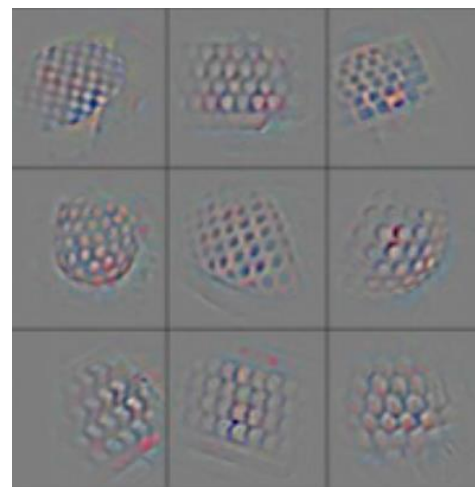
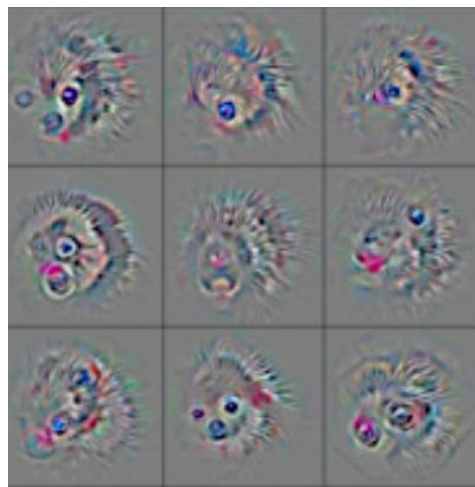
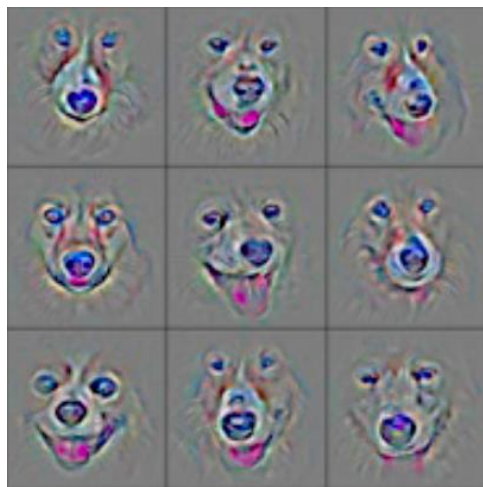


Deconvolutional nets

Визуализация свёрточных сетей

Антон Осокин

ВМК МГУ



План доклада

- Свёрточные сети
- Визуализация свёрточных сетей
- Анализ архитектуры сети

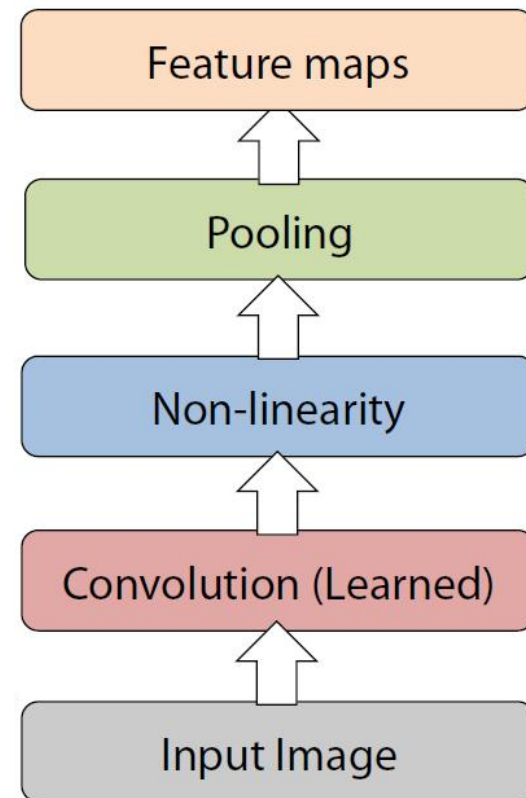
Материалы

- M. Zeiler, R. Fergus, Visualizing and Understanding Convolutional Networks, ECCV 2014
- M. Zeiler, G. Taylor, R. Fergus, Adaptive deconvolutional networks for mid and high level feature learning, ICCV 2011
- R. Fergus, Deep Learning for Computer Vision, NIPS 2013 tutorial

Большинство слайдов взяты у Matthew Zeiler и Rob Fergus, за что им большое спасибо!

Convolutional Networks (LeCun et al., 89)

- Сеть прямого распространения
- Операции:
 - Свёртка
 - Нелинейность
 - Pooling
- Обучение с учителем
- Стохастические градиентные методы
- Градиент вычисляется методом обратного распространения ошибок (back propagation)



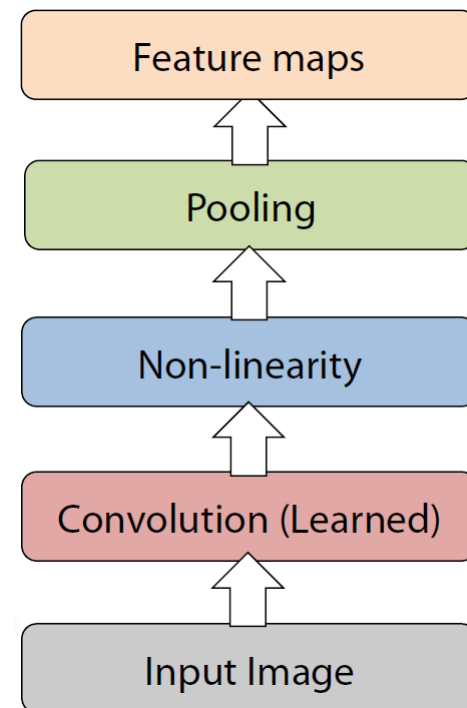
Структура слоя сети

- Вход: K_i -канальное изображение z
- Выход: K_{i+1} -канальное изображение y
- Свёртка (convolution)

$$y_c := \sum_{k=1}^{K_i} g_{k,c}(z_k \oplus f_{k,c})$$

- Нелинейность
 - rectified linear $y_c := \max(y_c, 0)$
 - локальная нормировка

- Pooling
 - 3D max-pooling $y_{c,p} := \max_{\substack{d \in \text{some layers} \\ q \in \text{some pixels}}} y_{d,q}$
 - subsampling

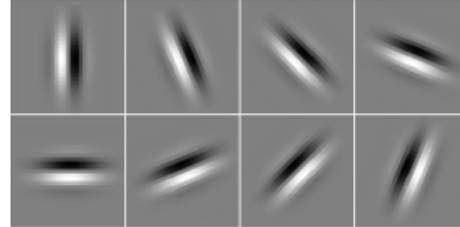


Compare: SIFT Descriptor

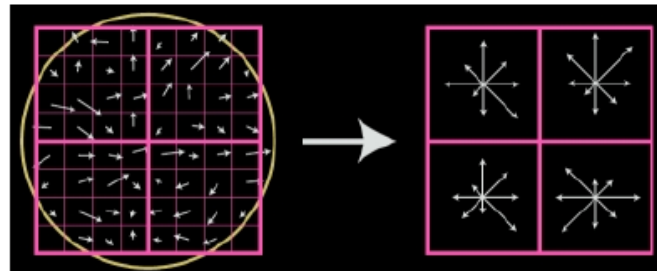
Image
Pixels



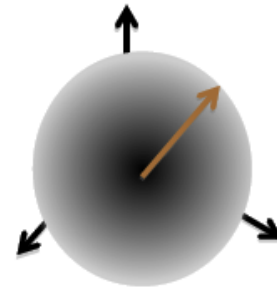
Apply
Gabor filters



Spatial pool
(Sum)



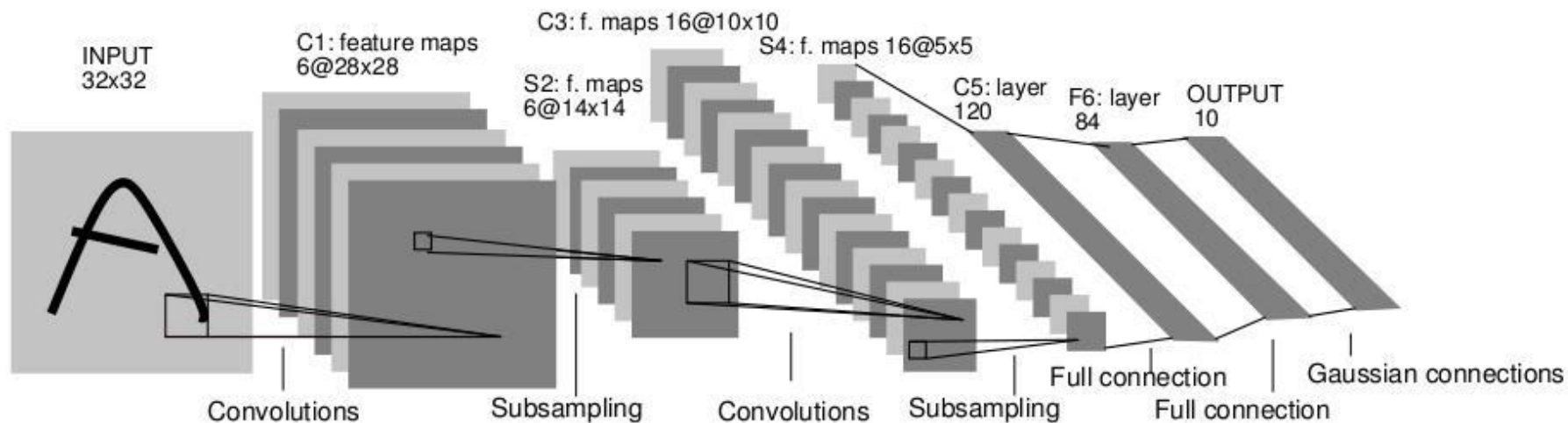
Normalize to
unit length



Feature
Vector

Архитектура для MNIST (1998)

- 60,000 объектов
- 28 x 28 пикселей
- 10 классов



Прорыв в 2012 г.

