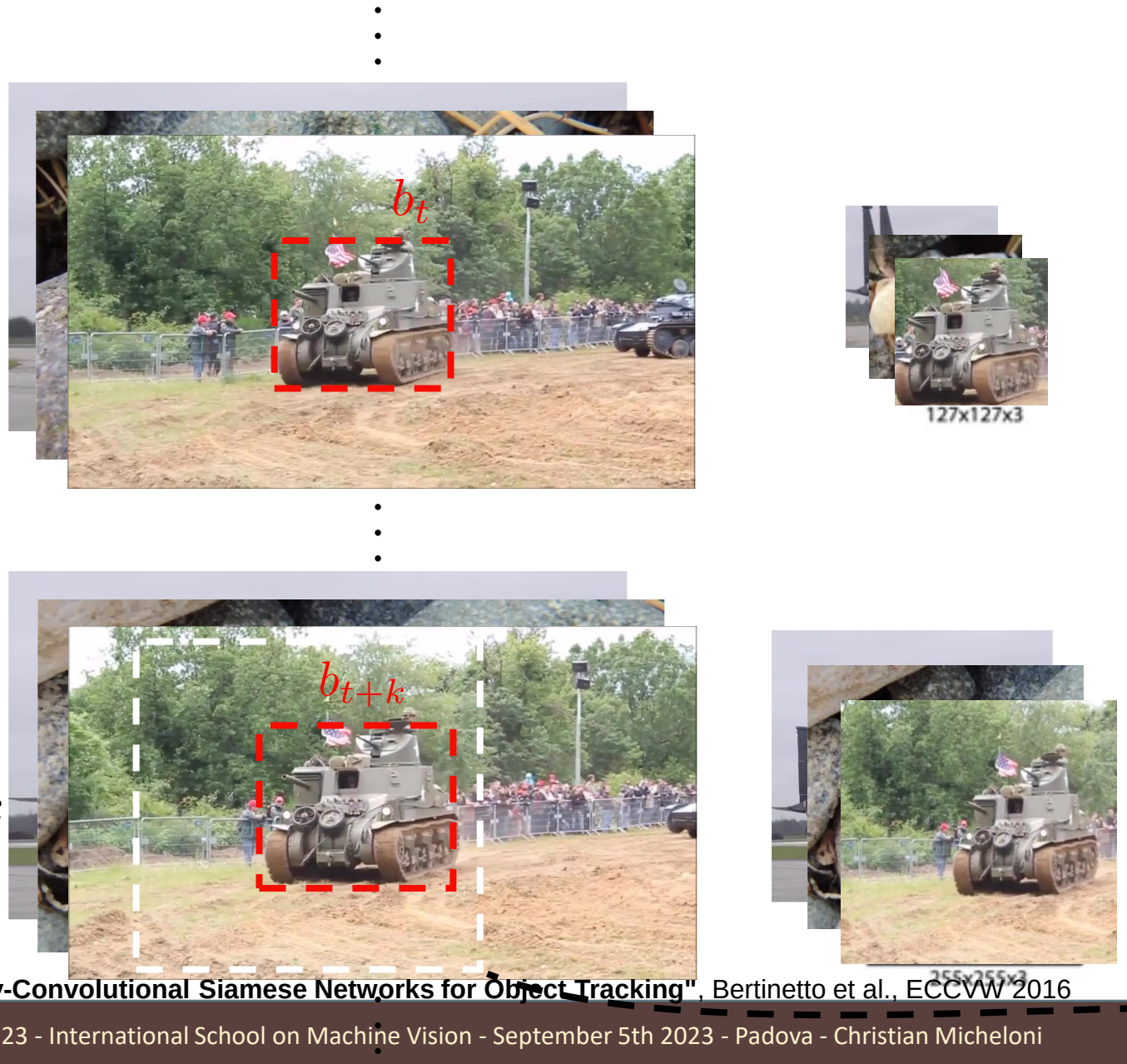
t



$t + k$

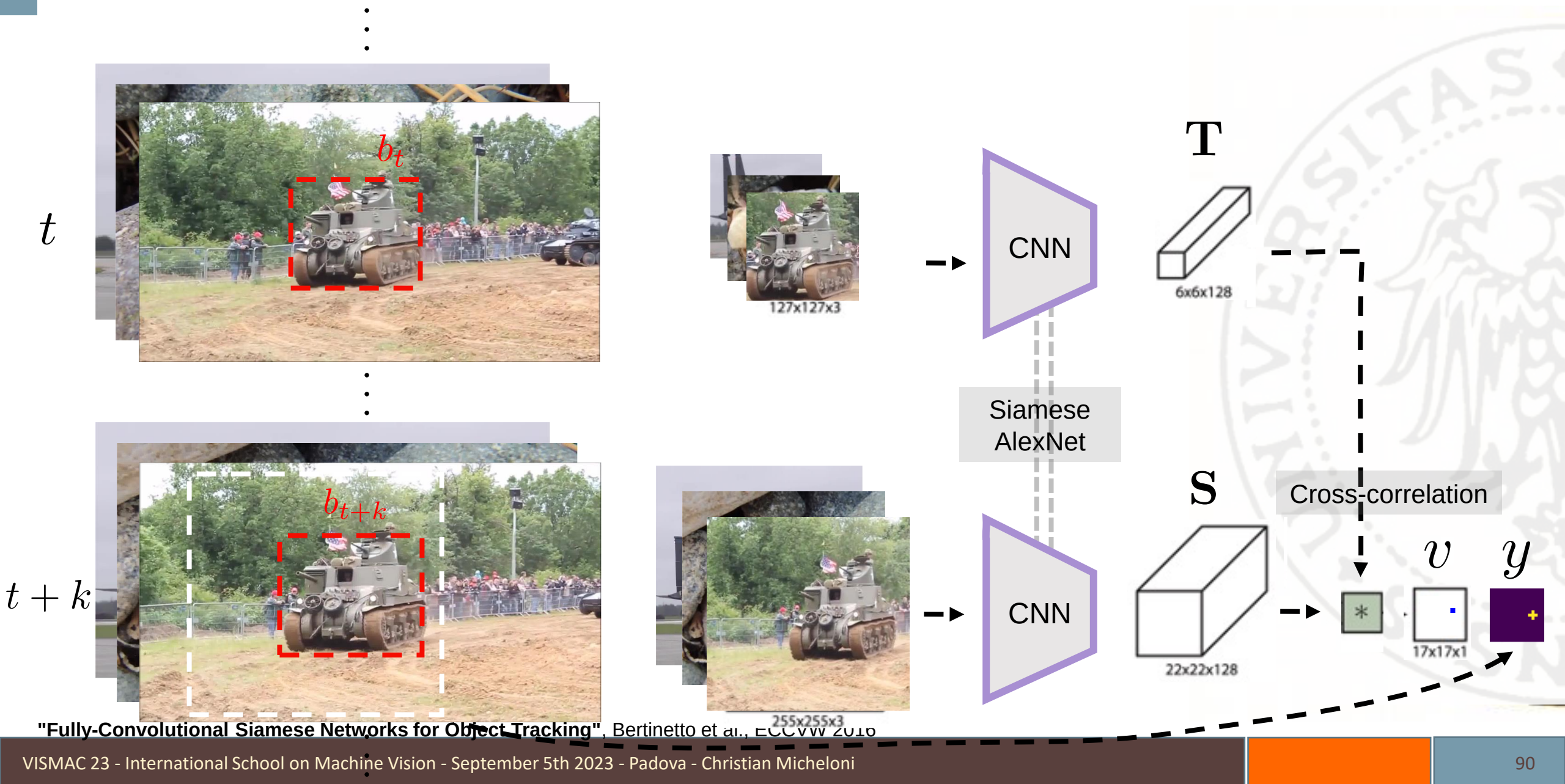
"Fully-Convolutional Siamese Networks for Object Tracking", Bertinetto et al., ECCVW 2016

SiamFC - Training



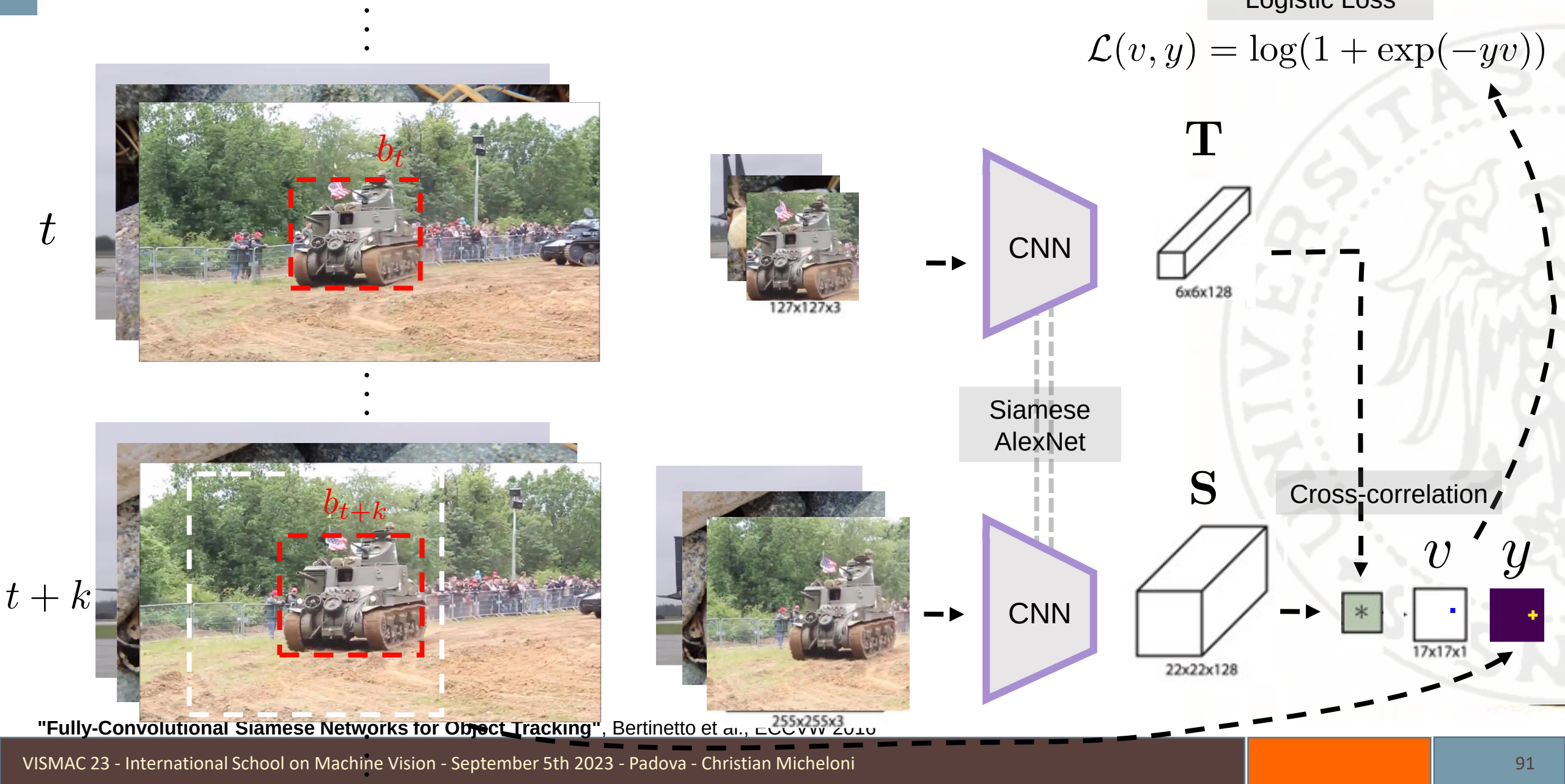
"Fully-Convolutional Siamese Networks for Object Tracking", Bertinetto et al., ECCVW 2016

SiamFC - Training



"Fully-Convolutional Siamese Networks for Object Tracking", Bertinetto et al., ECCVW 2016

