

SPP-net

Spatial Pyramid Pooling in Deep Convolutional Networks

Kaiming He

Xiangyu Zhang

Shaoqing Ren

Jian Sun

Microsoft Research Asia

Visual Computing Group

Highlights

- ILSVRC 2014 (all provided-data tracks)
 - DET - **2nd**
 - CLS - **3rd**
 - LOC - 5th
- ECCV 2014 paper
- Published 2 months ago (arXiv:1406.4729v1, June 18)
- Details disclosed (arXiv:1406.4729v2)

Overview

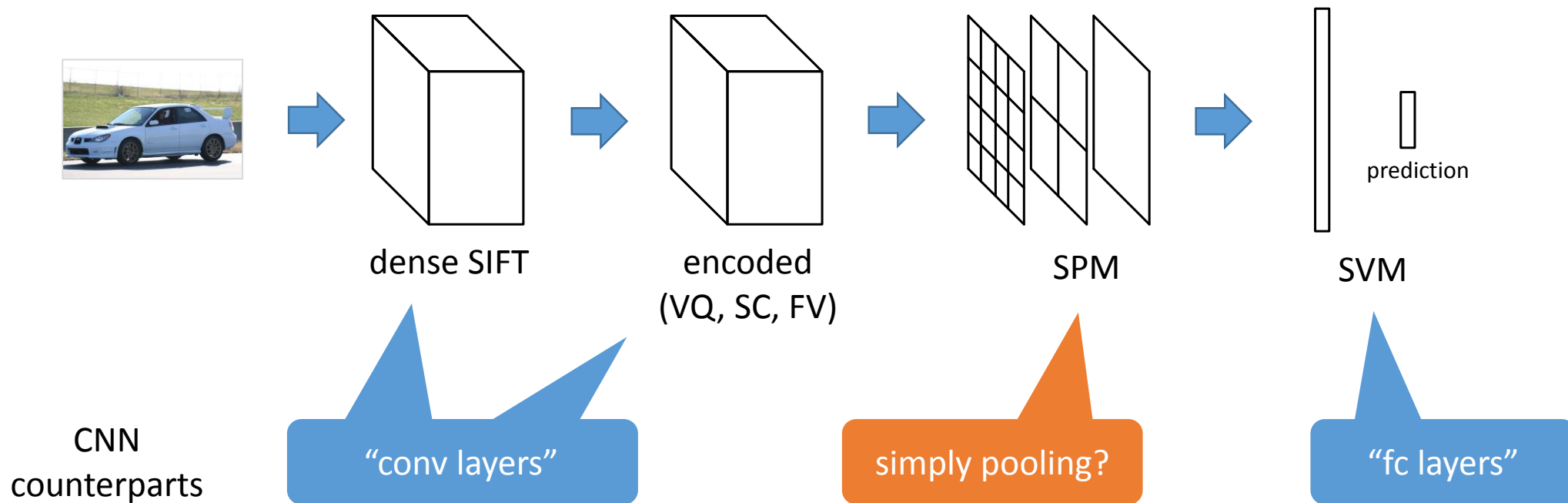
- SPP-net
 - a new network structure
- Classification
 - improves all CNNs
- Detection
 - 20-60x faster than R-CNN, as accurate

Spatial Pyramid Matching

- SPM: very successful in traditional computer vision

[Grauman & Darrell, ICCV 2005] "The Pyramid Match Kernel: Discriminative Classification with Sets of Image Features"

[Lazebnik *et al*, CVPR 2006] "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories"

