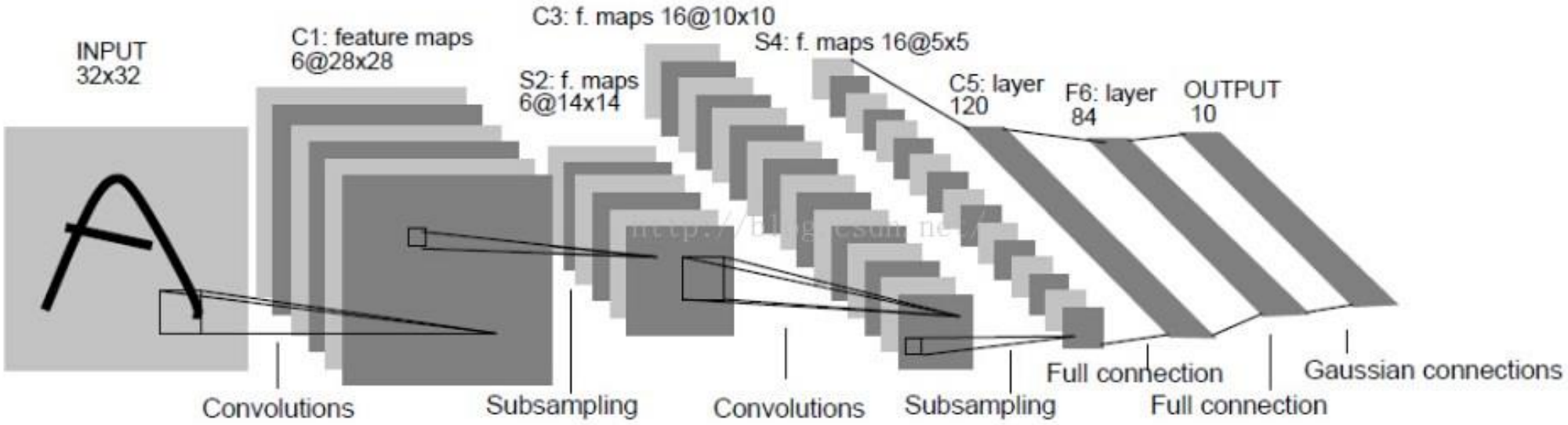
Convolutional Neural Networks

CNN = Convolutional Layer + Pooling Layer + Fully-connected Layer

- The starting,
  LeNet



- Convolution

1	0	1
0	1	0
1	0	1

Size: 3
x 3

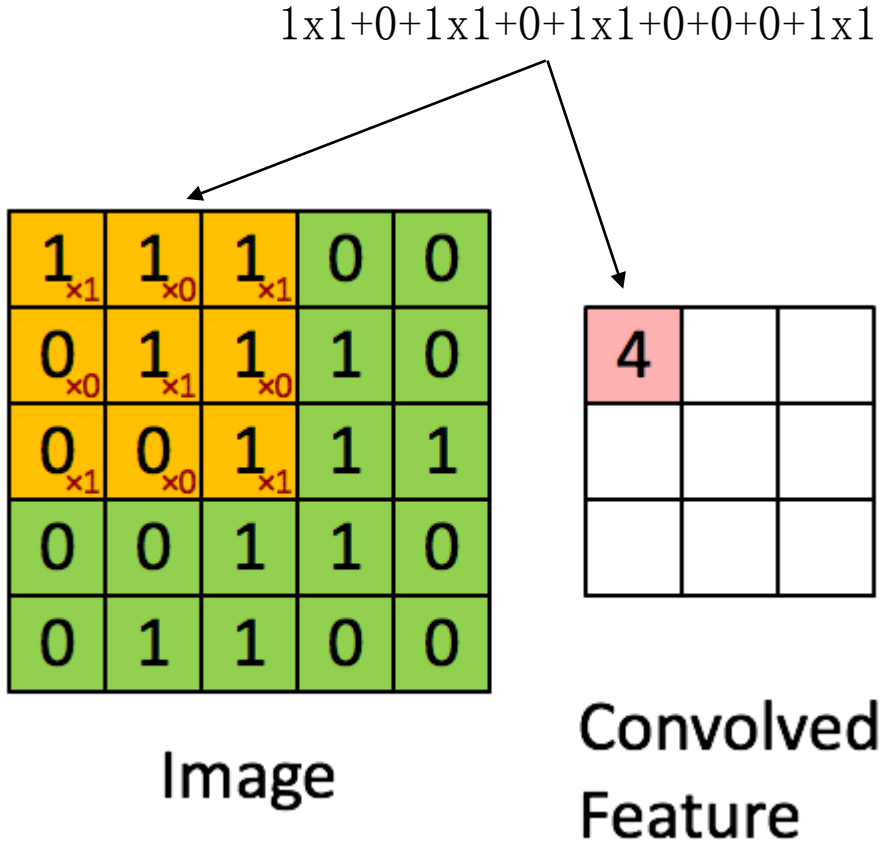


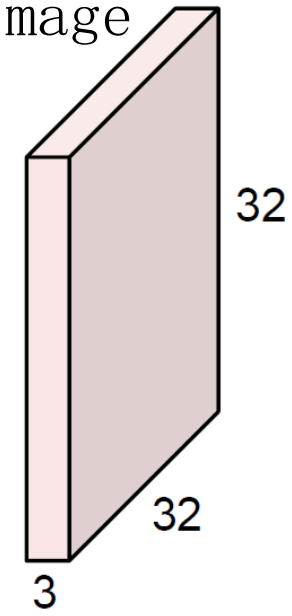
image size: 5
kernel size: 3
feature mapping size:
