Flexible 3D Neighbourhood Cascade Deformable Part Models for Object Detection

Hung Vu **Khoa Pho** Bac Le

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

Outline

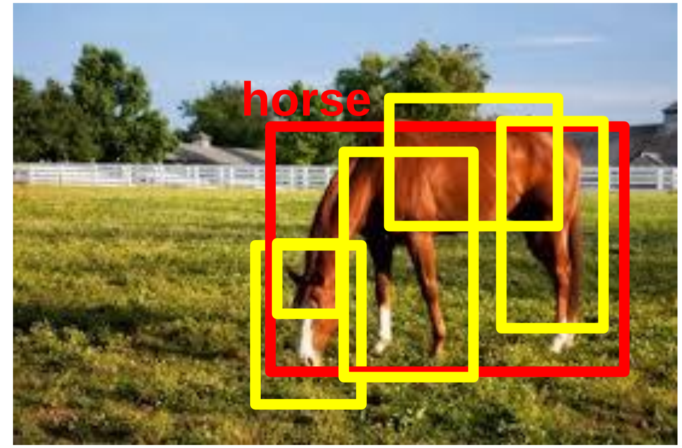
- Introduction
- Related Work
- Proposed Method
- Experiments
- Conclusion

Object Detection

Input image

Deformable Part Model (DPM) [1]

Output image



- DPMs describe the different views of an object via its components of parts.
- <u>**Drawback**</u>: For detection process, the templates of these parts are matched against all positions and scales → huge search space

  → **very slow**

[1] Pedro F. Felzenszwalb, Ross B. Girshick, David McAllester, and Deva Ramanan, “Object detection with discriminatively trained part-based models,” PAMI, vol. 32, no. 9, pp. 1627–1645, 2010.

Related Work of DPM speed-up

- **Reduce the cost of feature extraction**
  - Feature pyramid (Piotr Dollár et al., 2014)
  - Low-cost channel features (Piotr Dollár et al., 2009)
  - HOG with look-up tables (Junjie Yan et al., 2014)

  → not resolve the primary bottleneck of **massive cross-correlations**
- **Reduce cross-correlation cost**
  - FFT (Charles Dubout and François Fleuret, 2012)
  - Branch and Bound (Iasonas Kokkinos, 2011)
  - Root filters on low resolution image (Marco Pedersoli et al., 2015)
  - **Cascade DPMs** (Pedro F. Felzenszwalb et al., 2010; Junjie Yan et al., 2014; Tianfu Wu and Song-Chun Zhu, 2015)

Cascade framework

- Cascade frameworks have been popularized in Computer Vision community by the seminal work of Paul Viola and Michael J. Jones, 2004 and
- Other cascade frameworks (Dong Chen et al., 2016; Shuzhe Wu et al., 2017; Hakan Cevikalp and Bill Triggs, 2017)
  - Given a hypothesis/sub-window set of potential object position
  - Remove non-object hypothesis as much as possible after each cascade stage

  → All of them usually evaluate hypotheses individually
- Recent work investigate the dependency between hypotheses in 2D neighbourhood (NAC [2] and Crosstalk cascade [3]).

[2] Junjie Yan, Zhen Lei, Longyin Wen, and Stan Z. Li, "The fastest Deformable Part Model for object detection," in CVPR, 2014, pp. 2497–2504.

[3] Piotr Dollár, Ron Appel, and Wolf Kienzle, "Crosstalk cascades for frame-rate pedestrian detection," in ECCV, 2012, pp. 645–659.