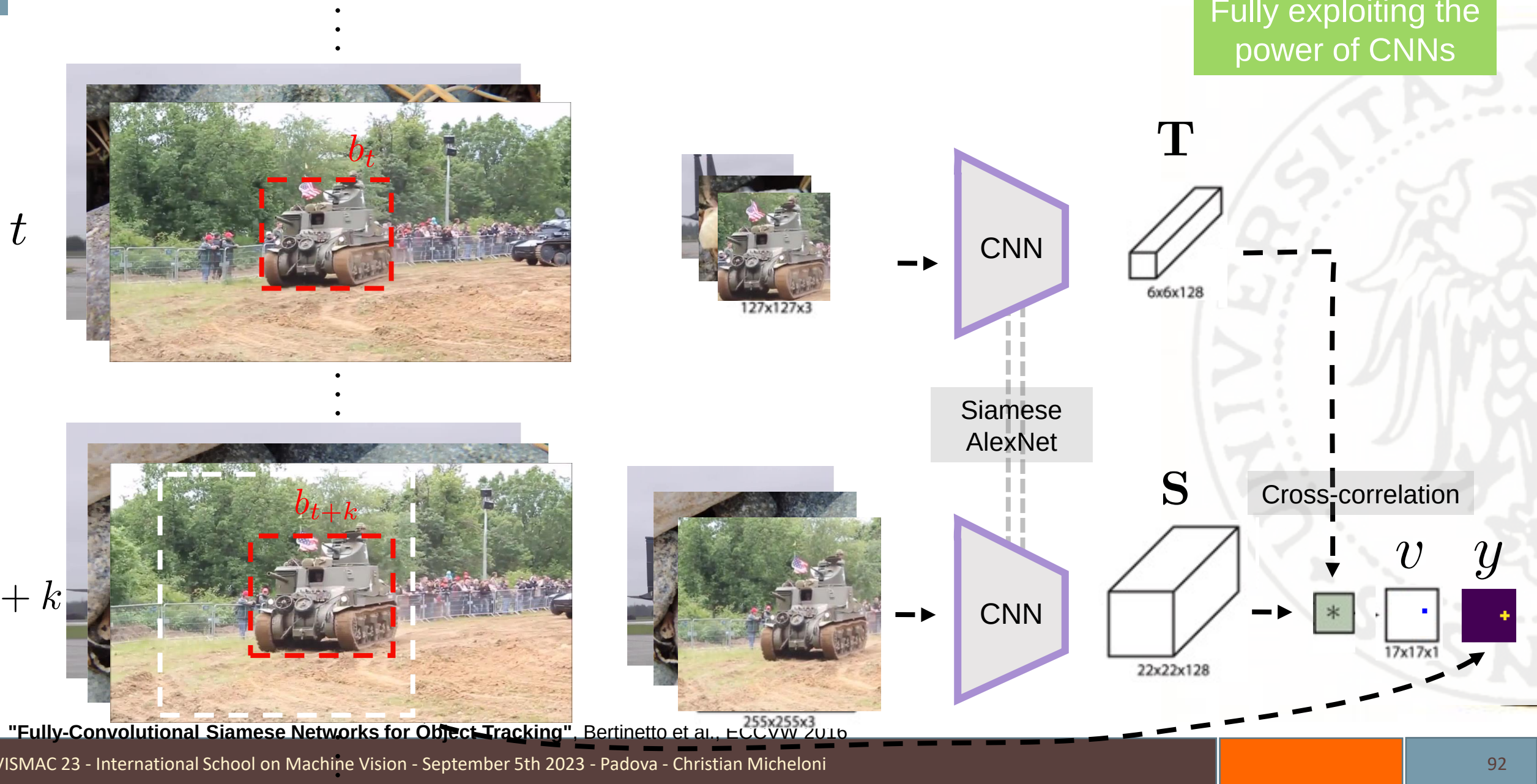
SiamFC - Training



"Fully-Convolutional Siamese Networks for Object Tracking", Bertinetto et al., ECCVW 2016

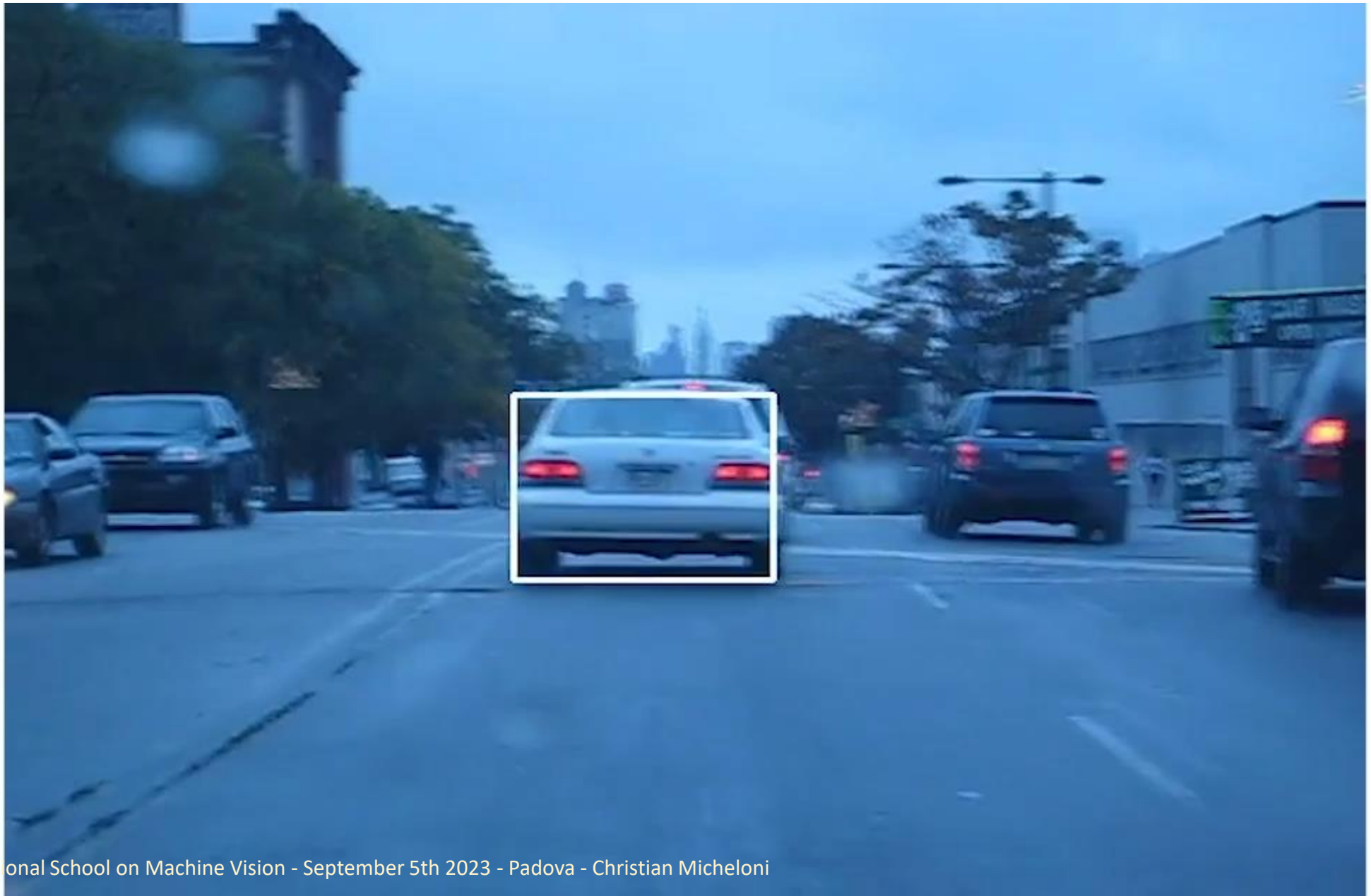
SiamFC - Training

Fully exploiting the power of CNNs



"Fully-Convolutional Siamese Networks for Object Tracking", Bertinetto et al., ECCVW 2016

SiamFC



STARK - Tracking

$t = 0$

$\vdots$

$t > 0$

$\vdots$

"Learning Spatio-Temporal Transformer for Visual Tracking", Yan et al., ICCV 2021

STARK - Tracking

$t = 0$

128 x 128 x 3

$t > 0$

"Learning Spatio-Temporal Transformer for Visual Tracking", Yan et al., ICCV 2021

STARK - Tracking

$t = 0$

128 x 128 x 3

CNN

T

Flatten

$t > 0$

"Learning Spatio-Temporal Transformer for Visual Tracking", Yan et al., ICCV 2021

