Paper review

Zeiler, Matthew D., and Rob Fergus.
"Visualizing and understanding convolutional networks.“
European Conference on Computer Vision.
Springer International Publishing, 2014.

Comprehension of deep-learning

- Visualizing and Understanding Convolutional Networks

17.01.06 You Sung Min

Smart **U**biquitous
Healthcare Lab.
Hanyang University

Contents

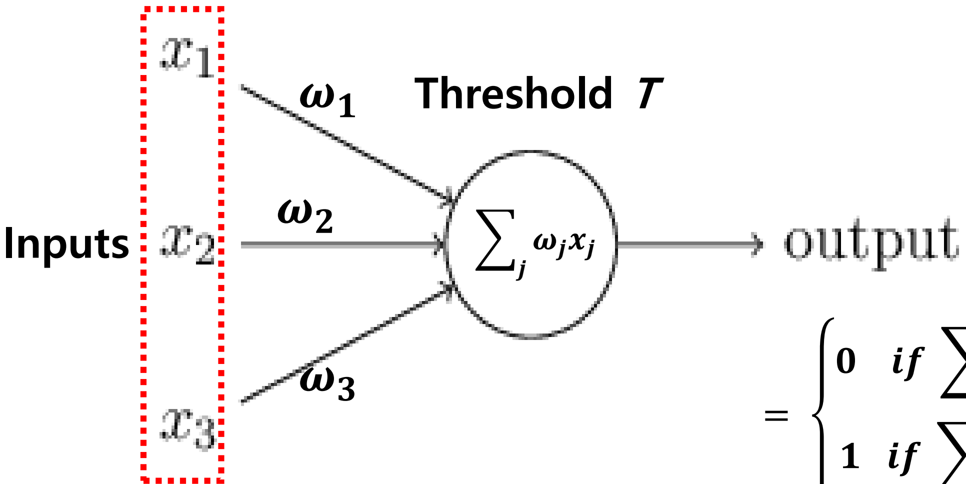
1. **Review of Deep learning (Convolutional Neural Network)**

2. **Visualization of CNN**

3. **Feature generalization (Transfer learning)**

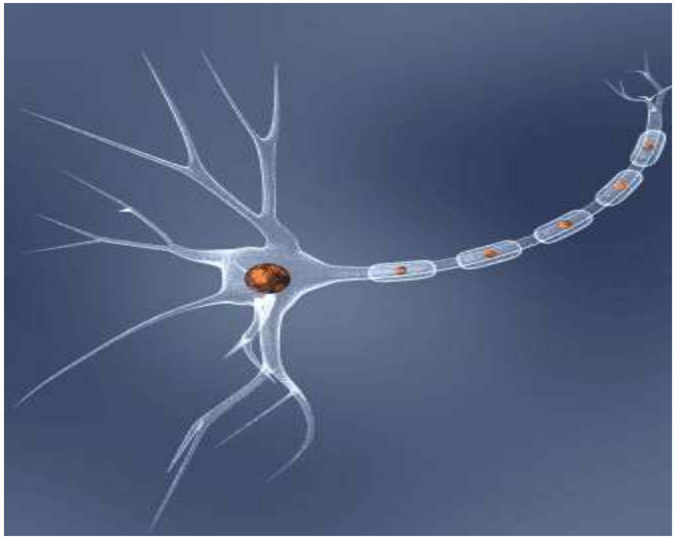
Review of Deep learning

❖Structure of Neural Networks



$$= \begin{cases} 0 \quad if \sum_j \omega_j x_j \le T \\ 1 \quad if \sum_j \omega_j x_j > T \end{cases}$$

Perceptron (1950)

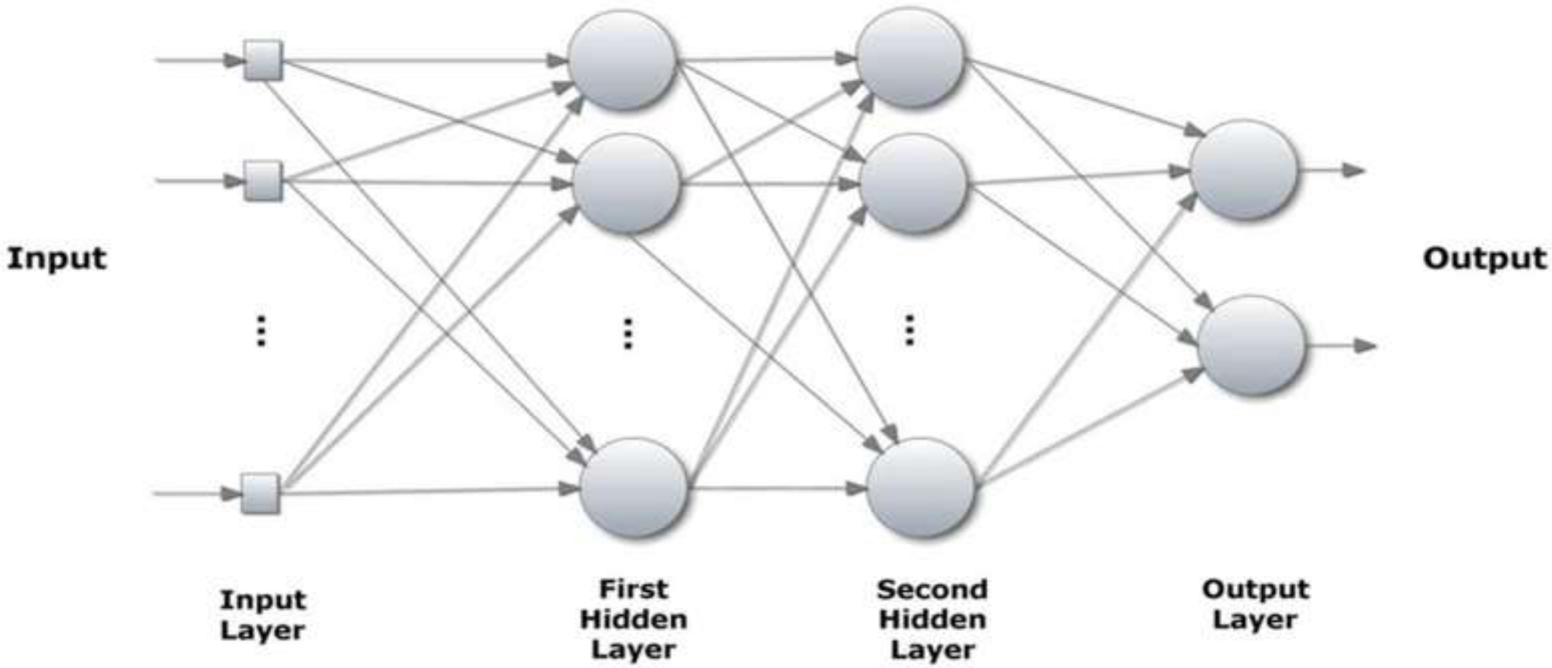


Neuron

- A simple model to **emulate a single neuron**
- This model **produces a binary output**

Review of Deep learning

❖ Multilayer Perceptron (MLP)



- A network model **consists of perceptrons**
- This model **produces vectorized outputs**

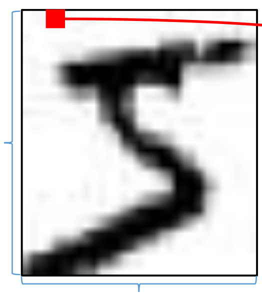
Review of Deep learning

❖Multilayer Perceptron (MLP)

Desired output for “5”

$y(x) = (0, 0, 0, 0, 1, 0, 0, 0, 0)^T$

Binary Input
(Intensity of a pixel)

