NIN, Good! (您好)

(Network in Network)

Jian DONG, Min LIN, Yunchao WEI, Qiang CHEN*, Wei, XIA, Hanjiang LAI, Shuicheng YAN

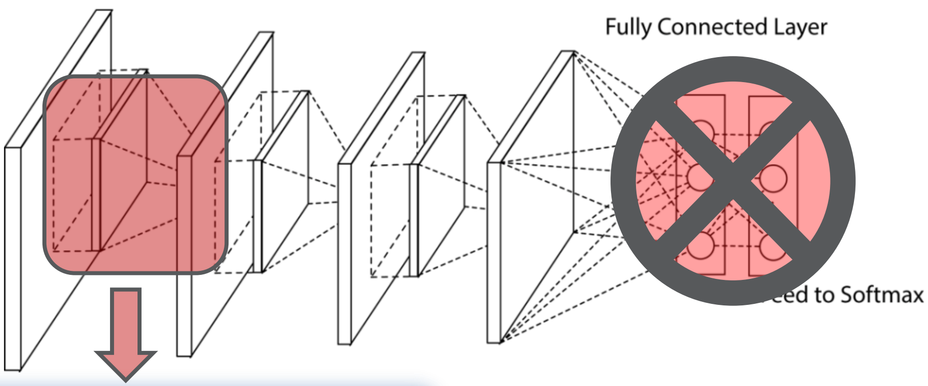
eleyans@nus.edu.sg
National University of Singapore
* IBM Research Australia

NUS
National University
of Singapore

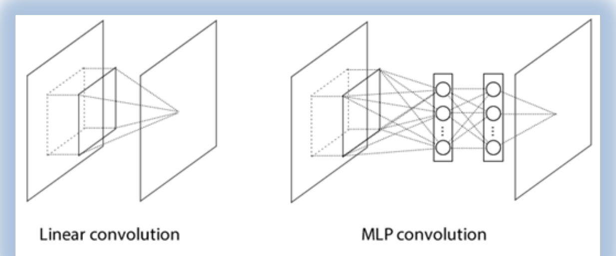
Episode-1: Network in Network, ILSVRC-2013

“Network in Network” (NIN)