STARK - Tracking

$t = 0$

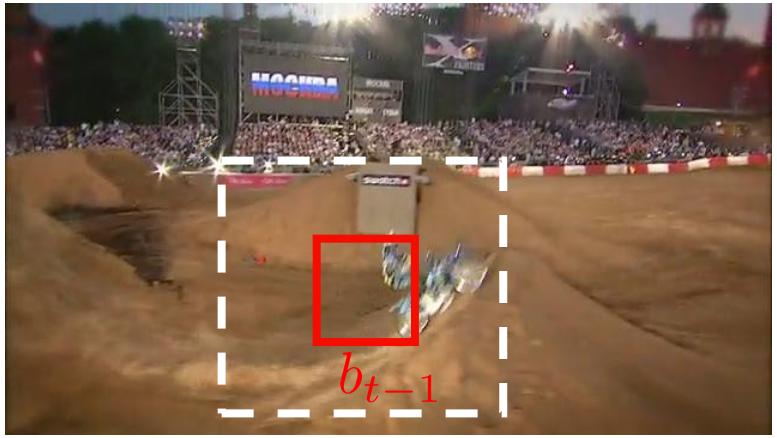
128 x 128 x 3

CNN

T

Flatten

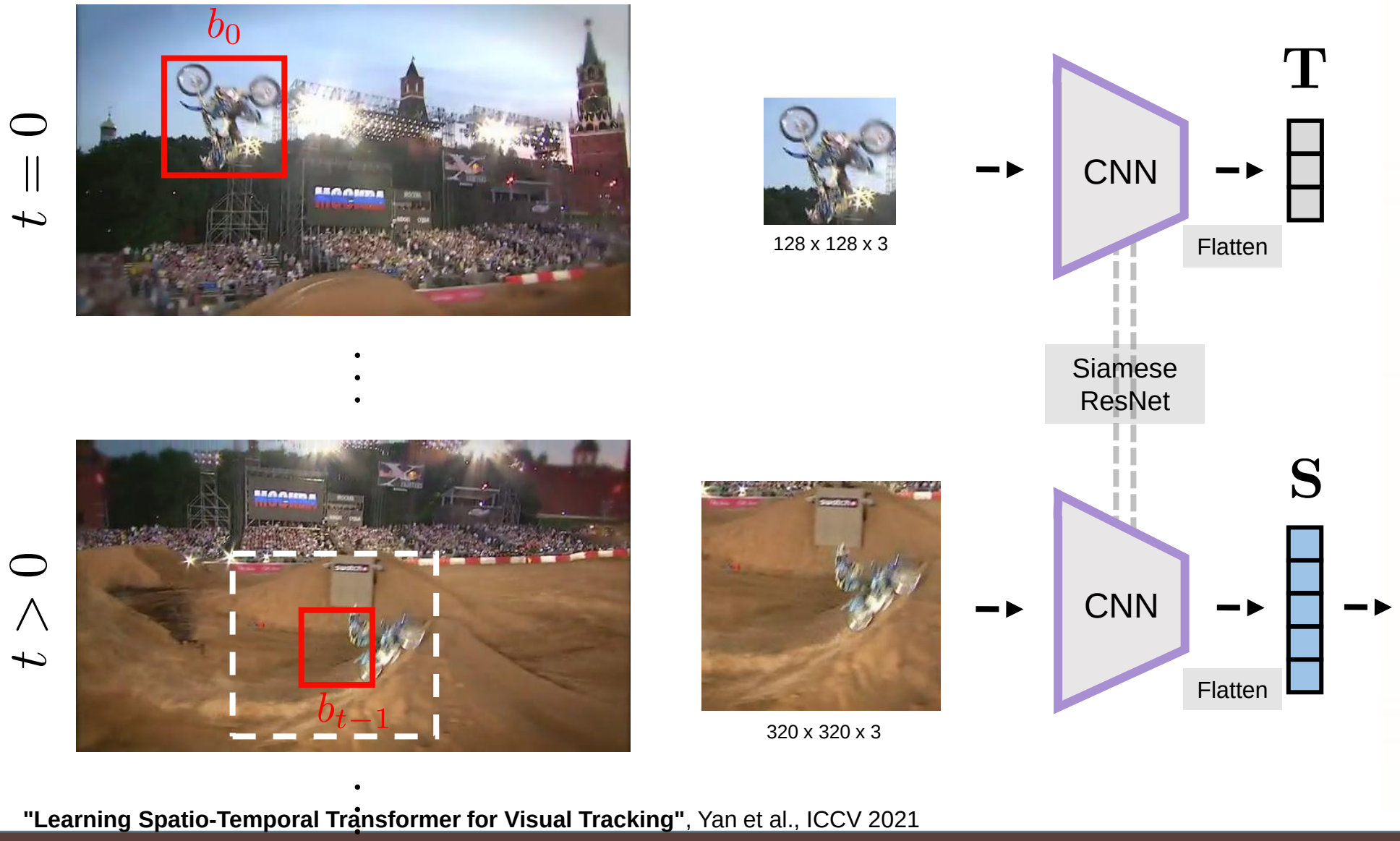
$t > 0$



320 x 320 x 3

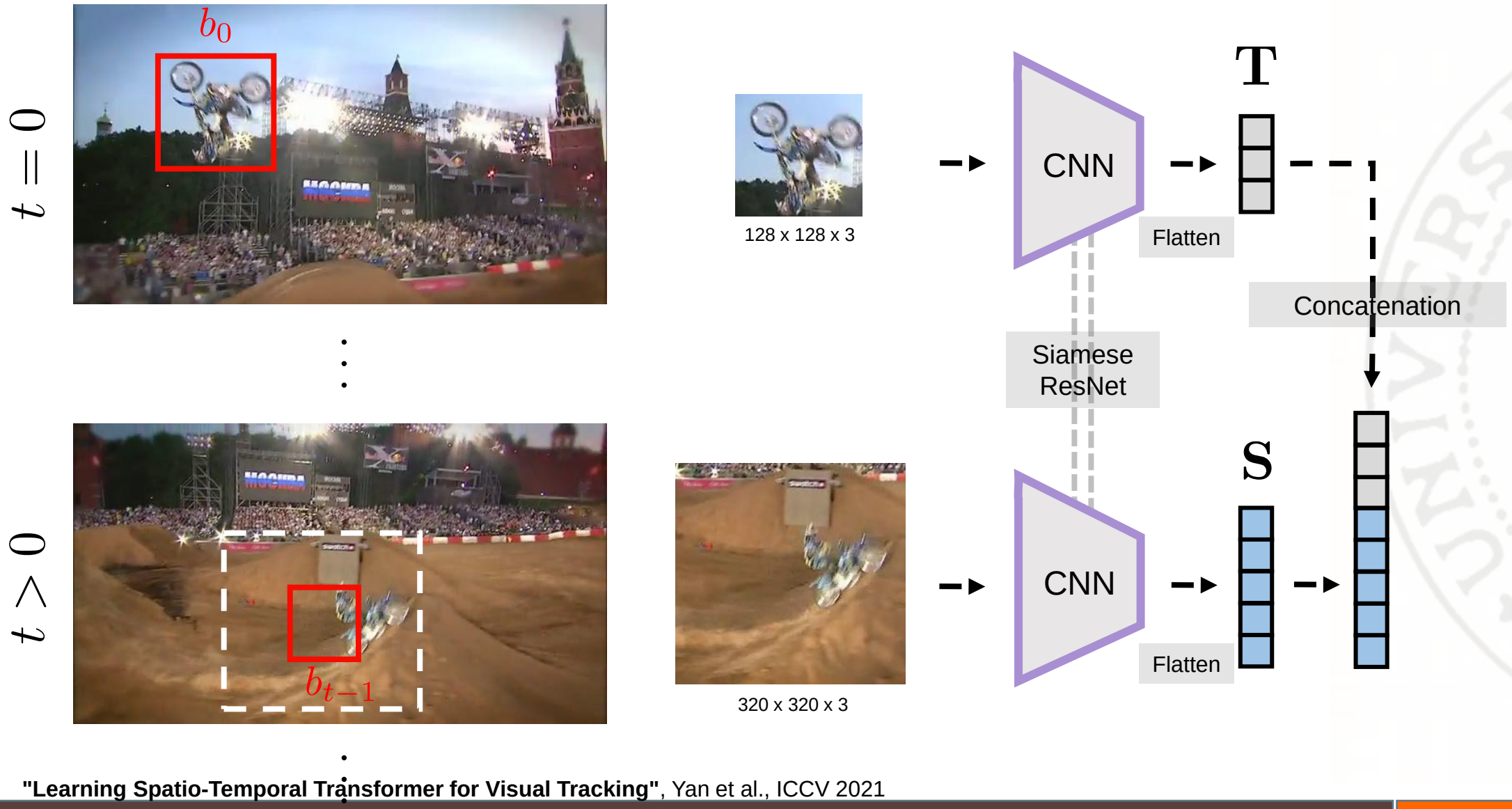
"Learning Spatio-Temporal Transformer for Visual Tracking", Yan et al., ICCV 2021

STARK - Tracking



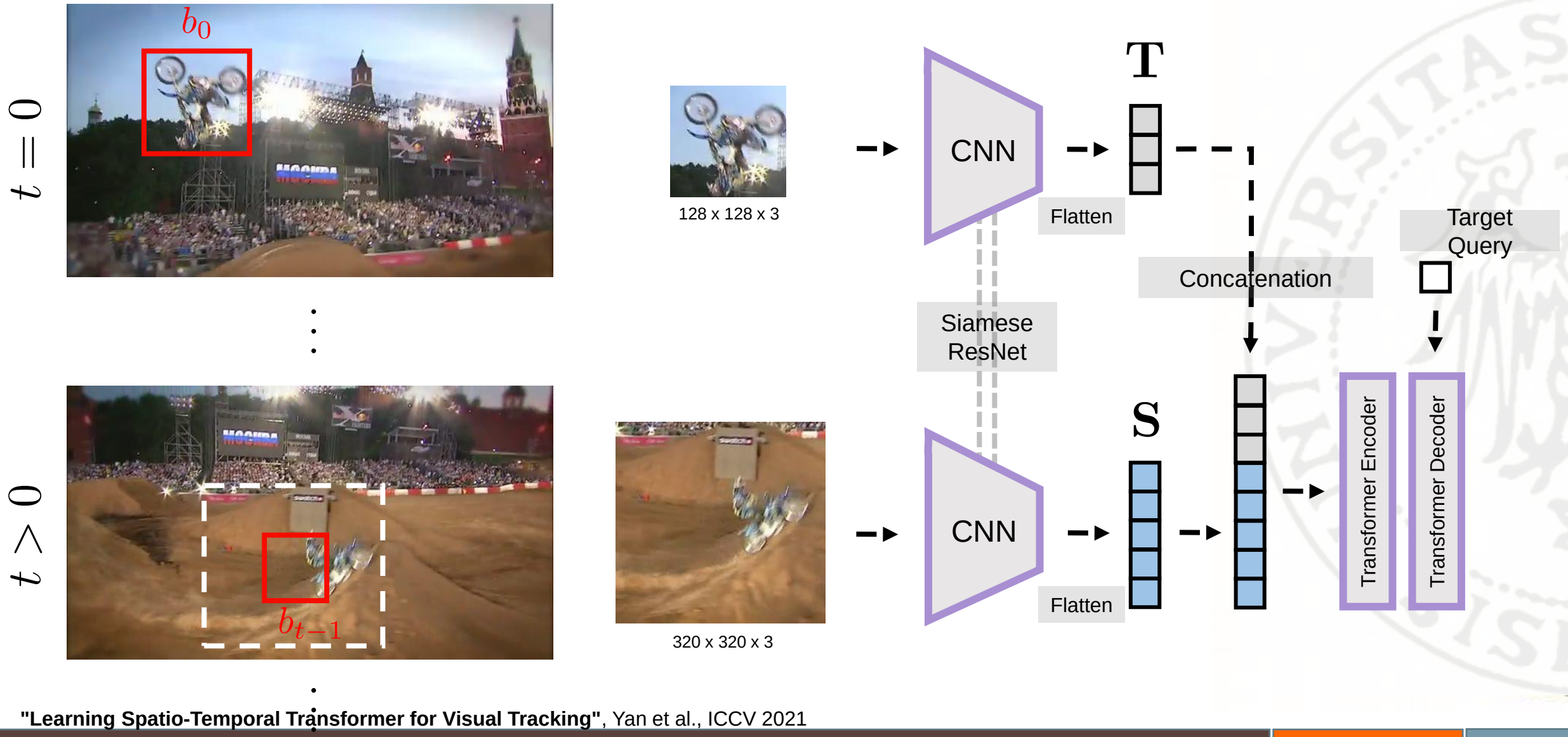
"Learning Spatio-Temporal Transformer for Visual Tracking", Yan et al., ICCV 2021

STARK - Tracking



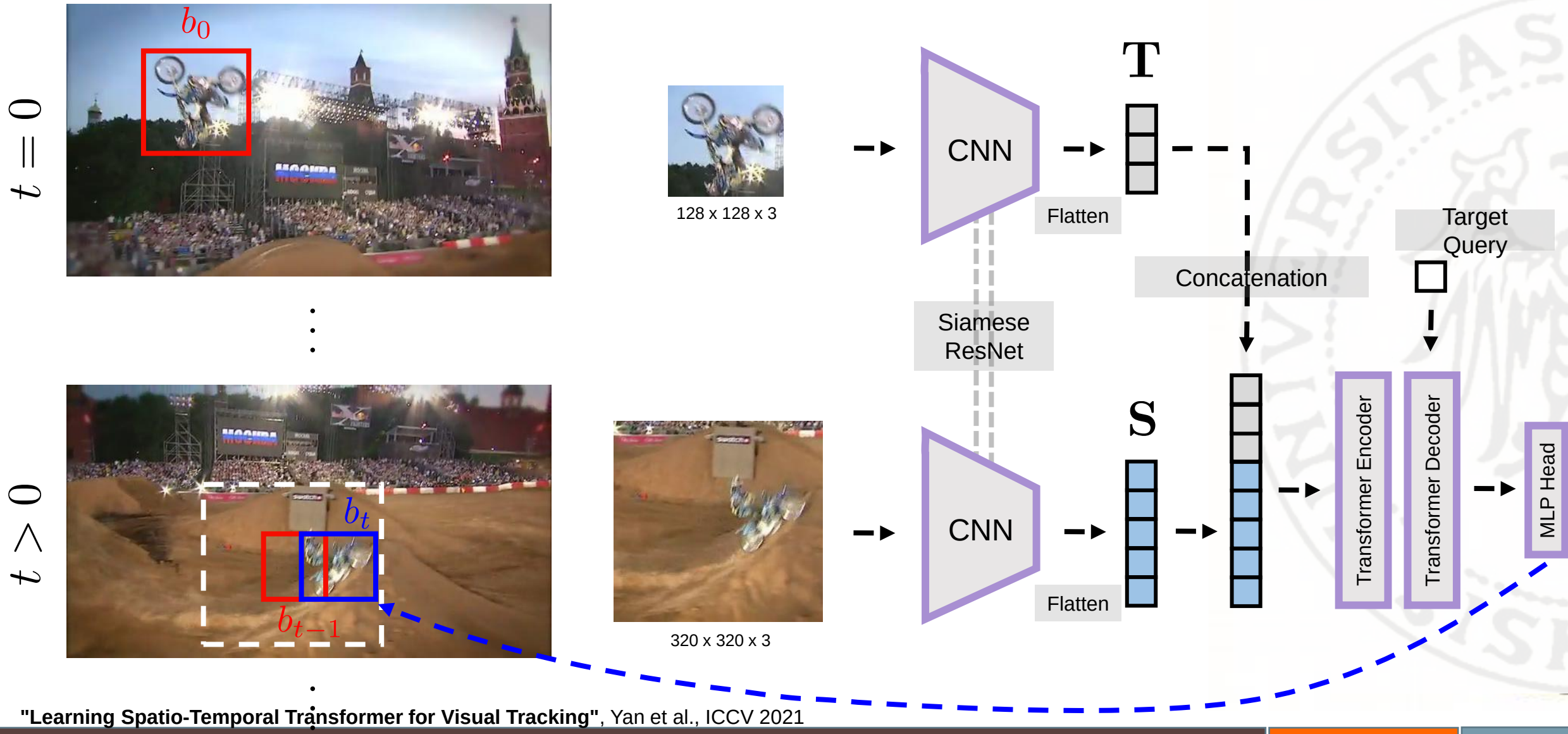
"Learning Spatio-Temporal Transformer for Visual Tracking", Yan et al., ICCV 2021

STARK - Tracking



"Learning Spatio-Temporal Transformer for Visual Tracking", Yan et al., ICCV 2021