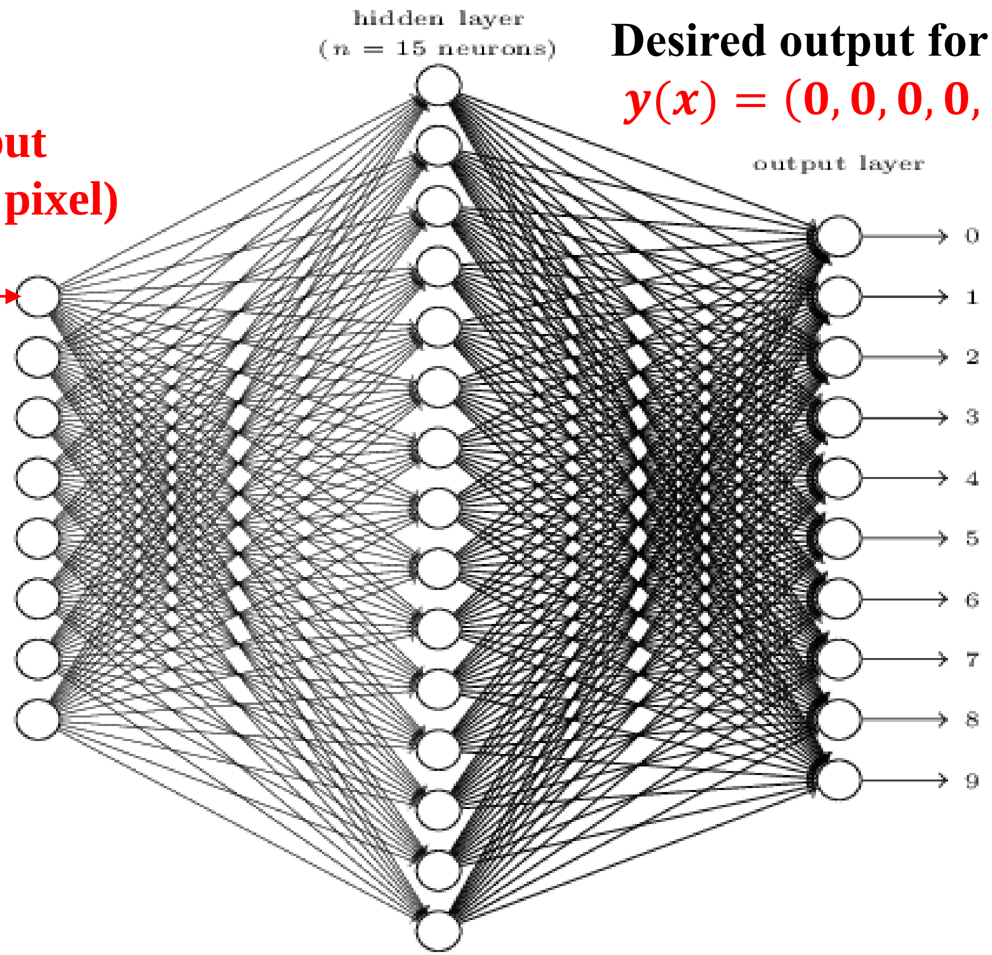


28

28

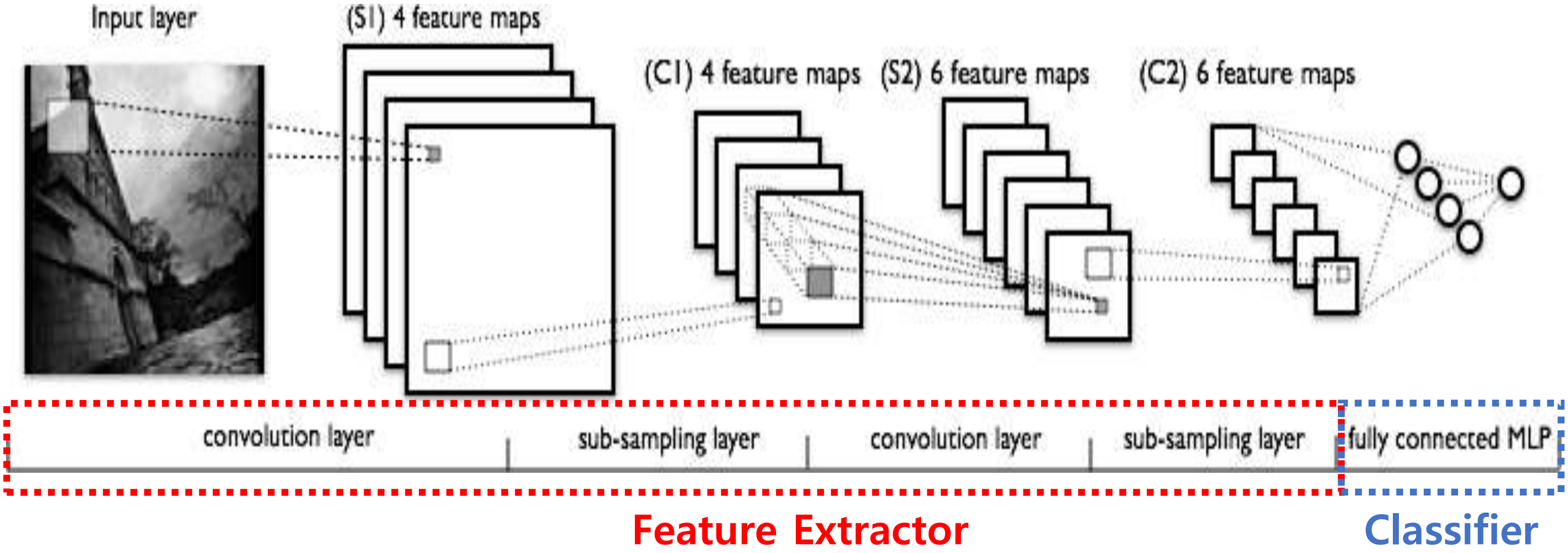
Input
(784)

Handwritten digit with
28 by 28 pixel image



Review of Deep learning

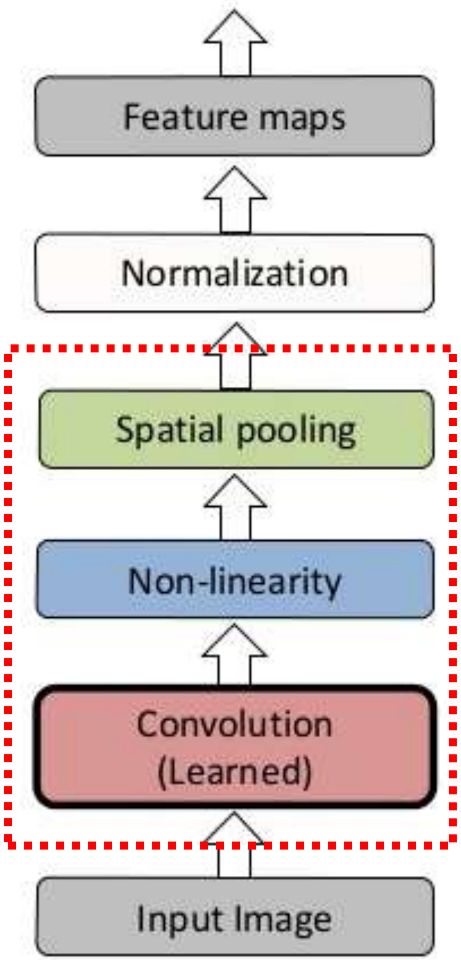
❖Convolutional Neural Network

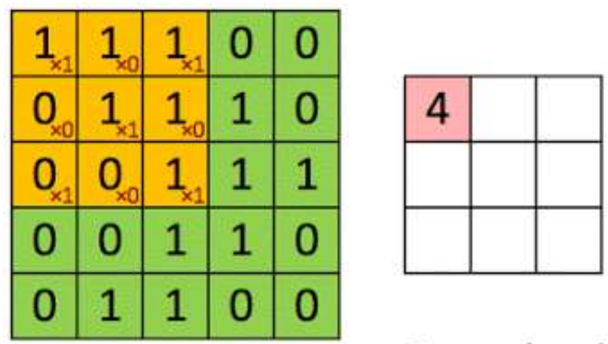


- Convolution layer
- Subsampling (Pooling) layer
- Rectified Linear Unit(ReLU)