ImageNet Classification with Deep Convolutional Neural Networks

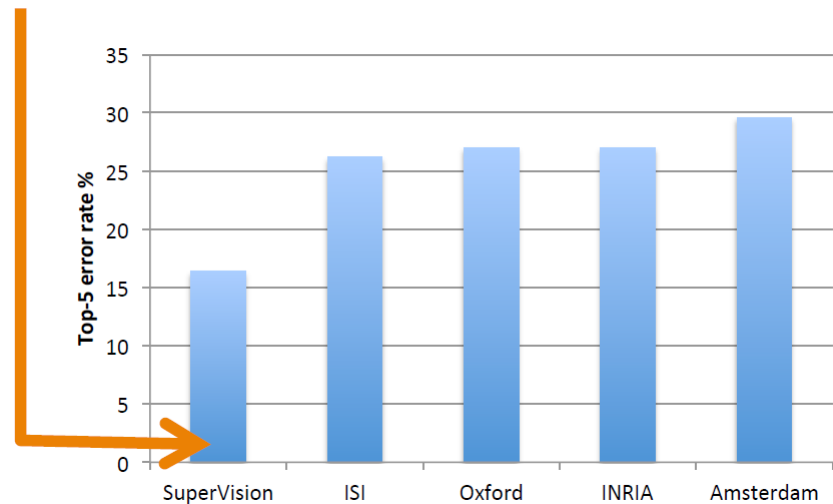
Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

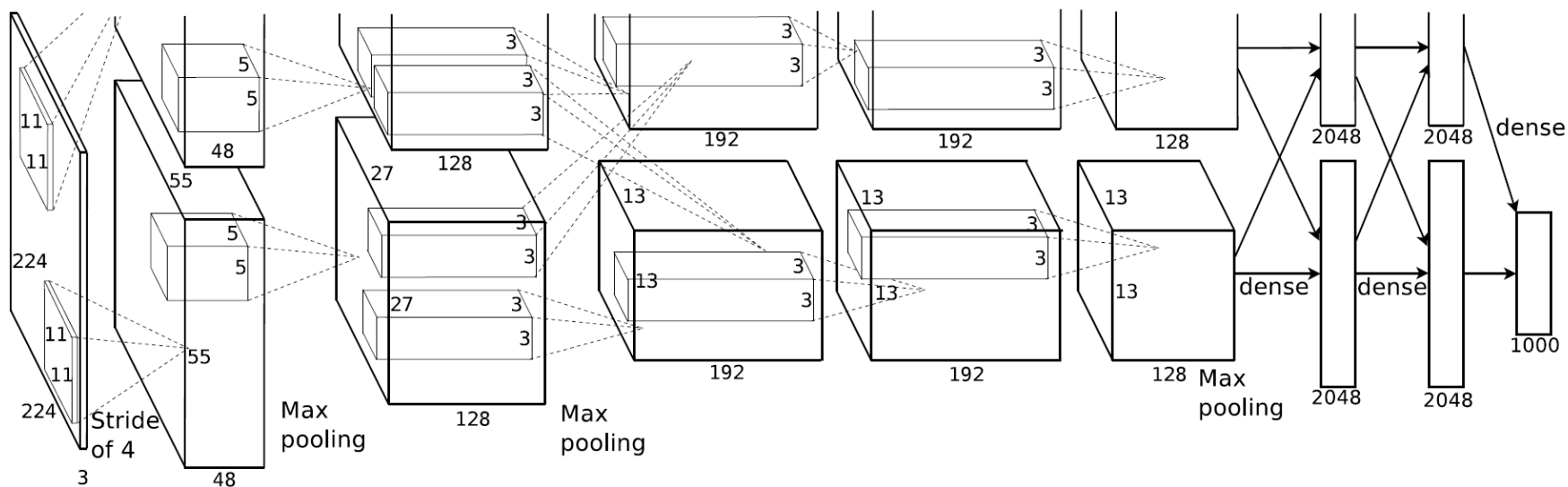
Классификация изображений
Базы ImageNet 2012

Отрыв от не нейросетевых методов
составляет около 10%



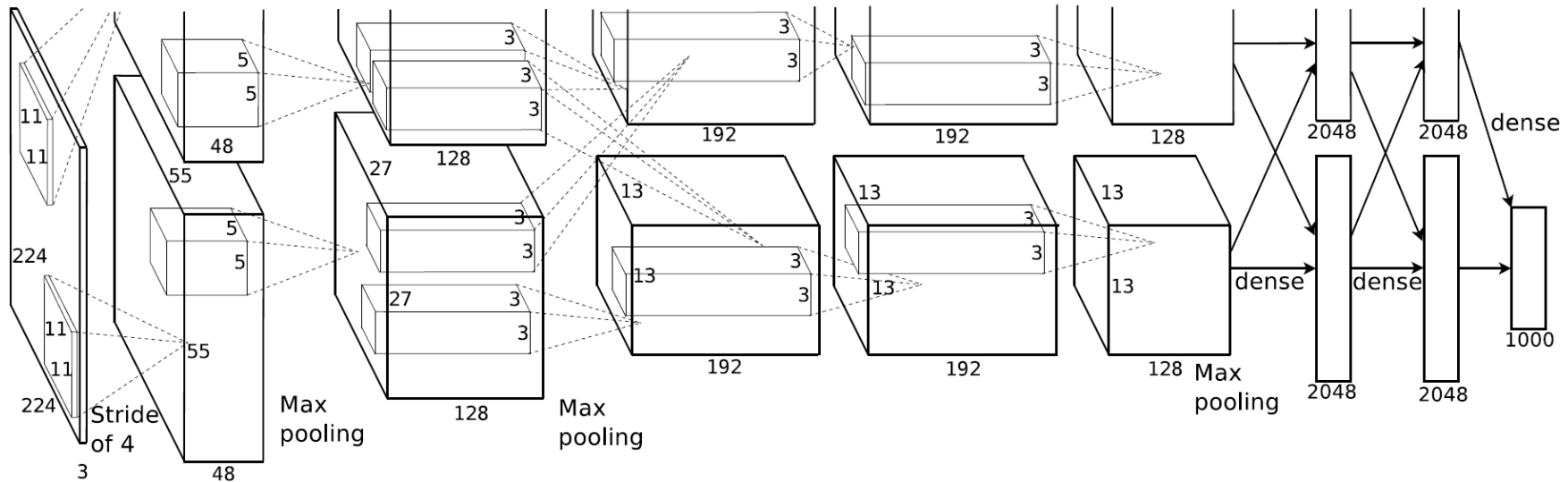
Архитектура для ImageNet (2012)

- 1,000,000 объектов
- 224 x 224 пикселей
- 1,000 классов



AlexNet (Krizhevsky et al., 2012)

- Увеличенная модель (8 слоев)
- Больше данных (10^6 vs. 10^4)
- Реализация на GPU
- Регуляризация (Data augmentation, DropOut)



650,000 узлов, 6,000,000 параметров
время обучения – 1 неделя на 2-х GPU

Training Big ConvNets

- Back-propagation of error
 - [Rumelhart, Hinton & Williams 1986] + many others
 - Chain rule
- Stochastic Gradient Descent
 - 2nd order methods expensive
 - L. Bottou “Stochastic Gradient Tricks” Neural Networks 2012
- Momentum
 - Nesterov variant [Sutskever et al. ICML 2012]
- Classification loss: cross-entropy
- GPU implementation

