Adaptive Cascade Threshold Learning from Negative Samples for Deformable Part Models

Khoa Pho Hung Vu Bac Le

VNU HCMC, University of Science,
Ho Chi Minh city, Vietnam

Outline

- Introduction of Cascade DPM
- Existing Methods of Training Thresholds
- Proposed Method
- Experiments
- Conclusion

Cascade DPM

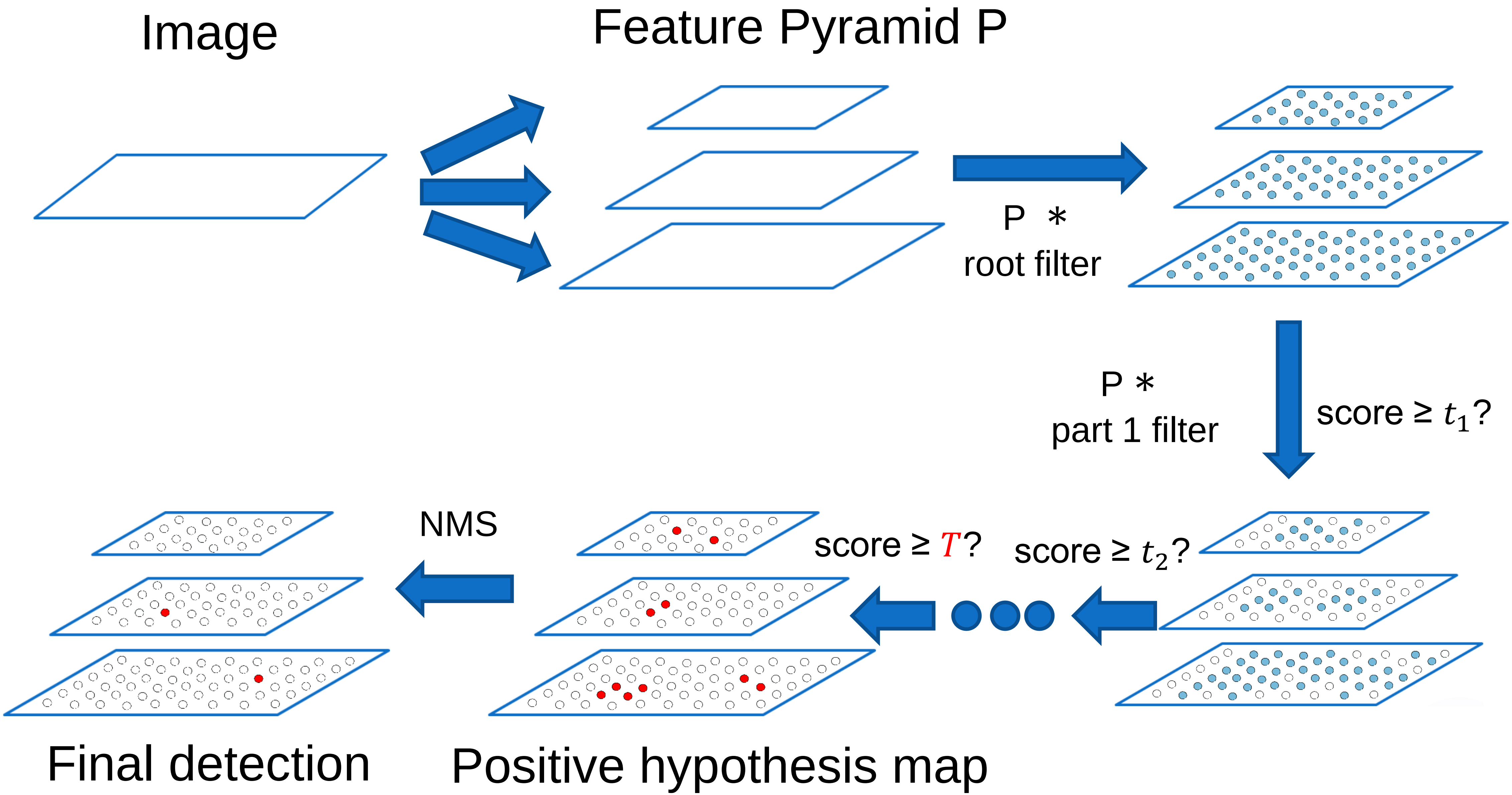


Image
Feature Pyramid P
P ∗
root filter
P ∗
part 1 filter
score ≥ t_1?
score ≥ t_2?
score ≥ T?
NMS
Final detection
Positive hypothesis map