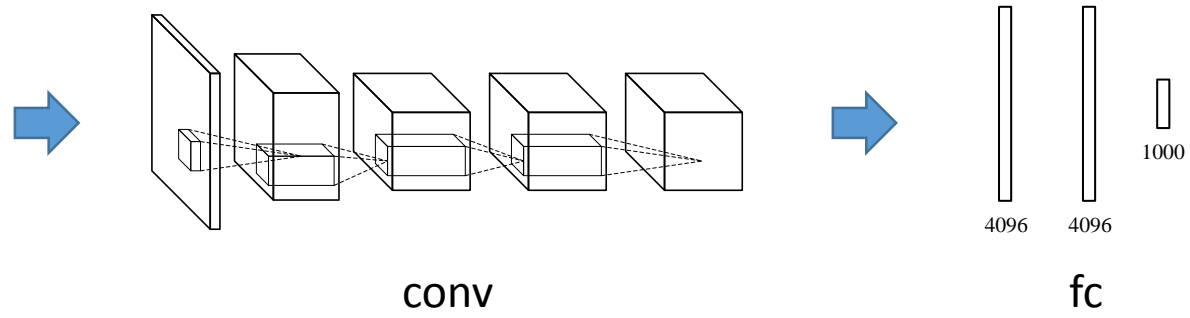


SPP-net: SPM in CNN

traditional
CNN



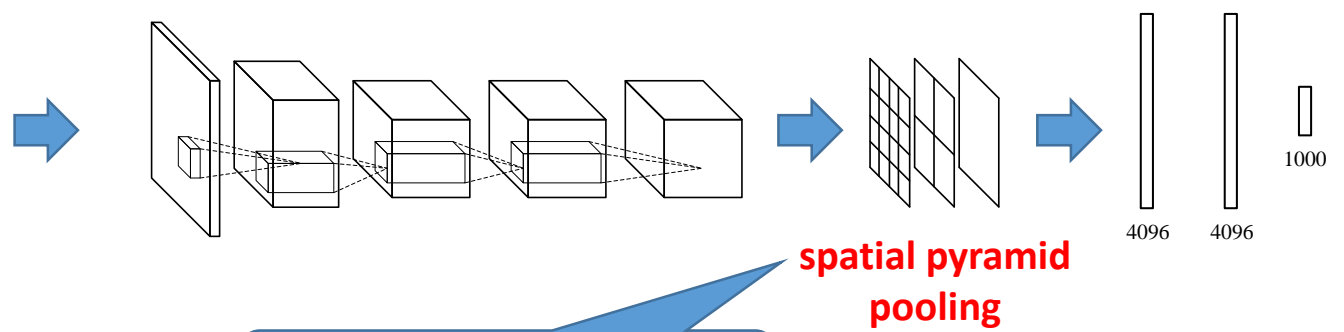
fixed size



SPP-net