- **NIN: CNN with non-linear filters, yet without fully-connected layers**

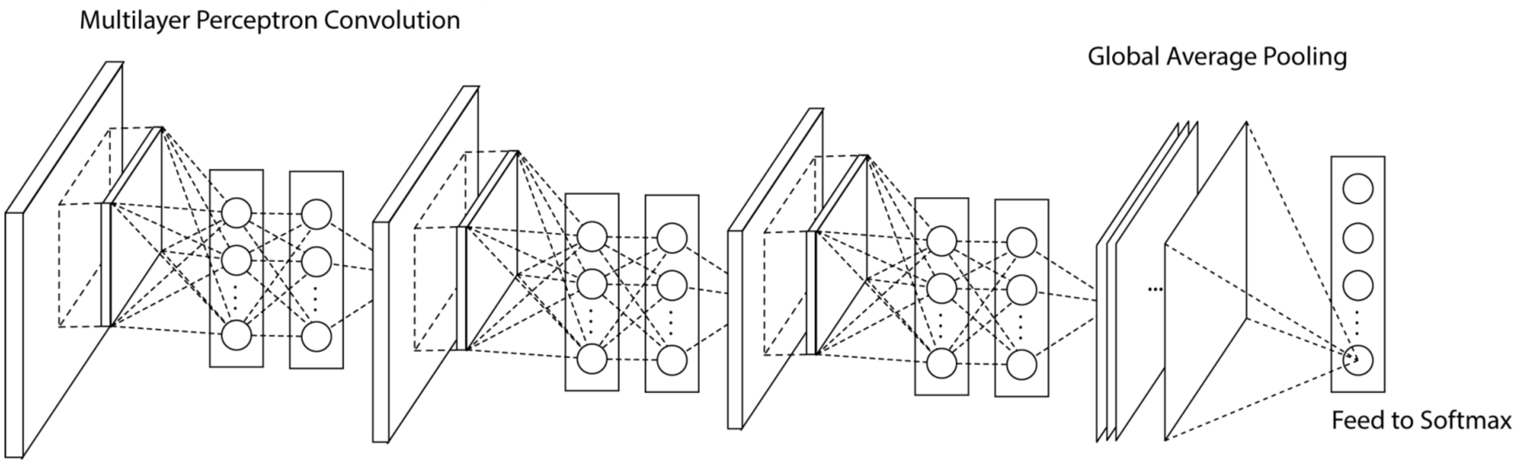


CNN



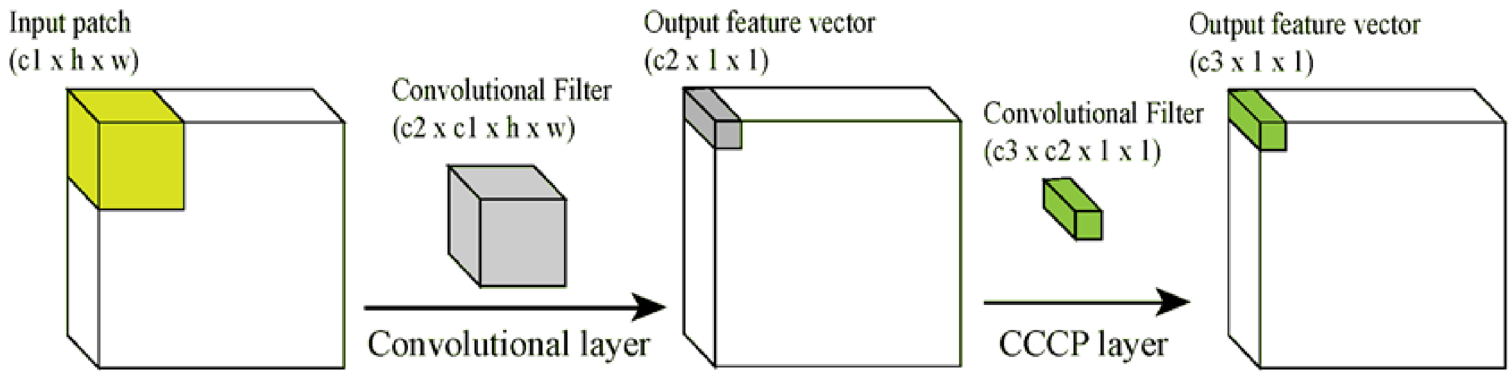
“Network in Network” (NIN)

