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
DPMs describe the different views of an object via its components of parts.

**Drawback**: For detection process, the templates of these parts are matched against all positions and scales → huge search space → very slow

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]).


Cascade DPM

Image

Feature Pyramid P

Positive hypothesis map

Final detection

NMS

score ≥ $T$?

score ≥ $t_2$?

score ≥ $t_1$?

P ∗ part 1 filter

P ∗ root filter
We extend the idea of neighbourhood cascade to the 3\textsuperscript{rd} dimension of scale to prune the hypotheses more aggressively.

This work introduces two techniques of **3D neighbourhood pruning** and **scale pruning**.

- 3D neighbourhood pruning

![Diagram showing 2D and 3D pruning]

2D pruning

3D pruning
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

The same location, different scales

\[ -9.24 \rightarrow -7.31 \rightarrow -6.18 \rightarrow -3.10 \rightarrow -4.57 \]

discarded

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 $\tau$, keep $K$ top scales and prune the others $\rightarrow$ reduce a lot of negative hypotheses $\rightarrow$ speed up the system.
Early prediction of optimal scale

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

The difference between the true scale and maximal scale over stage
Proposed framework: Flexible 3D Neighbourhood Cascade DPM

Step 1: 3D pruning
- Survival
- Discard

Stage $t = 0, \ldots, \tau$

Step 2: Level pruning
- Linked list of survivals
- Discard

Stage $\tau$

Step 3: 1D pruning

Stage $t = \tau + 1, \ldots, 2n + 1$
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] \( \alpha_t^1 \):
  prune \( \gamma \) if \( g_t(\gamma) < \alpha_t^1 \)

- Deformation threshold pruning [1] \( \alpha_t^2 \):
  prune \( \gamma \) if \( g_t(\gamma) - d_t^\top \theta(l_t, l_0) < \alpha_t^2 \)

- Semi-positive threshold [4] \( \alpha_t^3 \):
  prune \( \gamma \) if \( \exists \gamma' \in N(\gamma), g_t(\gamma') - g_t(\gamma) > \alpha_t^3 \)

Step 1: 3D Neighbourhood Pruning

- 3D neighbour pruning operates in the first $\tau$ stages:
  
  \[
  \text{prune } N_{3D}(\gamma) \text{ if } g_t(\gamma) < \alpha_t^4
  \]

- $N_{3D}(\gamma)$ is a square pyramid
- Thresholds $\alpha_t^2$ and $\alpha_t^3$ are also applied.
Step 2: Scale Pruning

Whenever the stage $\tau$ ends:

- Project survival hypotheses into feature map of scale 0
- Keep $K$ hypotheses at the same locations
- Run $\text{NMS}_K$ (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 $t > \tau$, we use the hypothesis thresholds $\alpha_t^1$ and deformation thresholds $\alpha_t^2$.
- Pass the global threshold $T$.
- Run NMS to filter out the redundant detection results.
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

<table>
<thead>
<tr>
<th>Object Detection</th>
<th>DPM</th>
<th>Cascade</th>
<th>NAC</th>
<th>Flex3DNB</th>
</tr>
</thead>
<tbody>
<tr>
<td>mAP (%)</td>
<td>32.85</td>
<td>32.69</td>
<td>31.39</td>
<td>29.30</td>
</tr>
<tr>
<td>Detection Time (second)</td>
<td>1.14</td>
<td>0.60</td>
<td>0.30</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Mean AP and detection time in PASCAL VOC 2017

- Face detection

<table>
<thead>
<tr>
<th>Face Detection</th>
<th>TSPM</th>
<th>EDEL</th>
<th>DPM</th>
<th>Cascade</th>
<th>NAC</th>
<th>Flex3DNB</th>
</tr>
</thead>
<tbody>
<tr>
<td>mAP (%)</td>
<td>81.38</td>
<td>80.84</td>
<td>80.02</td>
<td>80.03</td>
<td>80.11</td>
<td>80.58</td>
</tr>
<tr>
<td>Detection Time (second)</td>
<td>42.26</td>
<td>23.29</td>
<td>14.98</td>
<td>4.53</td>
<td>3.20</td>
<td>2.02</td>
</tr>
</tbody>
</table>

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