@DocXavi

Master in Computer Vision Barcelona

[http://pagines.uab.cat/mcv/]

Xavier Giró-i-Nieto

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Department of Signal Theory and Communications
Image Processing Group

Module 6

Deep Learning for Video: Object Detection & Segmentation

22nd March 2018

Deep Learning online courses by UPC:

- [MSc course](#) (2017)
- [BSc course](#) (2018)

Next edition Autumn 2018

- [1st edition](#) (2016)
- [2nd edition](#) (2017)
- [3rd edition](#) (2018)

Summer School (late June 2018)

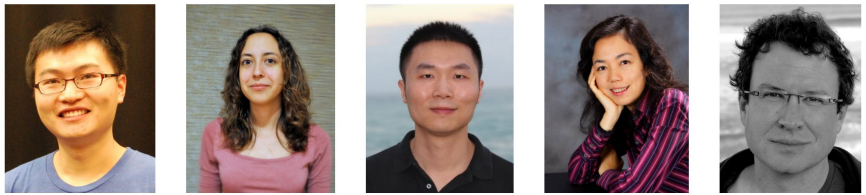
- [1st edition](#) (2017)
- [2nd edition](#) (2018)

Next edition Winter/Spring 2019

Video Object Detection

IMAGENET Large Scale Visual Recognition Challenge (ILSVRC) 2015

Object Detection from Video (VID)



Wei Liu, UNC Chapel Hill
Olga Russakovsky, CMU
Jia Deng, Univ. of Michigan
Fei-Fei Li, Stanford
Alex Berg, UNC Chapel Hill

[ILSVRC 2015 Slides and videos]

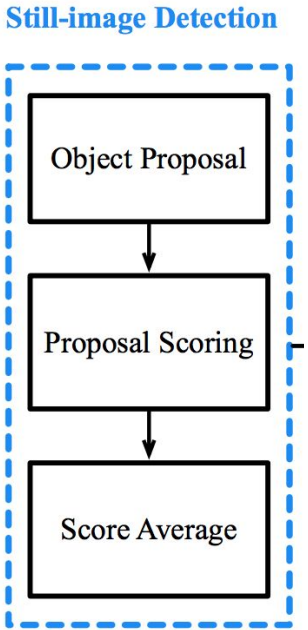
kaggle

[Challenge in Kaggle]

Objects: Object Detection

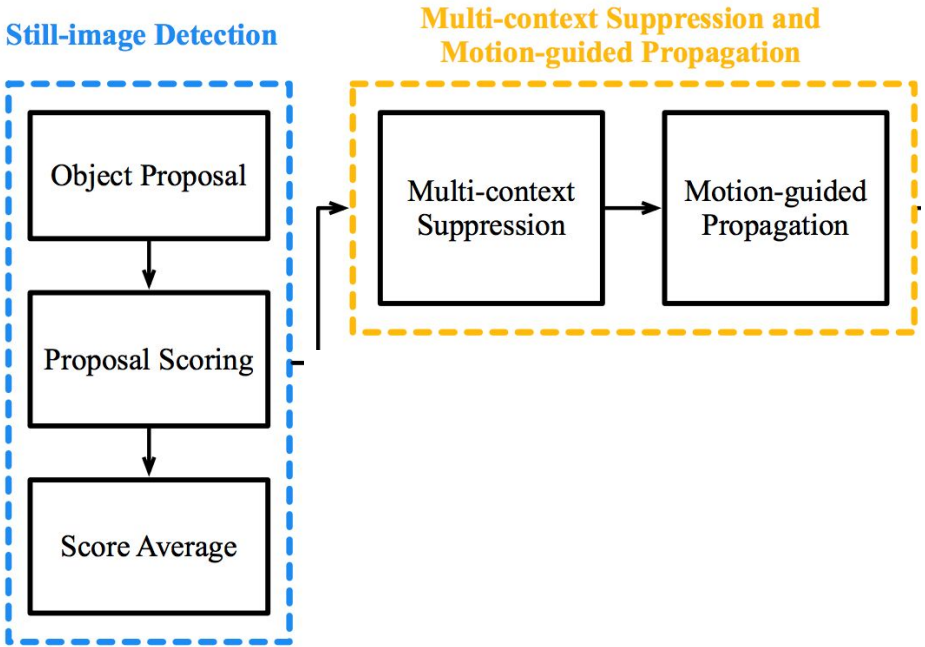
Video Object Detection = Intra-frame Localization + Inter-frame tracking