Improving Generalization

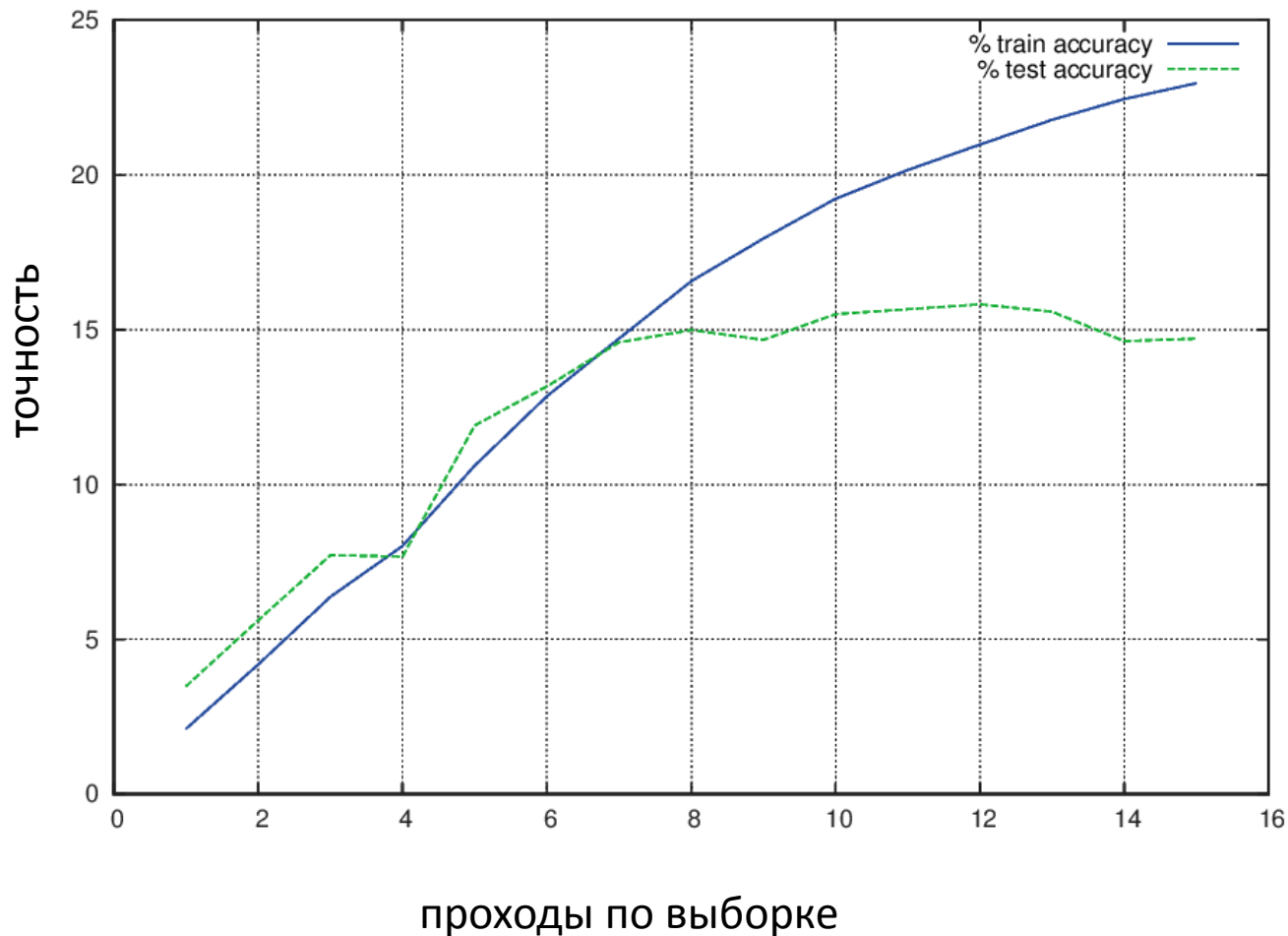
- Data augmentation (crops/flips etc. of images)
- Weight decay (L1 or L2 penalty on weights)
- Inject Noise into network
 - DropOut [Hinton et al. 2012]
 - DropConnect [Wan et al. ICML 2012]
 - Stochastic Pooling [Zeiler & Fergus ICLR'13]

Реализации

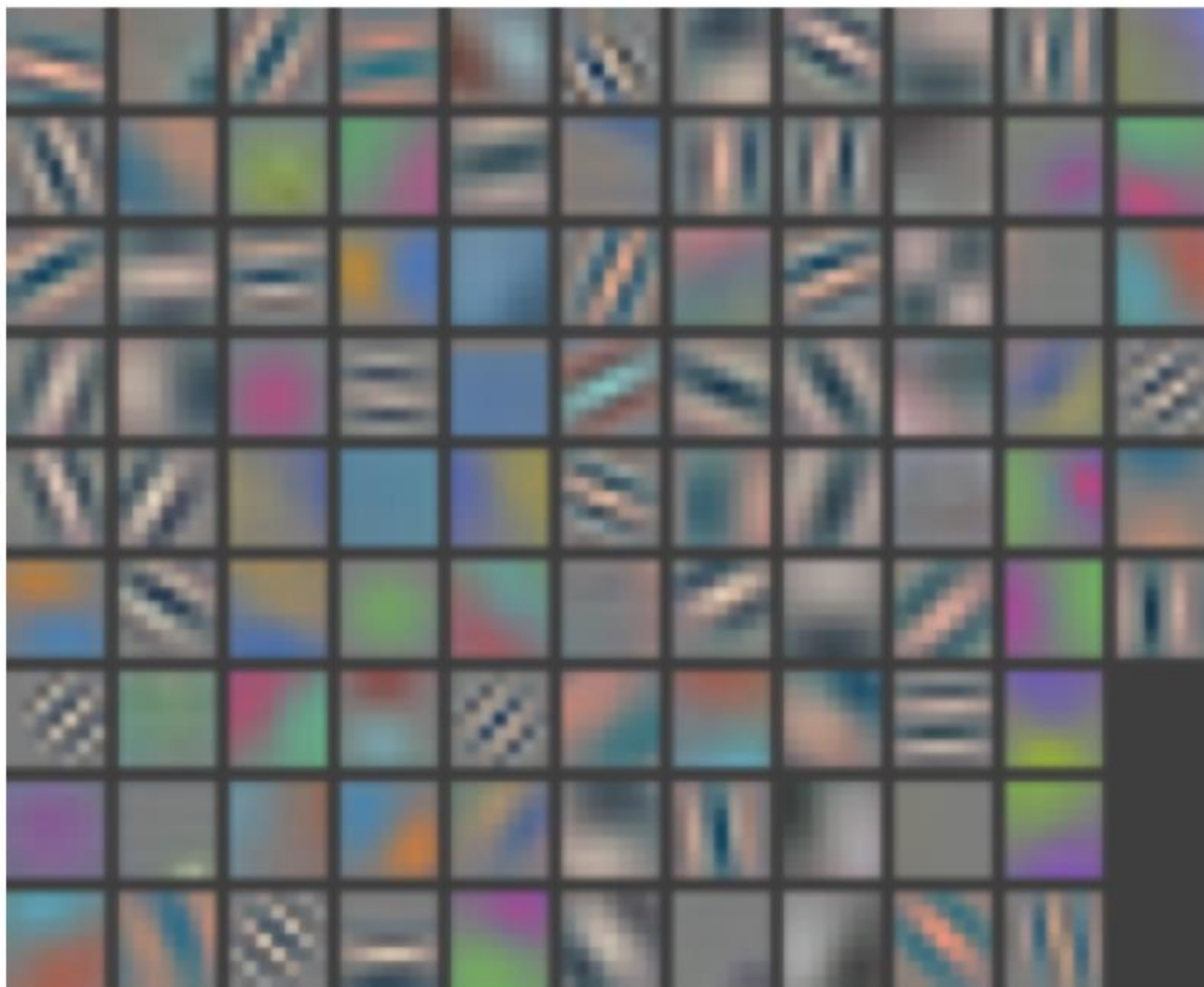
- cuda-convnet (Alex Krizhevsky et al., Google)
- Torch7 (Yann LeCun et al., Facebook)
- Caffe (Yangqing Jia, UC Berkeley)

http://deeplearning.net/software_links/

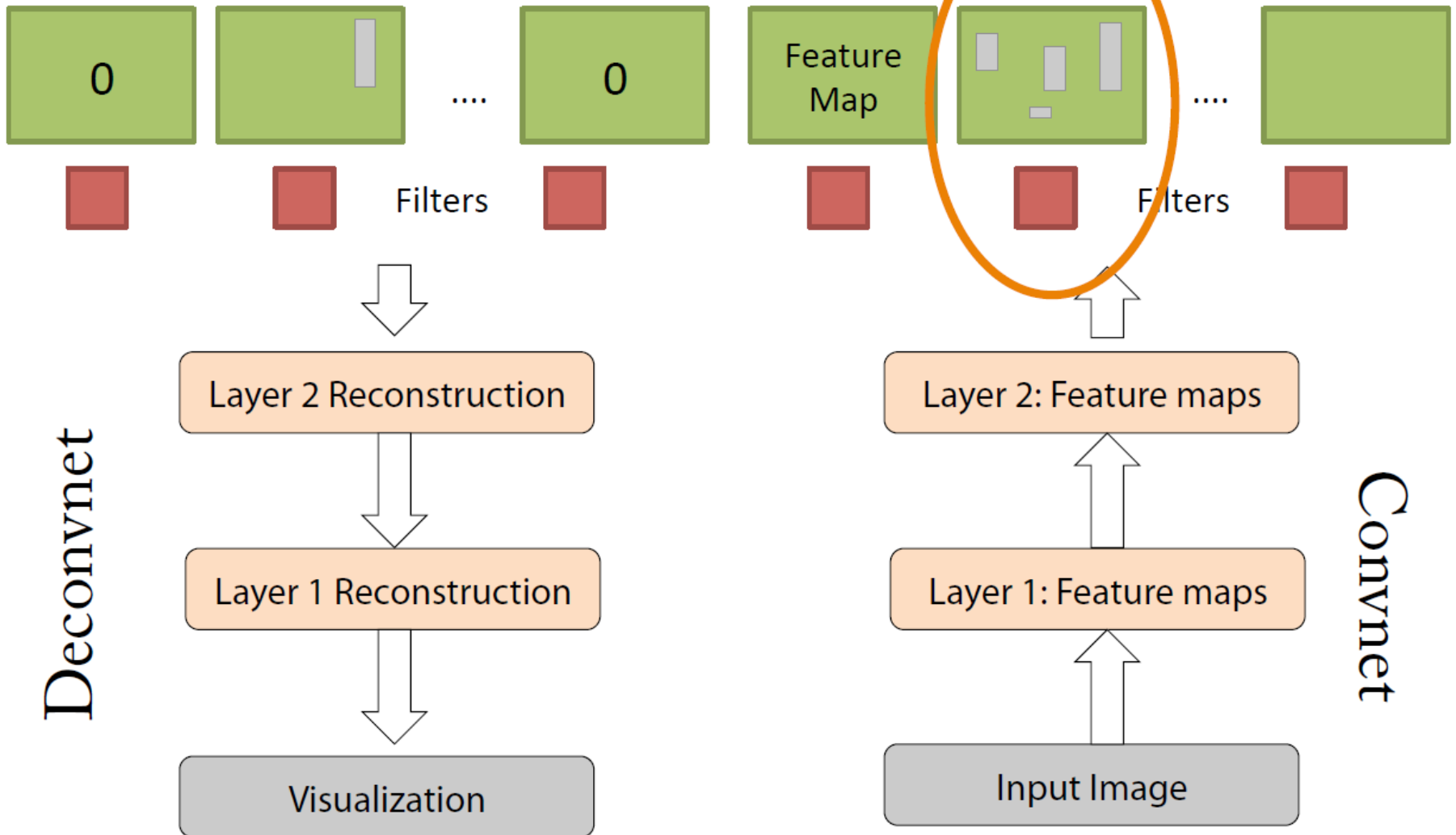
Как понять, что сеть обучается?