Thresholds in Cascade framework

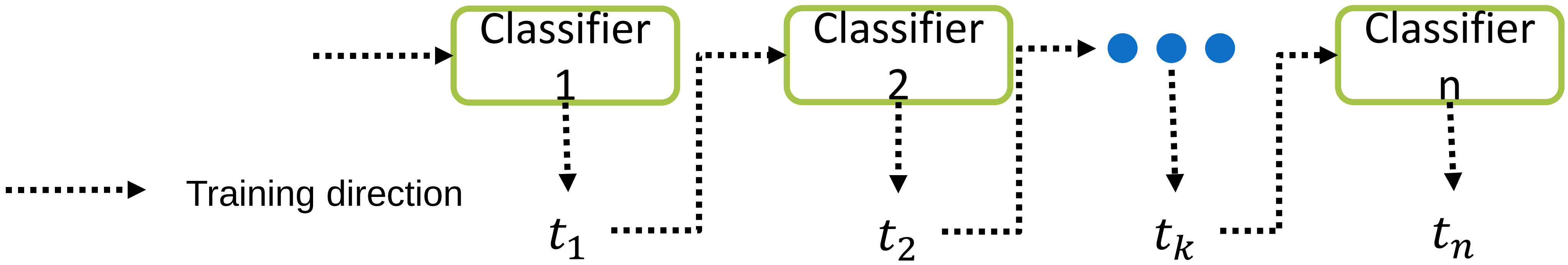
- Good threshold
  - speed up (22x [1])
  - reduce false positive hypothesis rate ([1])

- Higher threshold → remove potential objects
- Lower threshold → harm the speed

→ appropriate threshold → need training data

Existing Methods of Training Threshold

Approach 1: Learn classifers and thresholds together



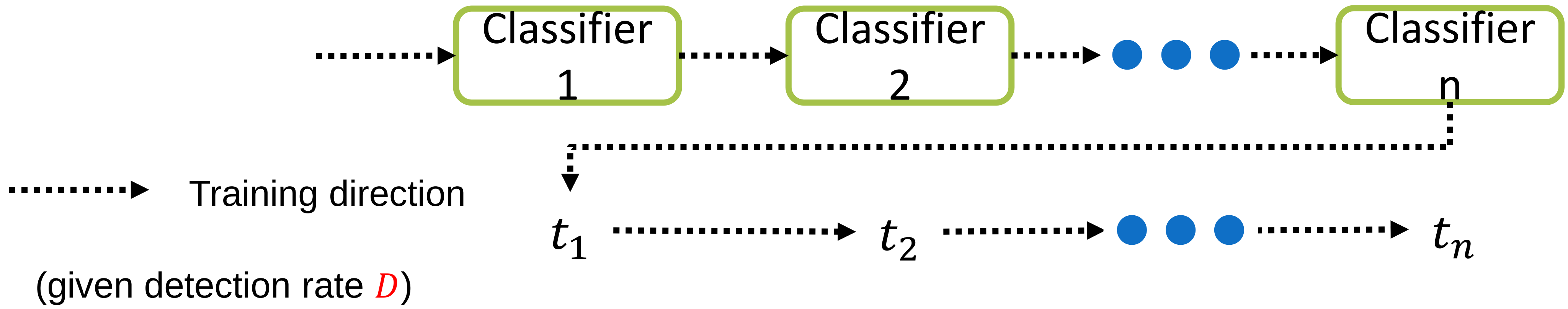
(given detection rate D)

$$D = \prod_{i=1}^{n} d_i$$

- Brubaker at el. [2]: fix number of classifiers
- Viola and Jones [3]: keep training and adding classifiers until achieve D

Existing Methods of Training Threshold

Approach 2: Learn classifers and thresholds separately



- Bourdev [4], Yang[5]: maximum threshold
- Lou [6]: greedy search on ROC curve
- Felzenswab [1]: the minimum positive hypothesis score