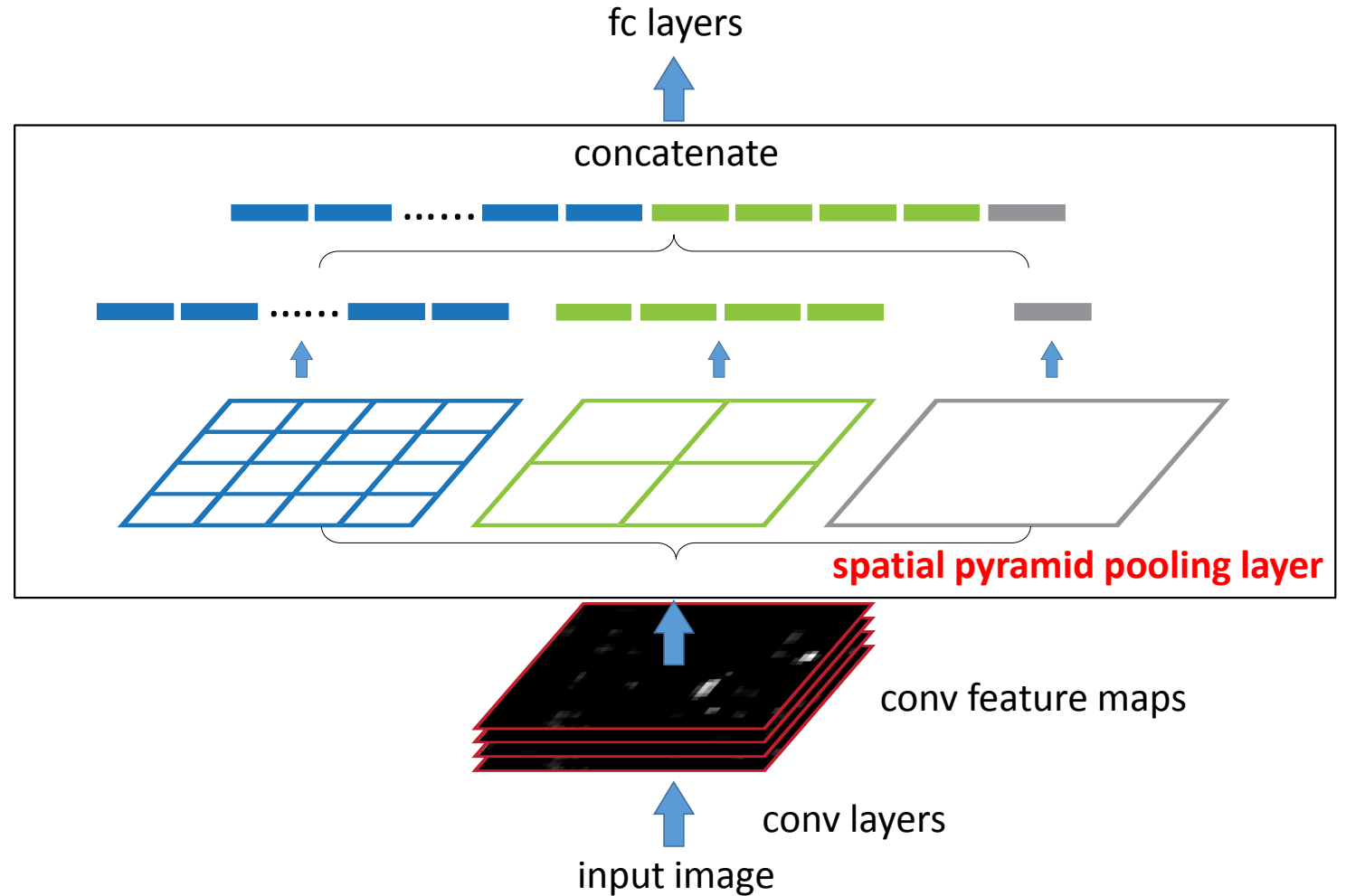
any size



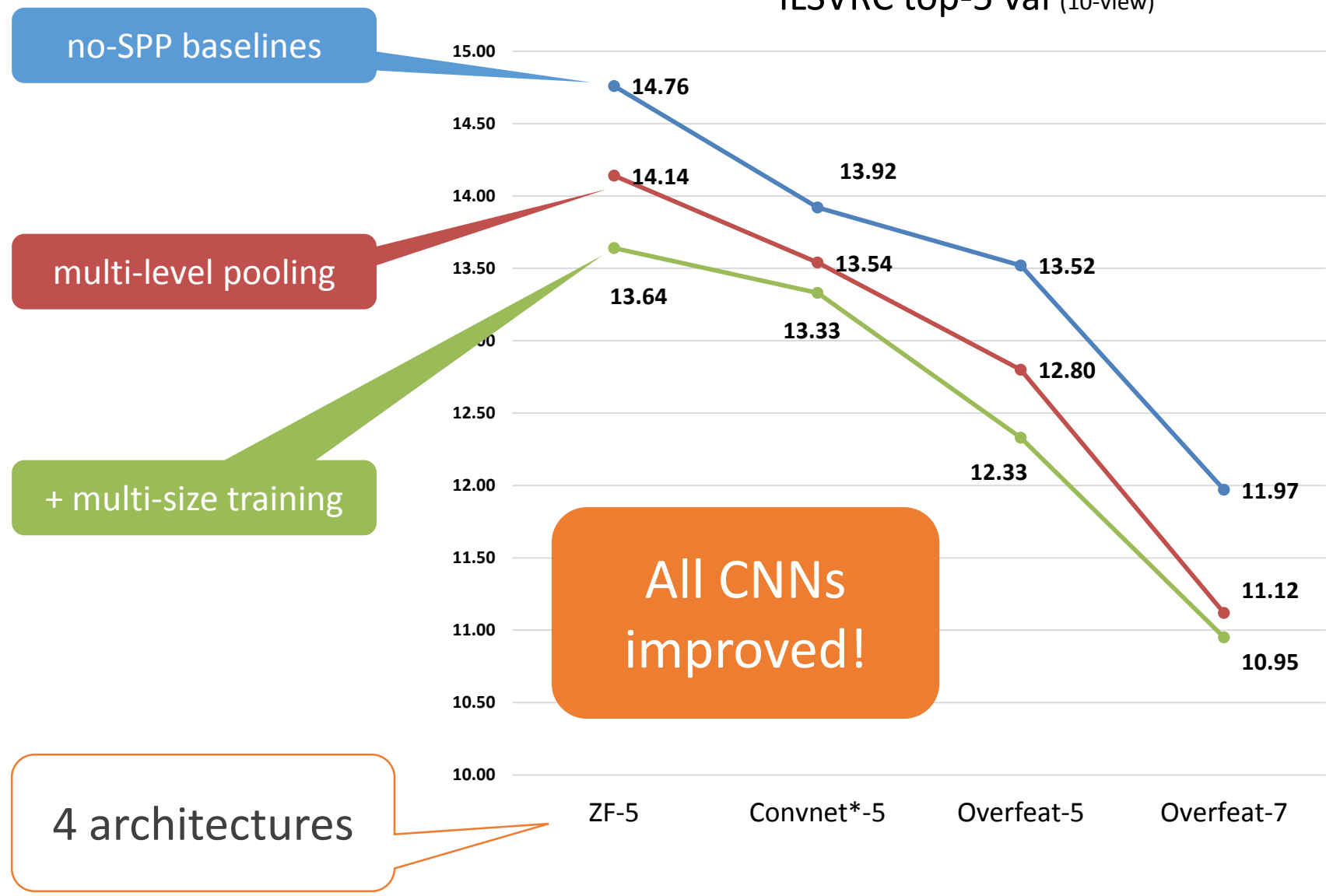
- Fix bin numbers
- **DO NOT** fix bin size