$(5-3+1)$

Given a $m \times n$ image and $a \times b$ filter, we got a $(m-a+1) \times (n-b+1)$ feature mapping after convolving.

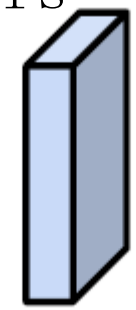
- Convolution Layer

More than 1 filter...

Original image



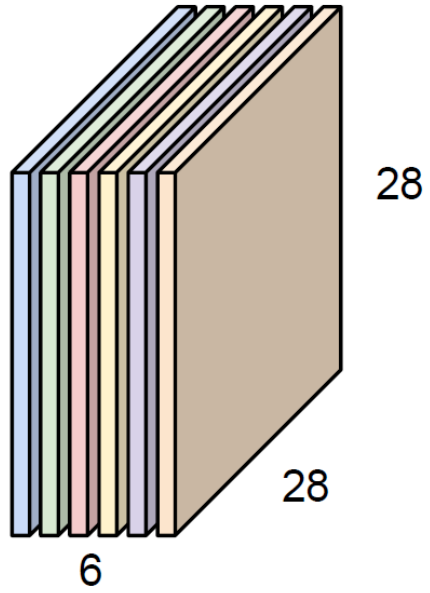
6 $5 \times 5 \times 3$ filters



Convolution Layer:

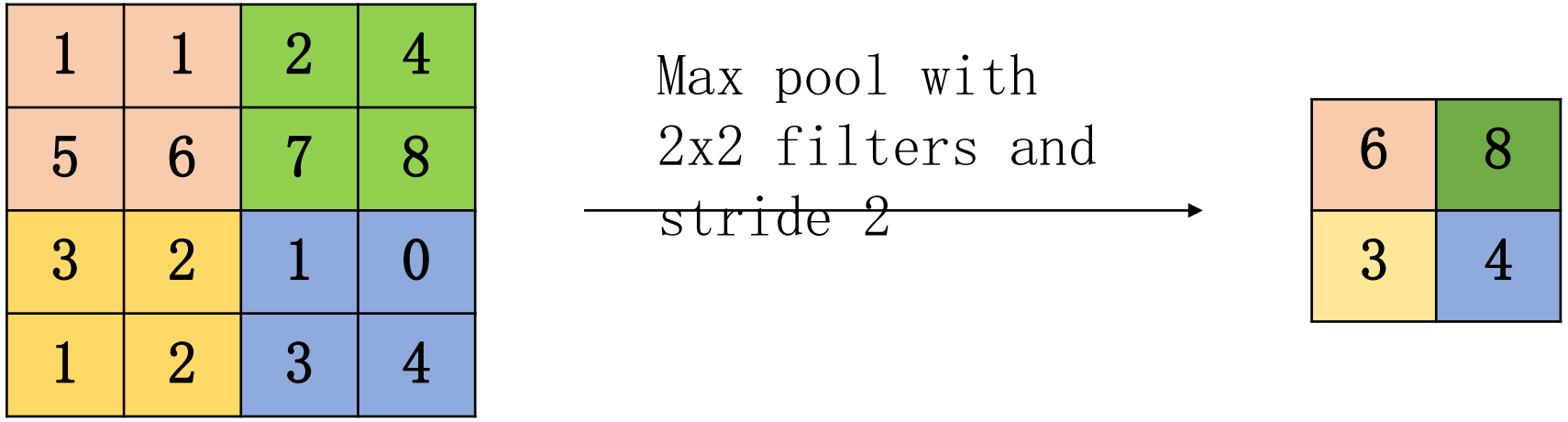
Conv + ReLu

Feature mapping



- Pooling

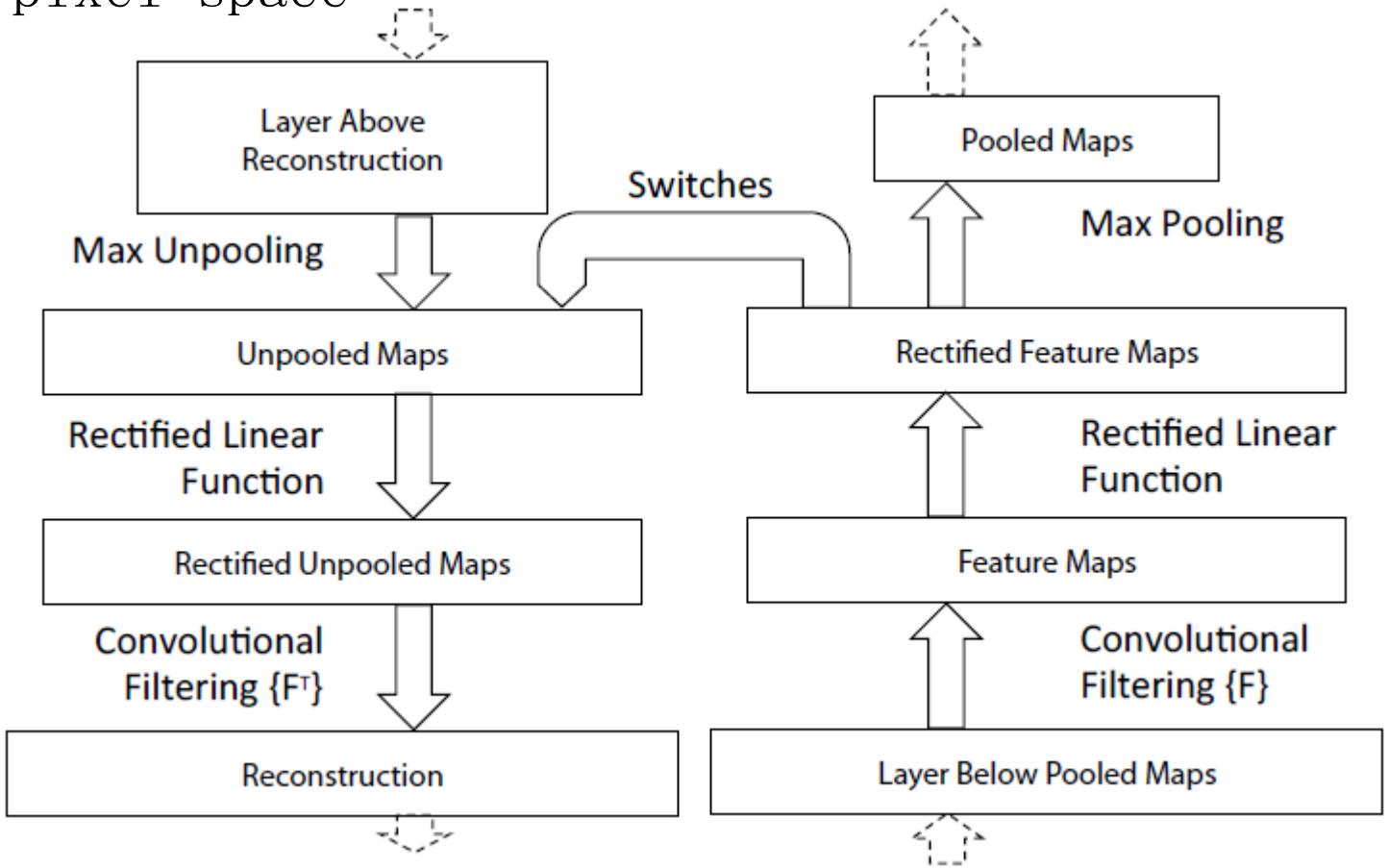