Review of Deep learning

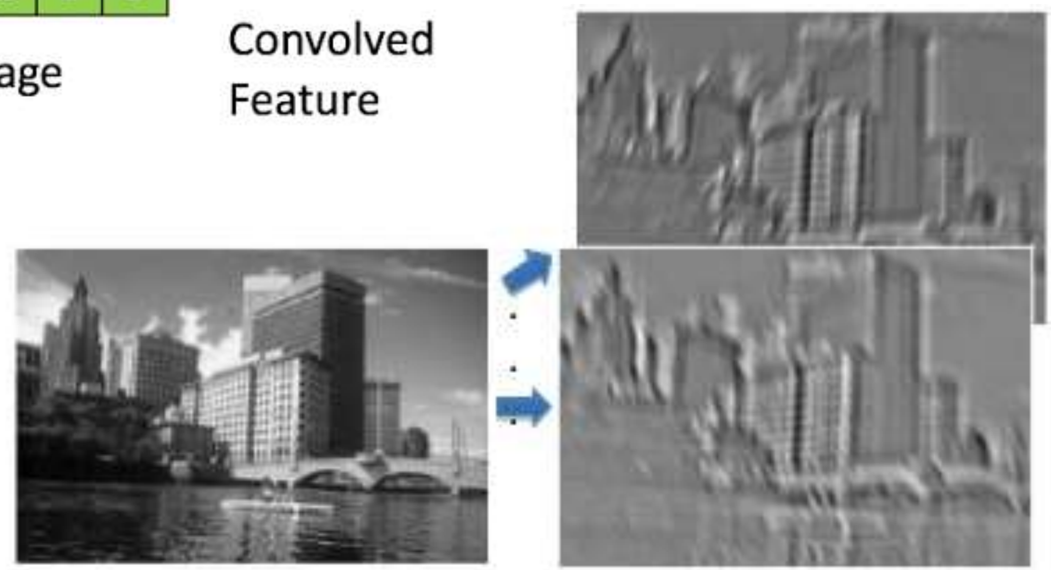
❖Convolutional Neural Network





Image

Convolved Feature

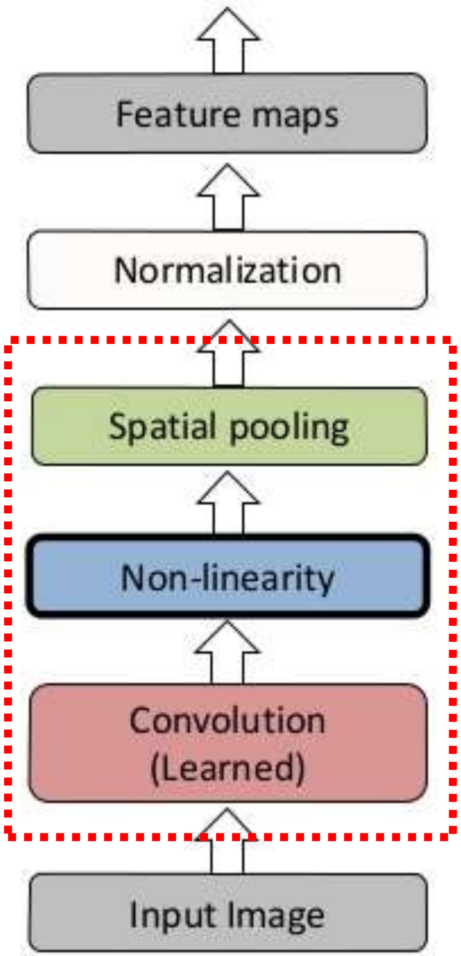


Input

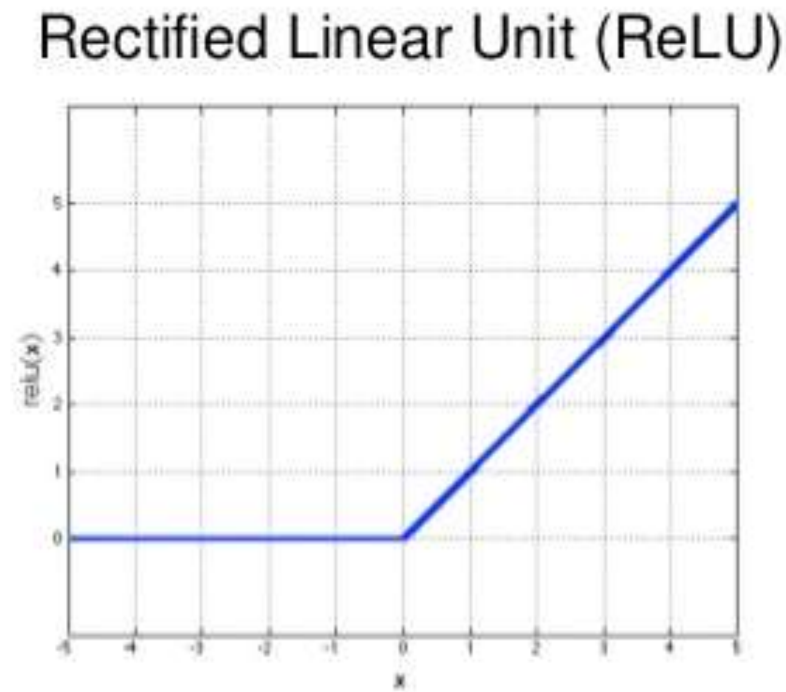
Feature Map

Review of Deep learning

❖Convolutional Neural Network

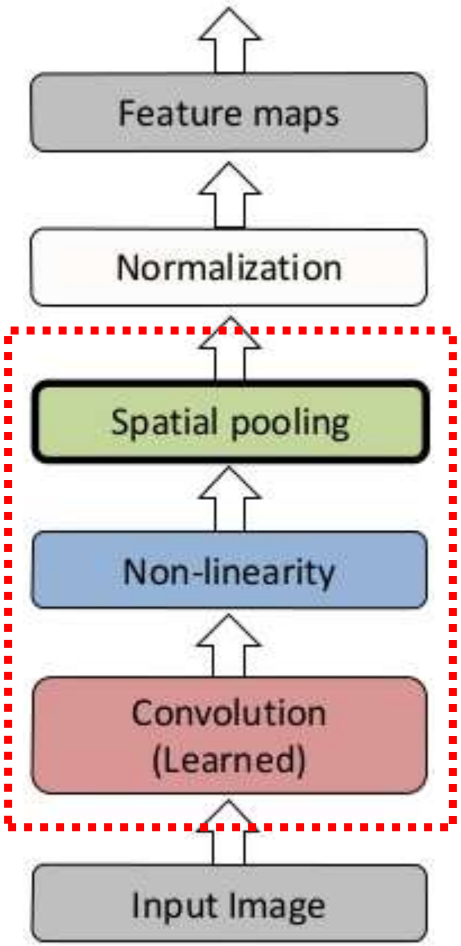


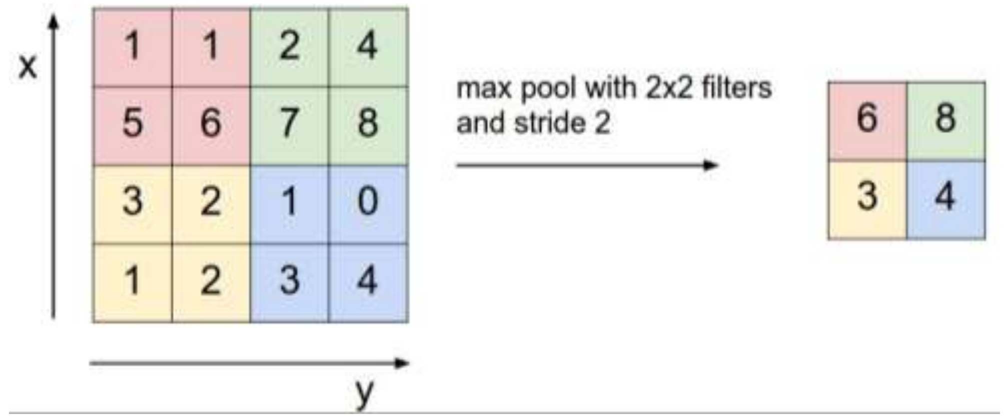
$$y = \max(x,0)$$

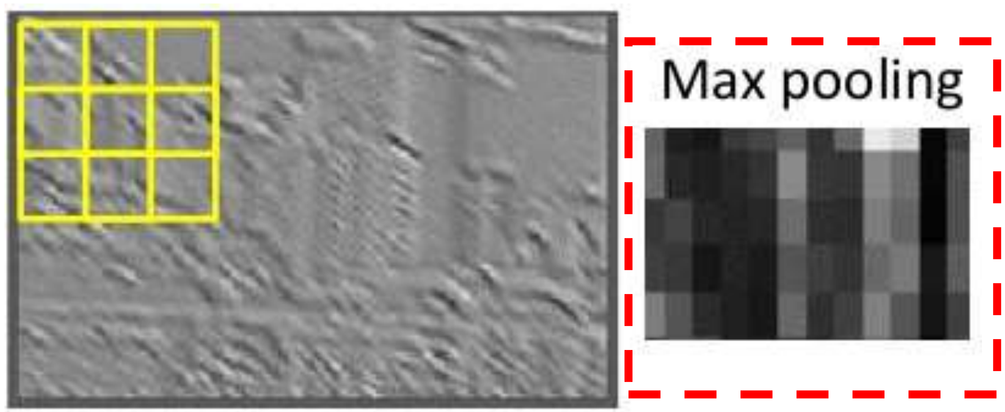


Review of Deep learning

❖Convolutional Neural Network



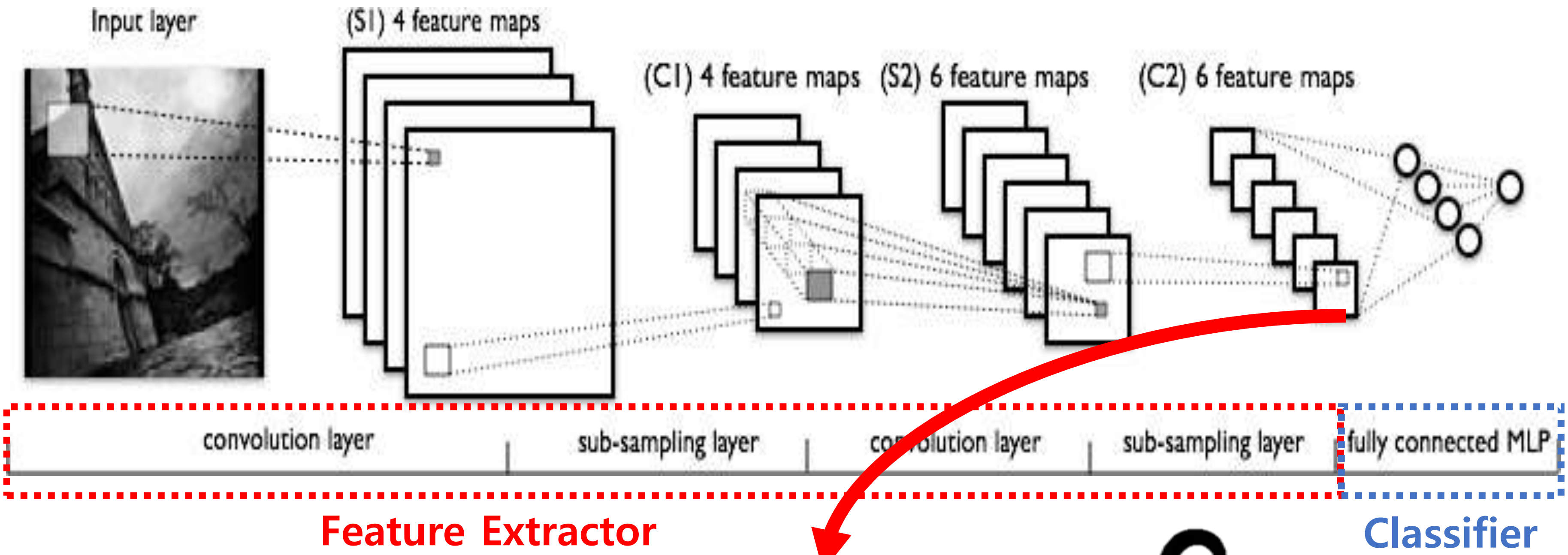