- **NIN: CNN with non-linear filters, yet without fully-connected layers**



NIN

Better Local Abstraction ≈ Cascaded 1x1 Convolution

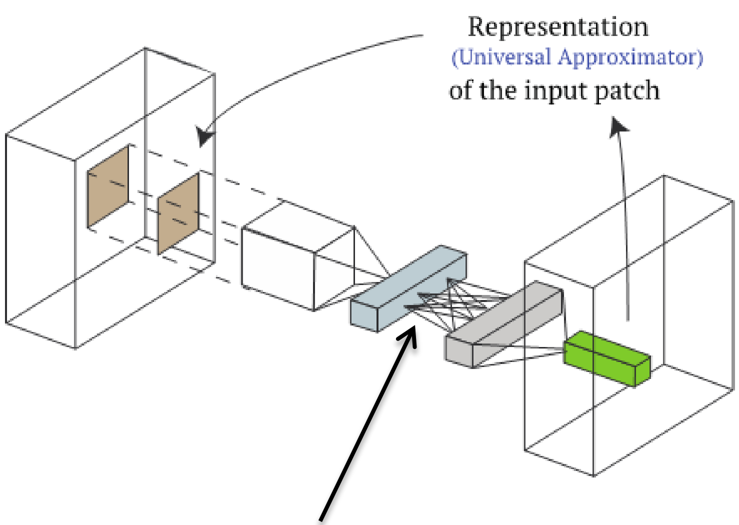


Efficient implementation of CCCP

Local patch is projected to its value in a feature map using **a small network**

$$y_i = \phi(w_i^T y_{i-1} + b_i)$$

$$y_0 = x$$

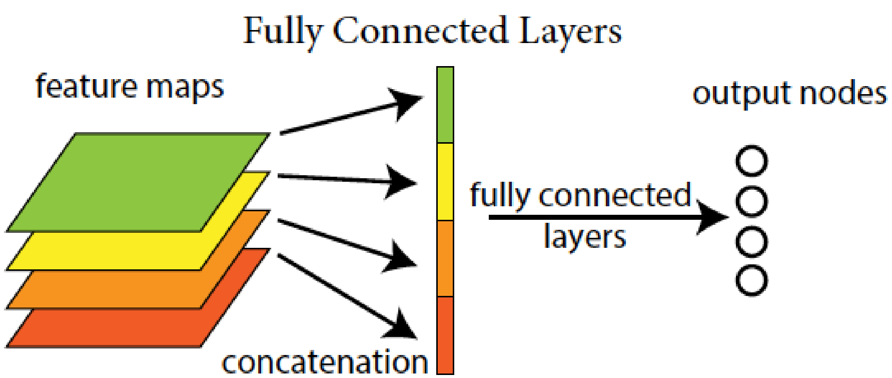


Cascaded Cross Channel Parametric Pooling (CCCP)

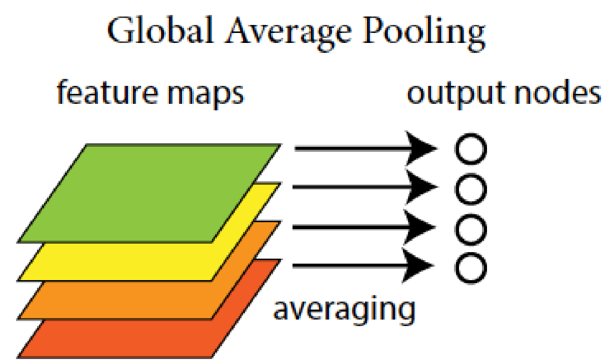
Lin, Min, Qiang Chen, and Shuicheng Yan. "Network In Network." arXiv preprint arXiv:1312.4400 (2013) & ICLR-2014

Global Average Pooling

CNN



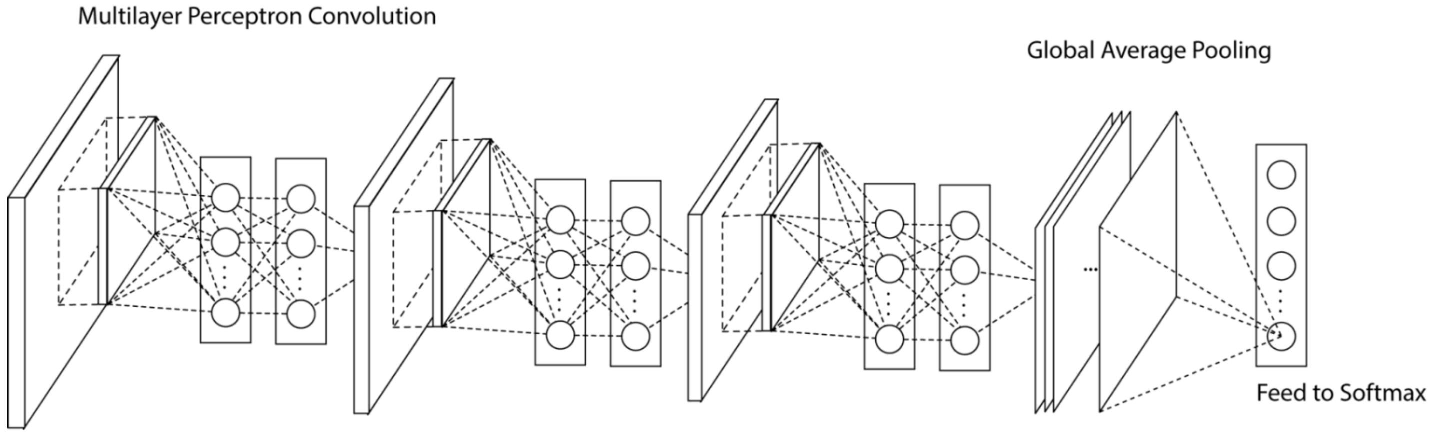
NIN



Explicitly confidence map of each category

Save a large portion of parameters

"Network in Network" (NIN) - Summary



NIN

- **Better** local abstraction, **less** global overfitting, and **much less** parameters