Cascade DPM

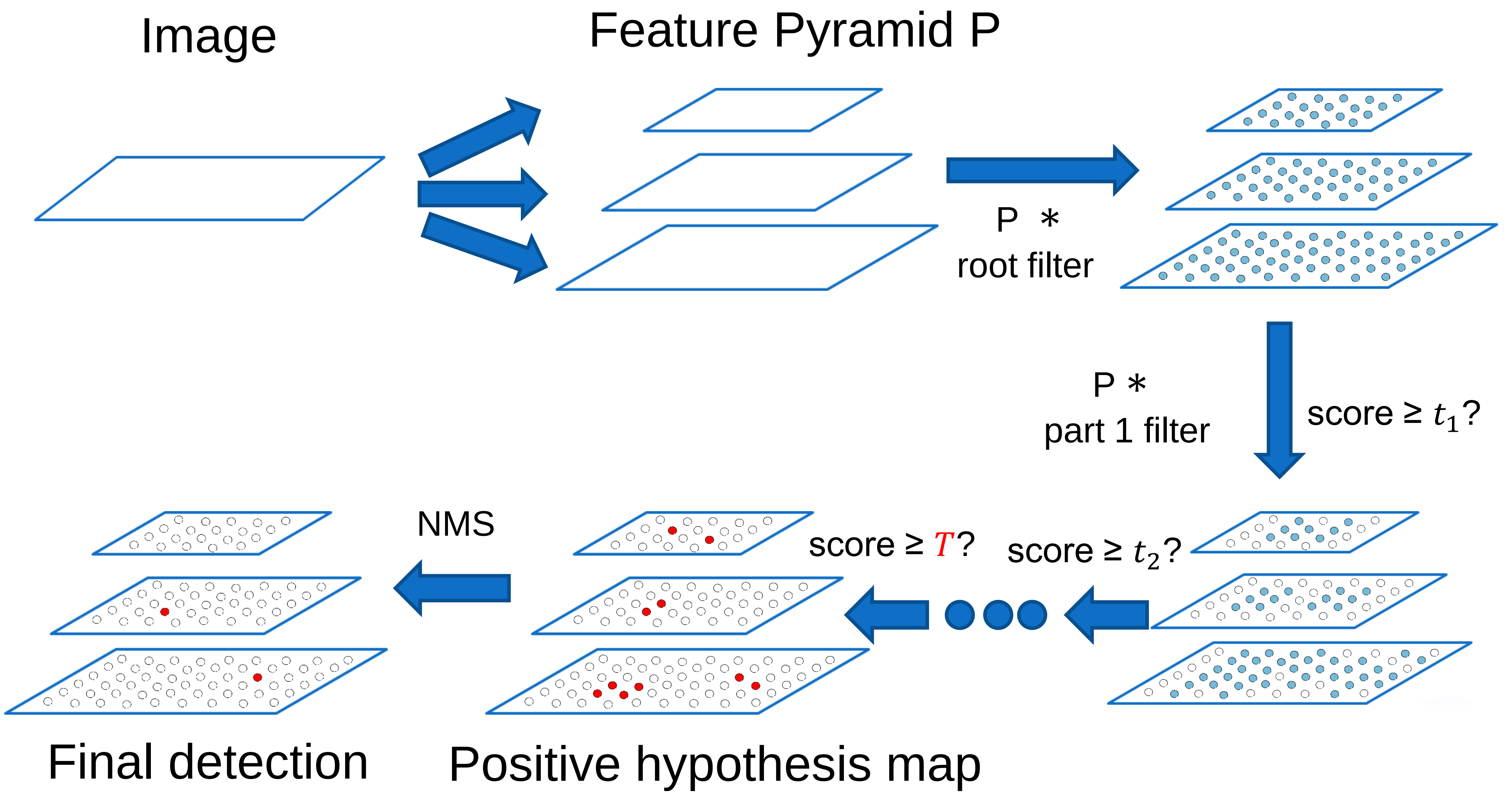
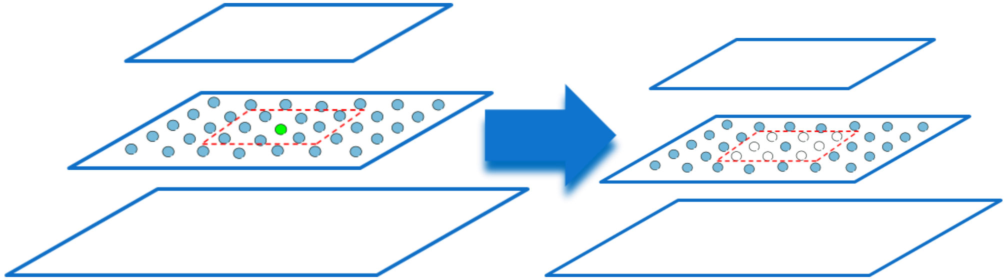


Image
Feature Pyramid P
P ∗
root filter
P ∗
part 1 filter
score ≥ t_1?
score ≥ t_2?
score ≥ T?
NMS
Final detection
Positive hypothesis map