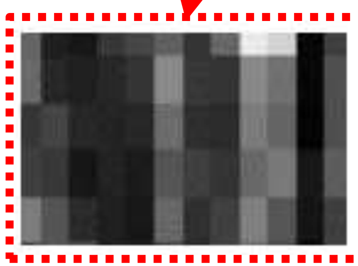


Feature map

Review of Deep learning

❖Convolutional Neural Network





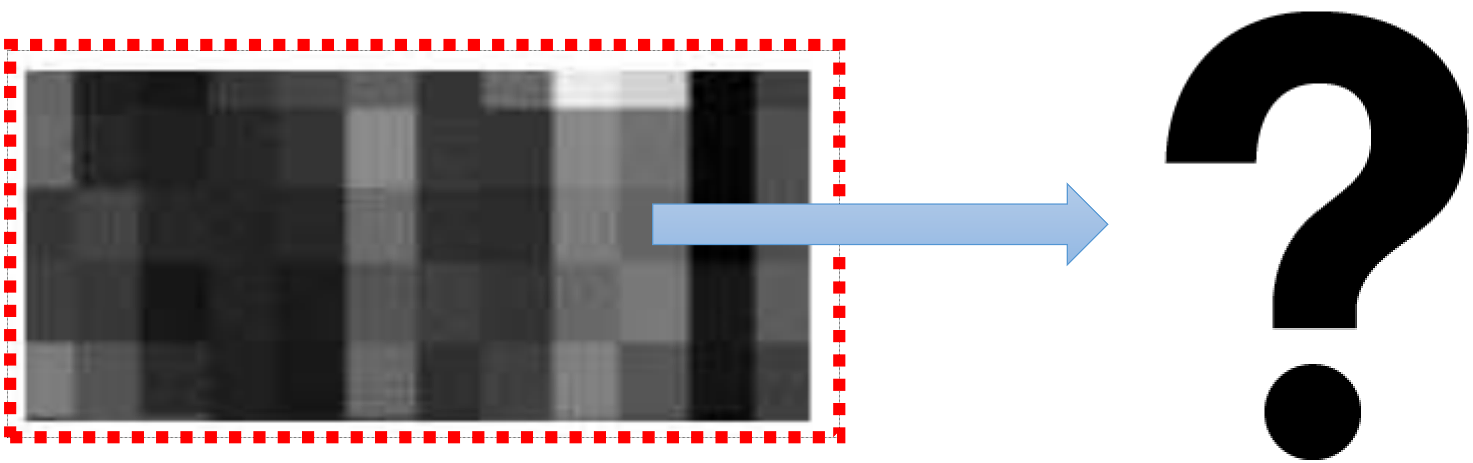
Feature map

Visualization of CNN

❖Deconvnet (Deconvolutional Network)

- Mapping the **activations** back to **the input pixel space**
- **What input pattern caused activation** in the feature map

→ **Reconstruct input space with feature map**

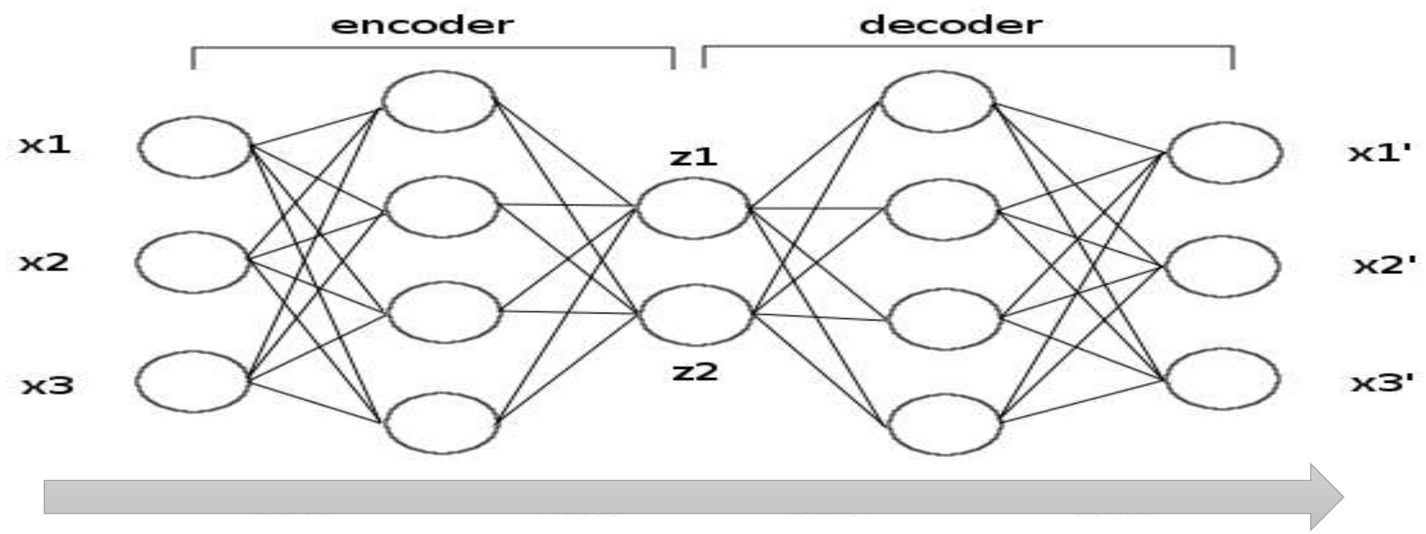


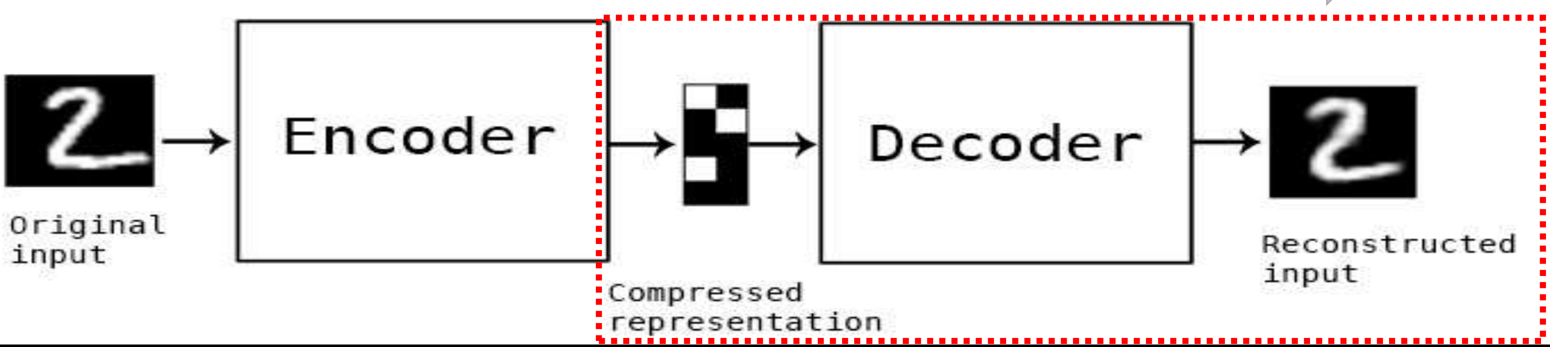
Feature map

Visualization of CNN

❖Stacked-Autoencoder (SAE)