Max pooling



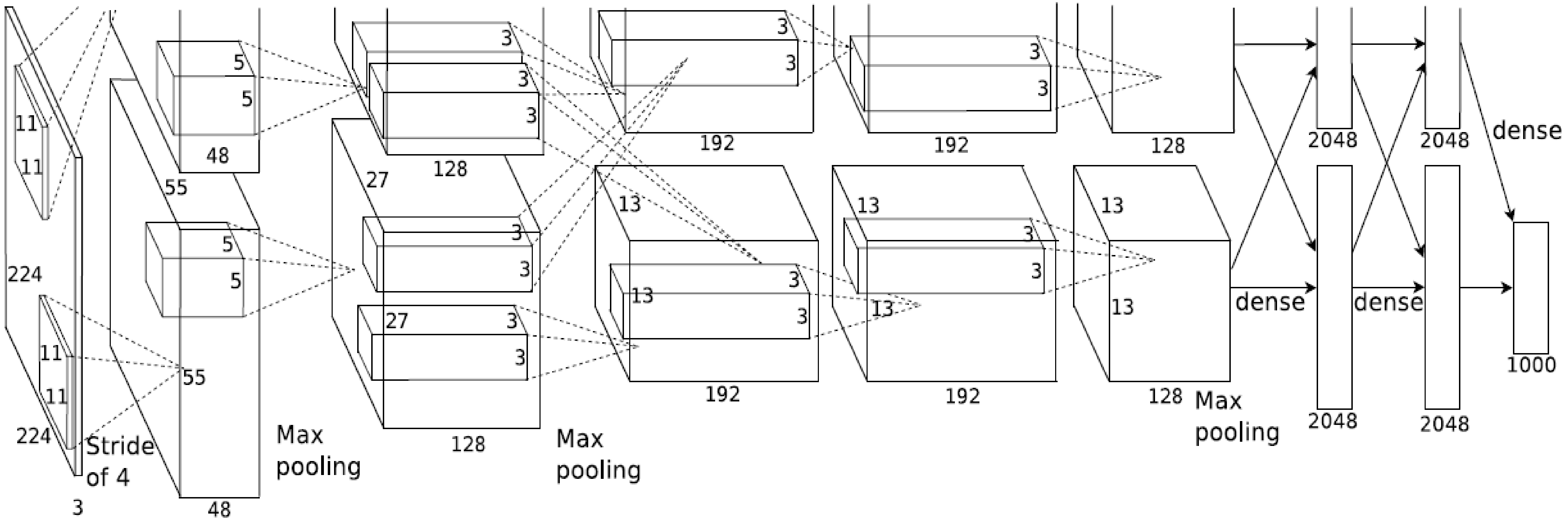
- Pooling layer makes the representations smaller and more computable