	Cifar-10	Cifar-100
Previous Best performance (Maxout) [1]	11.68%	38.57%
Our method	10.41%	36.30%

With less parameter #

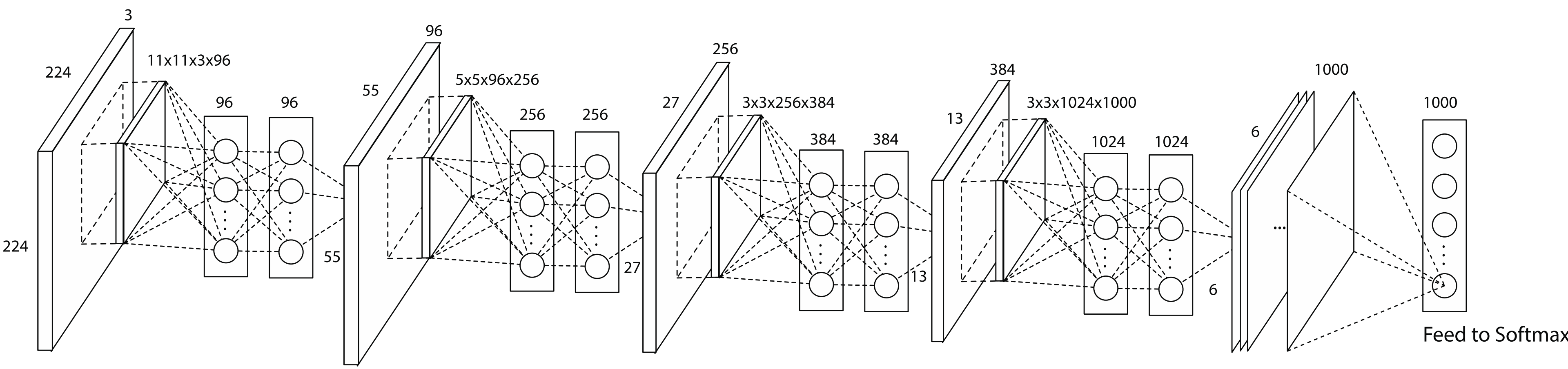
[1] Ian J. Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron C. Courville, Yoshua Bengio: Maxout Networks. ICML (3) 2013: 1319-1327

Episode-2: Network in Network, ILSVRC-2014

ILSVRC-2014: Classification + Detection

NIN for ImageNet Object Classification

A simple 4 layer NIN + Global Average Pooling:



	Parameter Number	Performance	Time to train (GTX Titan)
AlexNet	60 Million (230 Megabytes)	40.7% (Top 1)	8 days
NIN	7.5 Million (**29 Megabytes**)	39.2% (Top 1)	4 days

NIN for ImageNet Object **Classification**

To avoid hyper-parameter tuning, we put cccp layer directly on convolution layers of ZFNet (Network in ZFNet)

layer	details
Conv1	Stride = 2, kernel = 7x7, channel_out = 96
Conv2	Stride = 2, kernel = 5x5, channel_out = 256
Conv3	Stride = 1, kernel = 3x3, channel_out = 512
Conv4	Stride = 1, kernel = 3x3, channel_out = 1024
Conv5	Stride = 1, k[illegible] chan[illegible]
Fc1	Output [illegible]
Fc2	Output = 4096
Fc3	Output = 1000

Not deep enough!