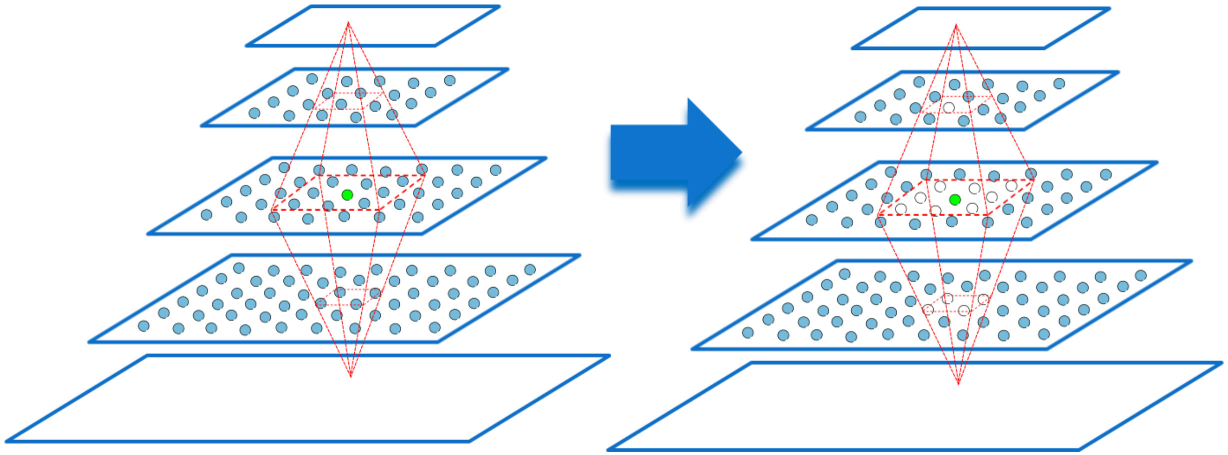
Idea

- We extend the idea of neighbourhood cascade to the 3rd dimension of scale to prune the hypotheses more aggressively.
- This work introduces two techniques of **3D neighbourhood pruning** and **scale pruning**.

❑ 3D neighbourhood pruning



2D pruning

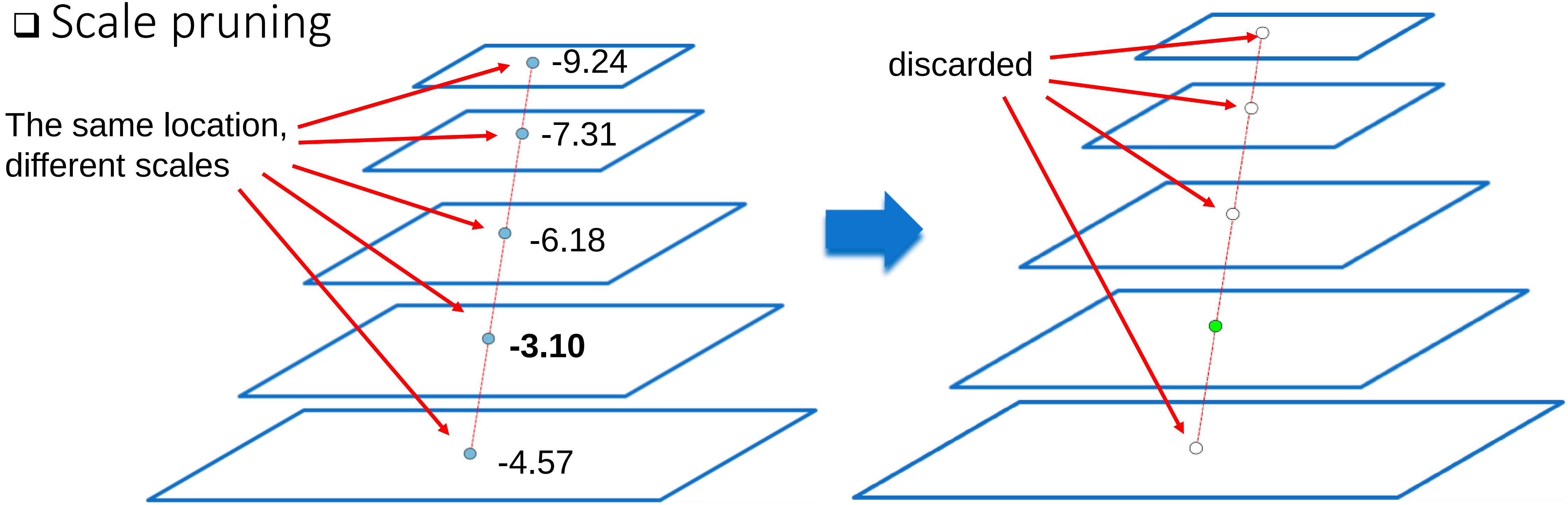


3D pruning

Idea

- We extend the idea of neighbourhood cascade to the 3rd dimension of scale to prune the hypotheses more aggressively.
- This work introduces two techniques of **3D neighbourhood pruning** and **scale pruning**.