SPP-net

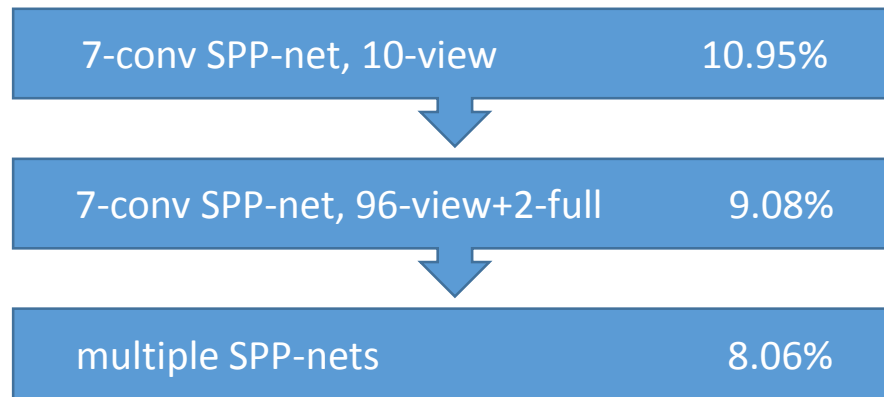
- variable input size/scale
 - multi-size training
 - multi-scale testing
 - full-image view
- multi-level pooling
 - robust to deformation
- operates on feature maps
 - pooling in regions



ILSVRC top-5 val (10-view)