Object Detection: T-CNN



Kang, Kai, Hongsheng Li, Junjie Yan, Xingyu Zeng, Bin Yang, Tong Xiao, Cong Zhang et al. "T-CNN: Tubelets with convolutional neural networks for object detection from videos." TCSVT 2017. [code]

Object Detection: T-CNN



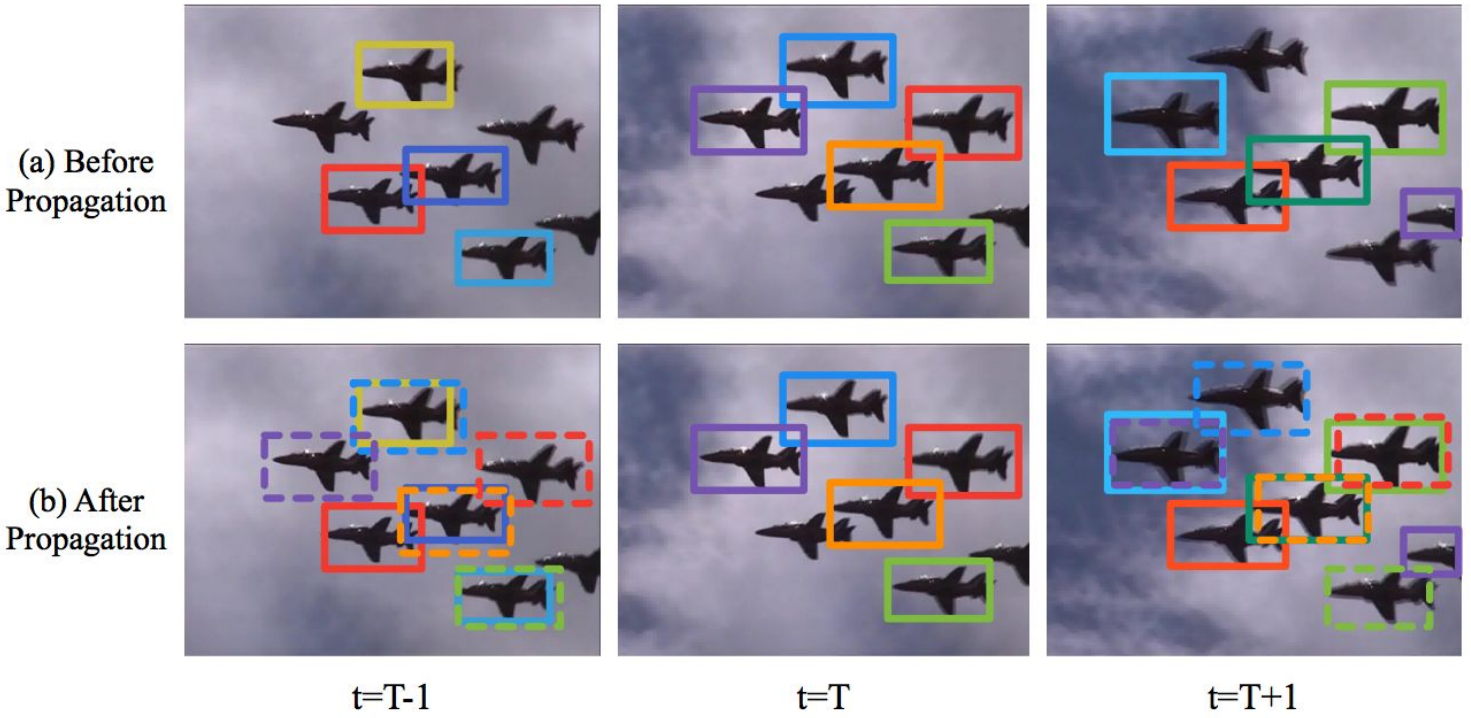
Kang, Kai, Hongsheng Li, Junjie Yan, Xingyu Zeng, Bin Yang, Tong Xiao, Cong Zhang et al. "T-CNN: Tubelets with convolutional neural networks for object detection from videos." TCSVT 2017. [code]

Object Detection: T-CNN

Fig 3. Multi-context Suppression (MCS)