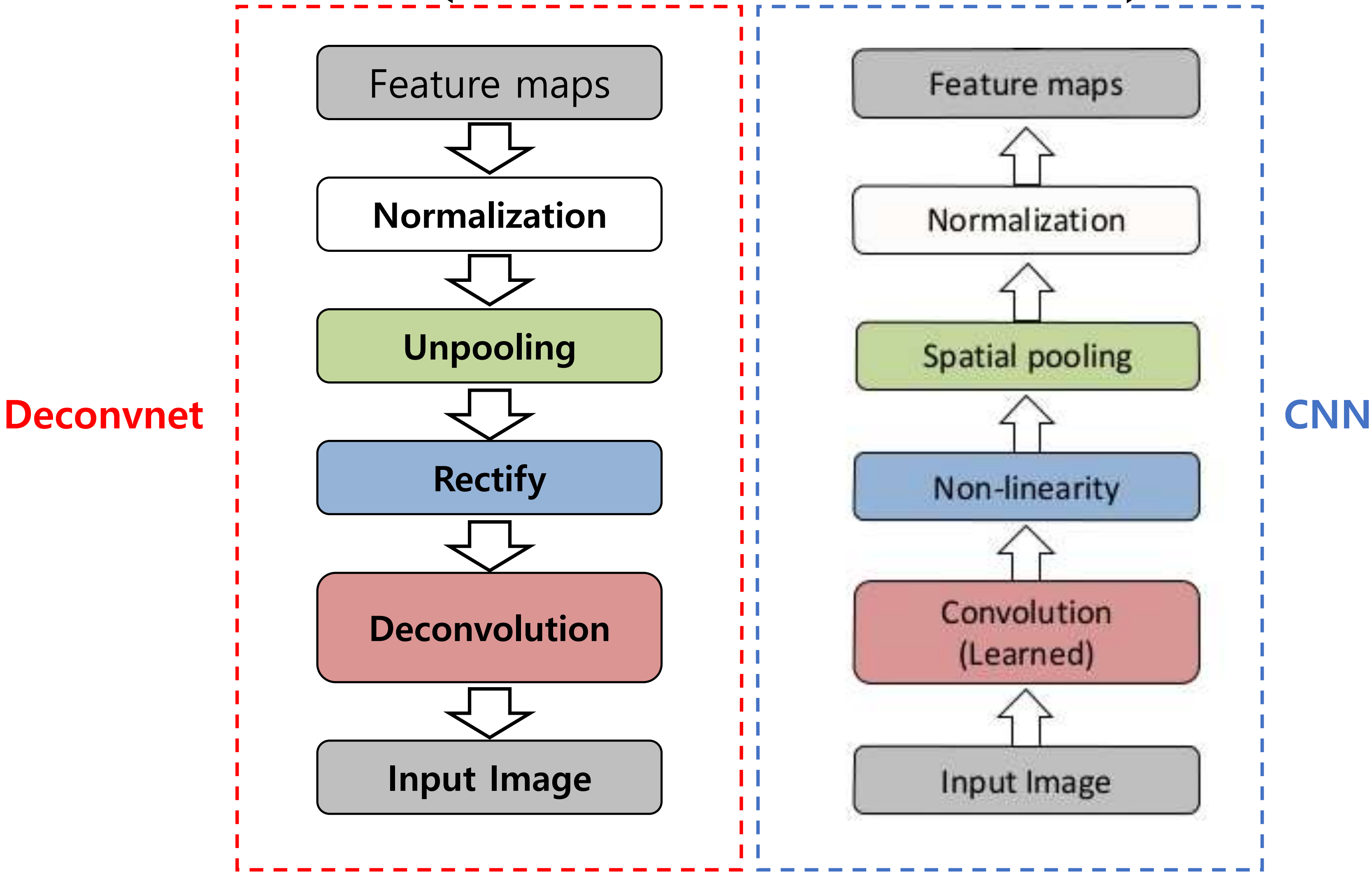
- Generative model with RBM
- Produce **same output with the input**





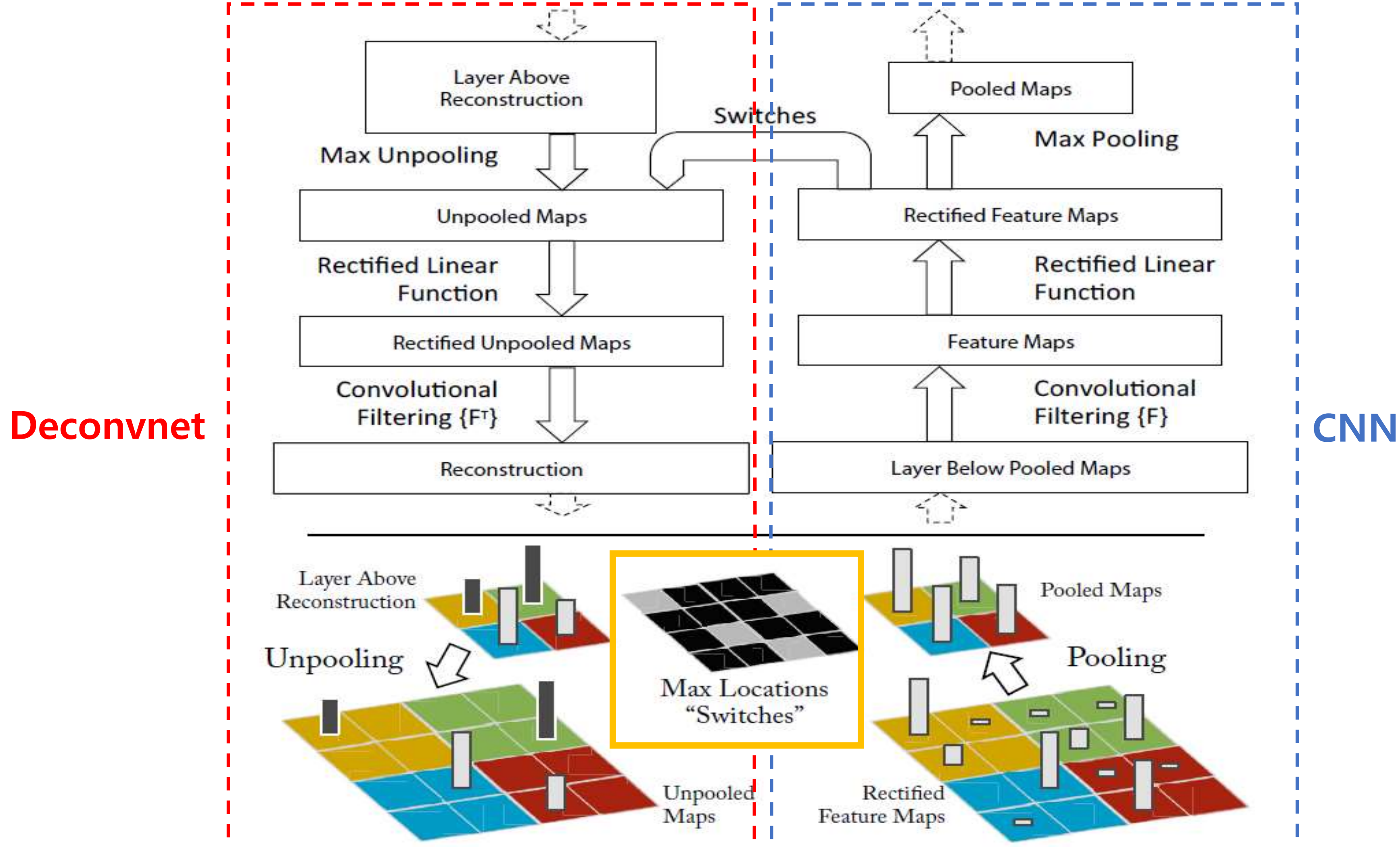
Visualization of CNN

❖Deconvnet (Deconvolutional Network)



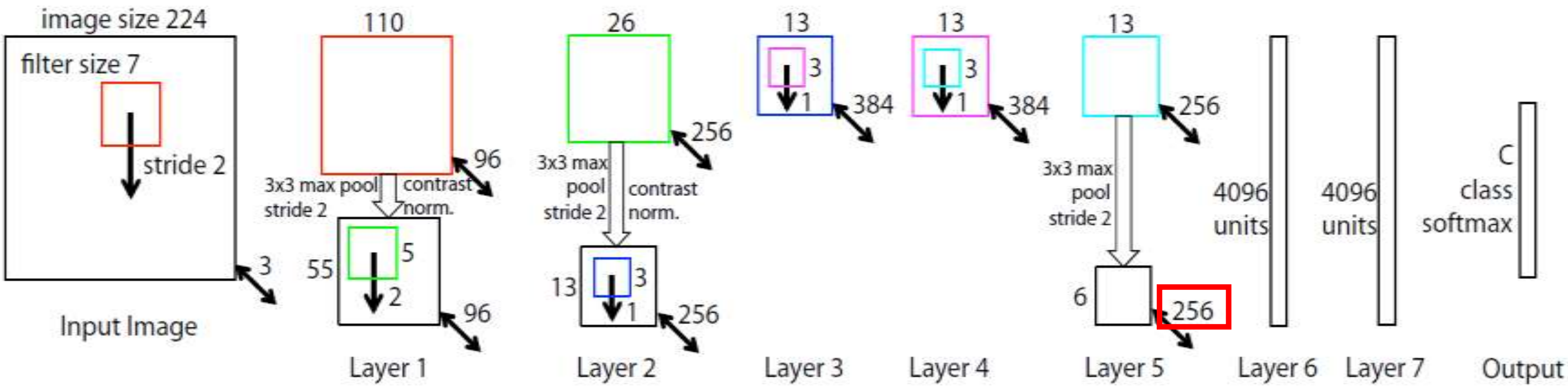
Visualization of CNN

❖Deconvnet (Deconvolutional Network)