Визуализация сети: слой 1



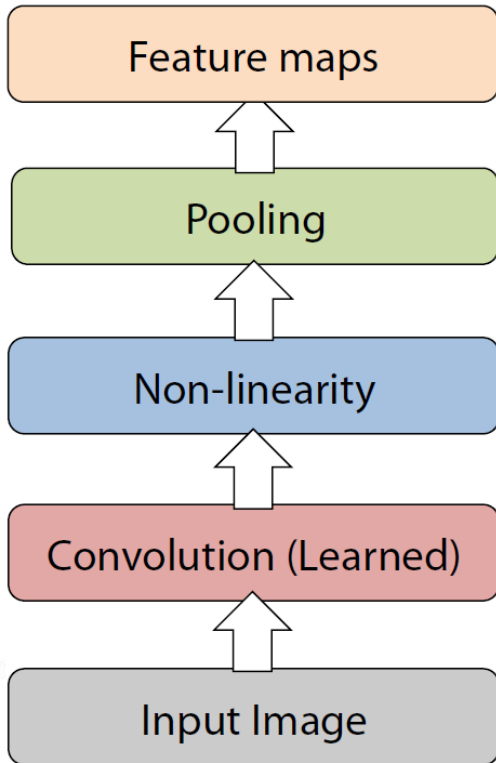
Projecting back from Higher Layers



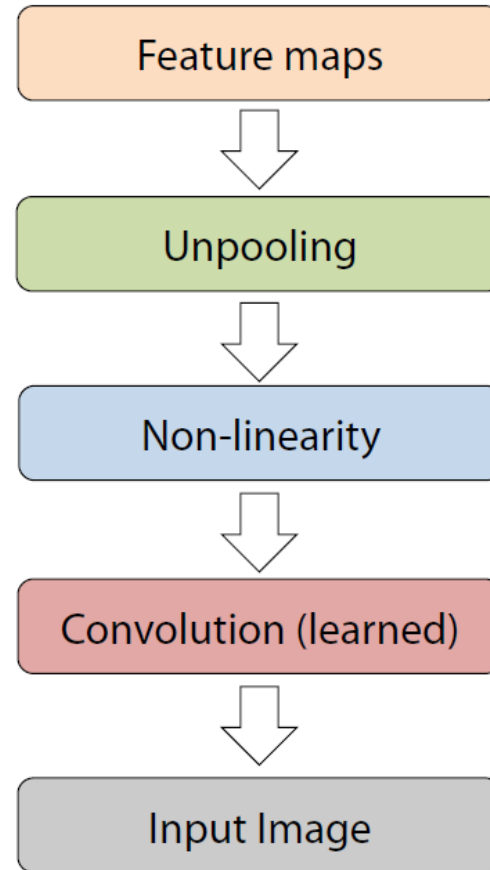
Визуализация внутренних слоев

- Пропускаем изображения валидационной выборки через сеть
- Выбираем внутренний узел
- Находим максимальные значения откликов
- Восстанавливаем вход, дающий эти отклики

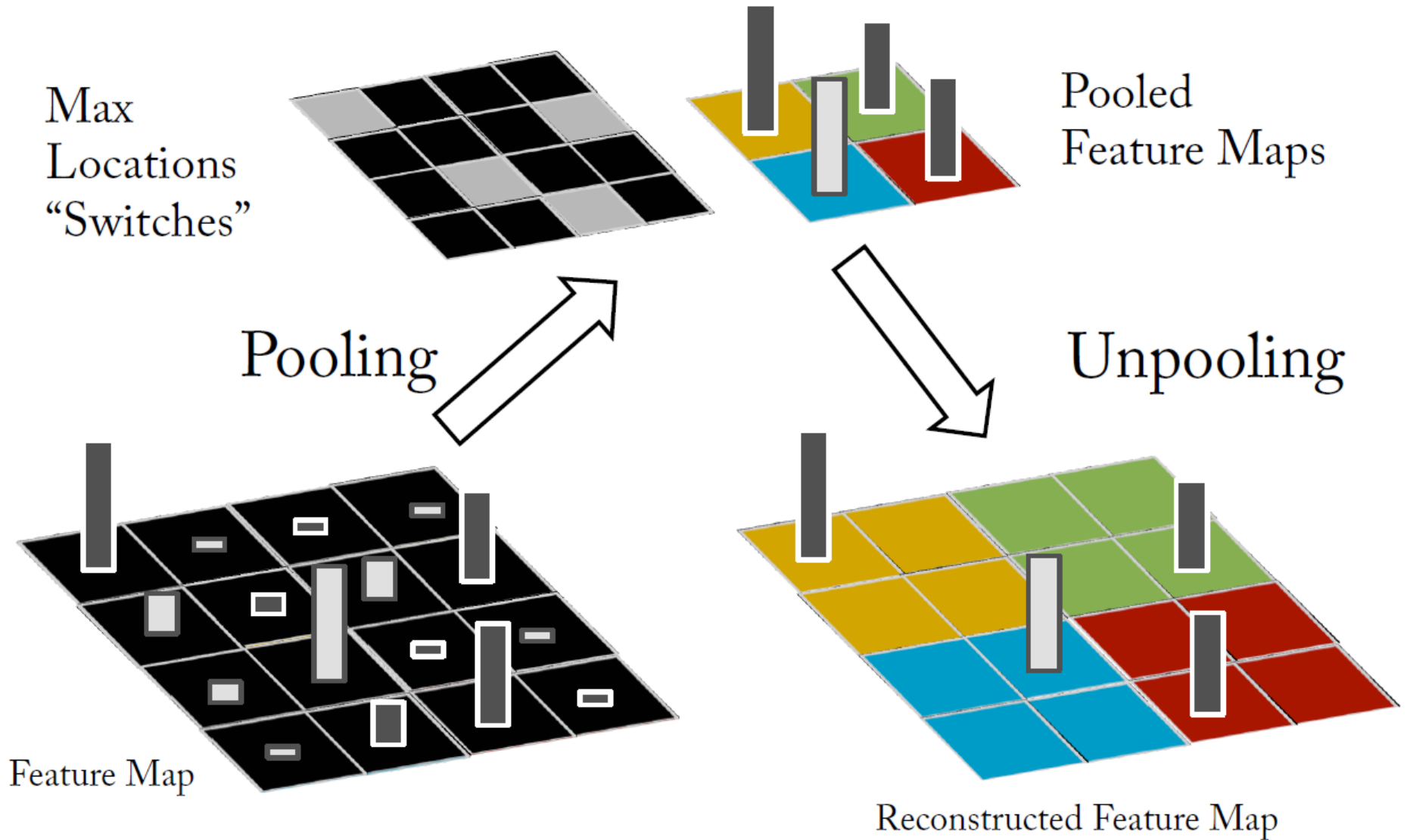
Convolutional layer



Deconvolutional layer



Reversible Max Pooling



Non-linearity

- локальная перенормировка с использованием коэффициентов из ConvNet
- rectified linear обеспечивает положительность откликов

Deconvolution

- Дано:
 - Отклики y (результат свёртки)
 - Фильтры f (обучены в рамках ConvNet)

- Найти:

- Исходный вход z :
$$y_c := \sum_{k=1}^{K_i} z_k \oplus f_{k,c}$$

- Задача:

$$\min_{\{z_k\}} \frac{\lambda}{2} \sum_c \left\| \sum_{k=1}^{K_i} z_k \oplus f_{k,c} - y_c \right\|_2^2 + \sum_k \|z_k\|_1$$

Deconvolution

- Задача:

$$\min_{\{z_k\}} \frac{\lambda}{2} \sum_c \left\| \sum_{k=1}^{K_i} z_k \oplus f_{k,c} - y_c \right\|_2^2 + \sum_k \|z_k\|_1$$