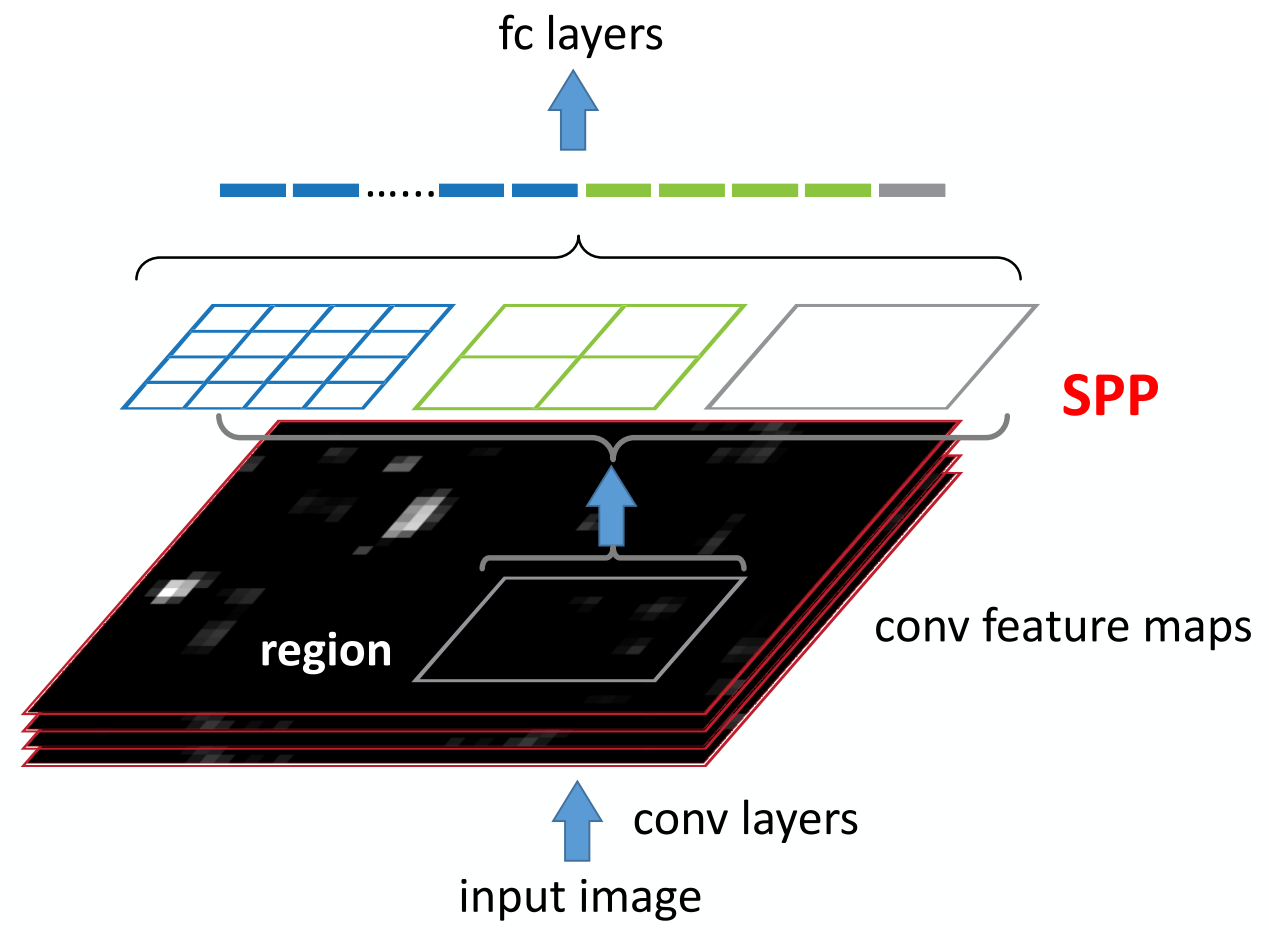
ILSVRC 2014 CLS Results



team	top-5 test
GoogLeNet	6.66
Oxford VGG	7.32
ours	8.06
Howard	8.11
DeeperVision	9.50
NUS-BST	9.79
TTIC_ECP	10.22
...	