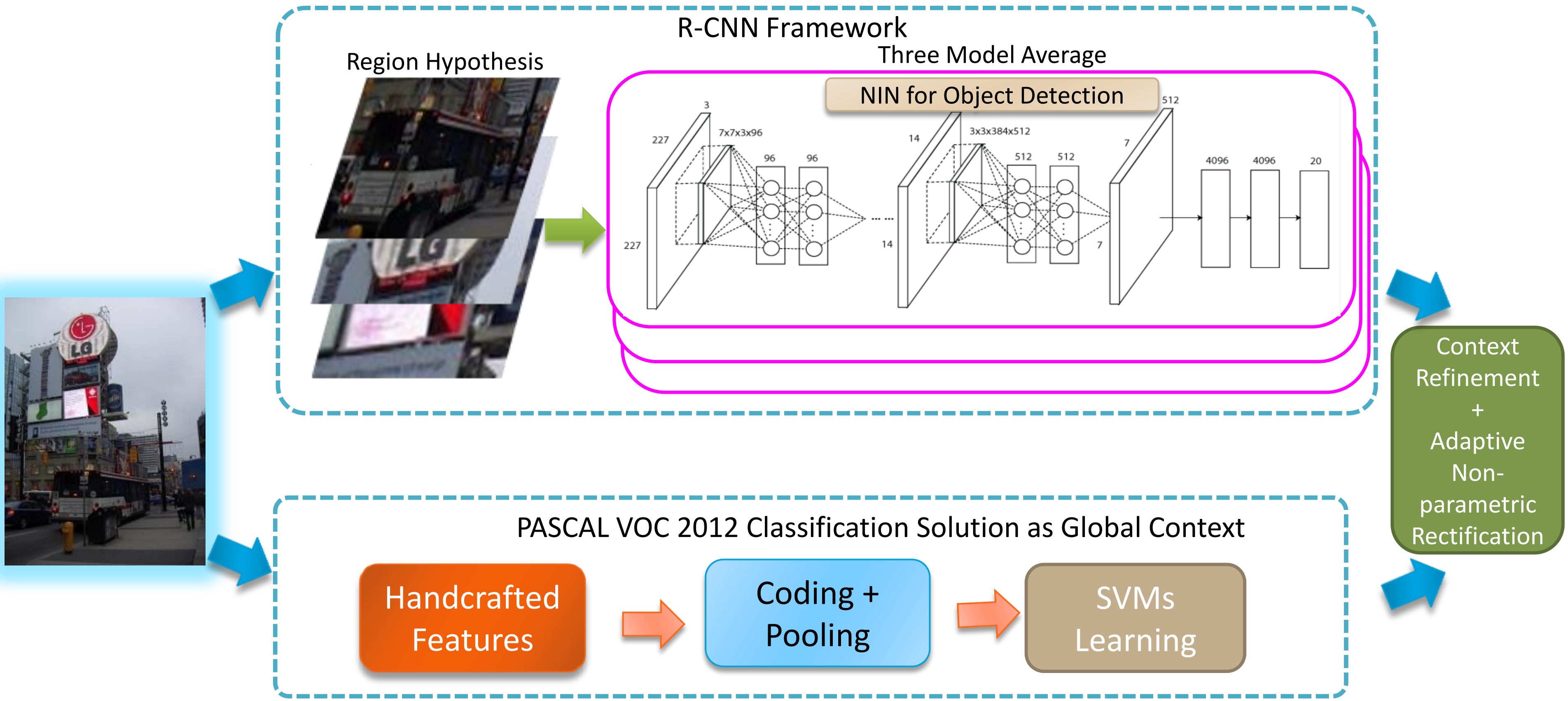
layer	details
Conv1	Stride = 2, kernel = 7x7, channel_out = 96
Cccp1	Output = 96
Conv2	Stride = 2, kernel = 5x5, channel_out = 256
Cccp2	Output = 256
Conv3	Stride = 1, kernel = 3x3, channel_out = 512
Cccp3	Output = 256
Conv4	Stride = 1, kernel = 3x3, channel_out = 1024
Cccp4	Output = 512
Cccp5	Output = 384
Conv5	Stride = 1, kernel = 3x3, channel_out = 512
Cccp6	Output = 256
Fc1	Output = 4096
Fc2	Output = 4096
Fc3	Output = 1000