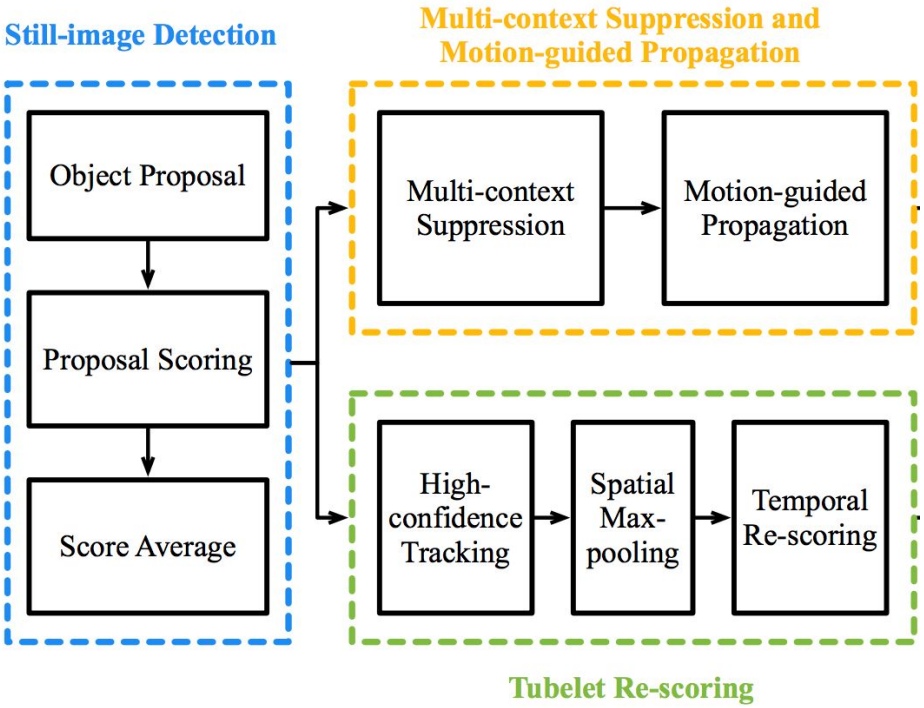
Kang, Kai, Hongsheng Li, Junjie Yan, Xingyu Zeng, Bin Yang, Tong Xiao, Cong Zhang et al. "T-CNN: Tubelets with convolutional neural networks for object detection from videos." TCSVT 2017. [code]

Object Detection: T-CNN



Kang, Kai, Hongsheng Li, Junjie Yan, Xingyu Zeng, Bin Yang, Tong Xiao, Cong Zhang et al. "T-CNN: Tubelets with convolutional neural networks for object detection from videos." TCSVT 2017. [code]

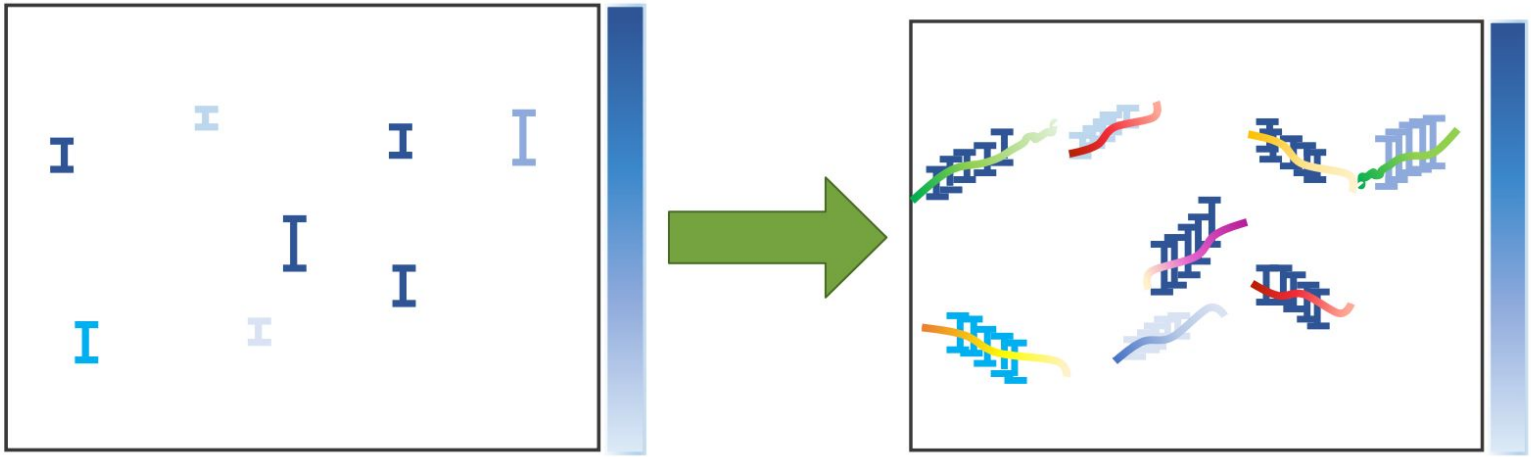
Object Detection: T-CNN



Kang, Kai, Hongsheng Li, Junjie Yan, Xingyu Zeng, Bin Yang, Tong Xiao, Cong Zhang et al. "T-CNN: Tubelets with convolutional neural networks for object detection from videos." TCSVT 2017. [code]