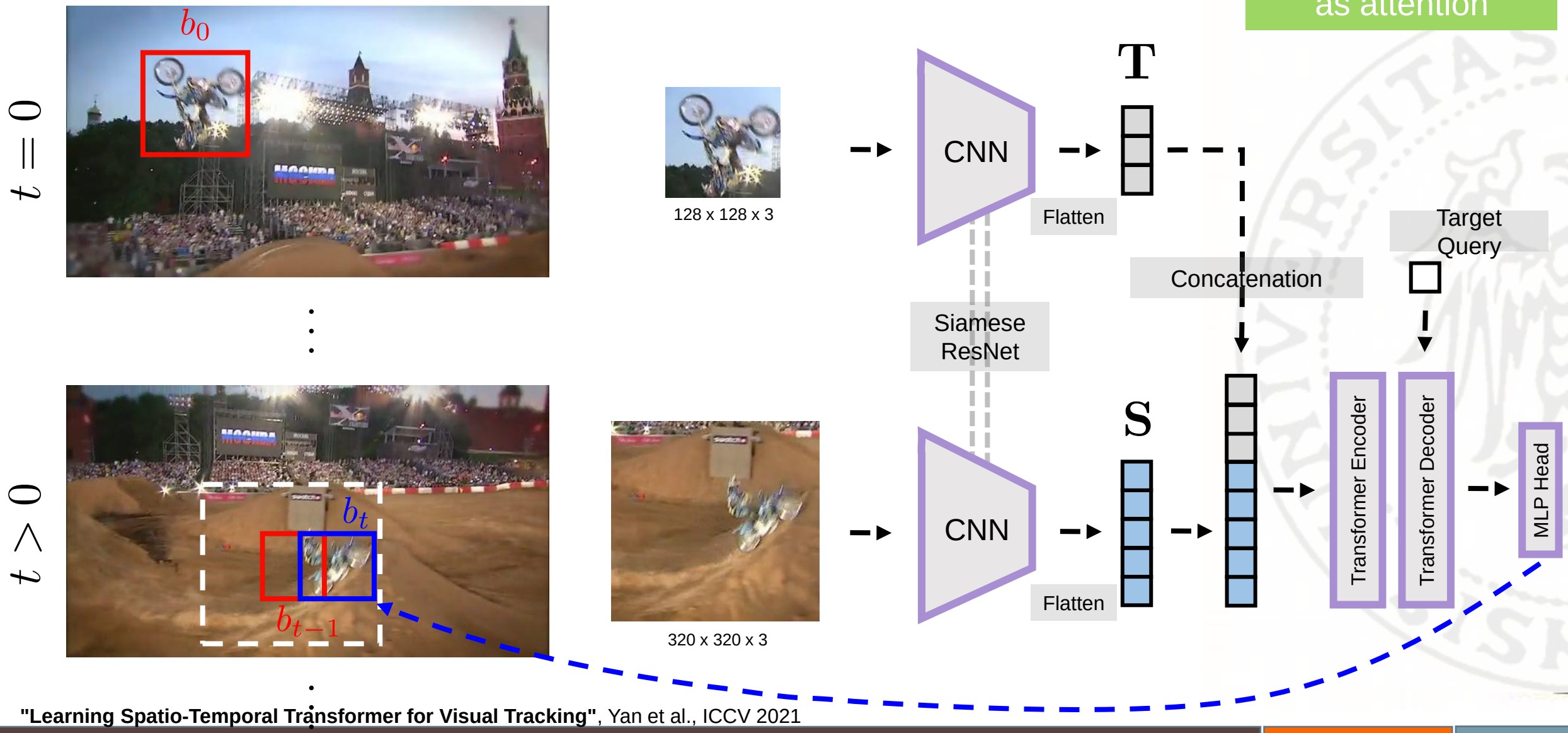
STARK - Tracking



"Learning Spatio-Temporal Transformer for Visual Tracking", Yan et al., ICCV 2021

STARK - Tracking

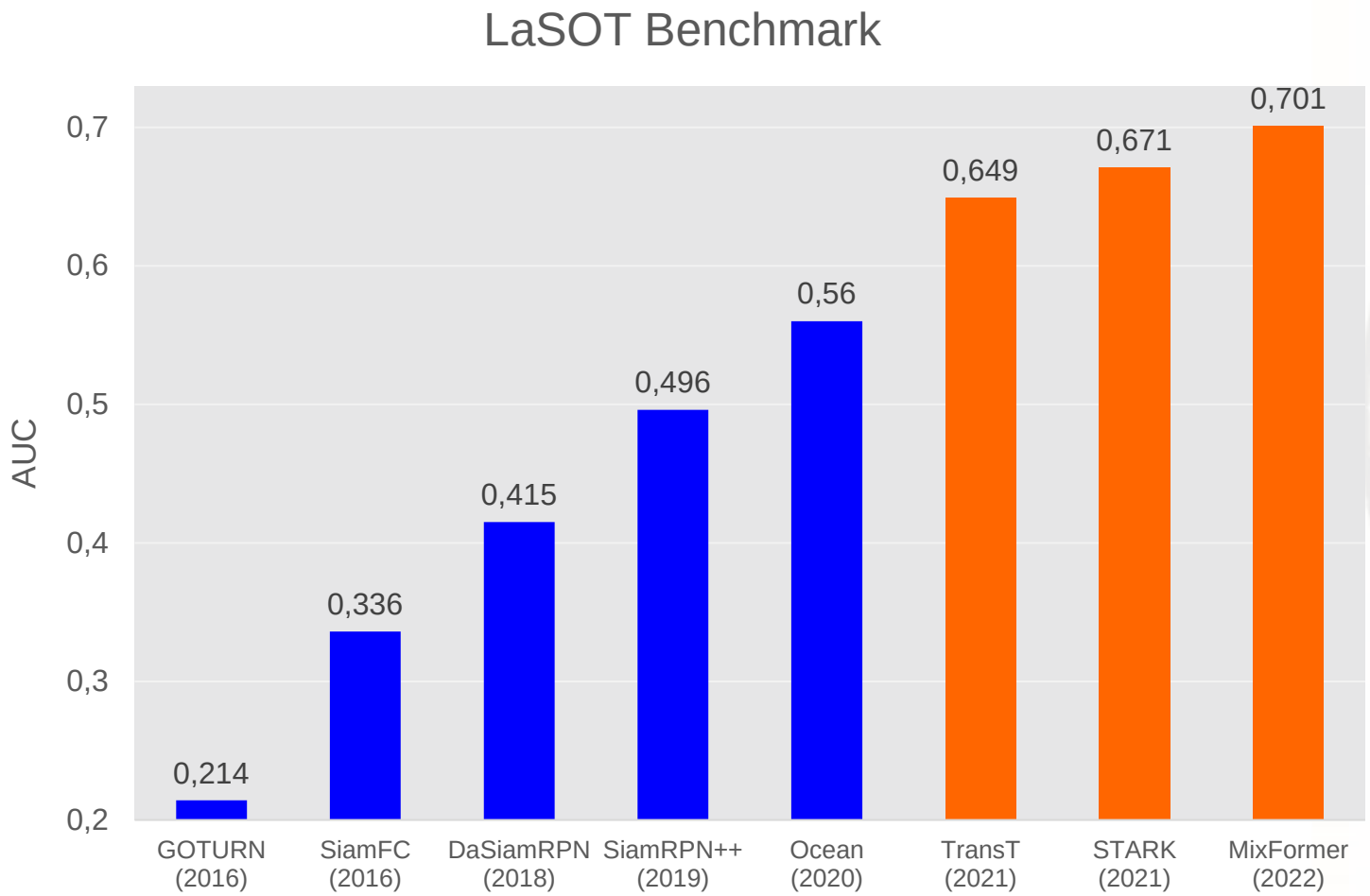
Target matching as attention



"Learning Spatio-Temporal Transformer for Visual Tracking", Yan et al., ICCV 2021

STARK

Transformer-based Trackers



Offline/Online Trackers

MDNet – Tracking – Online Training

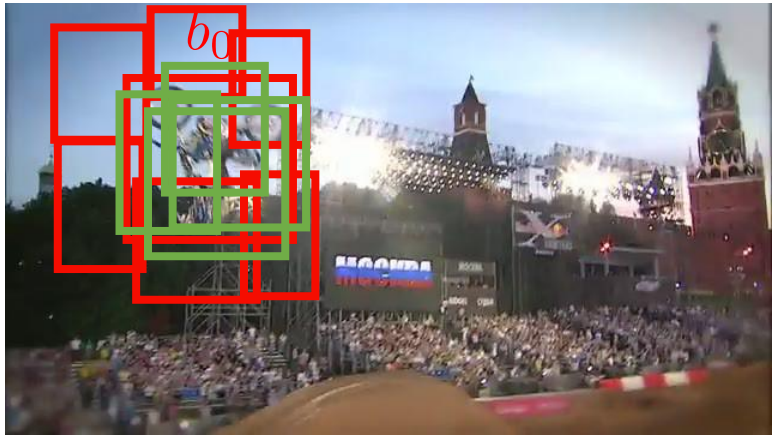
$t = 0$

$\vdots$

"Learning Multi-Domain Convolutional Neural Networks for Visual Tracking", Nam et al., CVPR 2016

MDNet – Tracking – Online Training

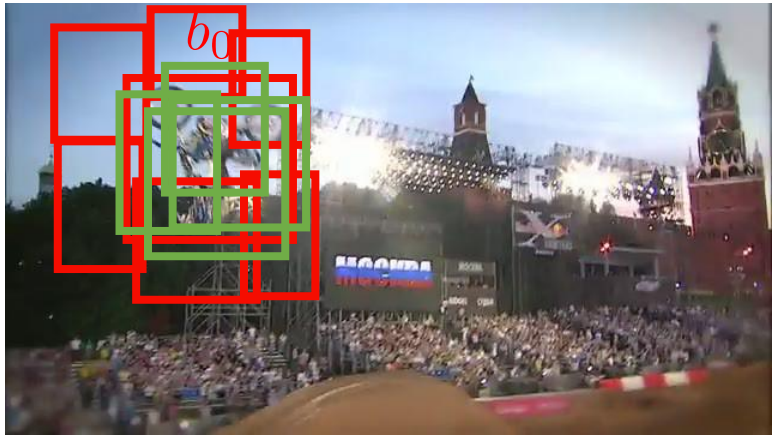
$t = 0$



"**Learning Multi-Domain Convolutional Neural Networks for Visual Tracking**", Nam et al., CVPR 2016

MDNet – Tracking – Online Training

$t = 0$

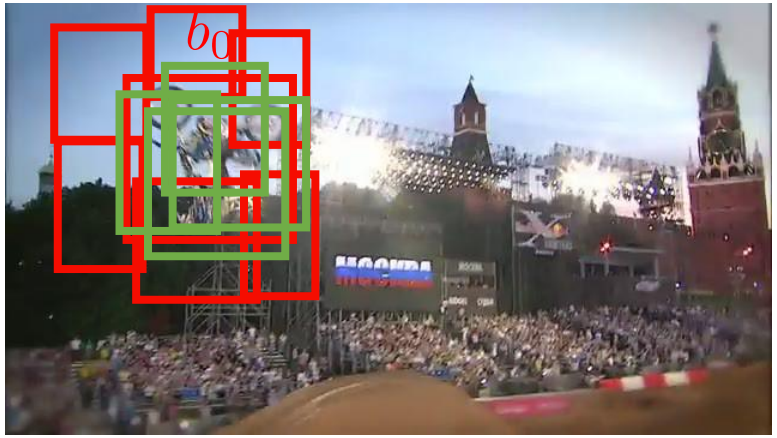


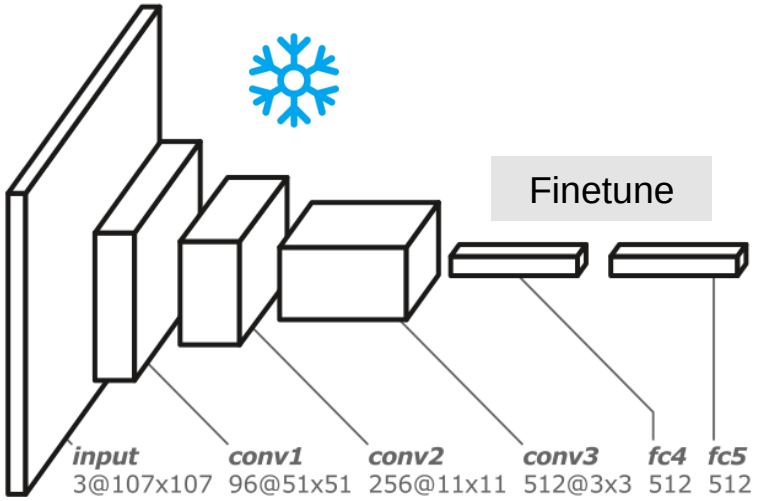
$0, \Delta b_0^{(1)}$
$0, \Delta b_0^{(2)}$
$1, \Delta b_0^{(3)}$
$1, \Delta b_0^{(4)}$
$0, \Delta b_0^{(5)}$

"Learning Multi-Domain Convolutional Neural Networks for Visual Tracking", Nam et al., CVPR 2016