(**10.91% for 1 model, 9.79% for 3 models**) with 256xN training and 3 view test

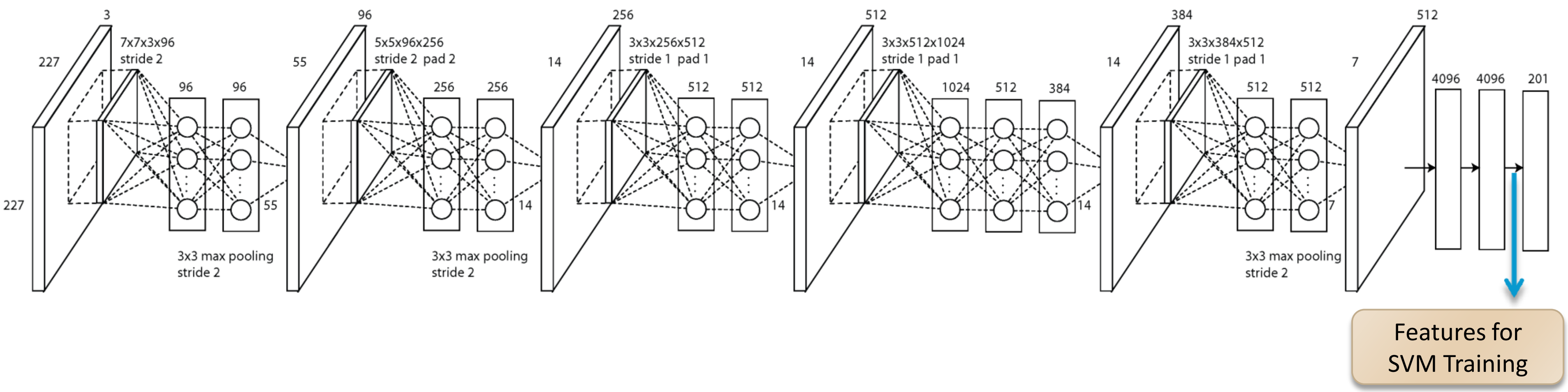
Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks." Computer Vision–ECCV 2014.

NUS National University of Singapore

NIN for ImageNet Object **Detection**



NIN for ImageNet Object Detection



- Pre-training the NIN on ILSVRC14 detection train set
- Fine-tuning on train (partial) + validation set
  - Discard the parameters of the last layer

NIN for ImageNet Object Detection

- Results on validation set (0.5:0.5 of val set for validation and testing)

Submission	Method	MAP
NIN	the baseline result by using NIN as feature extractor for RCNN	35.61%
3 NINs	Using concatenated features from multiple NIN as feature extractor for RCNN	36.52% (↑0.91%)
3 NINs + ctx	adaptive non-parametric rectification of outputs from both object detectors and global context	37.49% (↑0.97%)