- Свёртка – линейная операция:

$$\min_z \frac{\lambda}{2} \|Fz - y\|_2^2 + \|z\|_1$$

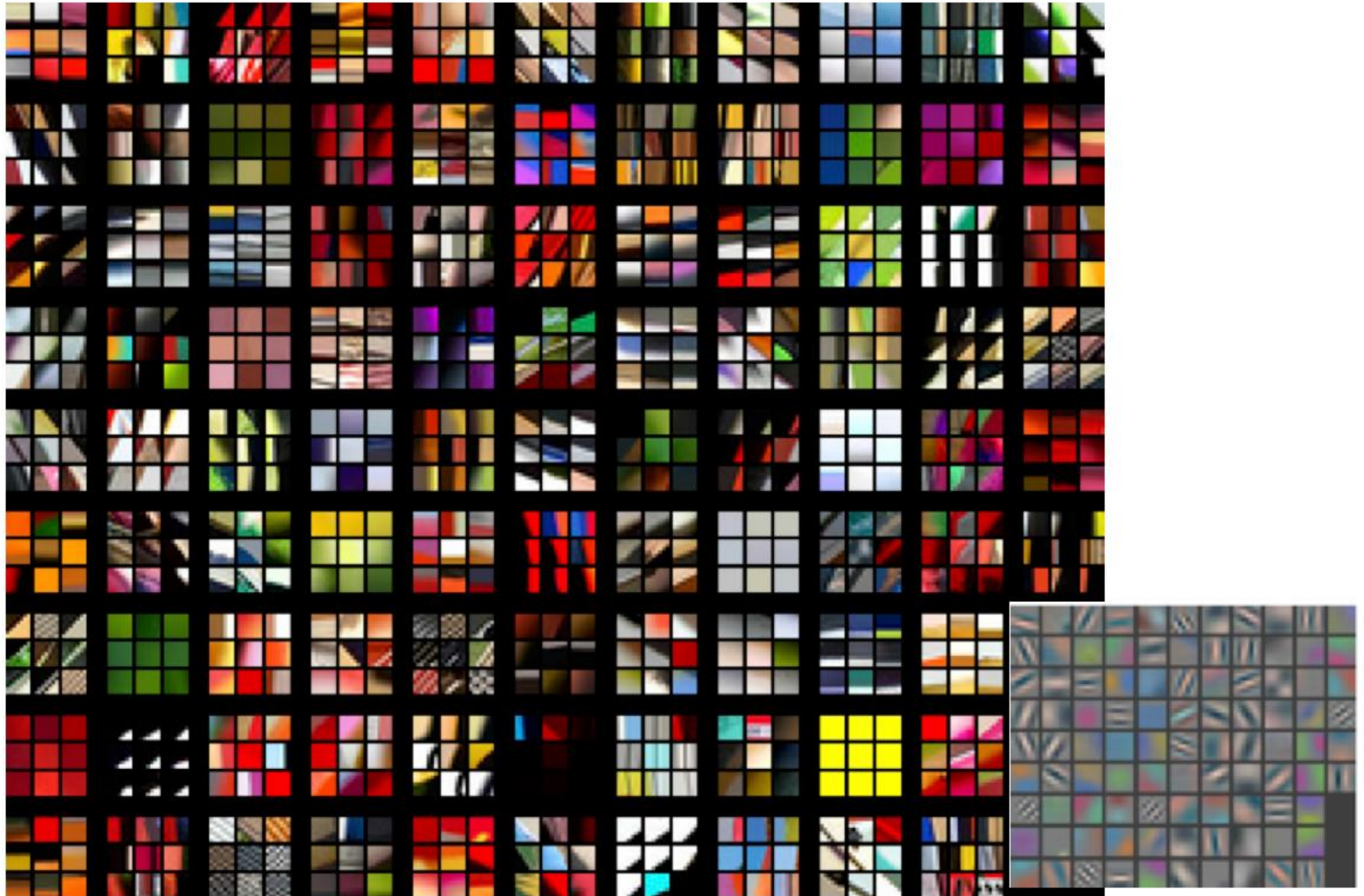
- Задача решается методом ISTA

(Beck&Teboulle, 2009)

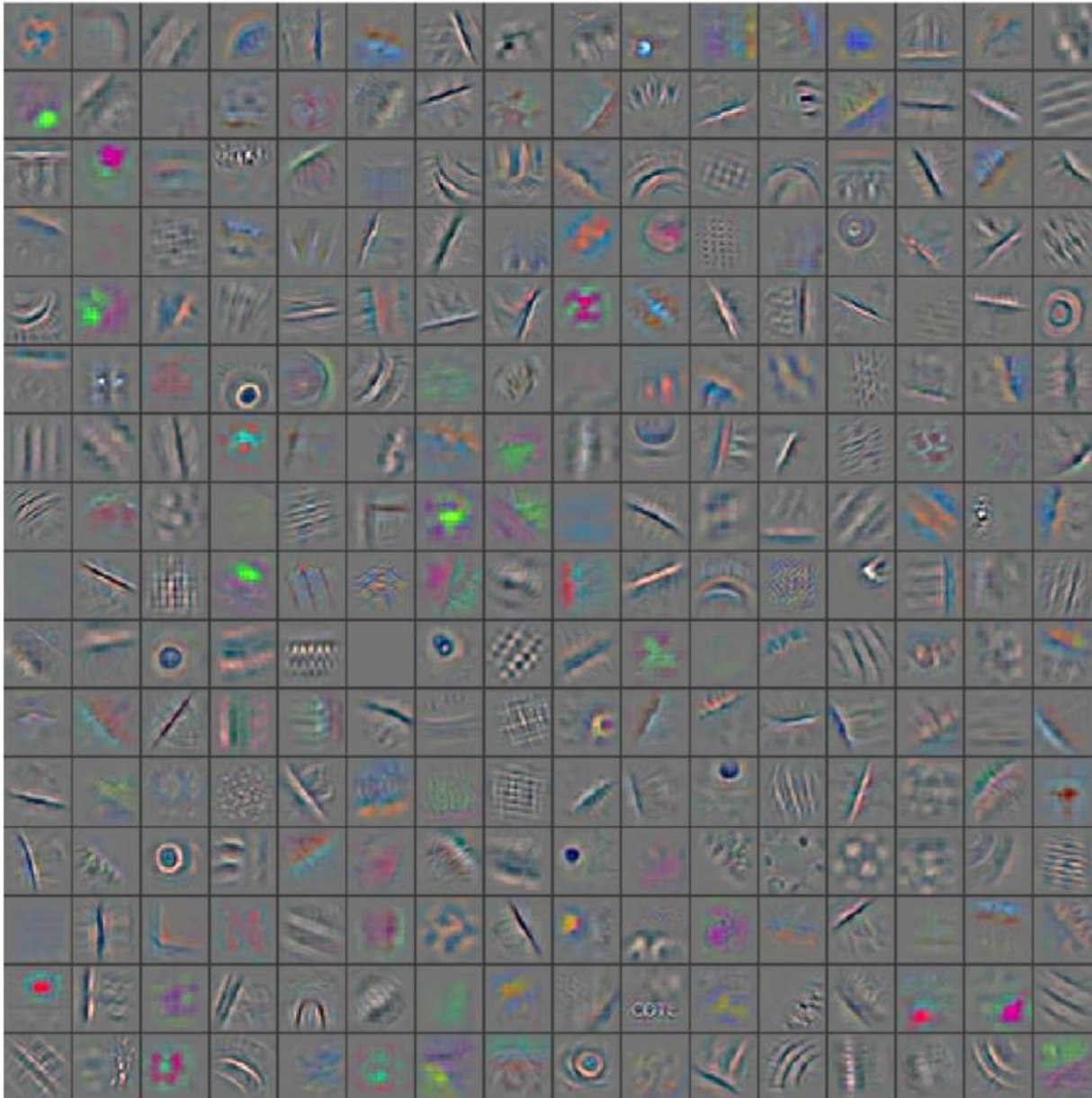
Алгоритм ISTA

- Задача: $\min_z \frac{\lambda}{2} \|Fz - \mathbf{y}\|_2^2 + \|z\|_1$
- Алгоритм:
 - Градиент гладкой части: $\mathbf{g} = \lambda F^T (Fz - \mathbf{y})$
 - Шаг по градиенту: $z = z - \beta \mathbf{g}$
 - Shrinkage step: $z := \max(|z| - \beta, 0) \text{sign}(z)$