Existing Methods of Training Threshold

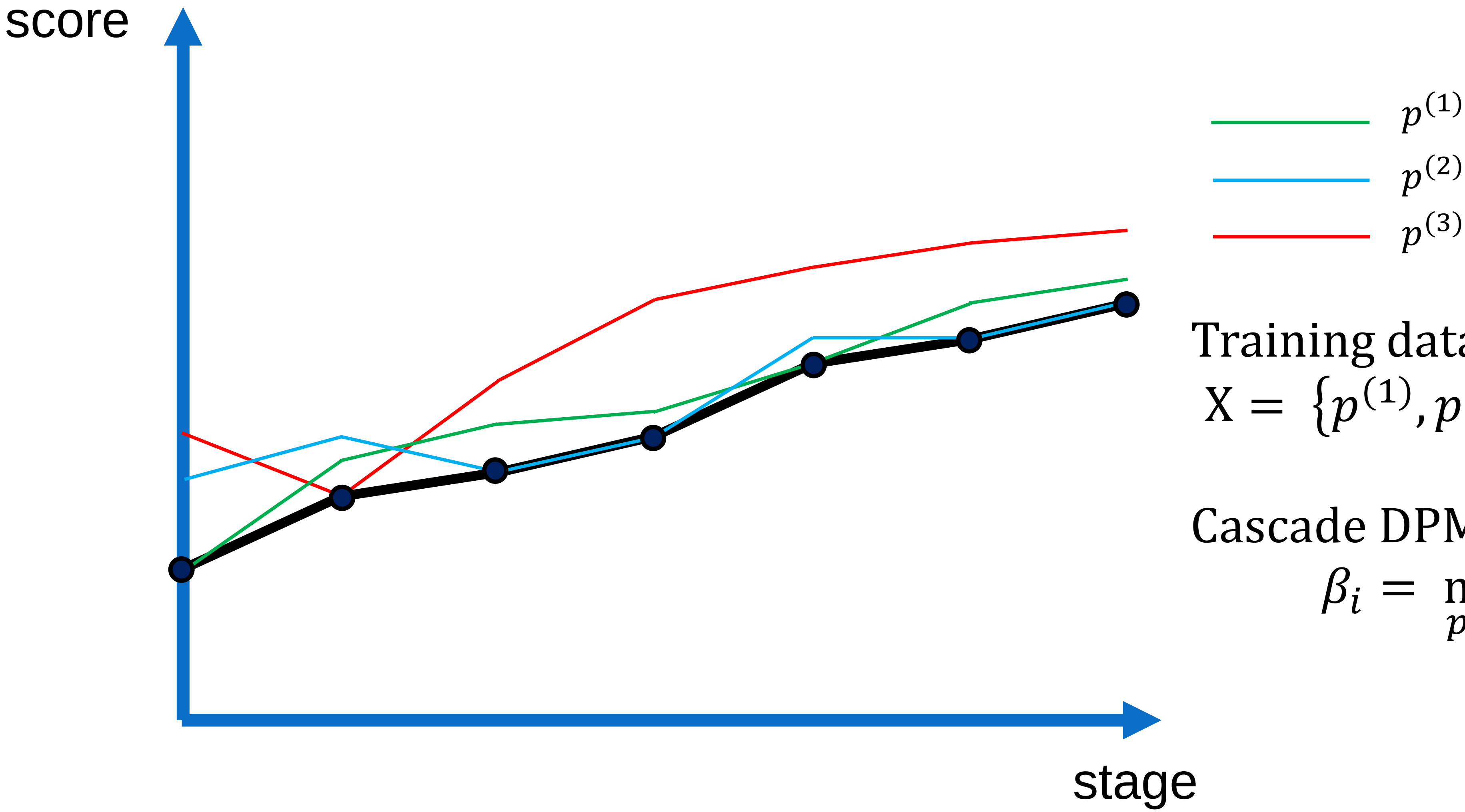
Drawbacks:

- a threshold learning step is required besides model training
- intensively human efforts for annotating object regions to learn high-quality thresholds

Idea

- Our idea is that, instead of collecting numerous training images with just several object samples per image, we learn thresholds from countless **negative samples** in **one image**.
- Advantages:
  - online threshold learning method with no labeled training data.
  - given an input image, our method is able to automatically estimate thresholds and detect objects of interest on the fly.