Visualization of CNN

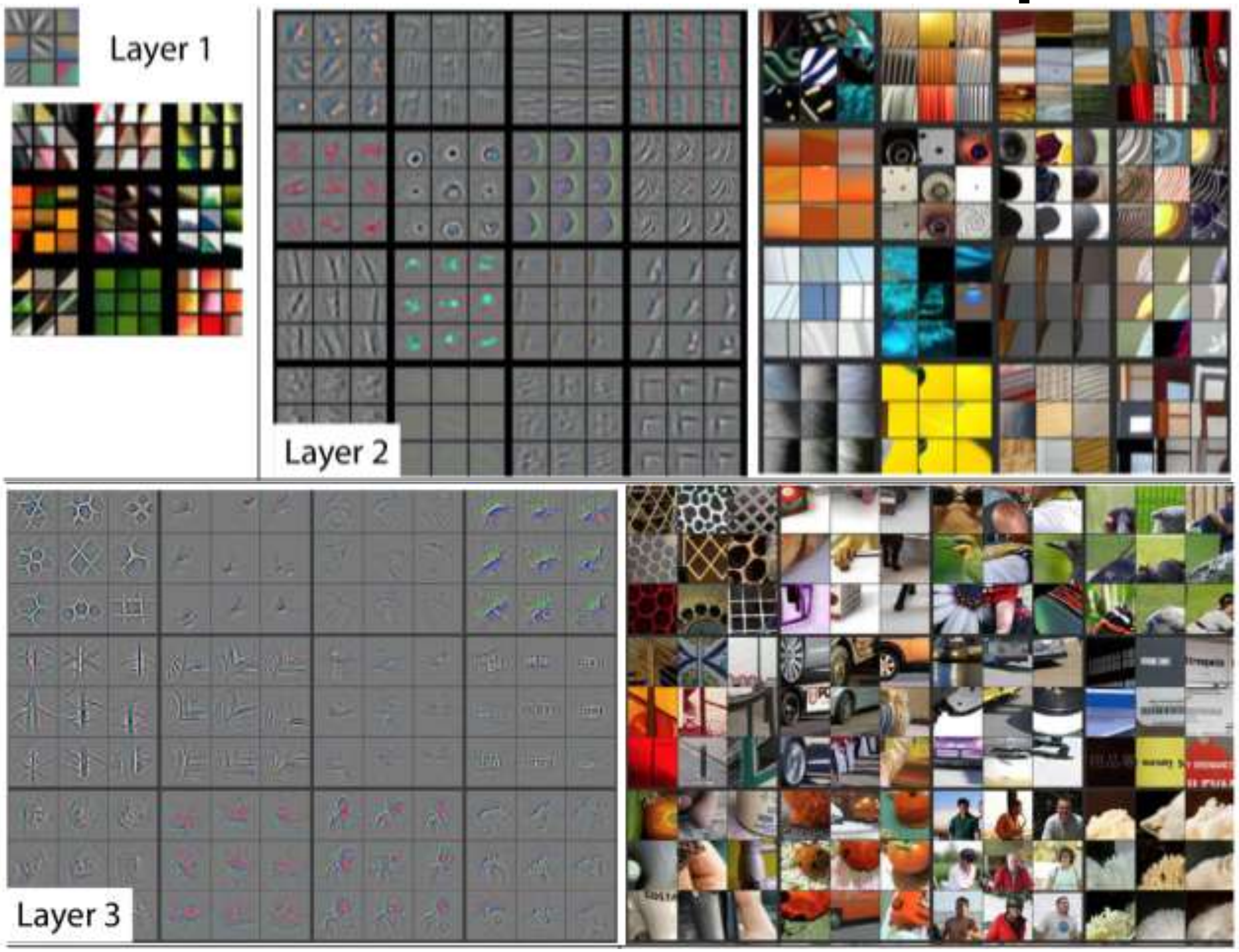
❖ Architecture of network



- CNN with 8 layers (5 as convolution, 3 for MLP)
- Trained with ImageNet 2012
  - 1.3 million images with 1000 classes
- Train took around 12 days with GTX 580