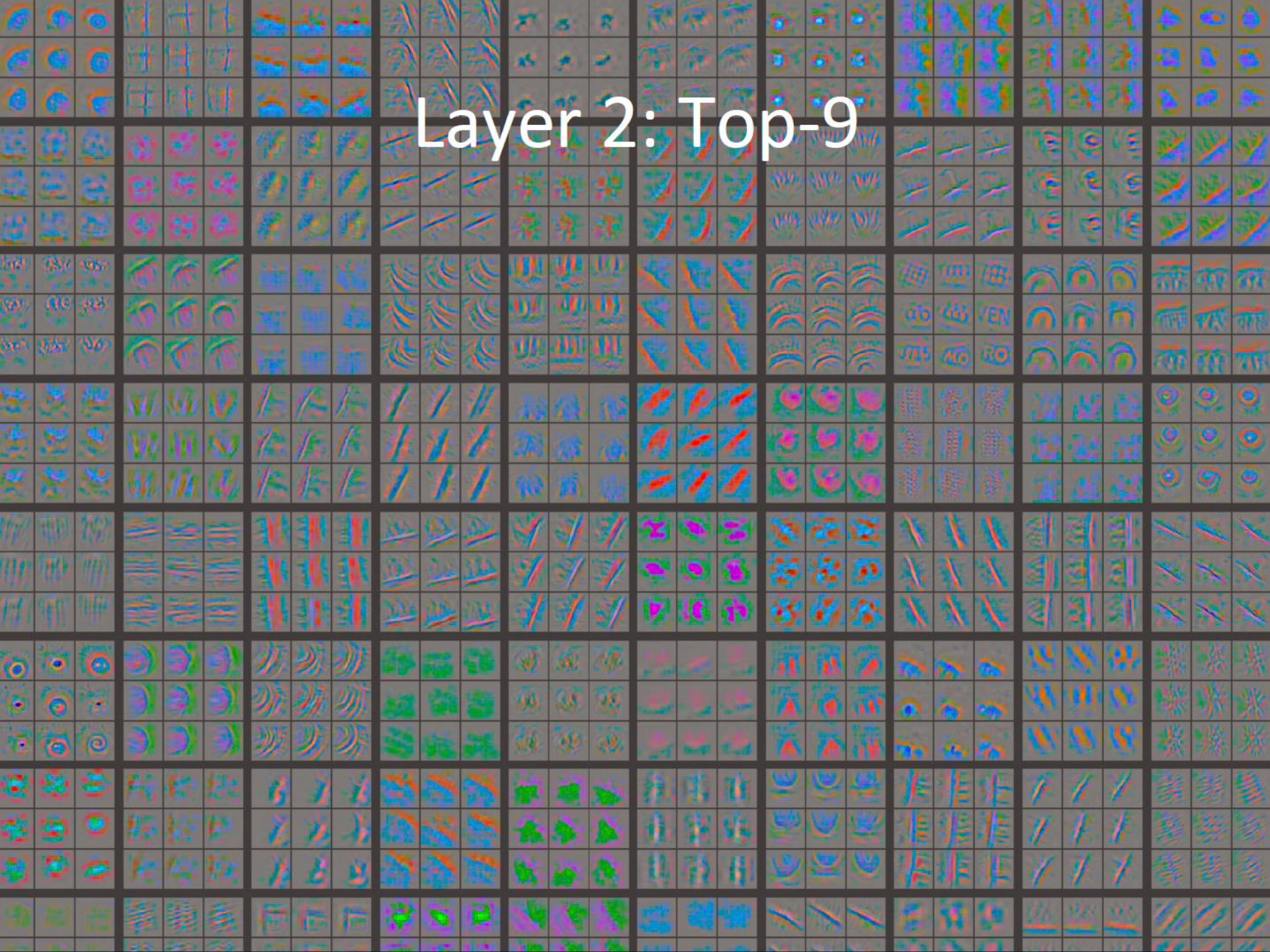
Layer 1: Top-9 Patches



Layer 2: Top-1



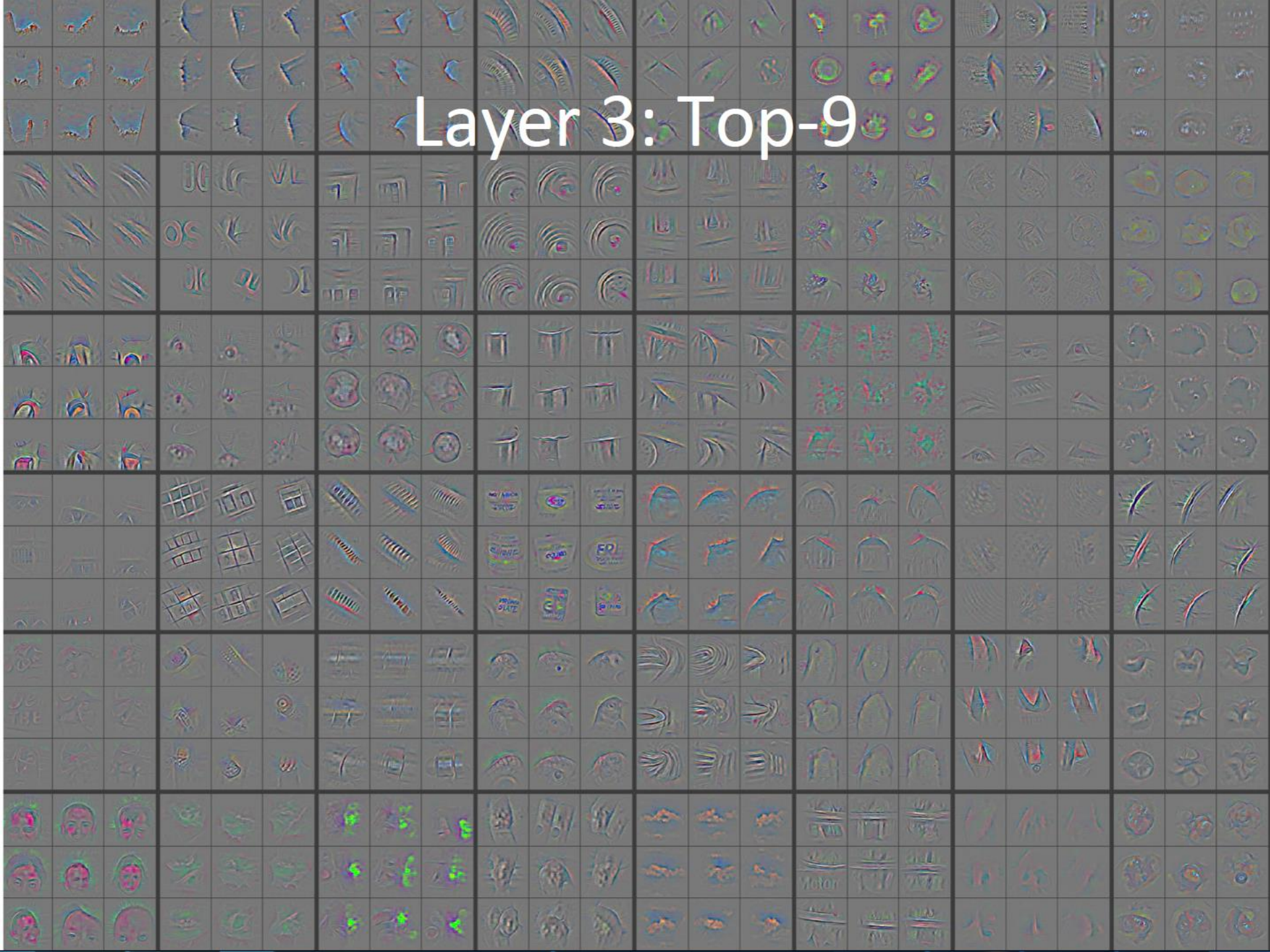
Layer 2: Top-9



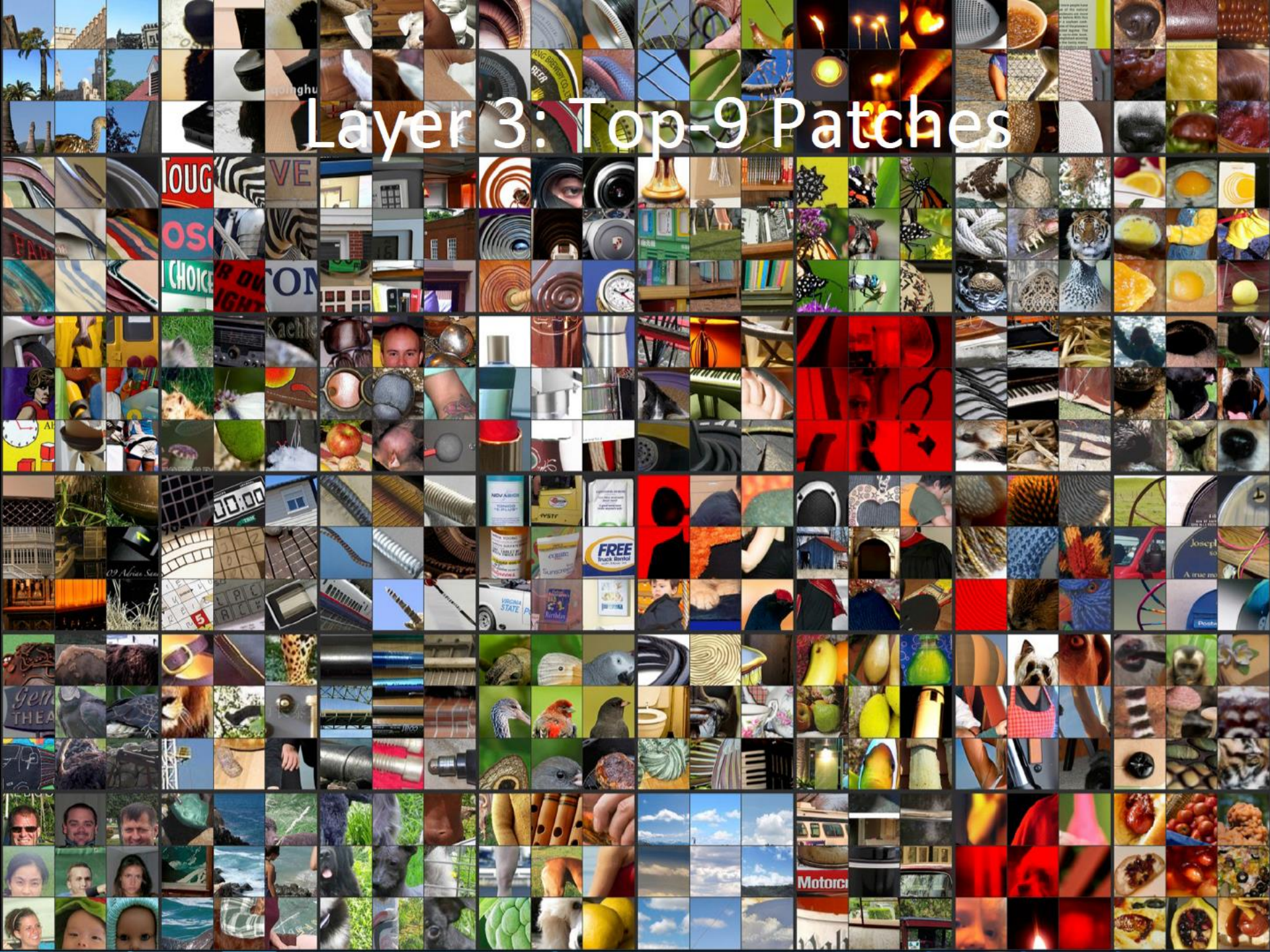
Layer 2: Top-9 Patches



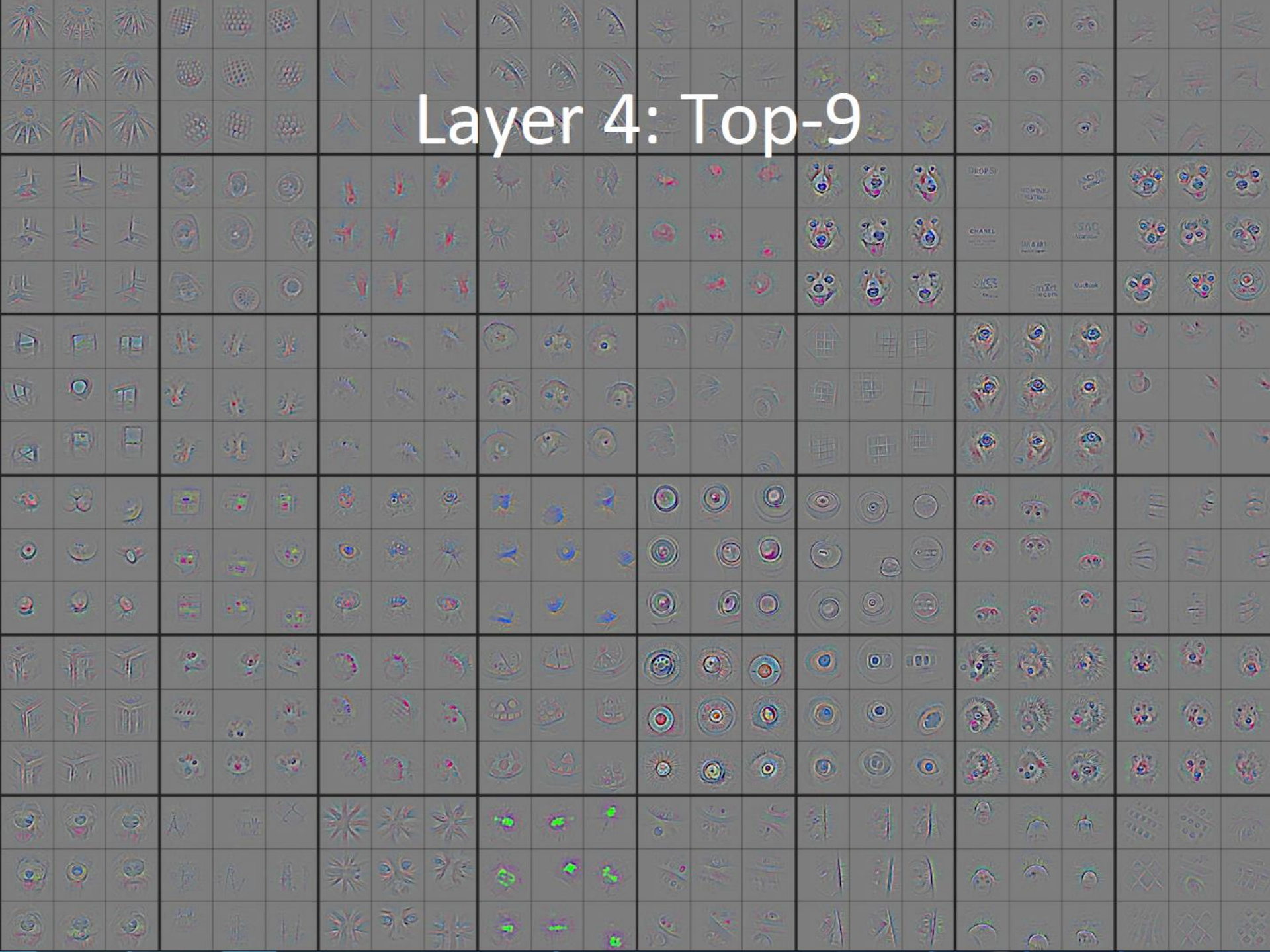
Layer 3: Top-9



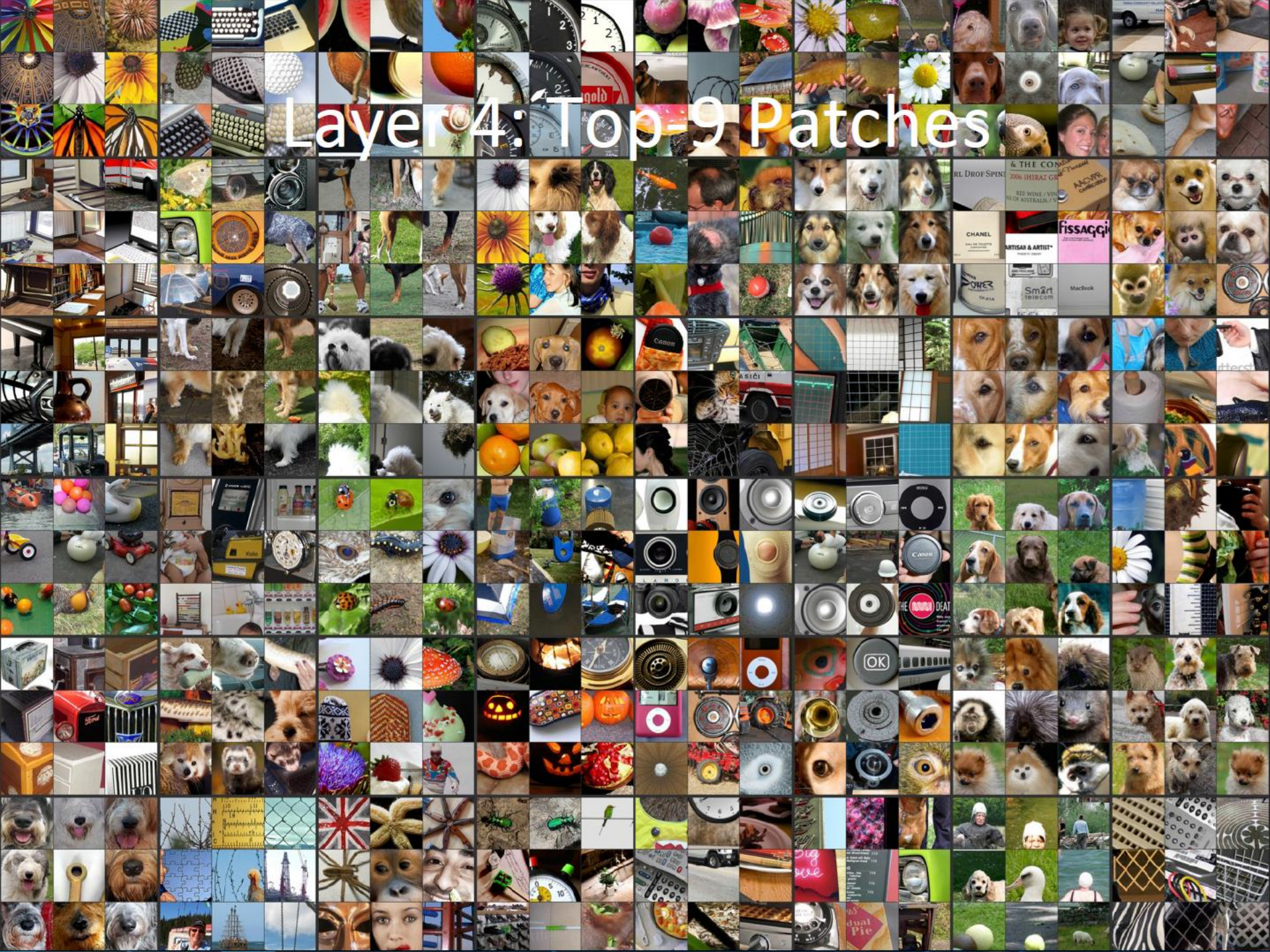