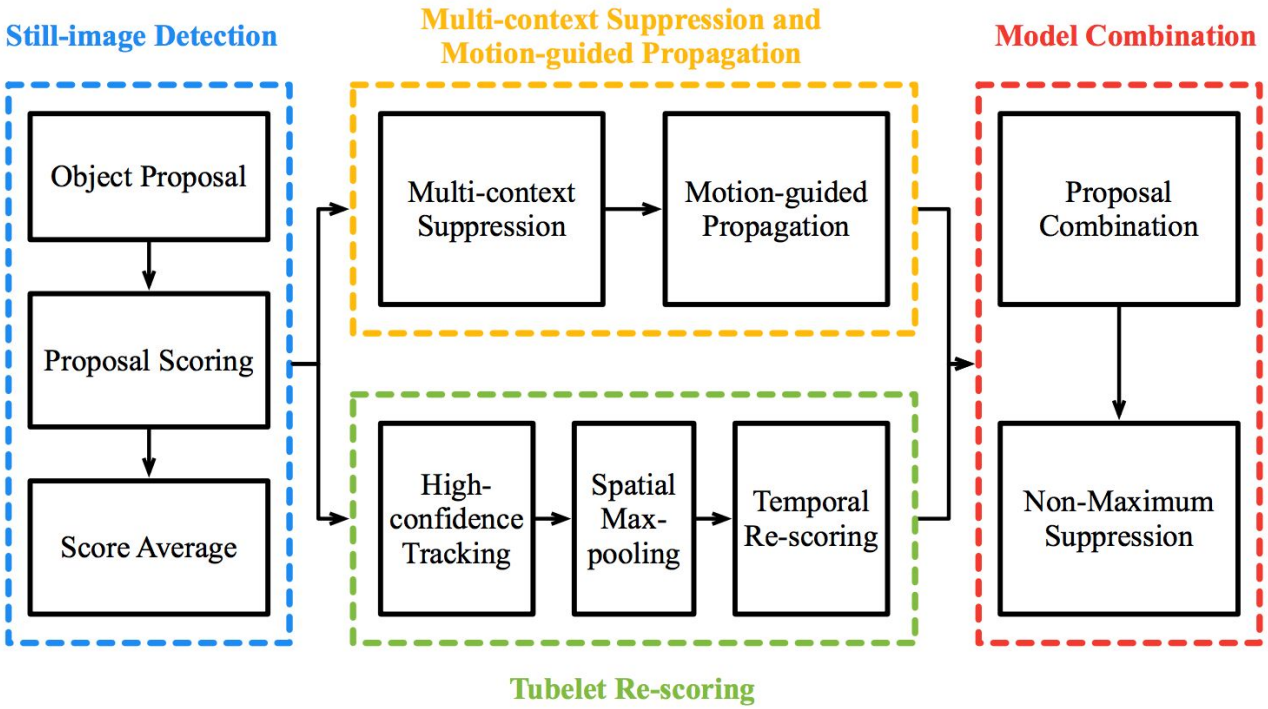
Object Detection: T-CNN

Long-term temporal consistency is obtained by running a tracking algorithm between short-term tubelets.



Kang, Kai, Hongsheng Li, Junjie Yan, Xingyu Zeng, Bin Yang, Tong Xiao, Cong Zhang et al. "T-CNN: Tubelets with convolutional neural networks for object detection from videos." TCSVT 2017. [code]