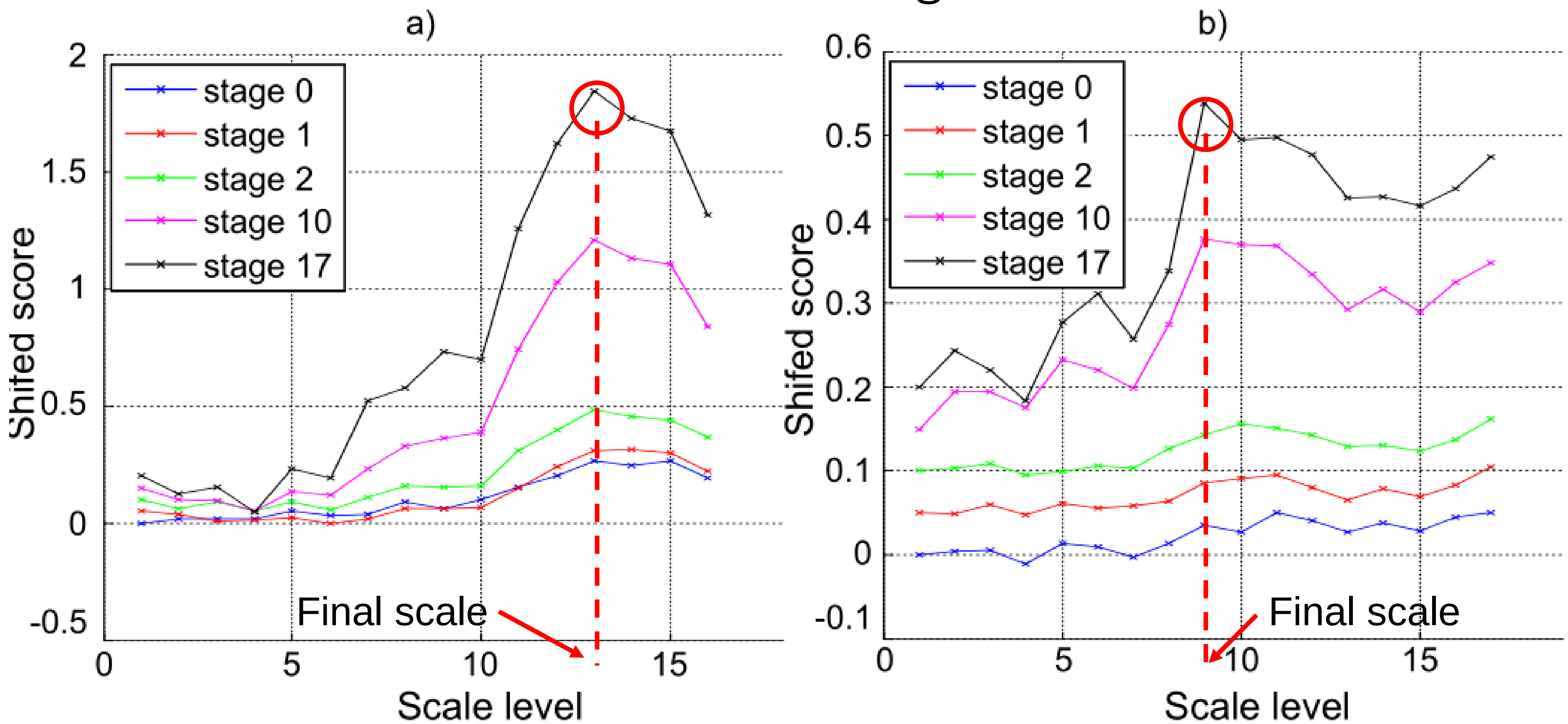
❑ Scale pruning



➔ Investigating the practicability of scale pruning

Can we prune the hypotheses over scale?