MDNet – Tracking – Online Training

$t = 0$



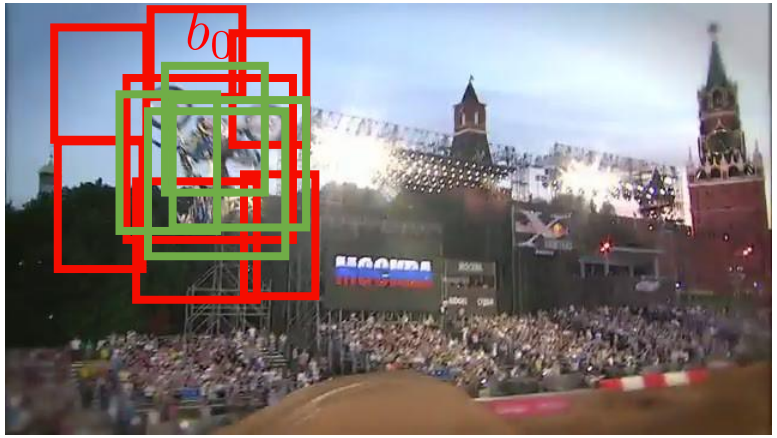


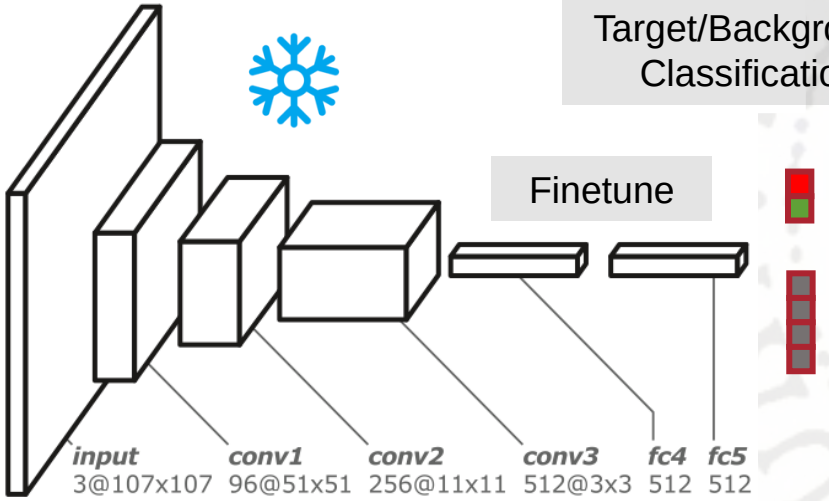
$0, \Delta b_0^{(1)}$
$0, \Delta b_0^{(2)}$
$1, \Delta b_0^{(3)}$
$1, \Delta b_0^{(4)}$
$0, \Delta b_0^{(5)}$

"Learning Multi-Domain Convolutional Neural Networks for Visual Tracking", Nam et al., CVPR 2016

MDNet – Tracking – Online Training

$t = 0$





$\mathcal{L}_{crossentropy}$

Target/Background Classification

Finetune

$0, \Delta b_0^{(1)}$
$0, \Delta b_0^{(2)}$
$1, \Delta b_0^{(3)}$
$1, \Delta b_0^{(4)}$
$0, \Delta b_0^{(5)}$

Box refinement

$\mathcal{L}_{mse}$

"Learning Multi-Domain Convolutional Neural Networks for Visual Tracking", Nam et al., CVPR 2016

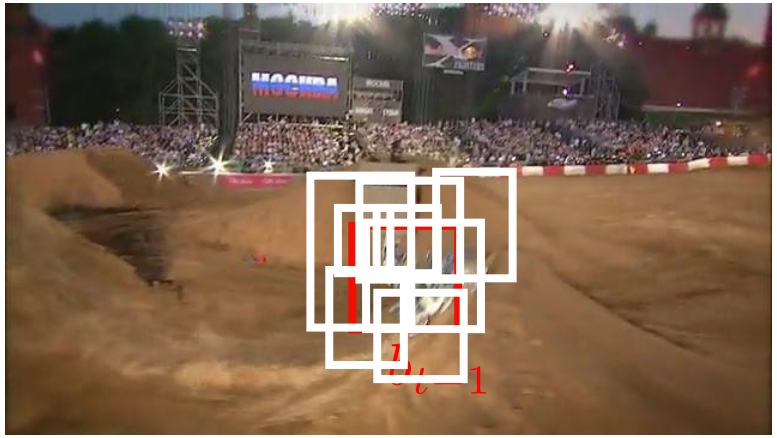
MDNet – Tracking

$t > 0$

"Learning Multi-Domain Convolutional Neural Networks for Visual Tracking", Nam et al., CVPR 2016

MDNet – Tracking

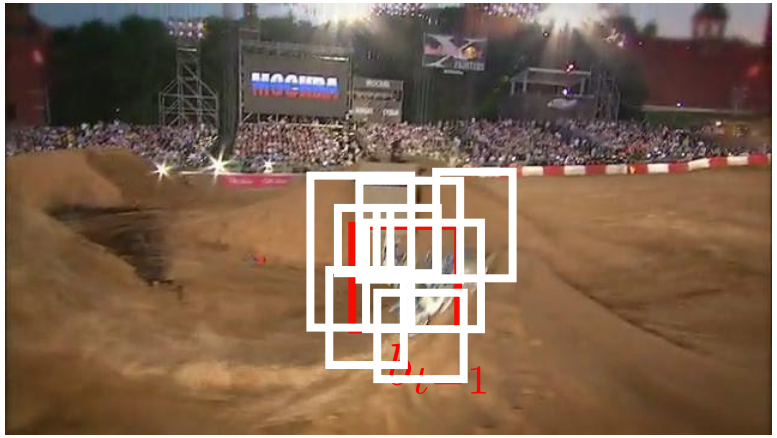
$t > 0$

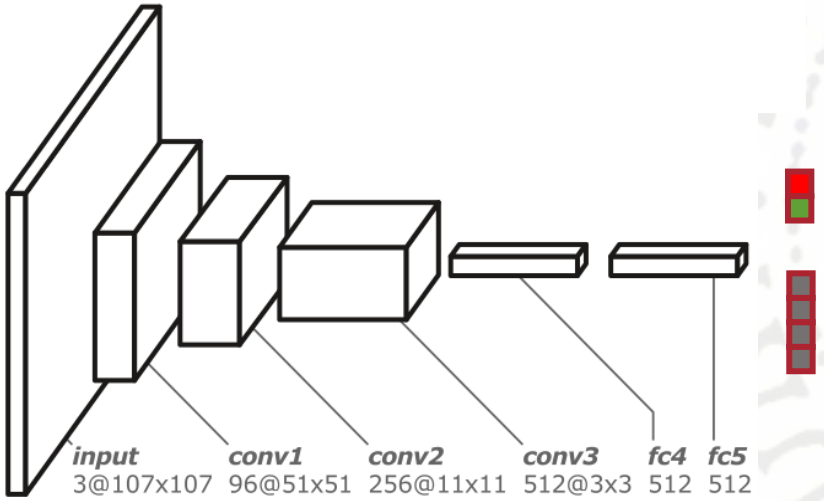


"Learning Multi-Domain Convolutional Neural Networks for Visual Tracking", Nam et al., CVPR 2016

MDNet – Tracking

$t > 0$

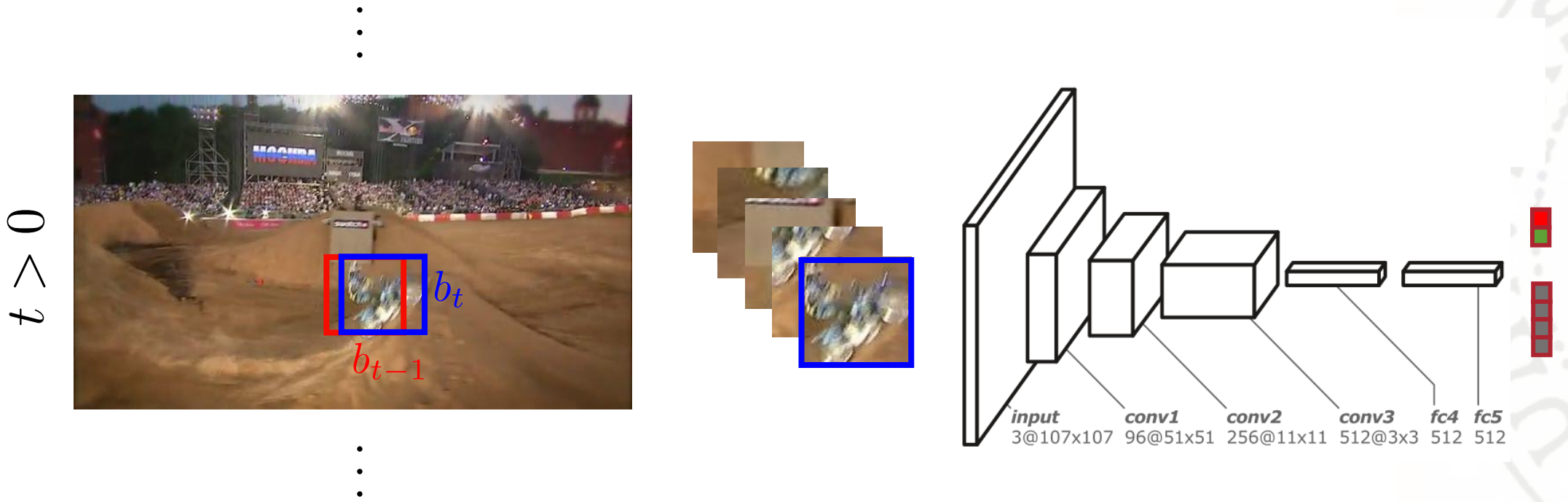




$0.1, \Delta b_t^{(1)}$
$0.2, \Delta b_t^{(2)}$
$0.0, \Delta b_t^{(3)}$
$0.3, \Delta b_t^{(4)}$
$0.9, \Delta b_t^{(5)}$

"Learning Multi-Domain Convolutional Neural Networks for Visual Tracking", Nam et al., CVPR 2016

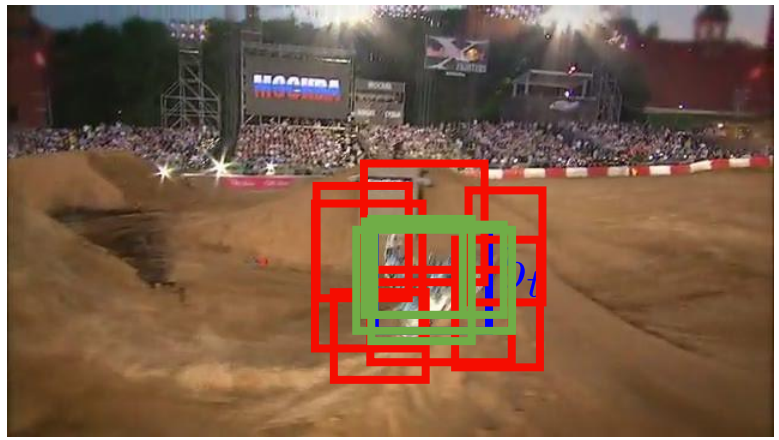
MDNet – Tracking



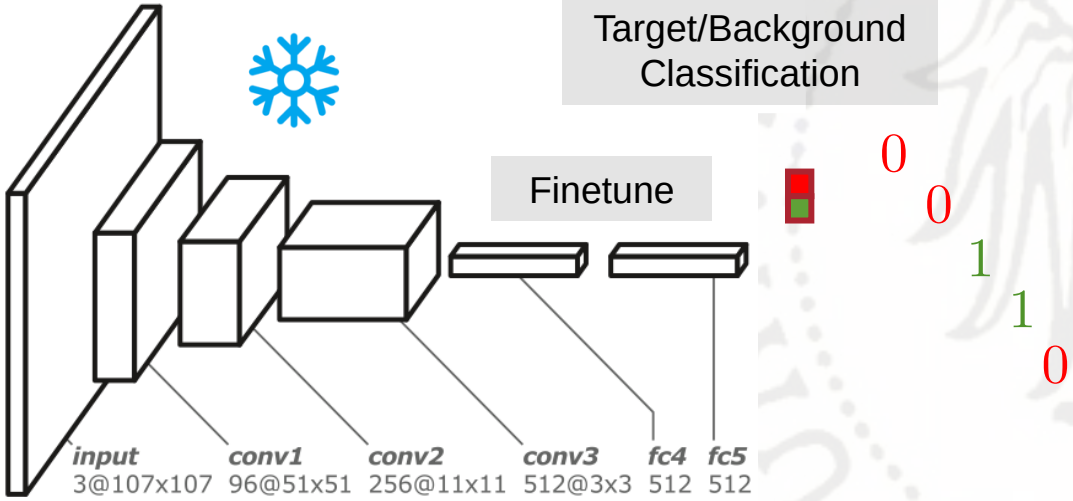
"Learning Multi-Domain Convolutional Neural Networks for Visual Tracking", Nam et al., CVPR 2016

MDNet – Tracking – Online Training

Tracking-by-Detection with Deep Learning



$t = 0$



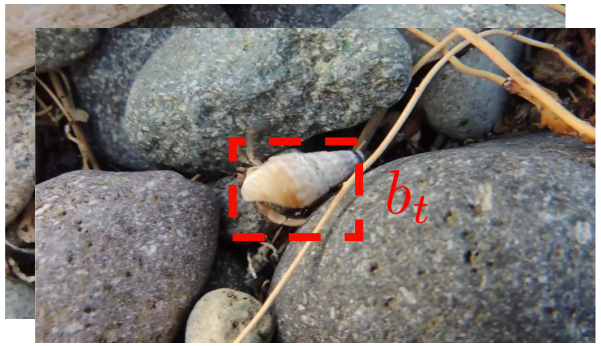
Repeat training as for the first frame

"Learning Multi-Domain Convolutional Neural Networks for Visual Tracking", Nam et al., CVPR 2016