Object Detection: T-CNN

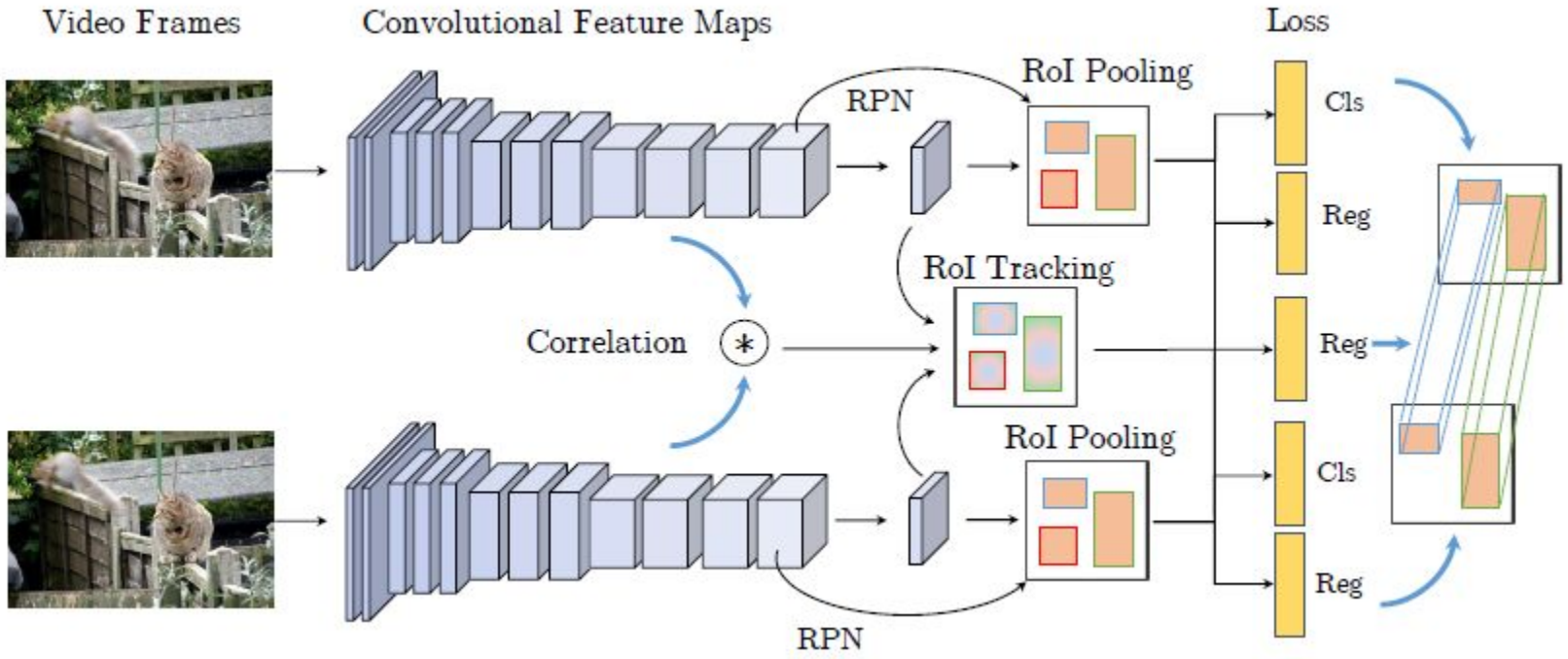


Kang, Kai, Hongsheng Li, Junjie Yan, Xingyu Zeng, Bin Yang, Tong Xiao, Cong Zhang et al. "T-CNN: Tubelets with convolutional neural networks for object detection from videos." TCSVT 2017. [code]

Feichtenhofer, Christoph, Axel Pinz, and Andrew Zisserman. "Detect to track and track to detect." ICCV 2017.