Cascade DPM threshold training



Training data

$$X = \{p^{(1)}, p^{(2)}, p^{(3)}\}$$

Cascade DPM threshold

$$\beta_i = \min_{p \in X} p_i$$

Proposed Method

Objective-like Negative hypothesis set:

$$\Omega_c = \{\gamma \in \Omega \mid \forall i < n, v_{2i}^{\gamma} \geq \beta_{2i}; v_{2n}^{\gamma} < \beta_{2n}\}$$

Property of Ω_c:

- F^{Ω_c} is close to and almost above $f(\beta)$
- Ω_c contains highest score negative hypotheses

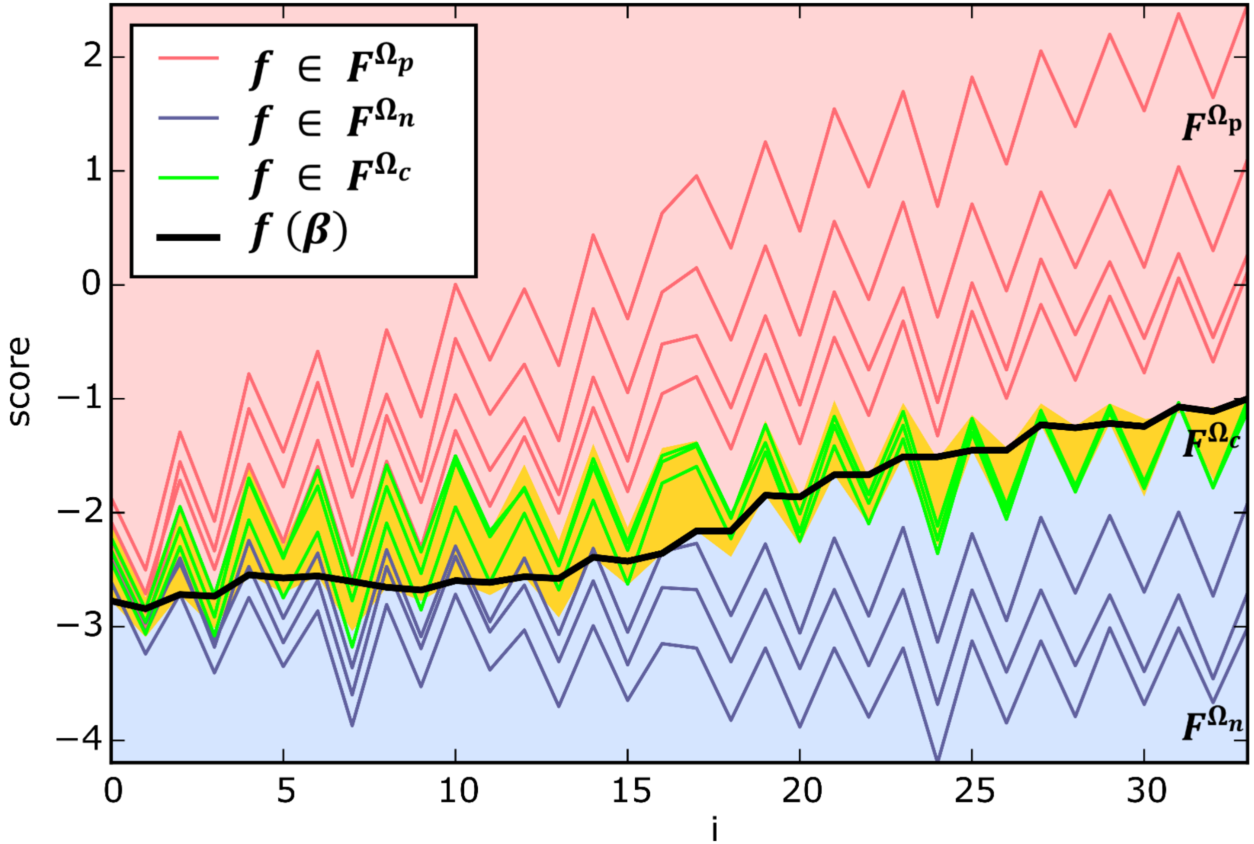
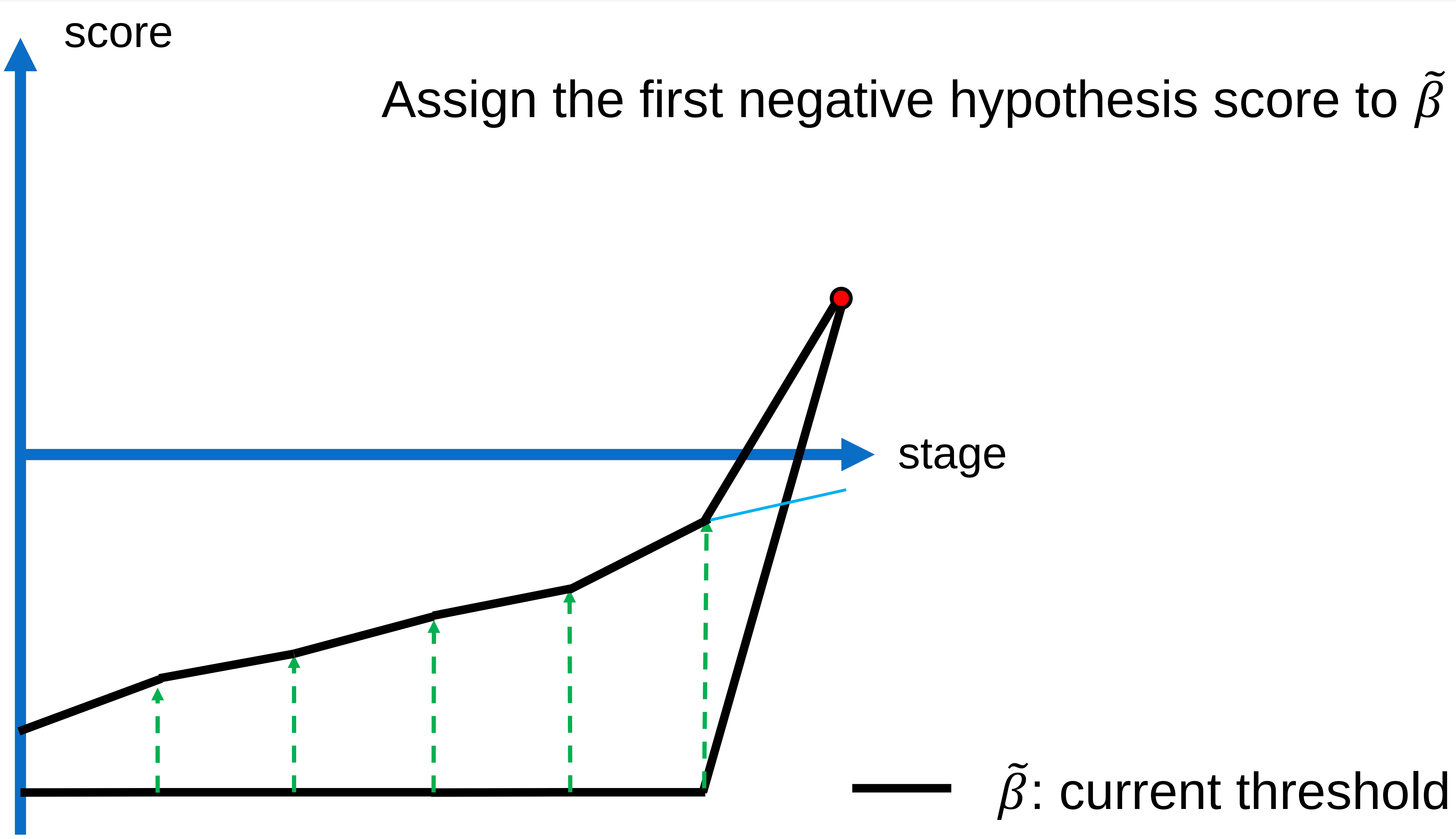