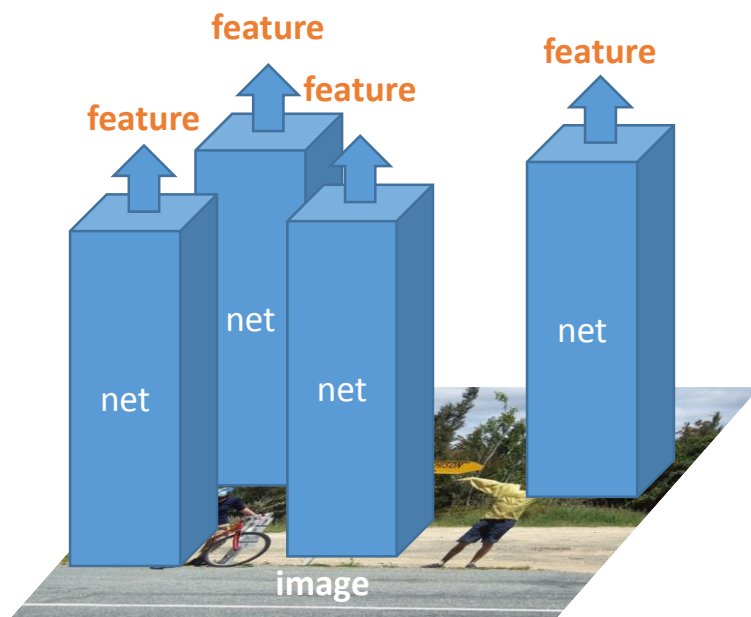
- “shallow”
 - 7-conv, 1 Titan GPU, 3 weeks
- but **potential**
 - SPP can improve deeper nets: >1% gain post-competition

Detection: SPP on Regions



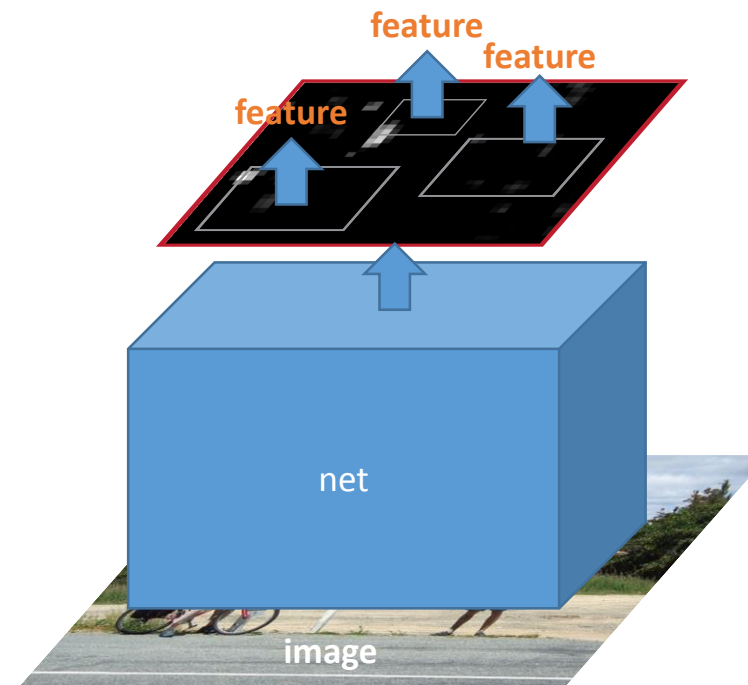
RCNN vs. SPP

- **image** regions vs. **feature map** regions



R-CNN

2000 nets on image regions



SPP-net

1 net on full image

- With regional features, we can do everything of RCNN
 - fine-tune, SVM, bbox regression...
 - **similar accuracy, much faster**

	SPP-net 1-scale	SPP-net 5-scale	RCNN
mAP	58.0	59.2	58.5
GPU time / img	0.14s	0.38s	9s
speed-up	64x	24x	-

VOC 2007

ILSVRC 2014 DET Results

	mAP
NUS	37.2
ours, multi SPP-nets	35.1
UvA	32.0
ours, 1 SPP-net	31.8
Southeast-CASIA	30.4
1-HKUST	28.8
CASIA_CRIPAC_2	28.6

“provided data” track

	SPP-net	RCNN
GPU time / img	0.6s	32s
40k test imgs	8 hours	15 days

cost of a single model

- **Conclusion**

- SPM in CNNs
- CLS: improve all CNNs in the literature
- DET: practical, fast, and accurate

- **Future work**

- SPP on advanced networks

- **Resources**

- code, config, tech report...

<http://research.microsoft.com/en-us/um/people/kahe/>

- **Acknowledgement**

- We thank NVIDIA for the GPU donation.