Object Detection: Detect & Track

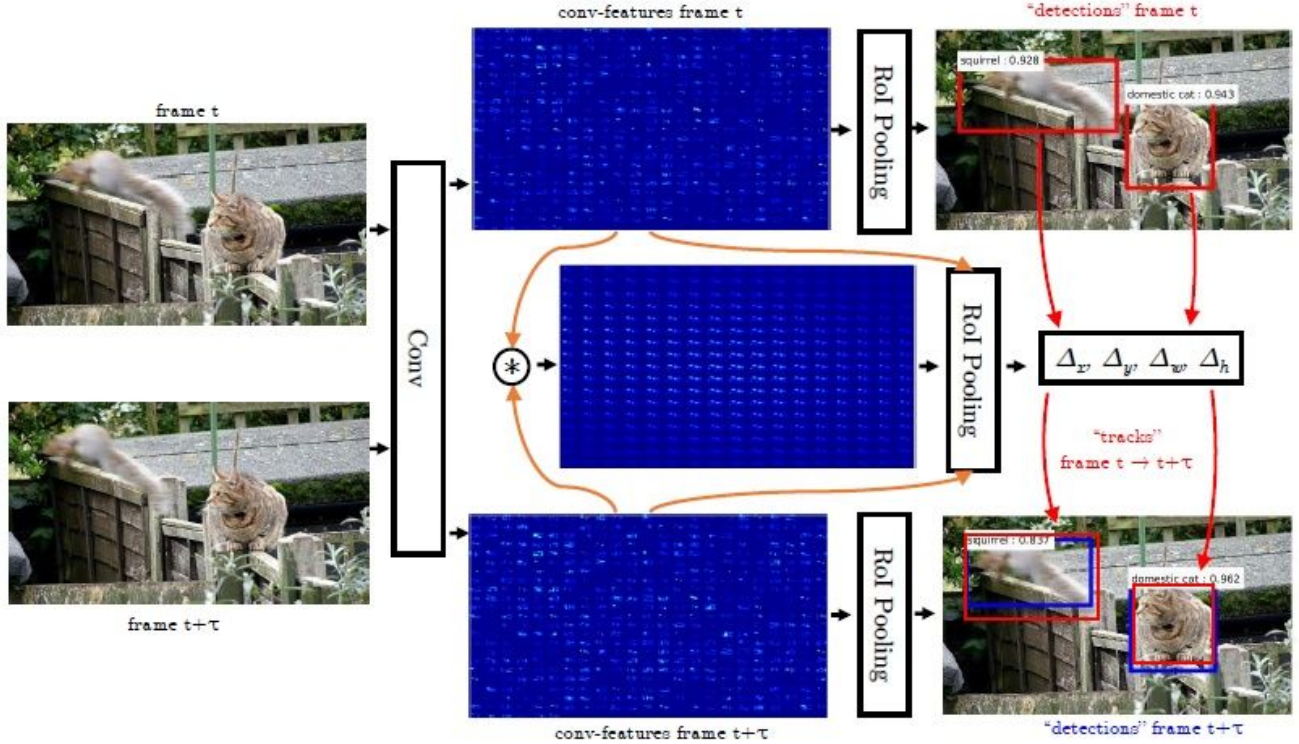
Convolutional Cross-Correlations between feature responses of adjacent frames.



Feichtenhofer, Christoph, Axel Pinz, and Andrew Zisserman. "Detect to track and track to detect." ICCV 2017. [talk]

Object Detection: Detect & Track

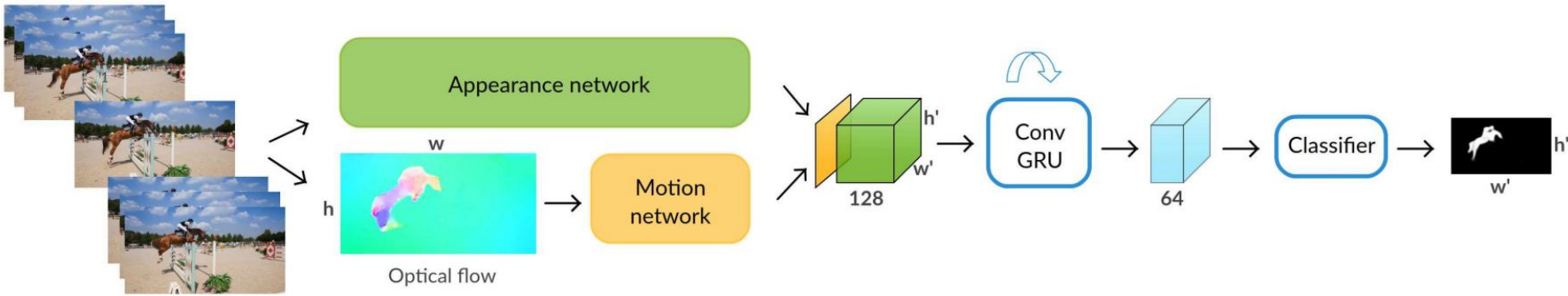
Convolutional Cross-Correlations between feature responses of adjacent frames.



Feichtenhofer, Christoph, Axel Pinz, and Andrew Zisserman. "Detect to track and track to detect." ICCV 2017. [talk]

Segmentation: Two-stream & Conv-GRU

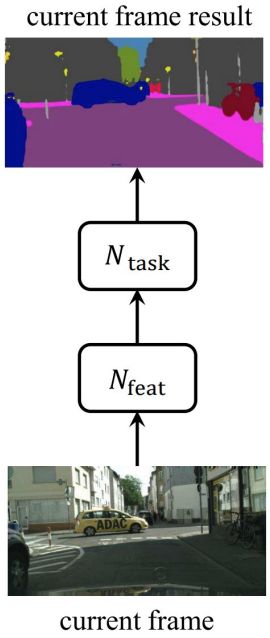
Temporal memory implemented with Conv-GRU.