MDNet – Offline Training

$\mathcal{V}_{i-1}$

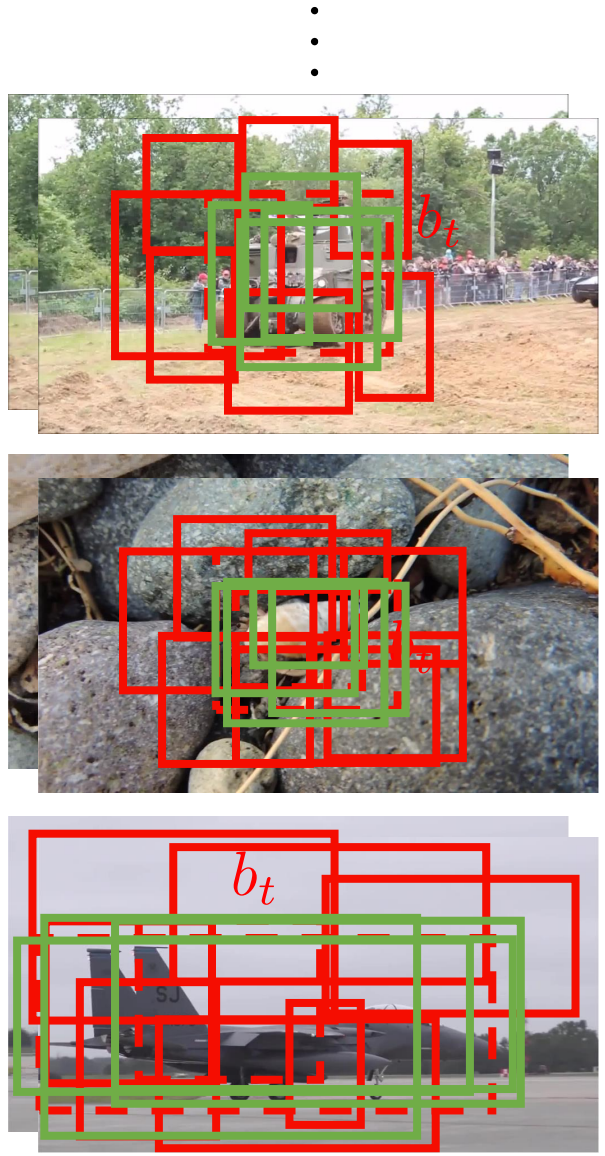
$\mathcal{V}_i$



$\mathcal{V}_{i+1}$

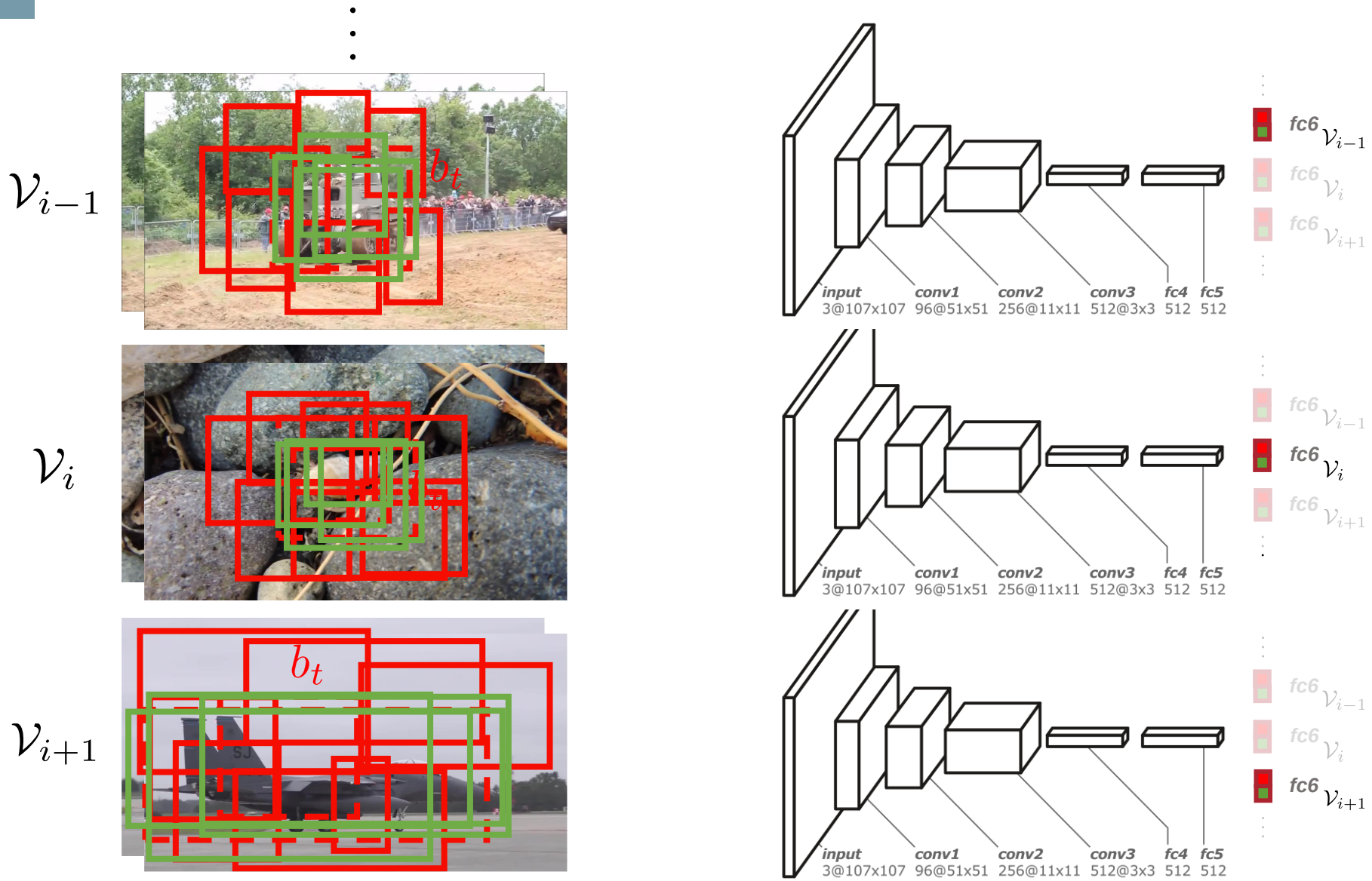
"Learning Multi-Domain Convolutional Neural Networks for Visual Tracking", Nam et al., CVPR 2016

MDNet – Offline Training



"Learning Multi-Domain Convolutional Neural Networks for Visual Tracking", Nam et al., CVPR 2016

MDNet – Offline Training



$\mathcal{L}_{crossentropy}$

Target/Background Classification

"Learning Multi-Domain Convolutional Neural Networks for Visual Tracking", Nam et al., CVPR 2016

DiMP - Tracking

$t = 0$

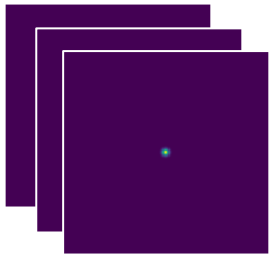
$\vdots$

$t > 0$

$\vdots$

"Learning Discriminative Model Prediction for Tracking", Bhat, Danelljan, et al., ICCV 2019

DiMP - Tracking



$t = 0$

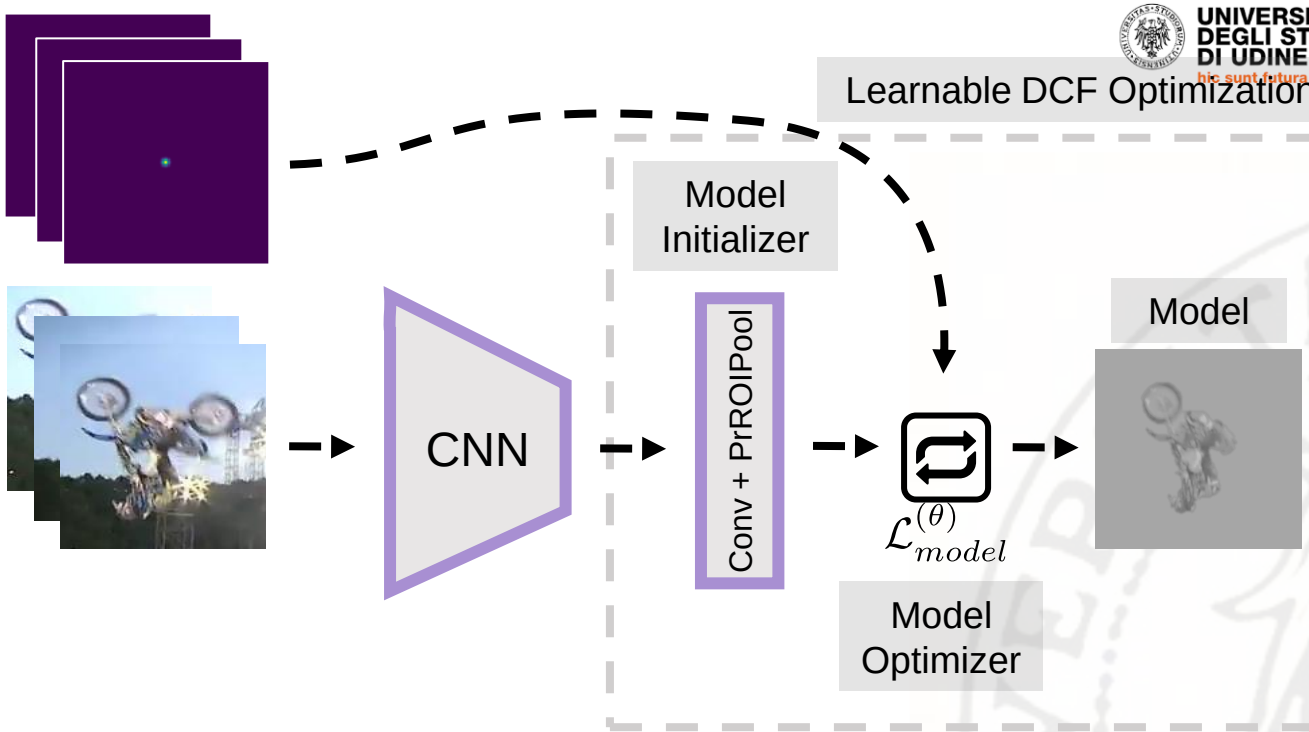
$\vdots$

$t > 0$

$\vdots$

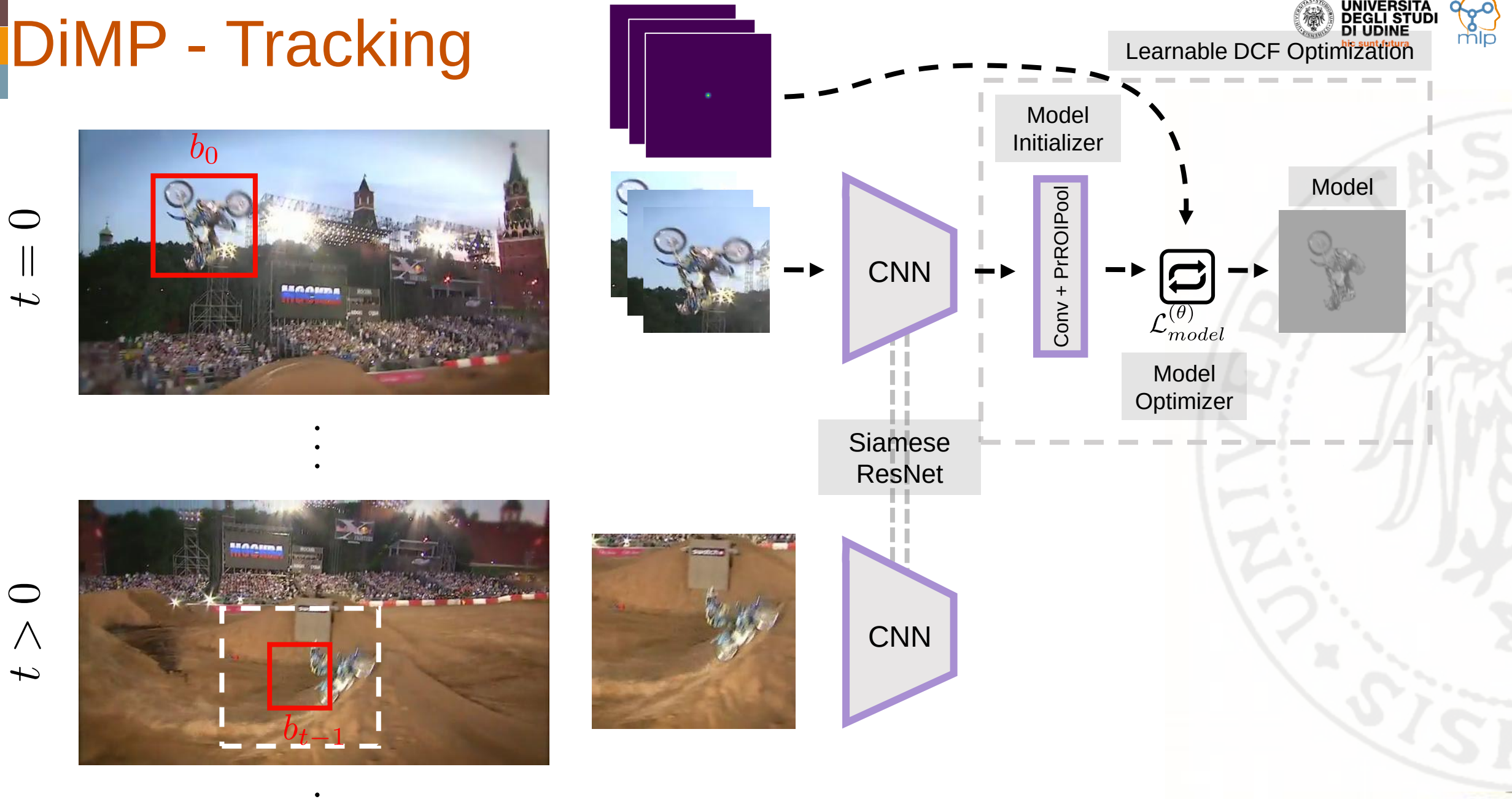
"Learning Discriminative Model Prediction for Tracking", Bhat, Danelljan, et al., ICCV 2019