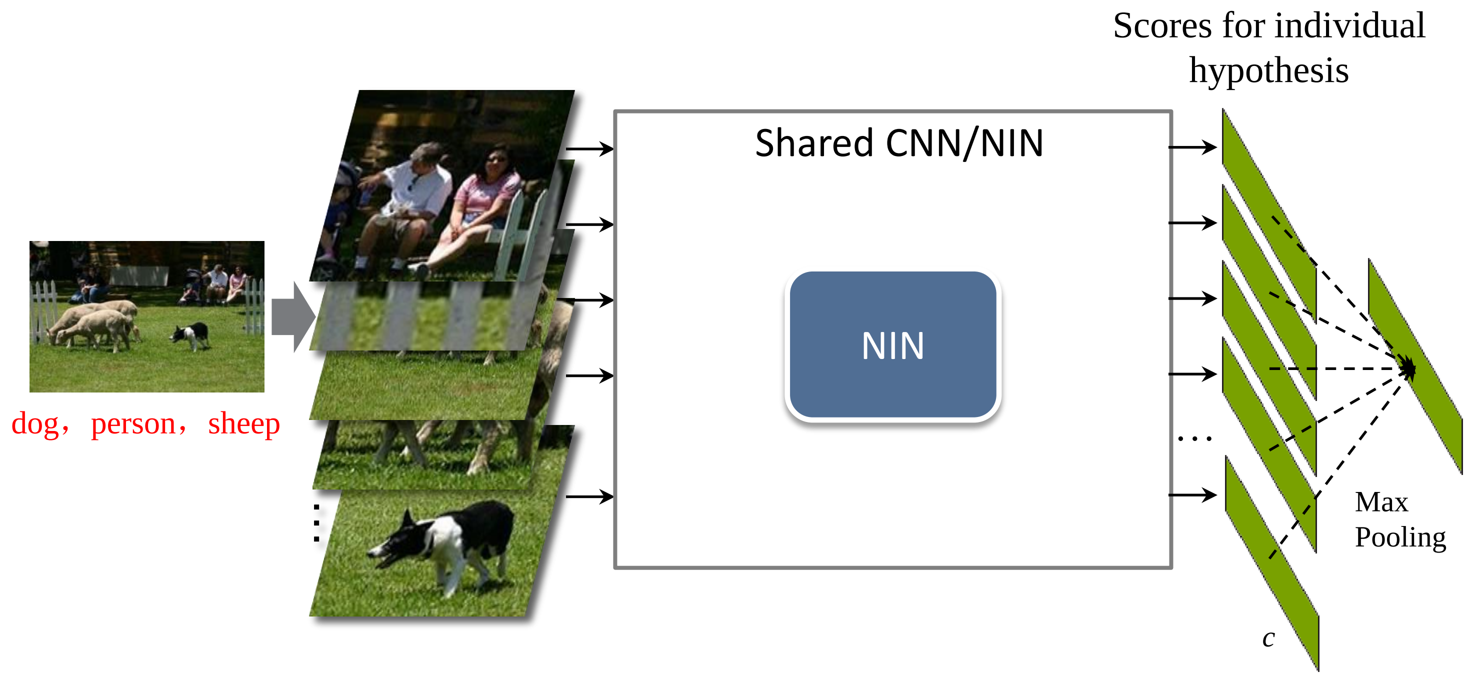
- Results on test set

3 NINs + ctx	adaptive non-parametric rectification of outputs from both object detectors and global context	37.212%

NUS National University of Singapore

PASCAL-VOC-2012: Multi-label Classification

NIN for Pascal VOC Multi-label Classification



- **HCP** = Hypotheses + CNN + Pooling
- **No ground-truth bounding box** is required for training
- The proposed infrastructure is **robust** to the **noisy and/or redundant hypotheses**
- **No explicit hypothesis label** is required for training
- The shared CNN can be well **pre-trained** with a large-scale single-label image dataset

Yunchao Wei, Shuicheng YAN. "CNN: single-label toi multi-label", http://arxiv.org/abs/1406.5726, Jun 22, 2014.

Classification State-of-the-arts on VOC 2012

Pre-trained with 1000 classes

Category	NUS-PSL[1]	PRE-1000C[2]	PRE-1512[2]	Chatfield *et al.*[3]	HCP	HCP+NUS-PSL
plane	97.3	93.5	94.6	96.8	98.4	**99.5**
bicycle	84.2	78.4	82.9	82.5	89.5	**93.7**
bird	80.8	87.7	88.2	91.5	96.2	**96.8**
boat	85.3	80.9	84.1	88.1	91.7	**94.0**
bottle	60.8	57.3	60.3	62.1	72.5	**77.7**
bus	89.9	85.0	89.0	88.3	91.1	**95.3**
car	86.8	81.6	84.4	81.9	87.2	**92.4**
cat	89.3	89.4	90.7	94.8	97.1	**98.2**
chair	75.4	66.9	72.1	70.3	73.0	**86.1**
cow	77.8	73.8	86.8	80.2	89.5	**91.3**
table	75.1	62.0	69.0	76.2	75.1	**83.5**
dog	83.0	89.5	92.1	92.9	96.3	**97.3**
horse	87.5	83.2	93.4	90.3	93.0	**96.8**
motor	90.1	87.6	88.6	89.3	90.5	**96.3**
person	95.0	95.8	**96.1**	95.2	94.8	95.8
plant	57.8	61.4	64.3	57.4	66.5	**72.2**
sheep	79.2	79.0	86.6	83.6	90.3	**91.5**
sofa	73.4	54.3	62.3	66.4	65.8	**81.1**
train	94.5	88.0	91.1	93.5	95.6	**97.6**
tv	80.7	78.3	79.8	81.9	82.0	**90.0**
MAP	82.2	78.7	82.8	83.2	86.8	91.4