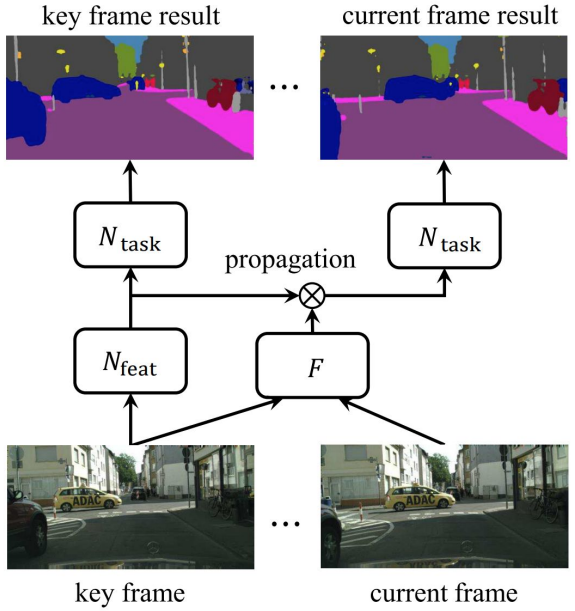
Tokmakov, Pavel, Karteek Alahari, and Cordelia Schmid. "Learning video object segmentation with visual memory." ICCV 2017.

Segmentation: Deep Feature Flow

Deep features from sparse frames are propagated to neighbouring frames by the optical flow estimated by lightweight network.