Visualization of CNN

❖Visualization of feature map



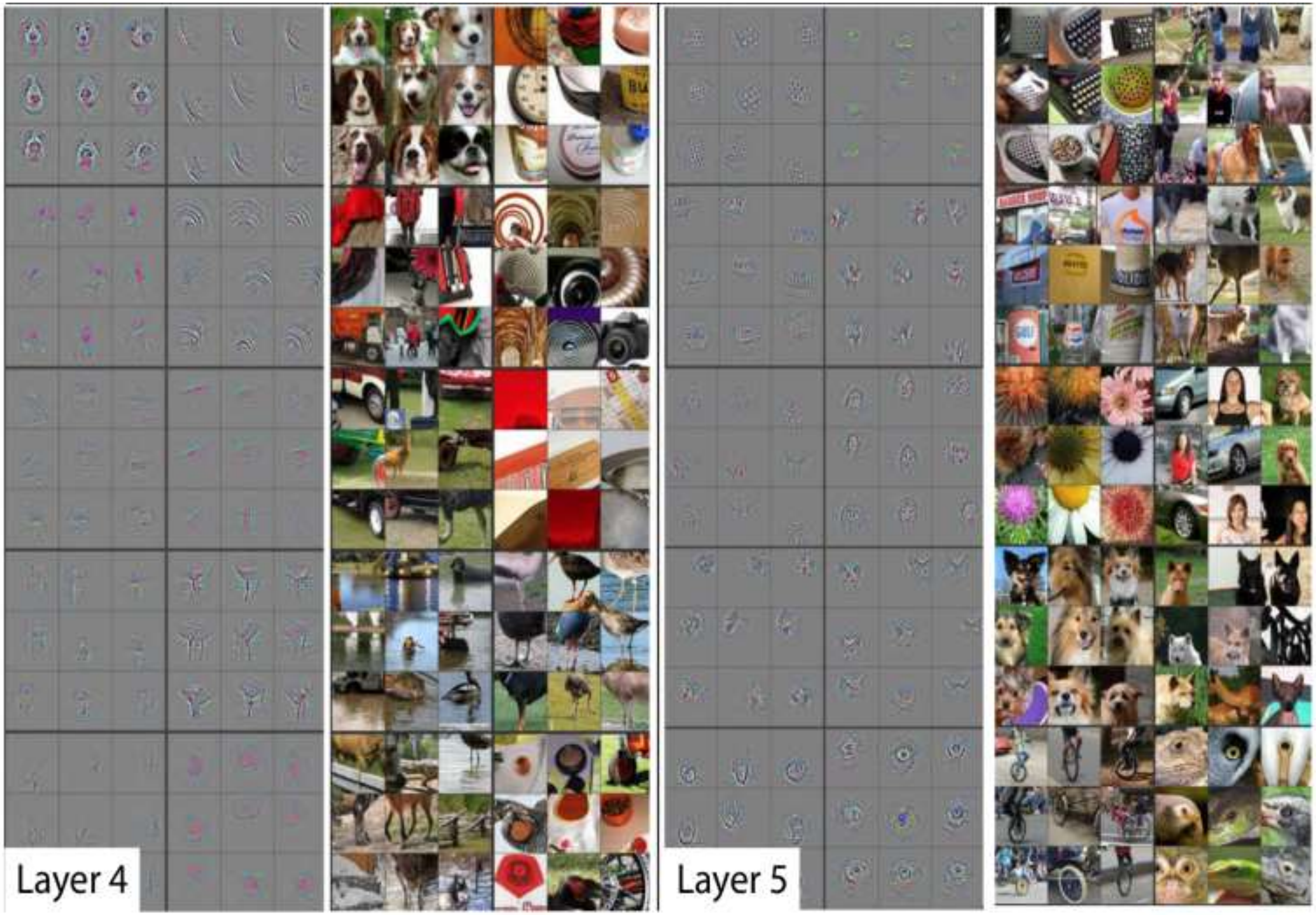
Layer 2
- Corner, Edge

Layer 3
- Texture, Text

Reconstructed Image **Corresponding input images**

Visualization of CNN

❖Visualization of feature map



Layer 4
- Object

Layer 5
- Object with pose variation

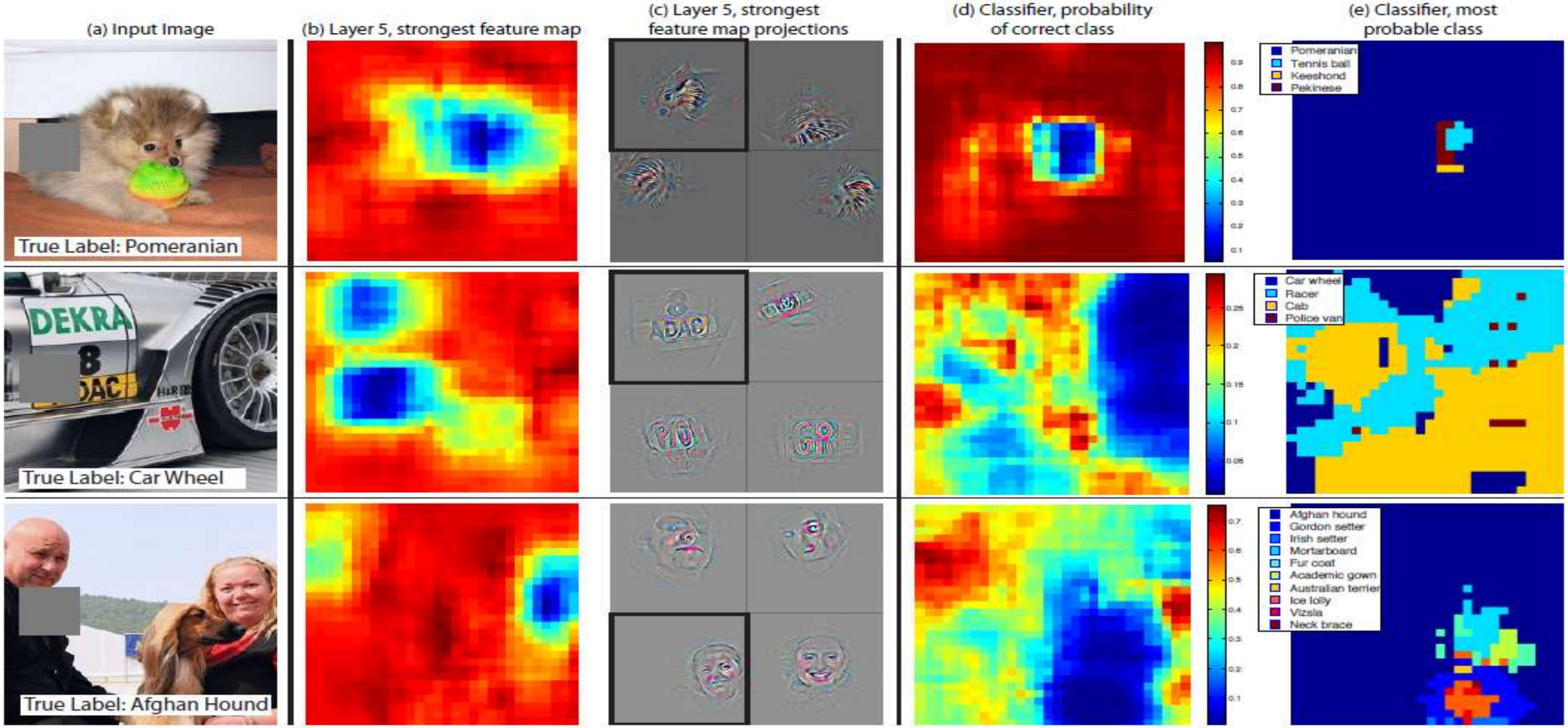
Visualization of CNN

❖Visualization of feature map

- The network is trained **discriminatively,**
  those features maps (strong activations) **shows which part of the input image are discriminative**

Visualization of CNN

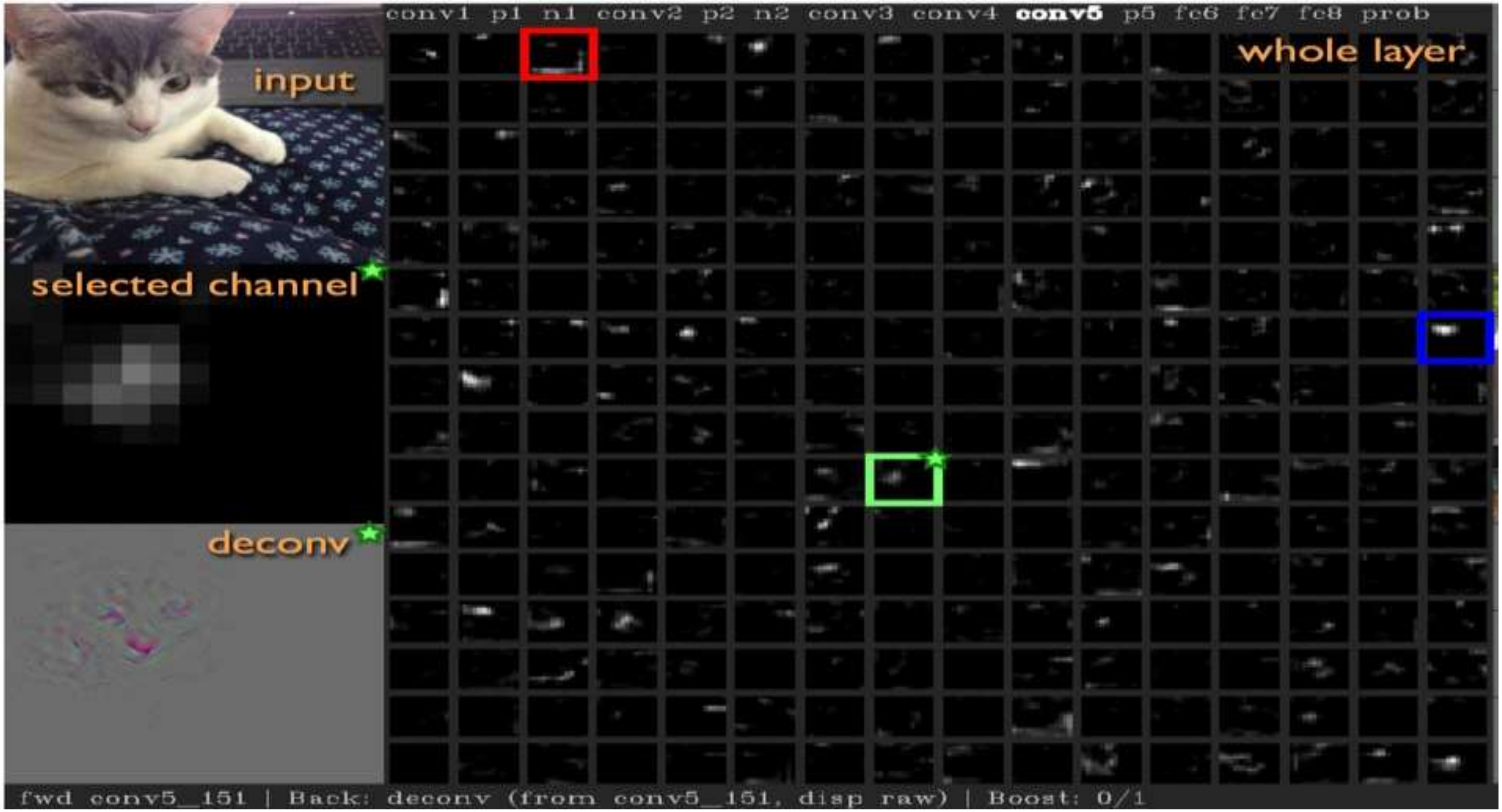
❖Effect of occlusion



- Changes in output and feature map with different portions of gray square

Visualization of CNN

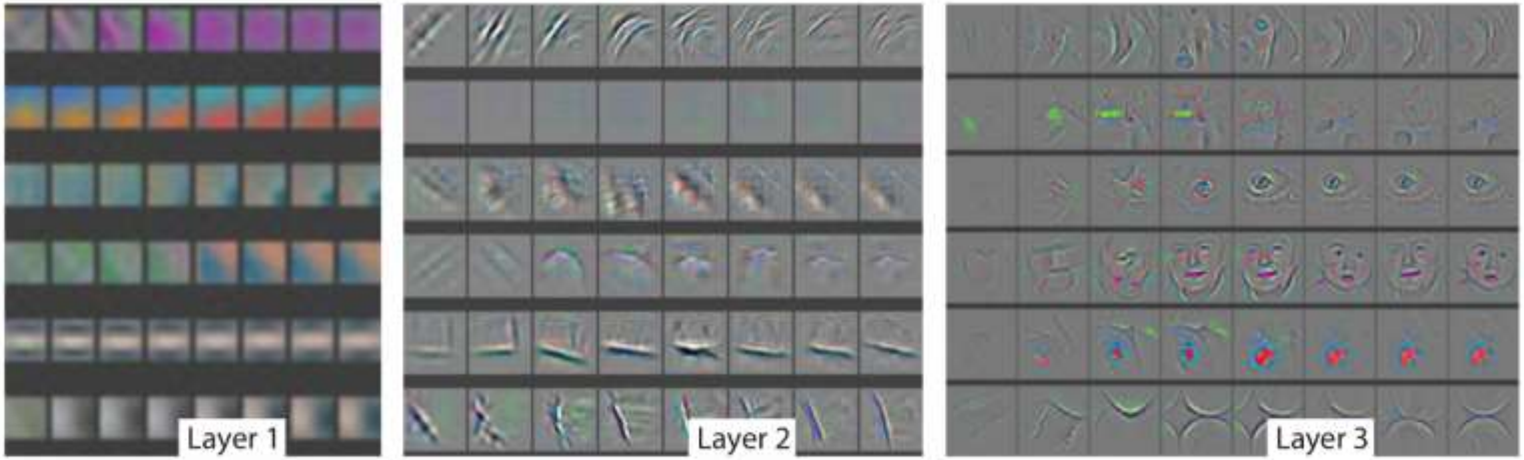
❖Visualization of feature map



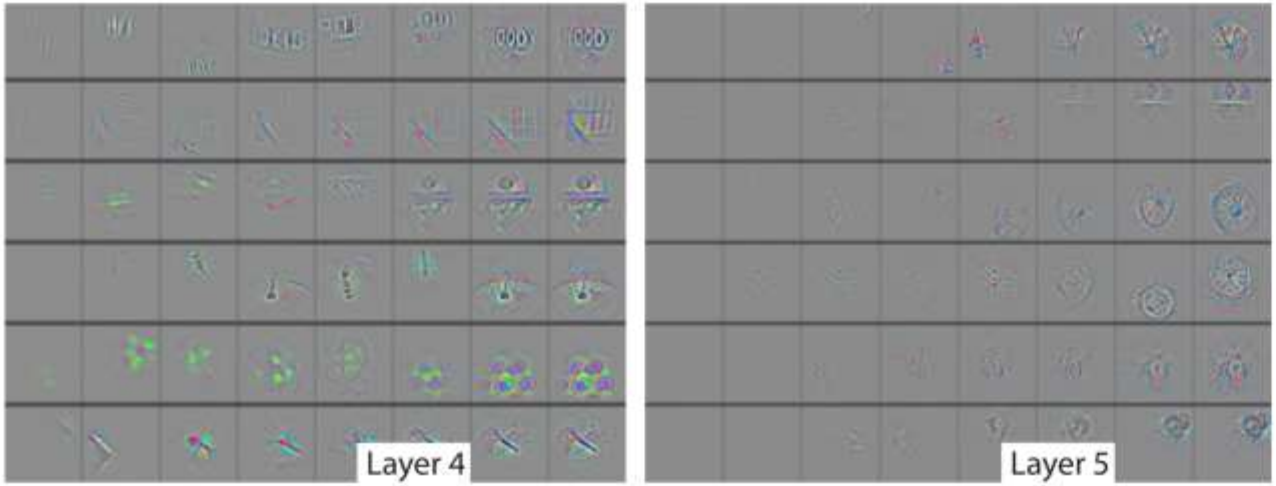
Yosinski, Jason, et al.
"Understanding neural networks through deep visualization."