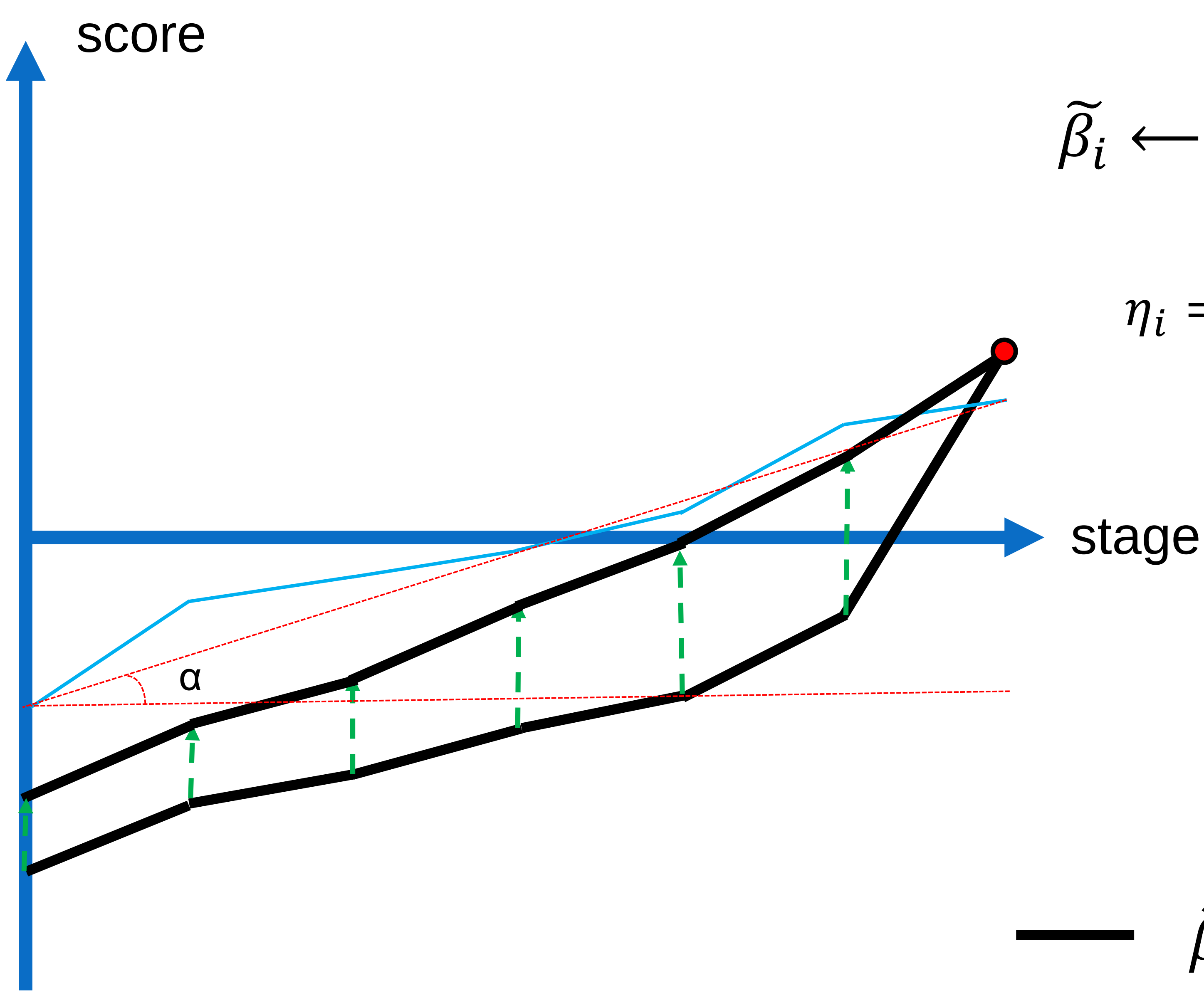
Fig 1. Human face Cascade DPM score functions