DiMP - Tracking



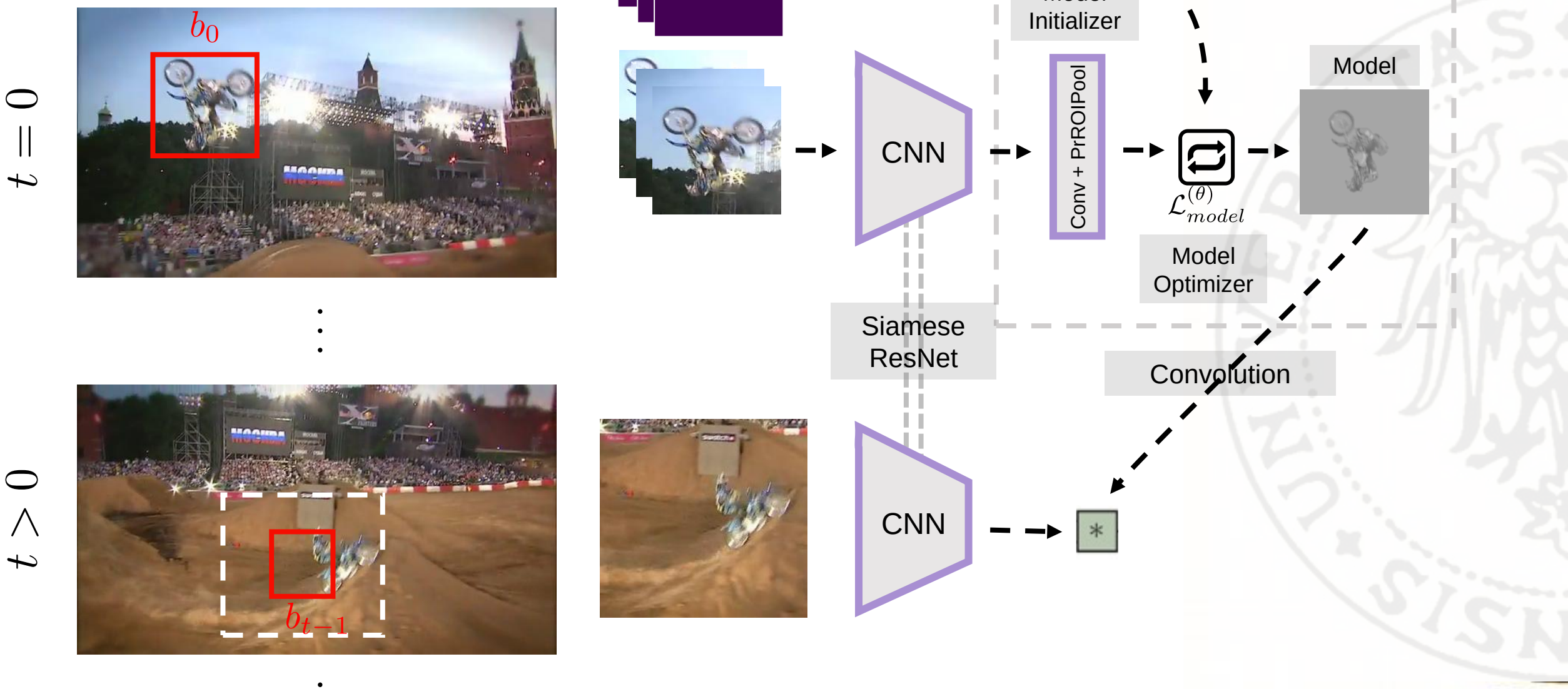
"Learning Discriminative Model Prediction for Tracking", Bhat, Danelljan, et al., ICCV 2019

DiMP - Tracking



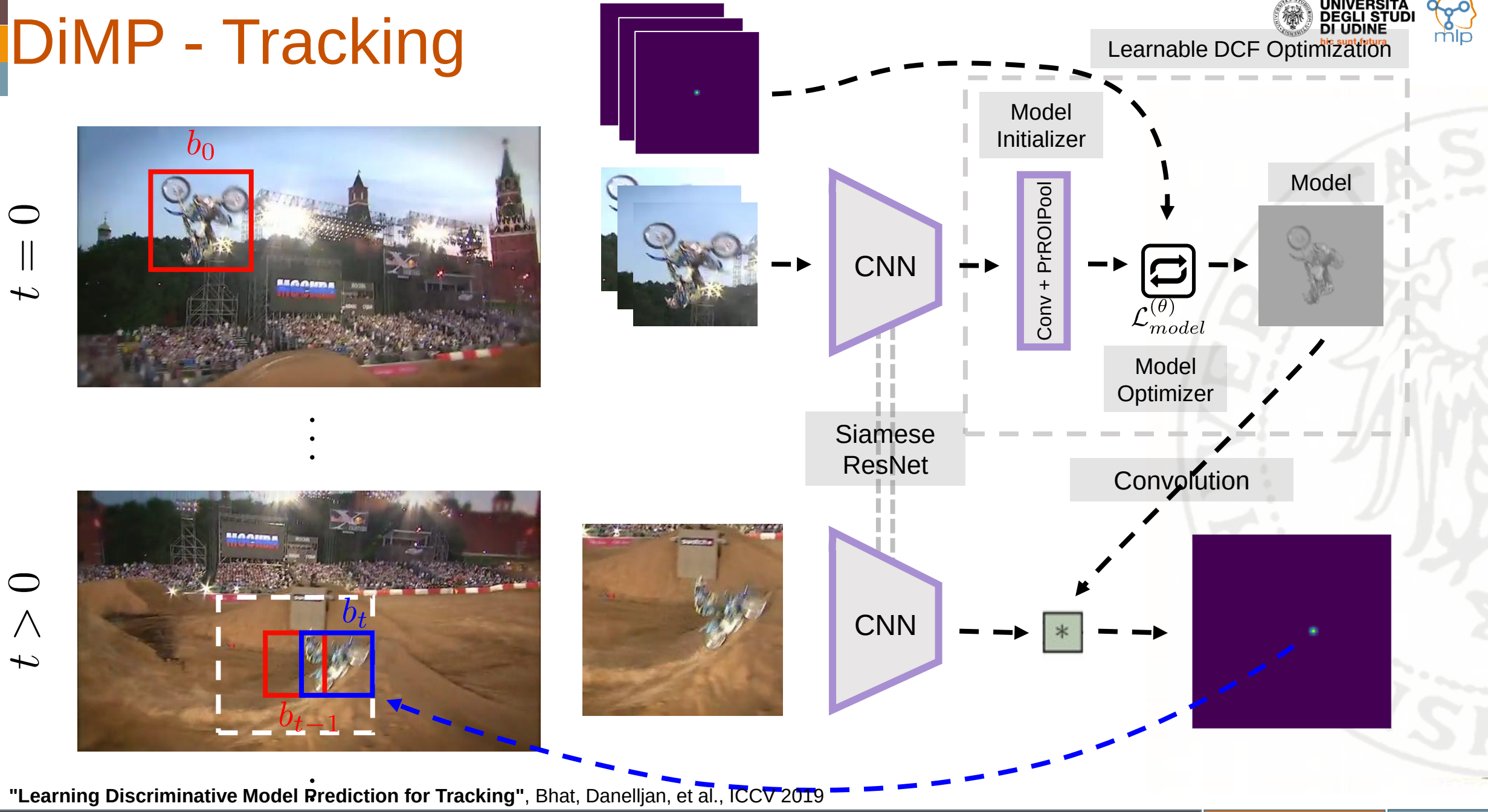
"Learning Discriminative Model Prediction for Tracking", Bhat, Danelljan, et al., ICCV 2019

DiMP - Tracking



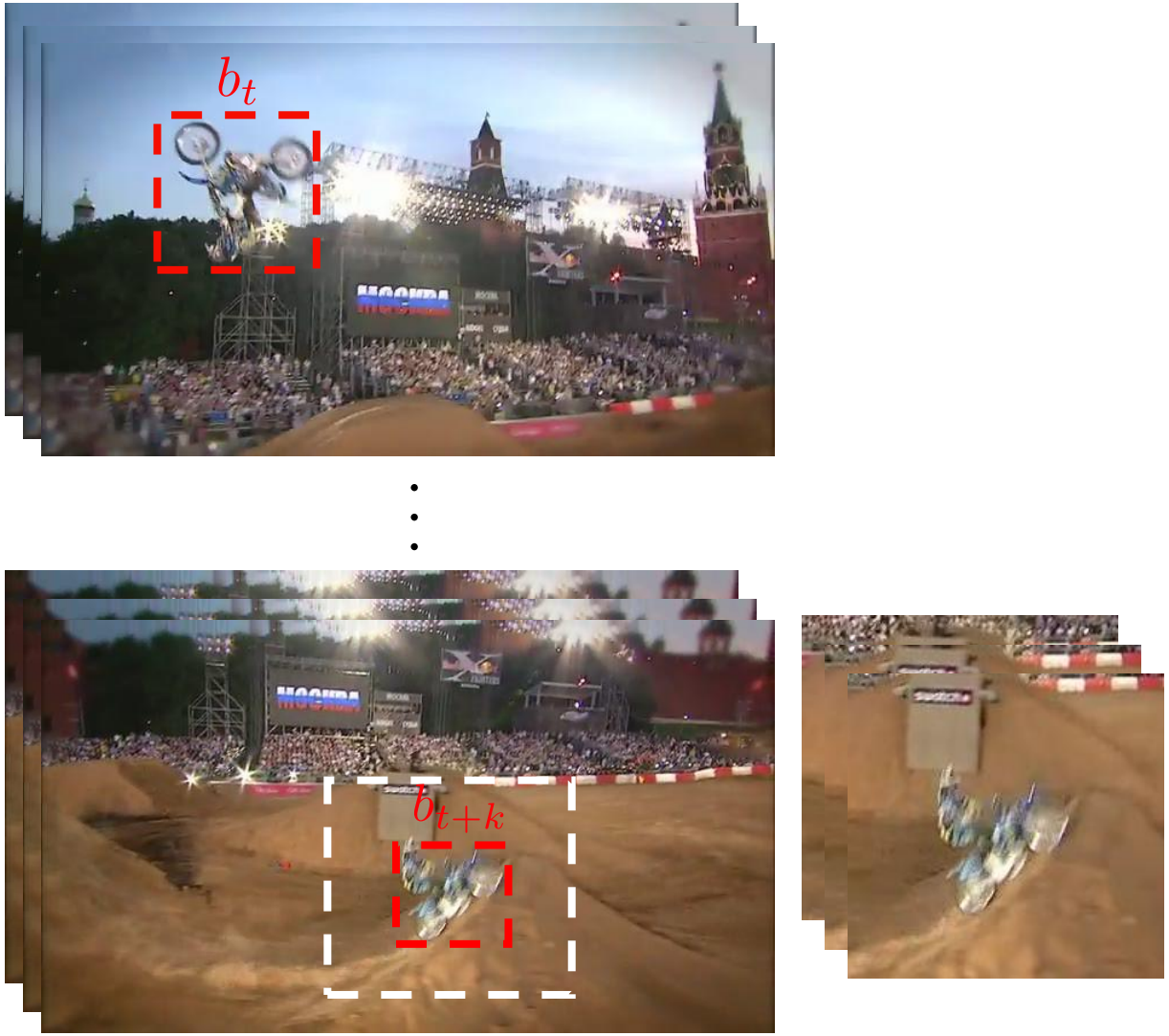
"Learning Discriminative Model Prediction for Tracking", Bhat, Danelljan, et al., ICCV 2019