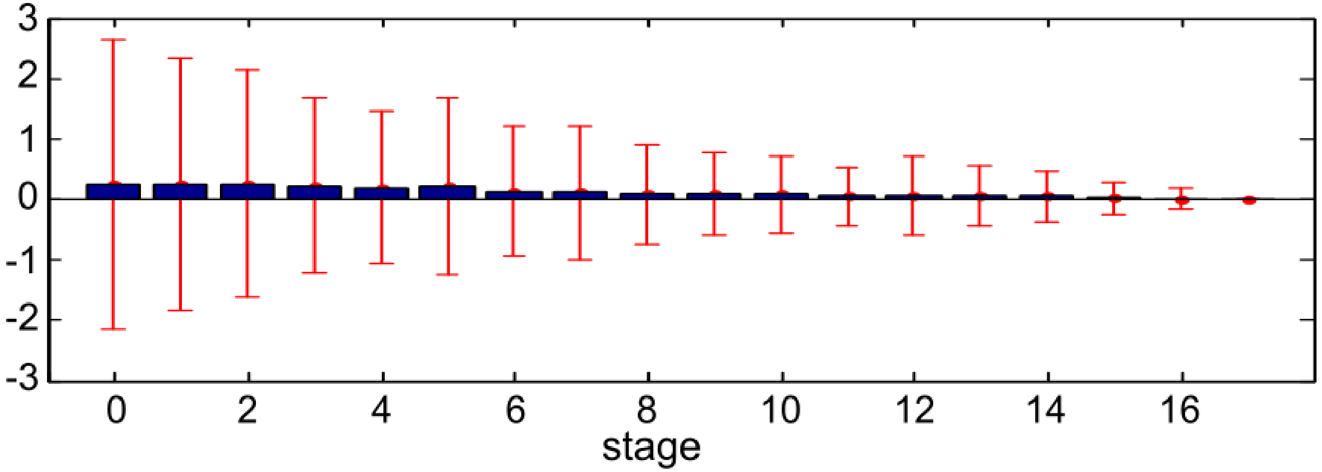
- We collected 1000 positive hypotheses randomly from 20 object classes in the PASCAL VOC 2007 training dataset.



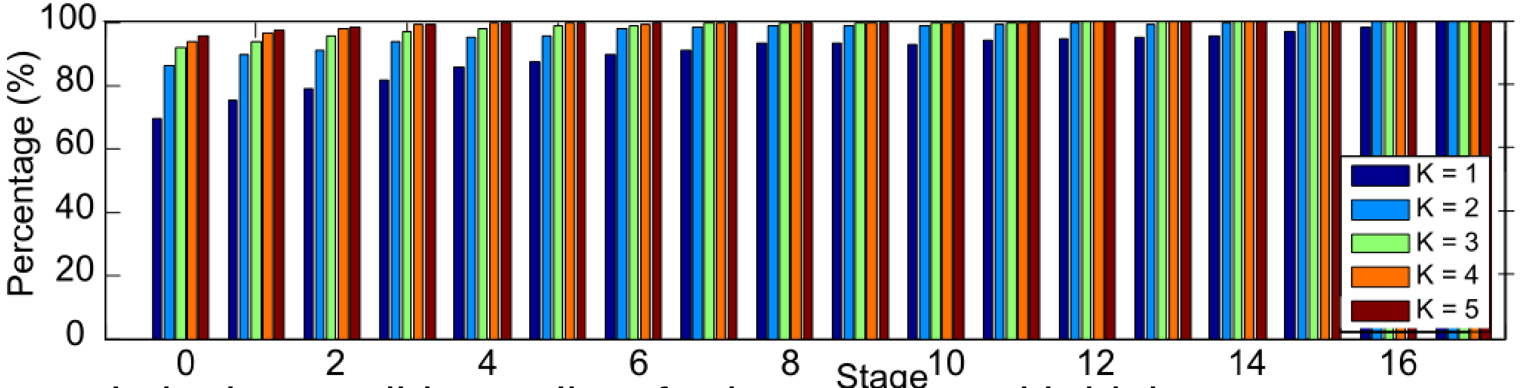
Score functions of hypotheses with respect to scale levels at different cascade stages

The scales with the highest scores over stages are almost the same (or very close)
Idea: At an early state τ, keep K top scales and prune the others → reduce a lot of negative hypotheses → speed up the system.