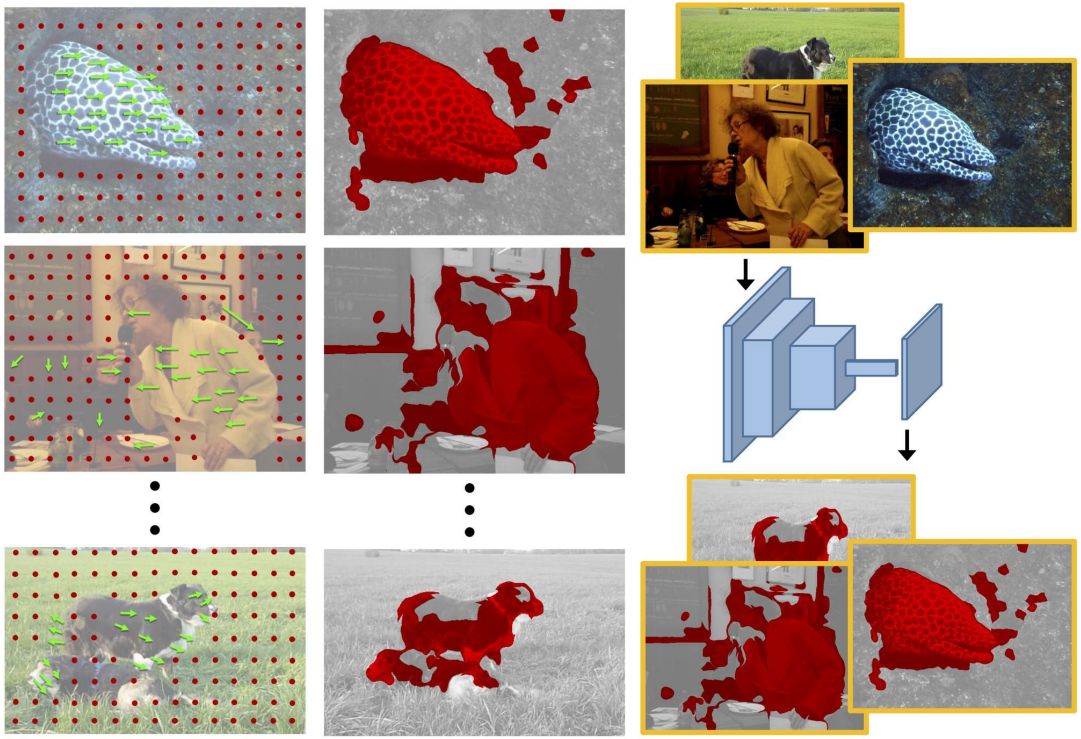
(a) per-frame network



(b) deep feature flow (DFF) network

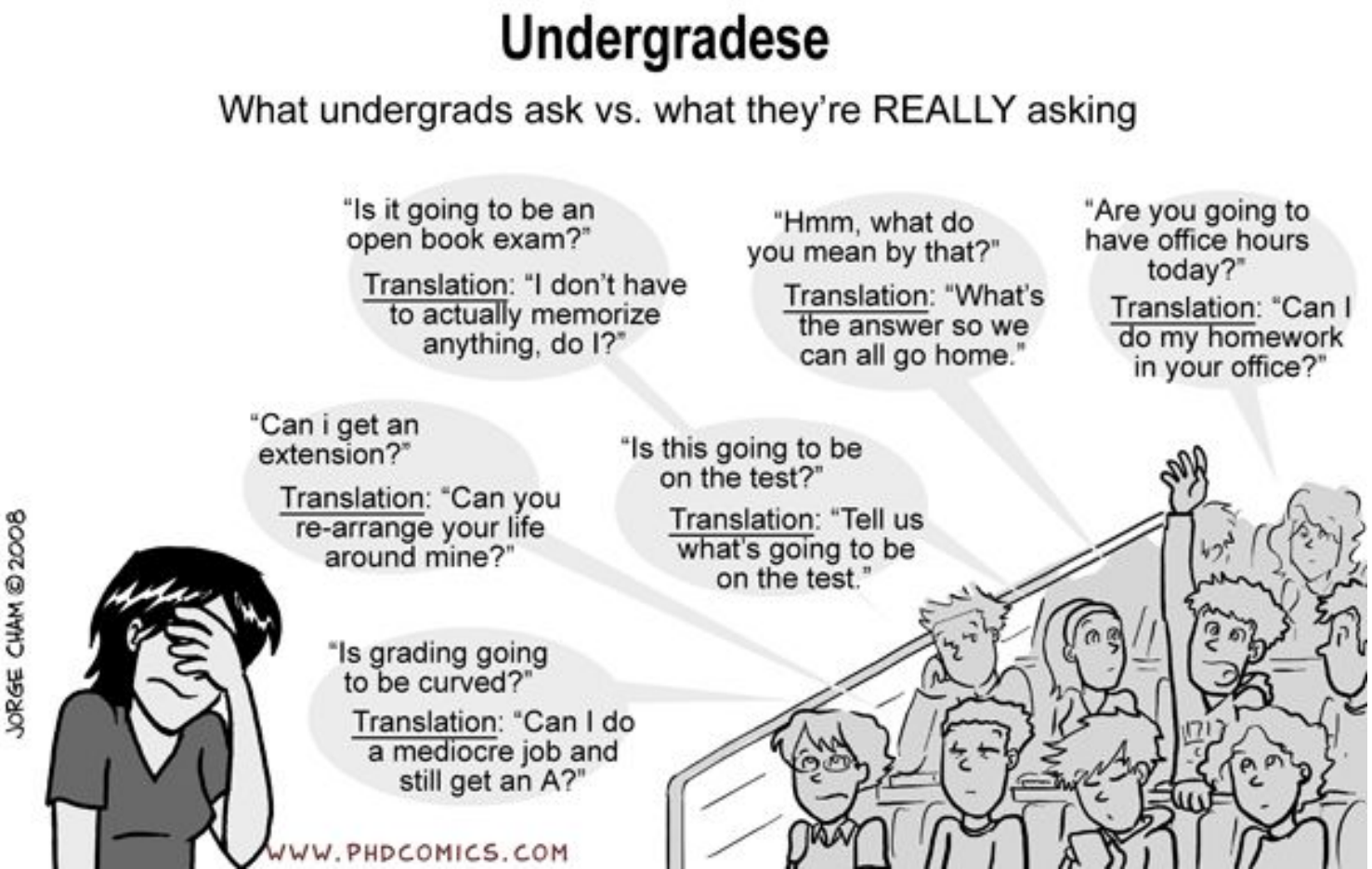
Zhu, Xizhou, Yuwen Xiong, Jifeng Dai, Lu Yuan, and Yichen Wei. "Deep feature flow for video recognition." CVPR 2017. [code]

Segmentation: Unsupervised



Pathak, Deepak, Ross Girshick, Piotr Dollár, Trevor Darrell, and Bharath Hariharan. "Learning features by watching objects move." CVPR 2017

Questions ?


Undergradese
What undergrads ask vs. what they're REALLY asking
"Is it going to be an open book exam?"
Translation: "I don't have to actually memorize anything, do I?"
"Hmm, what do you mean by that?"
Translation: "What's the answer so we can all go home."
"Are you going to have office hours today?"
Translation: "Can I do my homework in your office?"
"Can i get an extension?"
Translation: "Can you re-arrange your life around mine?"
"Is this going to be on the test?"
Translation: "Tell us what's going to be on the test."
"Is grading going to be curved?"
Translation: "Can I do a mediocre job and still get an A?"
JORGE CHAM © 2008
WWW.PHDCOMICS.COM