Proposed Method

Assign the first negative hypothesis score to $\tilde{\beta}$



Proposed Method



$$\tilde{\beta}_i \leftarrow \eta_i v_i^{\gamma} + (1-\eta_i)\tilde{\beta}_i$$

$$\eta_i = \alpha + (1-\alpha)\frac{i}{2n-1}$$

$\tilde{\beta}$: current threshold

Experiments

- Online ACTL
  - Thresholds are equivalent to Cascade DPM
  - The same performance
  - But our methods have more advantages
    - Online threshold learning
    - Faster
    - No training data for threshold learning
- Trn ACTL (learning from negative samples)
  - Offline threshold learning can obtain the same level of quality (showing the similar performance with Cascade DPM and online ACTL)

Experiments: Online ACTL

- Our learned thresholds are equivalent to Cascade DPM threshold
  - Average from 205 AFW images
  - Updated thresholds:
    - Converges
    - is close to β
  - 94.47 updates for 10^6 hypotheses

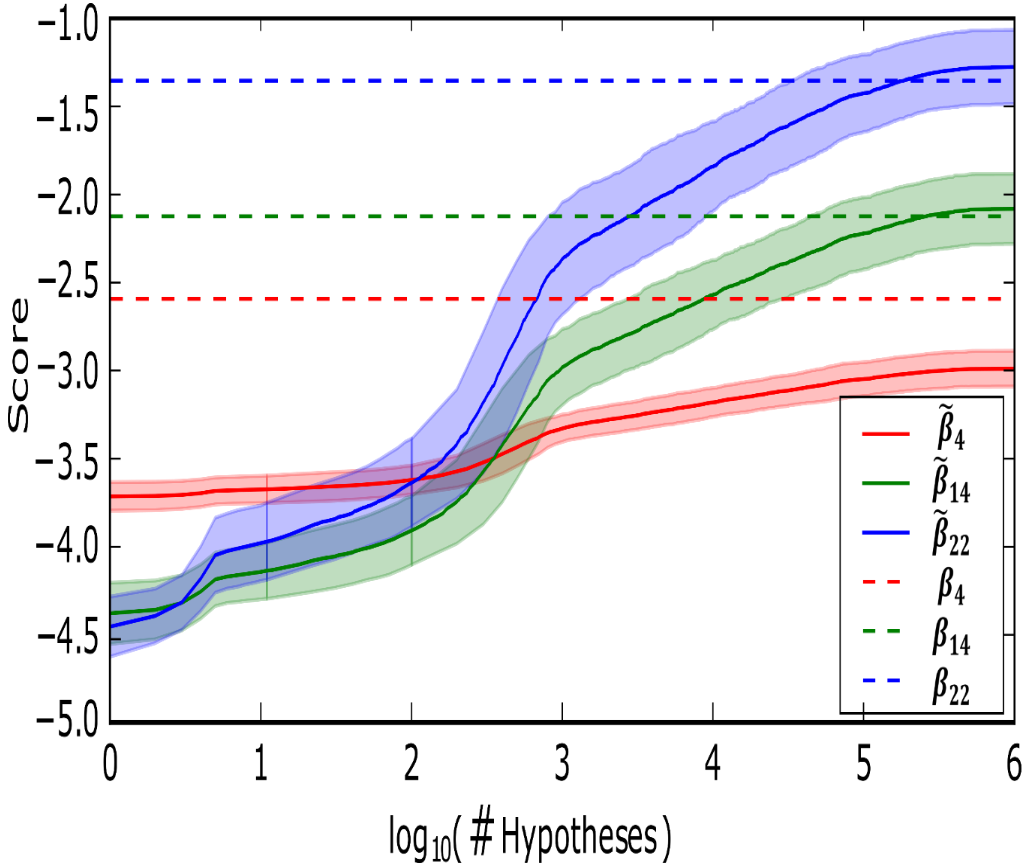


Fig 2. Updated threshold values w.r.t $\log_{10}(\#hypotheses)$

Experiments: Online ACTL

- Online ACTL achieves the same performance
  - Compare with:
    - Cascade DPM
    - Neighbor Awareness Cascade [7]

	Cascade	NAC	onl-ACTL
MAP	80.03	**80.11**	80.01
Time	4.53	**3.20**	4.02

Table 1. Results on AFW

MAP	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	**mean**
Cascade	33.48	**59.70**	**10.16**	**14.74**	**26.69**	**50.56**	52.23	**21.75**	**19.98**	**23.91**	26.26	12.95	**56.44**	**47.21**	**42.77**	13.43	**20.27**	**34.93**	**45.16**	**41.20**	**32.69**
NAC	**33.98**	58.69	9.71	12.31	25.10	48.43	**54.28**	19.43	17.95	23.43	22.77	11.69	55.19	46.39	40.51	11.89	19.11	31.12	44.89	40.99	31.39
onl-ACTL	32.41	59.36	9.27	9.09	23.30	46.60	52.69	21.12	13.98	23.62	**26.31**	**13.10**	56.22	47.17	41.19	**13.54**	10.51	33.77	45.09	41.19	31.00

Det Time	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	**mean**
Cascade	0.68	0.33	0.90	0.78	0.93	0.36	0.61	0.58	1.11	0.41	0.44	0.68	0.36	0.38	0.65	1.01	0.54	0.38	0.27	0.64	0.60
NAC	**0.32**	**0.29**	**0.32**	0.38	**0.27**	**0.32**	**0.30**	**0.27**	0.38	**0.26**	**0.25**	**0.34**	**0.26**	**0.23**	**0.31**	0.34	0.31	**0.27**	**0.21**	0.34	**0.30**
onl-ACTL	0.44	0.39	0.33	**0.26**	0.30	0.34	0.33	0.37	**0.25**	0.33	0.50	0.46	0.42	0.45	0.47	**0.28**	**0.20**	0.38	0.40	**0.29**	0.36