Early prediction of optimal scale



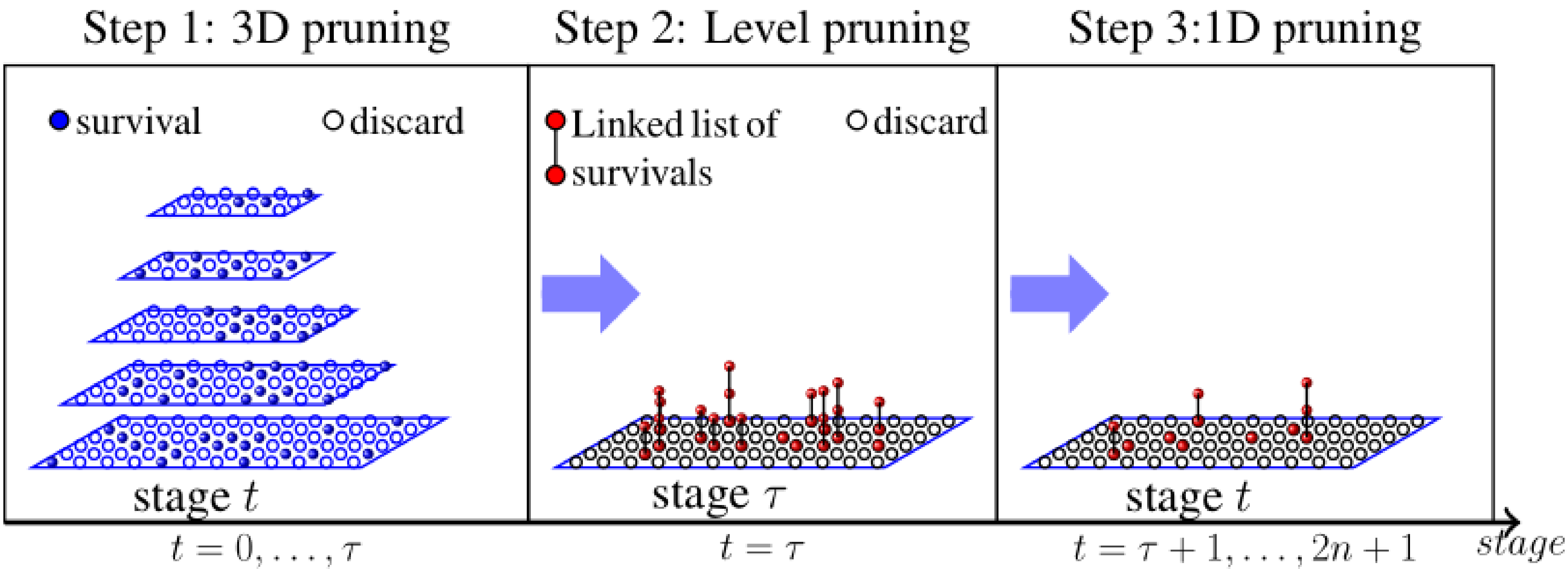
The difference between the true scale and maximal scale over stage



% true scale in the candidature list of K hypotheses with highest scores over state

- For example: $\tau = 2$ and $K \geq 2 \rightarrow 91\%$ true scale in the hypothesis list

Proposed framework: Flexible 3D Neighbourhood Cascade DPM



Existing DPM threshold pruning techniques

$$g_t(\gamma) = \omega_o^\top \phi(l_0, I) + \sum_{i=1}^{t} \omega_i^\top (l_i I) - d_i^\top \theta(l_i\,, l_0)$$

- Hypothesis threshold pruning [1] α_t^1:

  prune γ if $g_t(\gamma) < \alpha_t^1$

- Deformation threshold pruning [1] α_t^2:

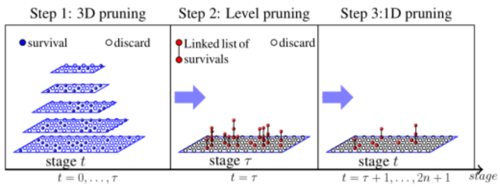
  prune γ if $g_t(\gamma) - d_t^\top \theta(l_t\,, l_0) < \alpha_t^2$

- Semi-positive threshold [4] α_t^3:

  prune γ if $\exists \gamma' \in N(\gamma), g_t(\gamma') - g_t(\gamma) > \alpha_t^3$

[1] Pedro F. Felzenszwalb, Ross B. Girshick, David McAllester, and Deva Ramanan, "Object detection with discriminatively trained part-based models," PAMI, vol. 32, no. 9, pp. 1627–1645, 2010.
[4] Junjie Yan, Zhen Lei, Longyin Wen, and Stan Z. Li, "The fastest Deformable Part Model for object detection," in CVPR, 2014, pp. 2497–2504.