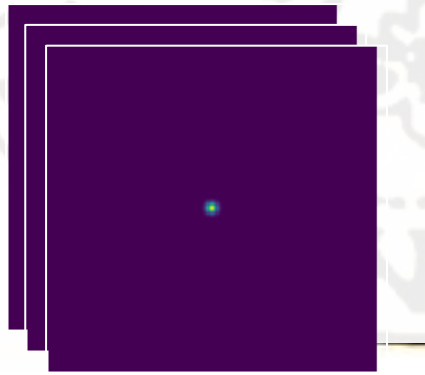
DiMP - Tracking



"Learning Discriminative Model Prediction for Tracking", Bhat, Danelljan, et al., ICCV 2019

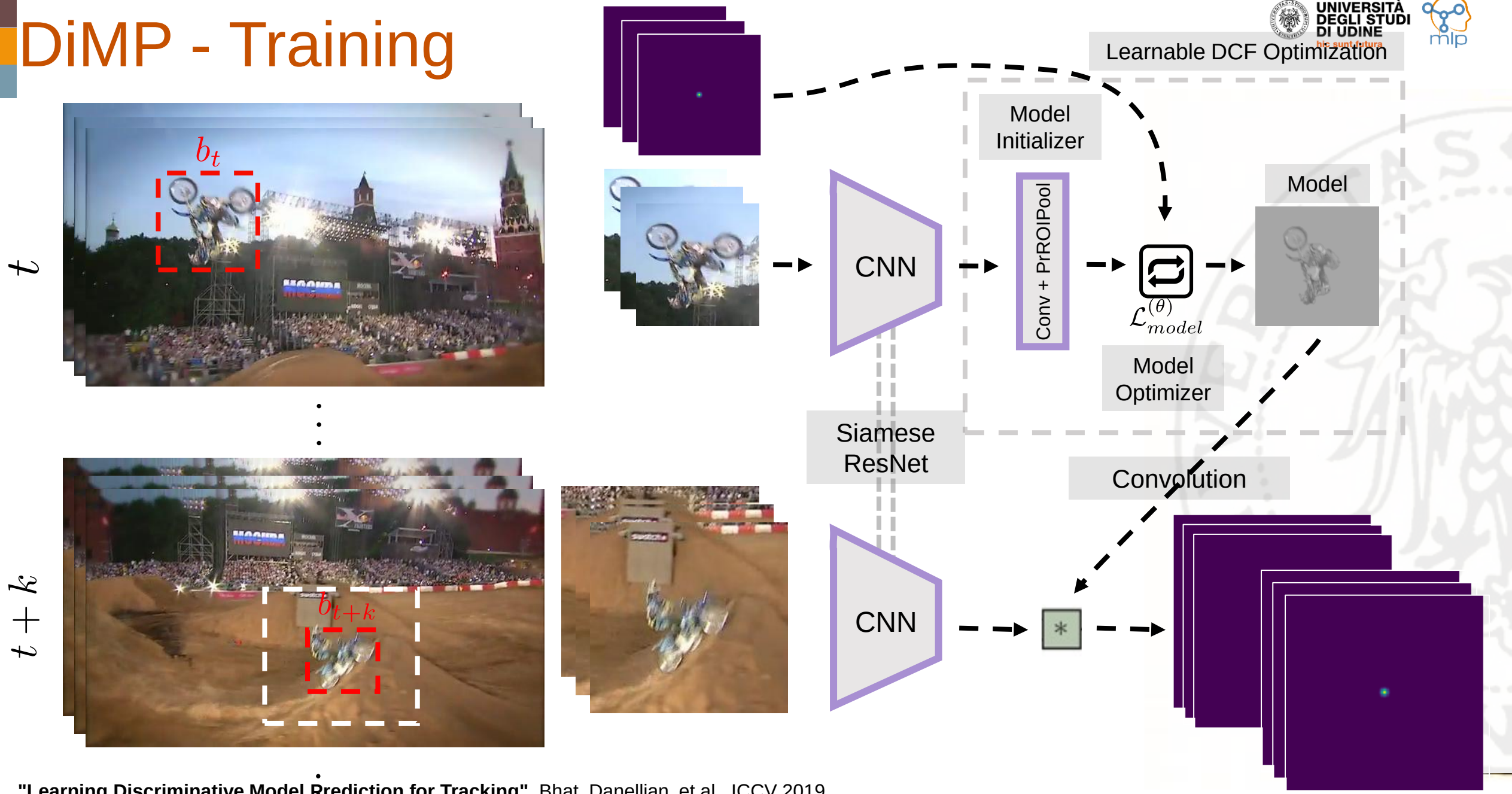
DiMP - Training





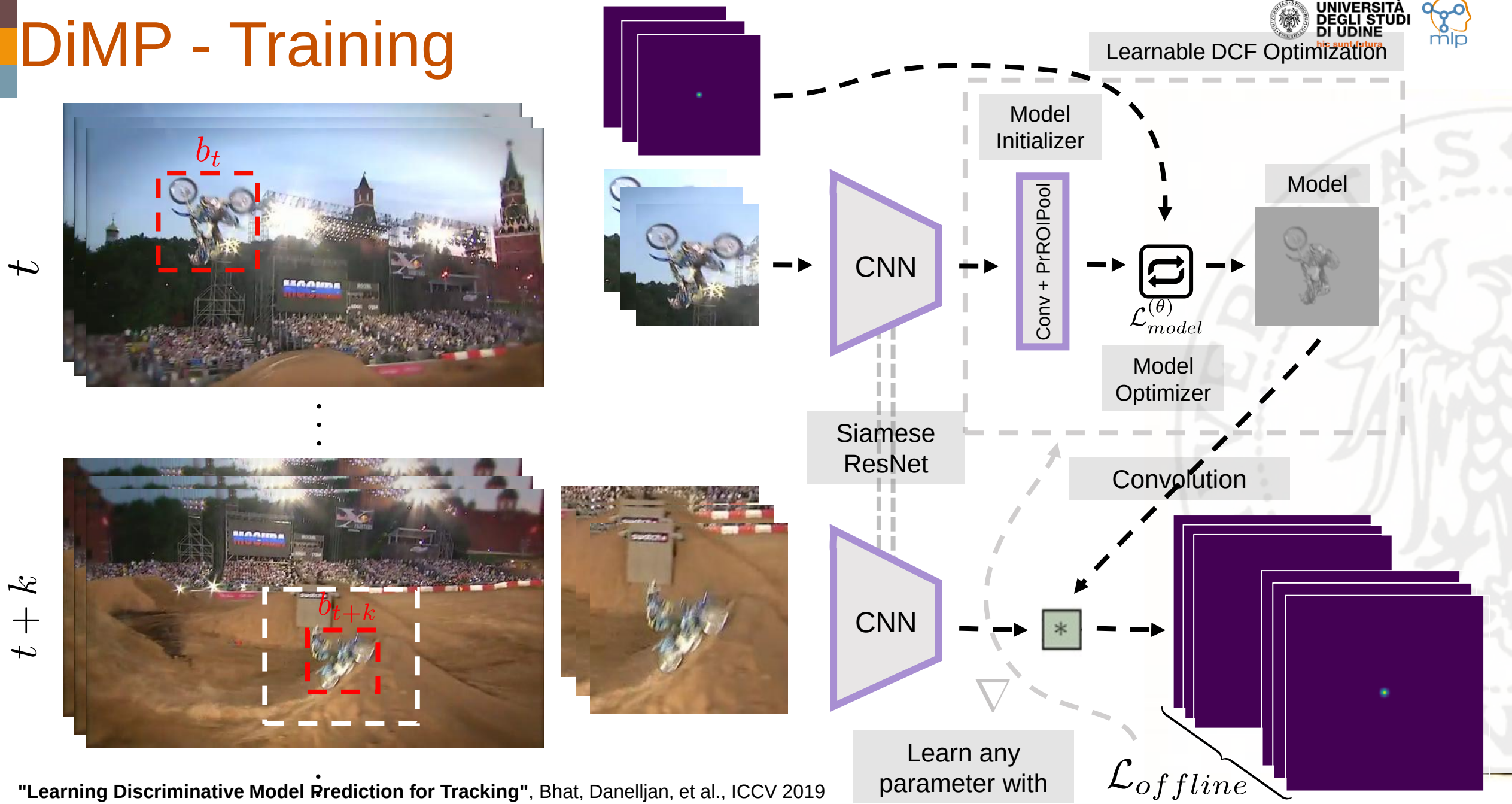
"Learning Discriminative Model Prediction for Tracking", Bhat, Danelljan, et al., ICCV 2019

DiMP - Training



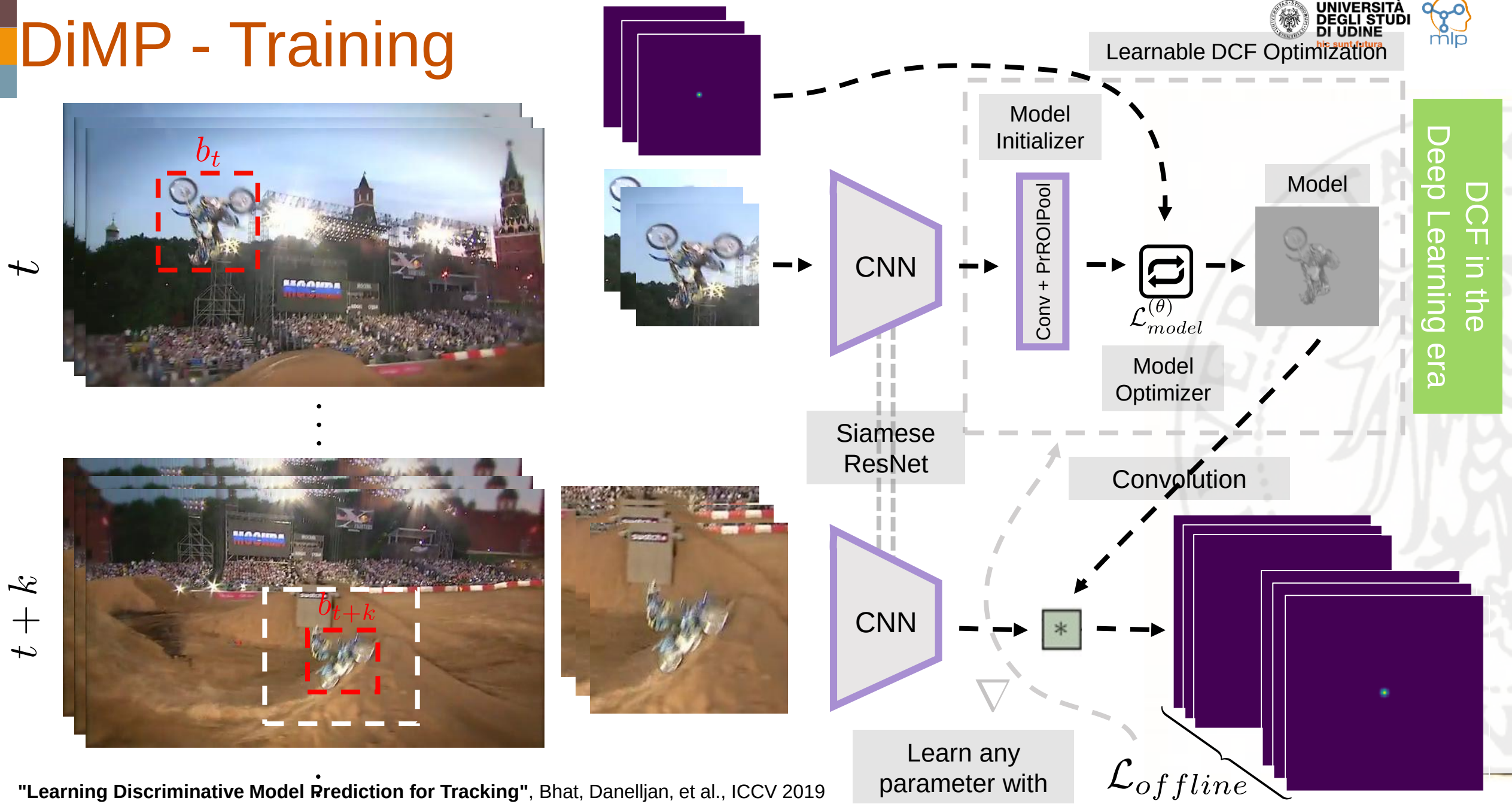
"Learning Discriminative Model Prediction for Tracking", Bhat, Danelljan, et al., ICCV 2019

DiMP - Training



"Learning Discriminative Model Prediction for Tracking", Bhat, Danelljan, et al., ICCV 2019

DiMP - Training



"Learning Discriminative Model Prediction for Tracking", Bhat, Danelljan, et al., ICCV 2019

Conclusions

Conclusions

- The deep learning revolution impacted visual object tracking too
- Transformer architectures are a promising wave
- Short-term and long-term trackers are going to merge
- Deep trackers are still subject to classical machine learning issues

Back to the future

Thank you!