Layer 3: Top-9 Patches



Layer 4: Top-9



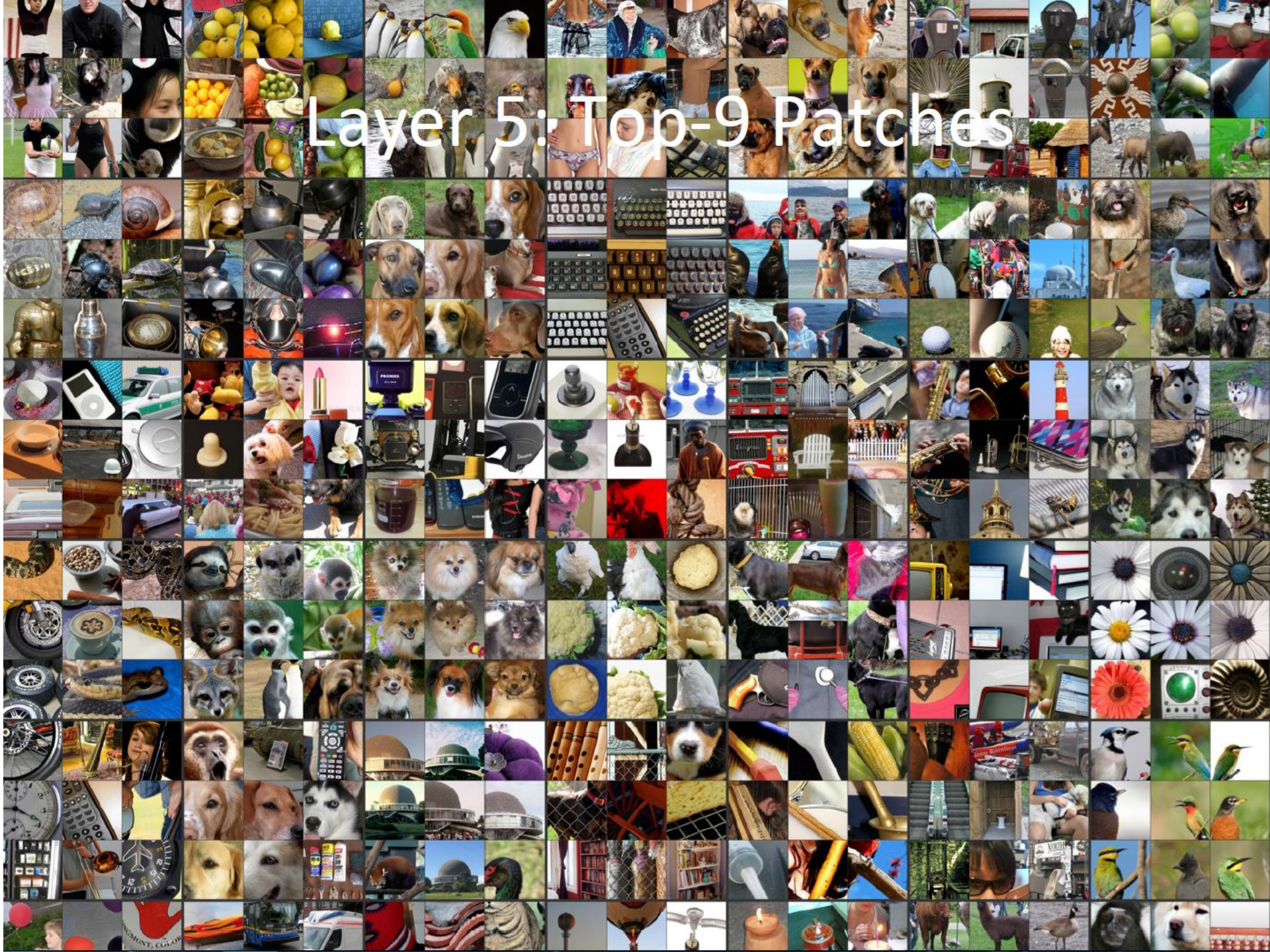
Layer 4: Top-9 Patches



Layer 5: Top-9



Layer 5: Top-9 Patches



Как использовать визуализацию?

- Понимание, что именно обучается в рамках сети
- Диагностика проблем, связанных с архитектурой сети.

Occlusion Experiment