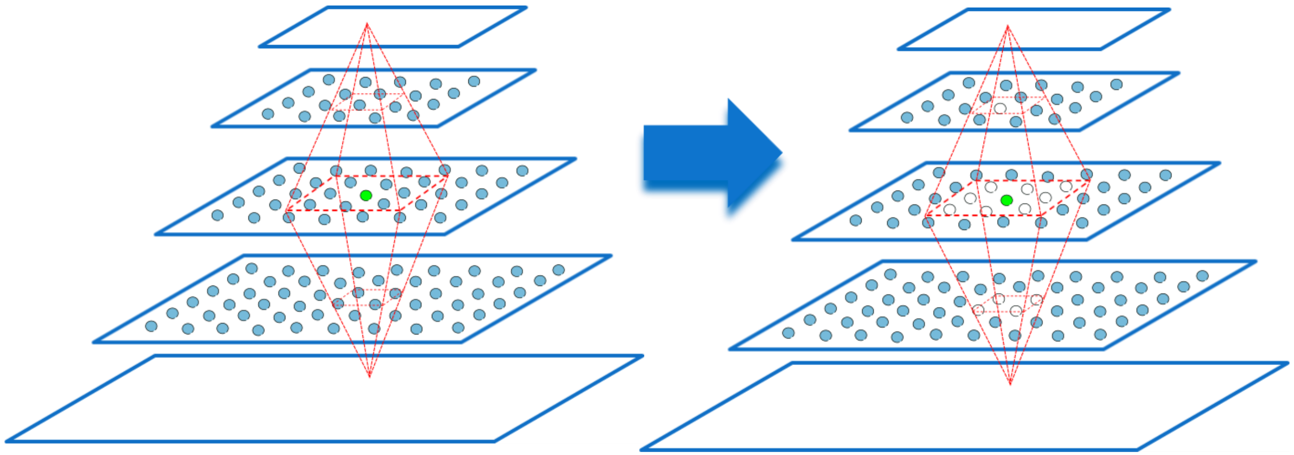
Step 1: 3D Neighbourhood Pruning



- 3D neighbour pruning operates in the first τ stages:

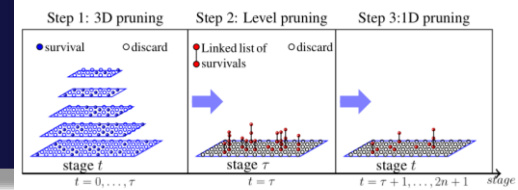
$$\boxed{\text{prune } \mathrm{N}_{3D}(\gamma)} \text{ if } g_t(\gamma) < \alpha_t^4$$

- $\mathrm{N}_{3D}(\gamma)$ is a square pyramid
- Thresholds α_t^2 and α_t^3 are also applied.