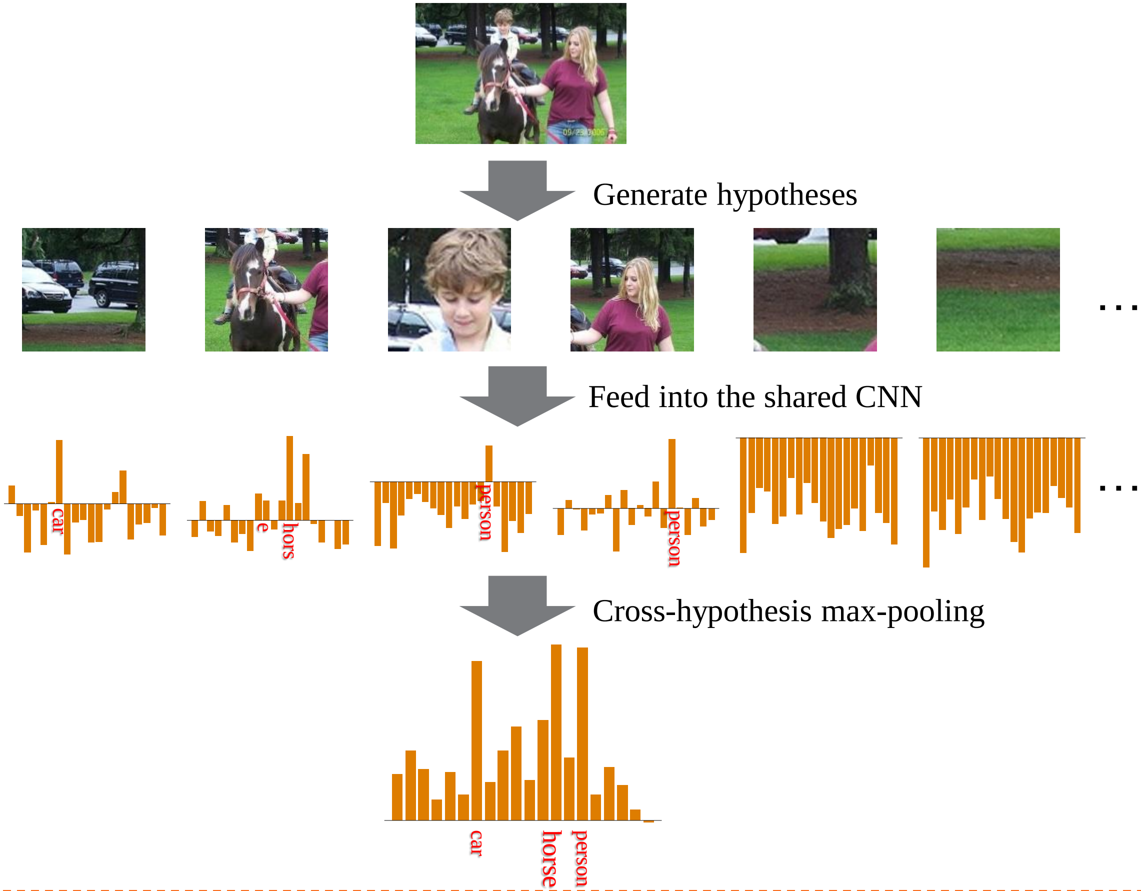
[1] S. Yan, J. Dong, Q. Chen, Z. Song, Y. Pan, W. Xia, H. Zhongyang, Y. Hua, and S. Shen. Generalized hierarchical matching for subcategory aware object classification. In Visual Recognition Challange workshop, ECCV, 2012.

[2] M. Oquab, L. Bottou, I. Laptev, and J. Sivic. Learning and transferring mid-level image representations using convolutional neural networks. CVPR, 2014.

[3] K. Chatfield, K. Simonyan, A. Vedaldi, A. Zisserman. Return of the Devil in the Details: Delving Deep into Convolutional Nets , BMVC, 2014

NIN for Pascal VOC Multi-label **Classification**

- Online demo with 1~1.5s per image



Further Work

- What's next?

Definitely: "Deeper NIN"

Great Appreciations to the Supports

NUS
National University
of Singapore