- Deconvnet

map feature activities back to the input pixel space



- AlexNet

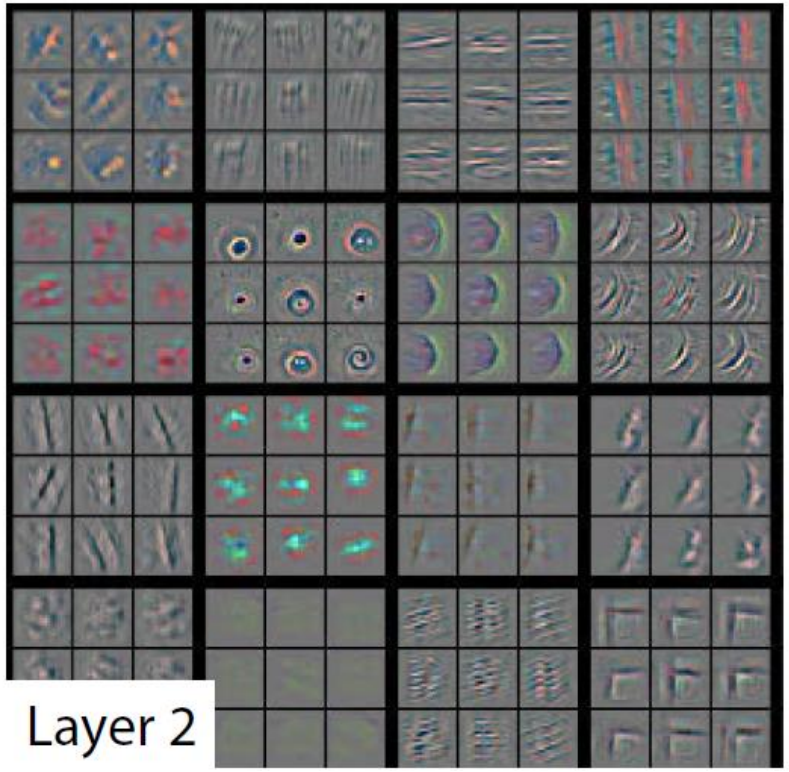


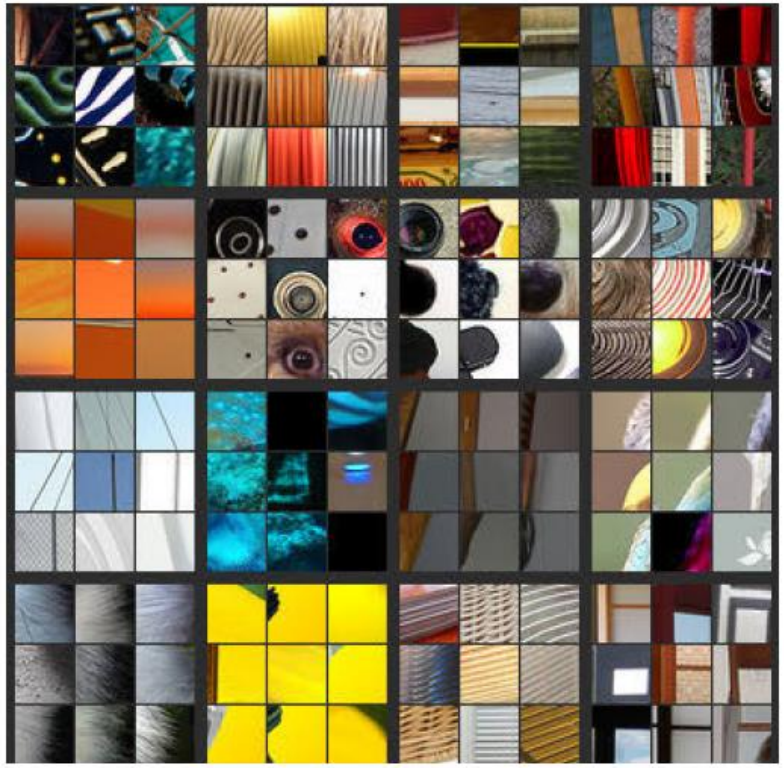
224
224
3
11
11
11
11
Stride of 4
55
55
48
48
5
5
5
5
Max pooling
27
27
128
128
3
3
3
3
Max pooling
13
13
192
192
3
3
13
13
192
192
3
3
13
13
128
128
Max pooling
dense
dense
dense
2048
2048
2048
2048
1000