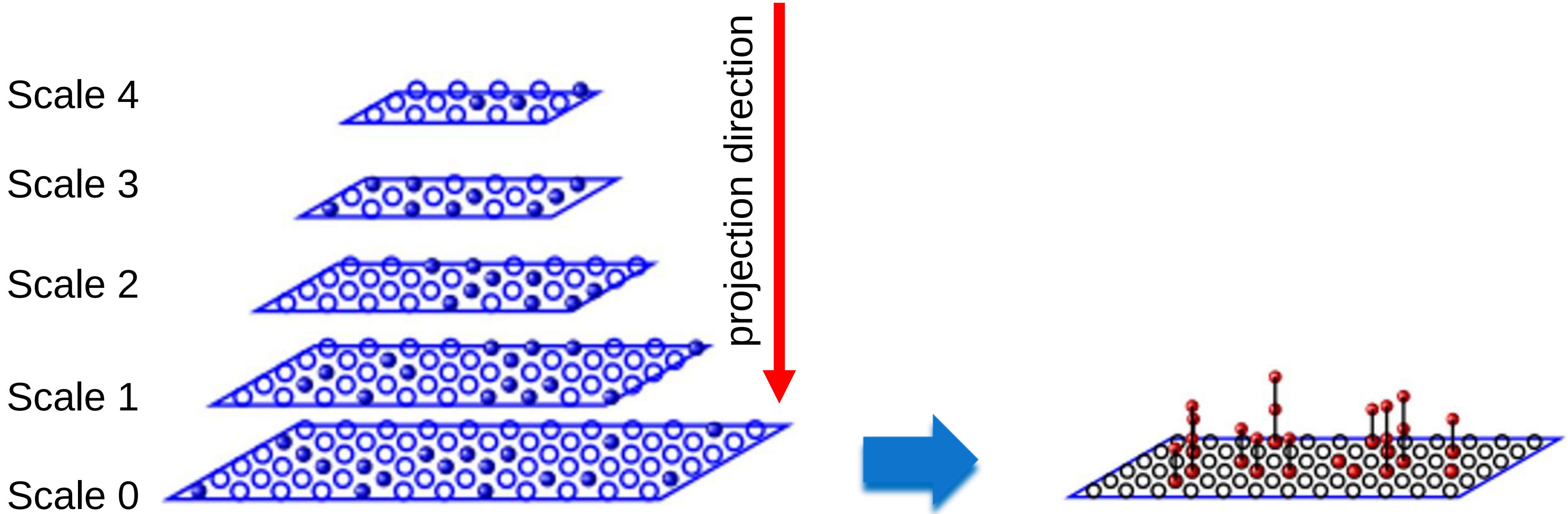
3D pruning

Step 2: Scale Pruning

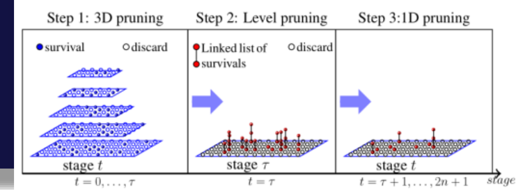


Whenever the stage τ ends:

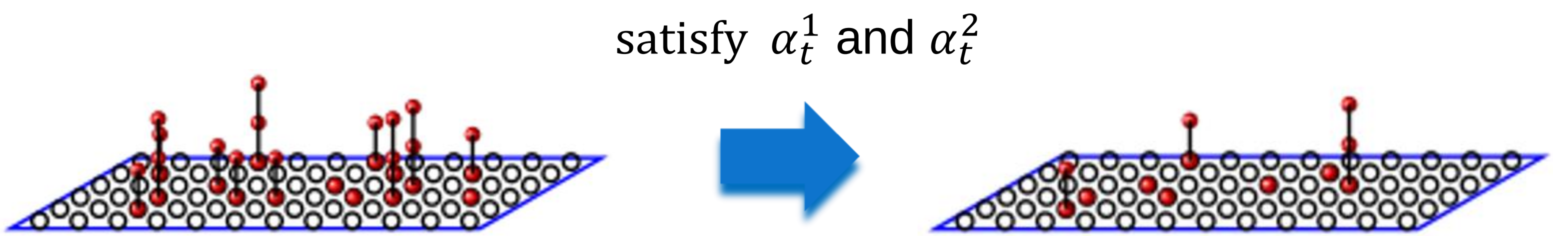
- Project survival hypotheses into feature map of scale 0
- Keep K hypotheses at the same locations
- Run $\mathrm{NMS}_{\mathrm{K}}$ (non-maximum suppression) to remove ones not in the top-K of the best hypotheses in its neighbourhood.



Step 1: 1D Pruning



- From the stage $\mathrm{t} > \tau$, we use the hypothesis thresholds α_t^1 and deformation thresholds α_t^2
- Pass the global threshold T
- Run NMS to filter out the redundant detection results



Experiments

- We tested our proposal on two problems of face detection and object detection.
- Hardware: Intel Core i7 2.6 GHz desktop with 20 GB memory.

❑ Object detection

Object Detection		DPM	Cascade	NAC	Flex3DNB
mAP (%)		**32.85**	32.69	31.39	29.30
Detection Time (second)		1.14	0.60	0.30	**0.19**

Mean AP and detection time in PASCAL VOC 2017

❑ Face detection

Face Detection	TSPM	EDEL	DPM	Cascade	NAC	Flex3DNB
mAP (%)	**81.38**	80.84	80.02	80.03	80.11	80.58
Detection Time (second)	42.26	23.29	14.98	4.53	3.20	**2.02**

Mean AP and detection time AFW

Conclusion

- This work investigated the capacity of integrating the 3D neighbourhood information into Cascade DBM framework.
- It allows to obtain more efficient performance (compared to Cascade DBM and 2D-neighbour Cascade DBM) but maintain the same level of accuracy.
- Main contributions of the paper include:
  - 3D neighbourhood cascade
  - Scale pruning technique
  - Flexible neighbourhood: The volume of the neighbourhood changes w.r.t. scores
  - Root score pruning (the first work to prune hypotheses at root stage)

THANK YOU

Question