Table 2. Results on PASCAL VOC 2007

Experiments: Trained ACTL

- Offline threshold learning can obtain the same level of quality
  - run ACTL through 20 negative images:
    - average updated thresholds
    - use average threshold instead of β for testing

	Cascade	NAC	onl-ACTL	trn-ACTL
MAP	80.03	**80.11**	80.01	80.04
Time	4.53	**3.20**	4.02	4.66

Table 3. Results on AFW

MAP	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	**mean**
Cascade	33.48	**59.70**	**10.16**	**14.74**	**26.69**	**50.56**	52.23	**21.75**	**19.98**	**23.91**	26.26	12.95	**56.44**	**47.21**	**42.77**	13.43	**20.27**	**34.93**	**45.16**	**41.20**	**32.69**
NAC	**33.98**	58.69	9.71	12.31	25.10	48.43	**54.28**	19.43	17.95	23.43	22.77	11.69	55.19	46.39	40.51	11.89	19.11	31.12	44.89	40.99	31.39
onl-ACTL	32.41	59.36	9.27	9.09	23.30	46.60	52.69	21.12	13.98	23.62	26.31	13.10	56.22	47.17	41.19	13.54	10.51	33.77	45.09	41.19	31.00
trn-ACTL	32.10	59.66	9.27	9.09	24.92	44.94	52.43	21.43	12.79	23.33	**26.37**	**13.16**	56.27	47.03	41.21	**13.84**	13.38	34.39	45.14	40.87	31.08
Det Time	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	**mean**
Cascade	0.68	0.33	0.90	0.78	0.93	0.36	0.61	0.58	1.11	0.41	0.44	0.68	0.36	0.38	0.65	1.01	0.54	0.38	0.27	0.64	0.60
NAC	0.32	0.29	0.32	0.38	0.27	0.32	0.30	**0.27**	0.38	**0.26**	**0.25**	**0.34**	**0.26**	**0.23**	**0.31**	0.34	0.31	**0.27**	**0.21**	0.34	0.30
onl-ACTL	0.44	0.39	0.33	0.26	0.30	0.34	0.33	0.37	0.25	0.33	0.50	0.46	0.42	0.45	0.47	0.28	0.20	0.38	0.40	0.29	0.36
trn-ACTL	**0.30**	**0.28**	**0.24**	**0.21**	**0.26**	**0.26**	**0.20**	0.28	**0.21**	0.27	0.36	**0.34**	0.28	0.31	0.37	**0.23**	**0.19**	0.29	0.25	**0.20**	**0.27**

Table 4. Results on PASCAL VOC 2007

Conclusion

- This work investigated the capacity of learning threshold from negative samples for Cascade DPM.
- It allows to remove the dependence on positive training data but still obtain more efficient performance (compared to Cascade DPM and 2D-neighbour Cascade DPM) but maintain the same level of accuracy.
- Main contributions of the paper include:
  - Online threshold learning during detection phase.
  - Offline threshold learning with several negative images.

Reference

[1] Pedro F. Felzenszwalb, “*Cascade object detection with deformable part models*”, CVPR 2010

[2] S Charles Brubaker, Jianxin Wu, Jie Sun, Matthew D Mullin, and James M Rehg, “On the design of cascades of boosted ensembles for face detection”, ICJV 2008

[3] Paul Viola and Michael Jones, “Rapid object detection using a boosted cascade of simple features”, CVPR 2001

[4] Lubomir Bourdev and Jonathan Brandt, “Robust object detection via soft cascade”, CVPR 2005

[5] Fan Yang, Wongun Choi, and Yuanqing Lin, “Exploit all the layers: Fast and accurate cnn object detector with scale dependent pooling and cascaded rejection classifiers”, *CVPR* 2016

[6] Huitao Luo, “Optimization design of cascaded classifiers”, *CVPR* 2005

[7] Junjie Yan, Zhen Lei, Longyin Wen, and Stan Z Li, ”The fastest deformable part model for object detection”, CVPR 2014

THANK YOU

Question