- Visualize Convolutional Networks

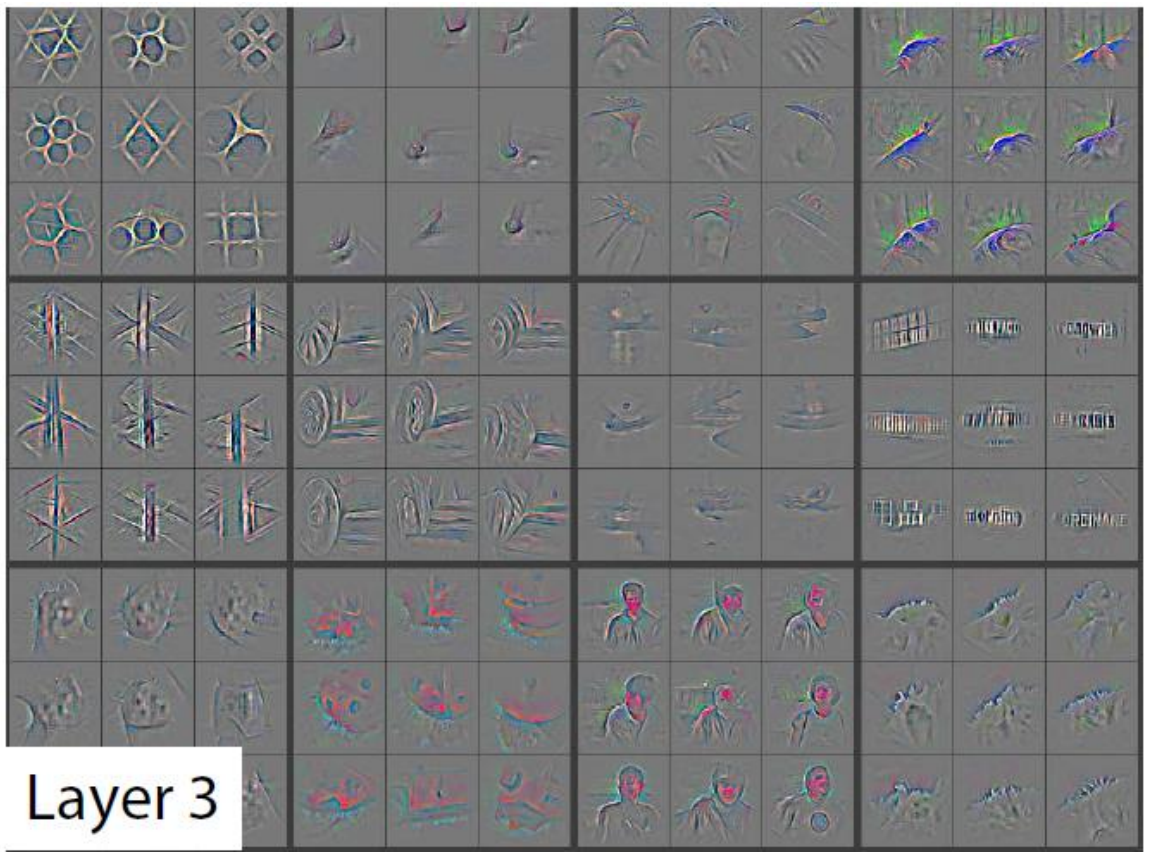
Layer 1

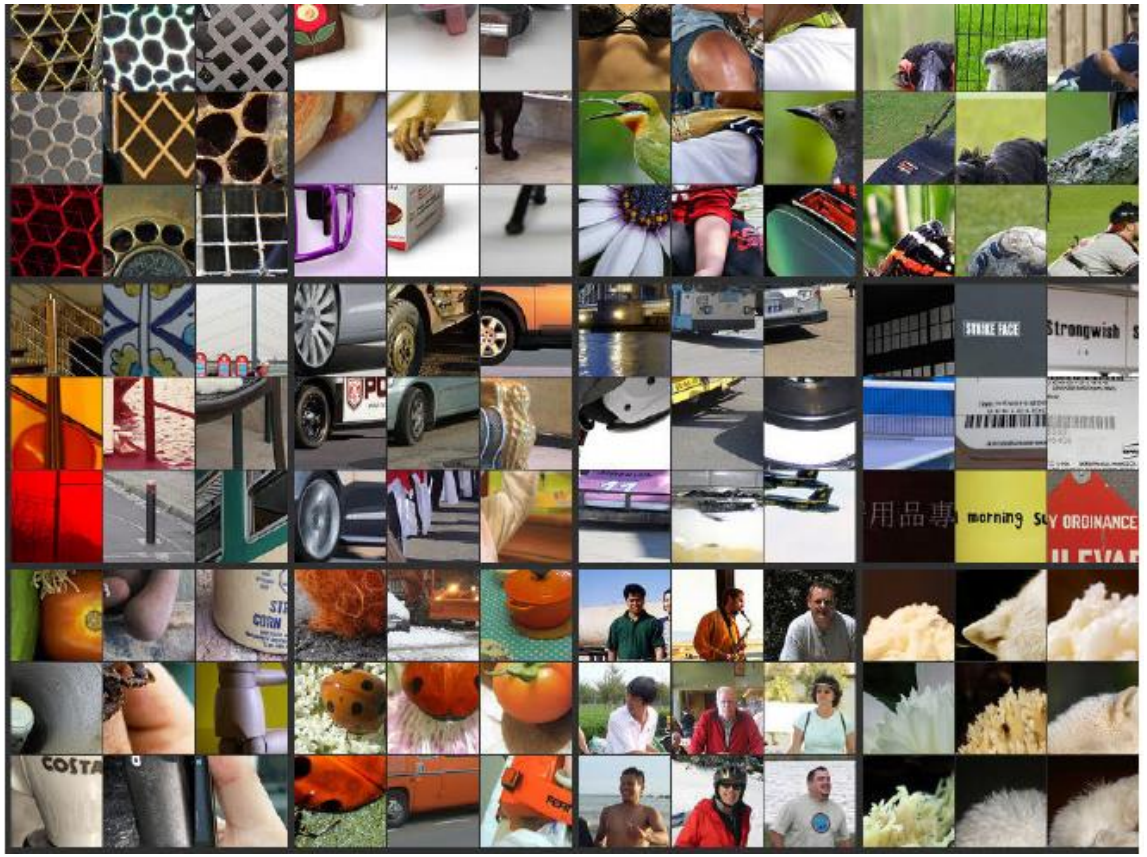
Only edges



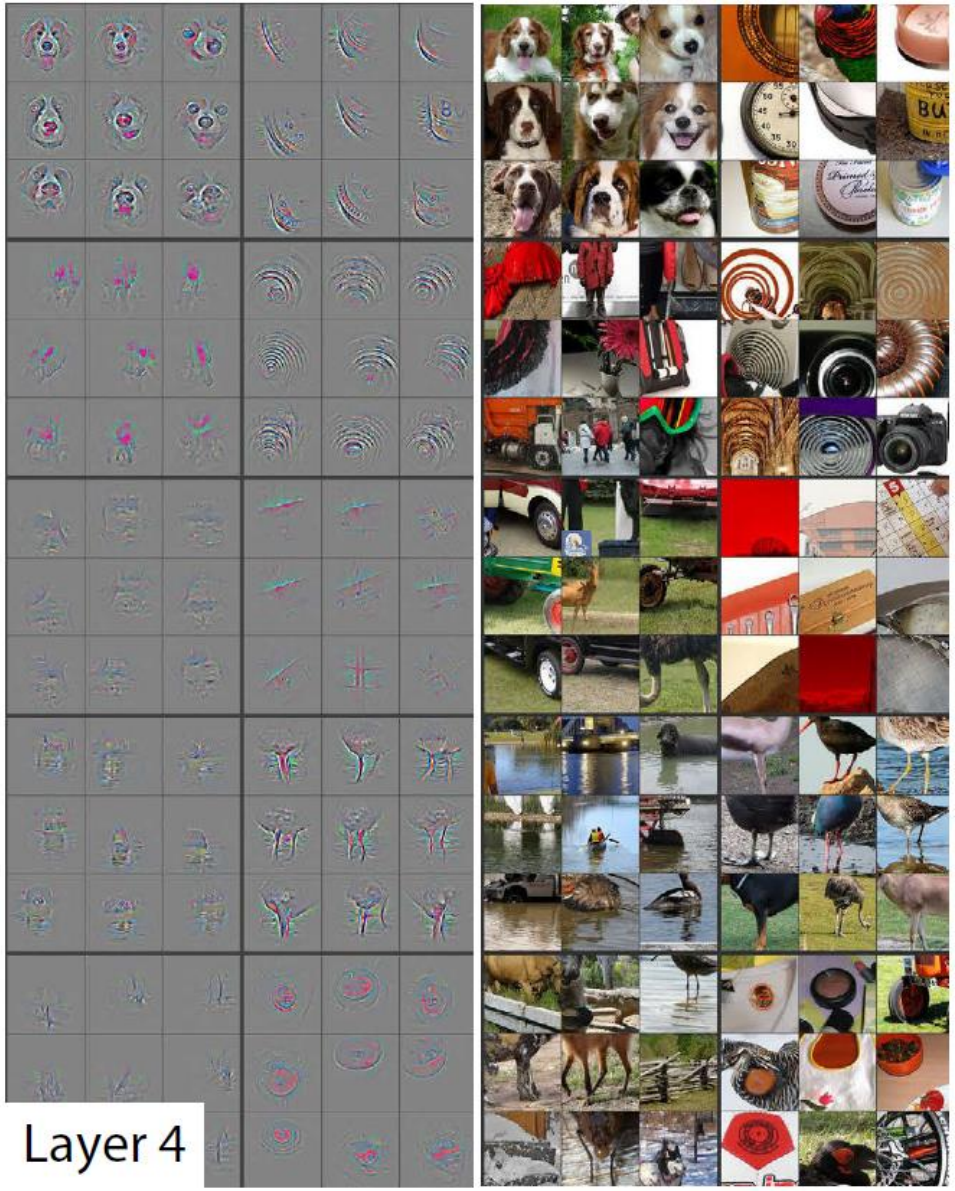


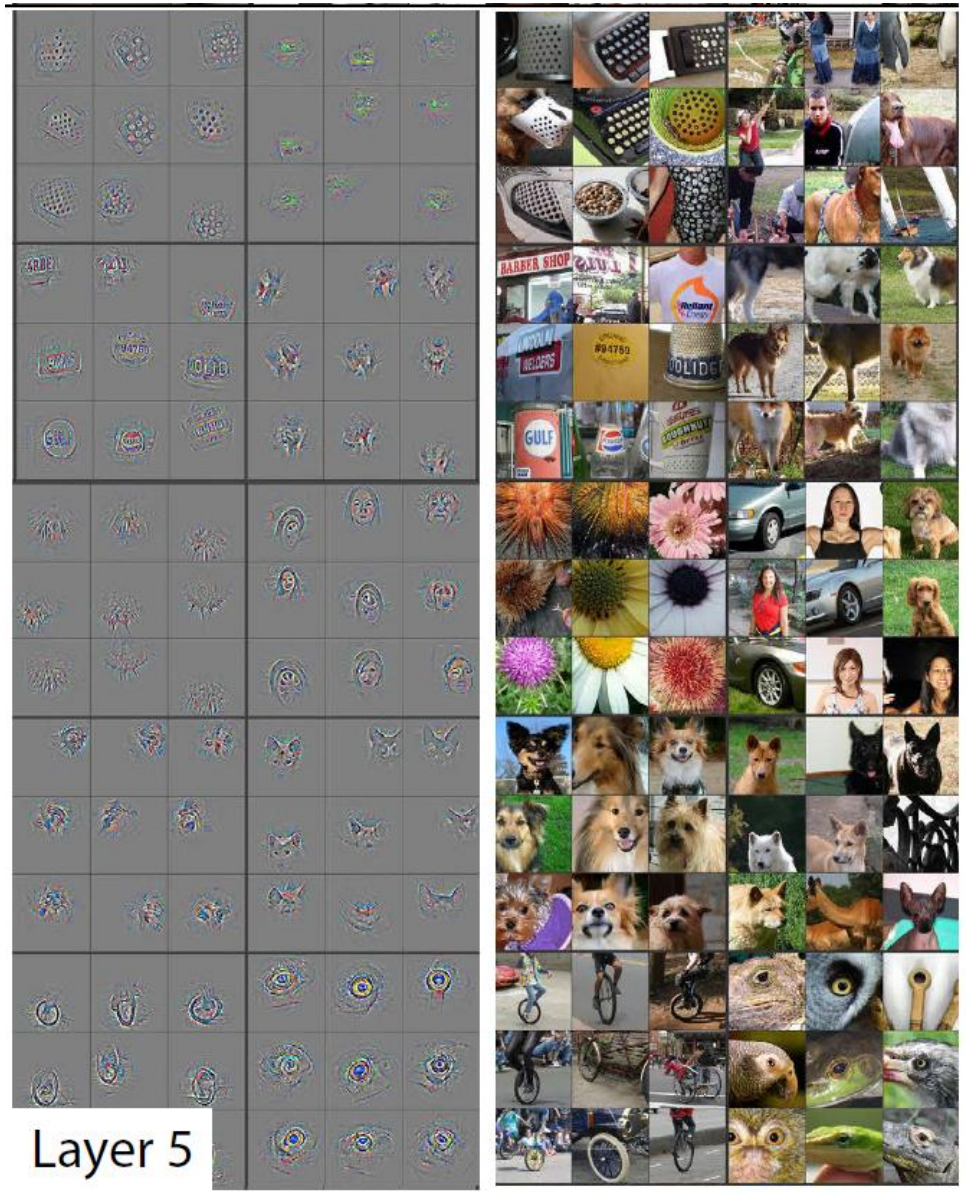
Combination of edges





Outlines of objects





Distinctive high-level features of objects

References

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.

Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks." *arXiv preprint arXiv:1311.2901* (2013).

Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, november 1998.

Andrew Ng. Unsupervised Feature Learning and Deep Learning Tutorial.