Visualization of CNN

❖Feature Evolution during Training



Epoch
=[1, 2, 5, 10, 20, 30, 40, 64]

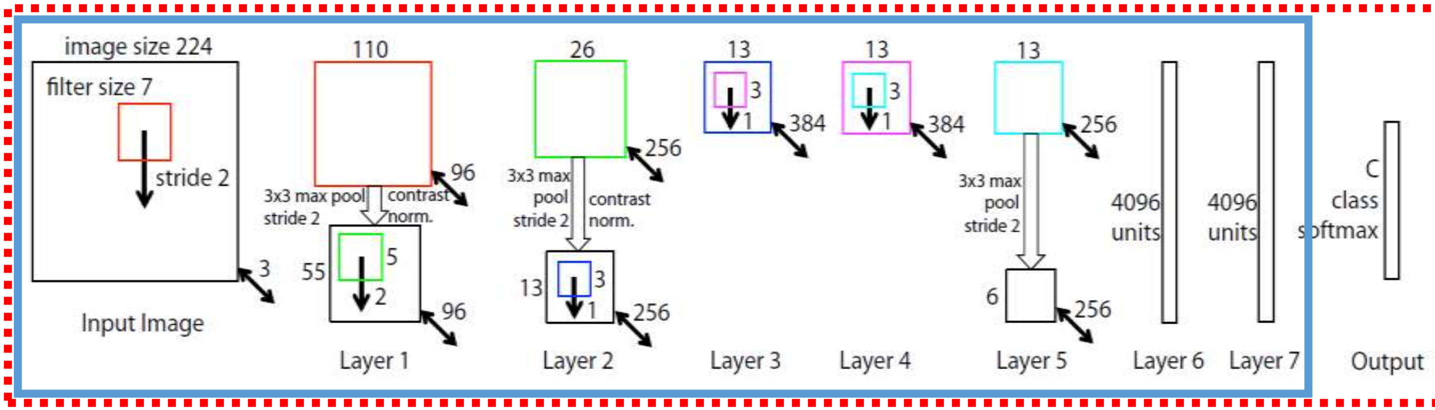


Feature generalization

❖Transfer learning

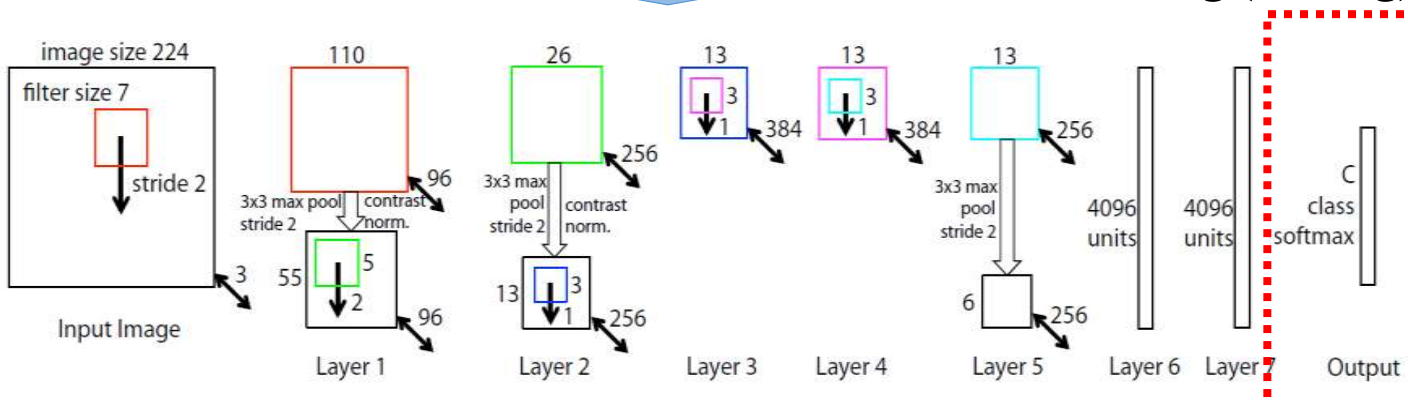
Training

ImageNet



Training (Tuning)

Caltech PASCAL



Feature generalization

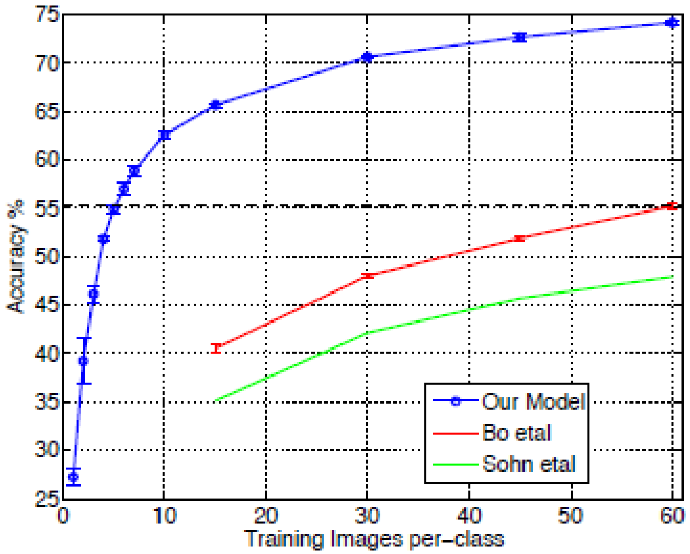
- **Caltech 101 classification accuracy**

# Train	Acc % 15/class	Acc % 30/class
Bo *et al.* [3]	–	81.4 ± 0.33
Yang *et al.* [17]	73.2	84.3
Non-pretrained convnet	22.8 ± 1.5	46.5 ± 1.7
ImageNet-pretrained convnet	$\mathbf{83.8 \pm 0.5}$	$\mathbf{86.5 \pm 0.5}$

Feature generalization

❖Caltech 256 classification accuracy

# Train	Acc % 15/class	Acc % 30/class	Acc % 45/class	Acc % 60/class
Sohn *et al.* [24]	35.1	42.1	45.7	47.9
Bo *et al.* [3]	40.5 ± 0.4	48.0 ± 0.2	51.9 ± 0.2	55.2 ± 0.3
Non-pretr.	9.0 ± 1.4	22.5 ± 0.7	31.2 ± 0.5	38.8 ± 1.4
ImageNet-pretr.	$\mathbf{65.7 \pm 0.2}$	$\mathbf{70.6 \pm 0.2}$	$\mathbf{72.7 \pm 0.4}$	$\mathbf{74.2 \pm 0.3}$



Feature generalization

❖PASCAL 2012 classification accuracy

Acc %	[22]	[27]	[21]	Ours	Acc %	[22]	[27]	[21]	Ours
Airplane	92.0	**97.3**	94.6	96.0	Dining table	63.2	**77.8**	69.0	67.7
Bicycle	74.2	**84.2**	82.9	77.1	Dog	68.9	83.0	**92.1**	87.8
Bird	73.0	80.8	88.2	**88.4**	Horse	78.2	87.5	**93.4**	86.0
Boat	77.5	85.3	60.3	**85.5**	Motorbike	81.0	**90.1**	88.6	85.1
Bottle	54.3	**60.8**	60.3	55.8	Person	91.6	95.0	**96.1**	90.9
Bus	85.2	**89.9**	89.0	85.8	Potted plant	55.9	57.8	**64.3**	52.2
Car	81.9	**86.8**	84.4	78.6	Sheep	69.4	79.2	**86.6**	83.6
Cat	76.4	89.3	90.7	**91.2**	Sofa	65.4	**73.4**	62.3	61.1
Chair	65.2	**75.4**	72.1	65.0	Train	86.7	**94.5**	91.1	91.8
Cow	63.2	77.8	**86.8**	74.4	Tv	77.4	**80.7**	79.8	76.1
Mean	74.3	82.2	**82.8**	79.0	# won	0	11	6	3

- **Due to the inequality of the dataset type**

References

- ❖ Image Source from https://deeplearning4j.org/convolutionalnets

- ❖ Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks.“ European Conference on Computer Vision, Springer International Publishing, 2014.

- ❖ Jia-Bin Huang, “Lecture 29 Convolutional Neural Networks”, Computer Vision Spring 2015

- ❖ Yosinski, Jason, et al. "Understanding neural networks through deep visualization."