- Mask parts of input with occluding square

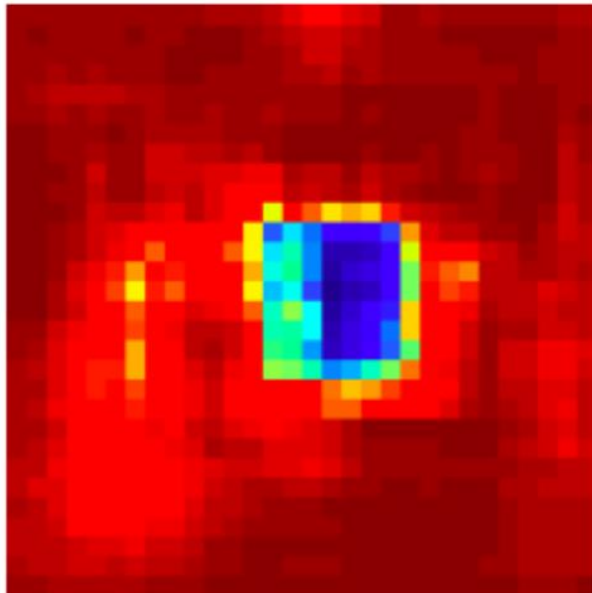


- Monitor output

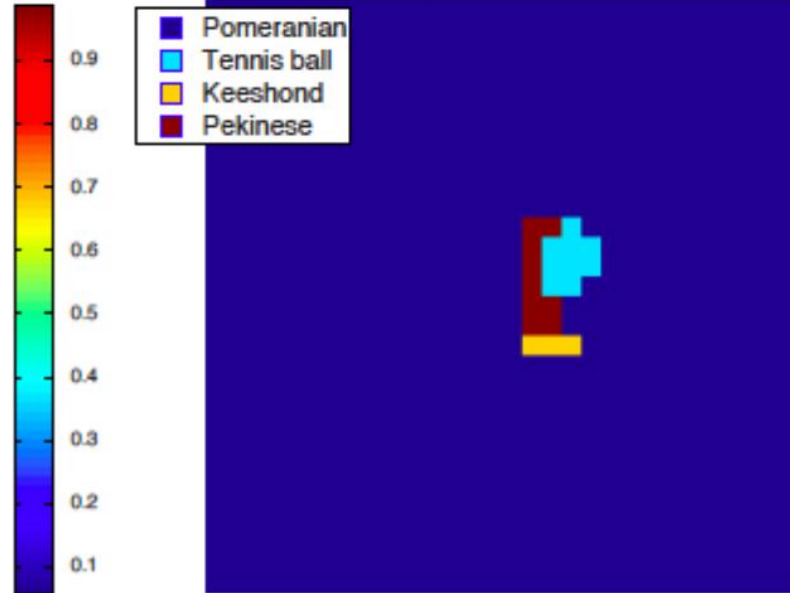
- Perhaps network using scene context?



$p(\text{True class})$

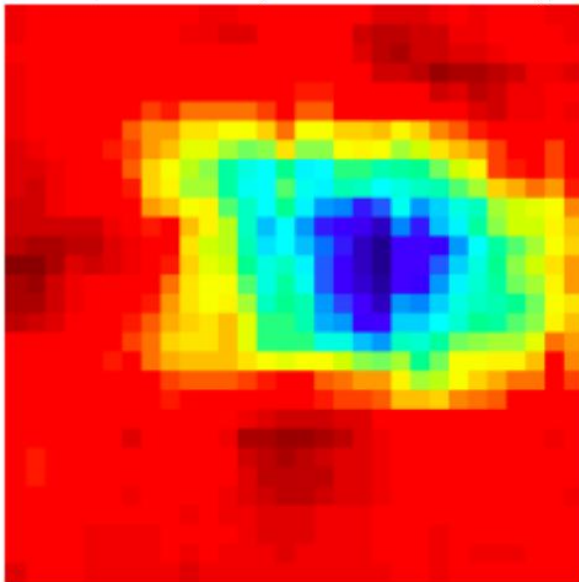


Most probable class

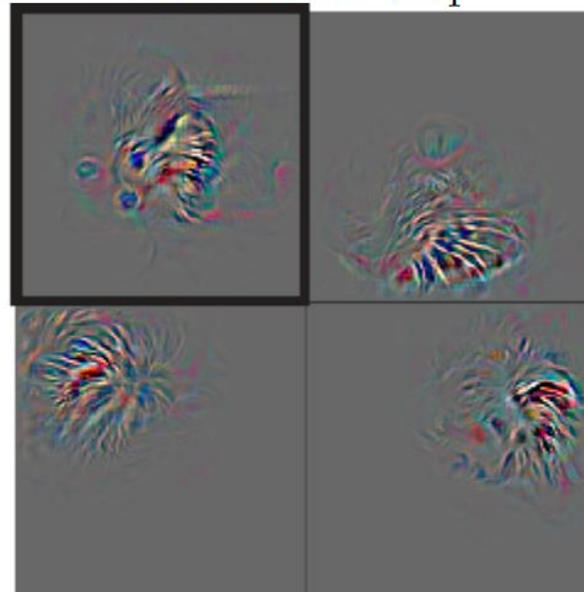


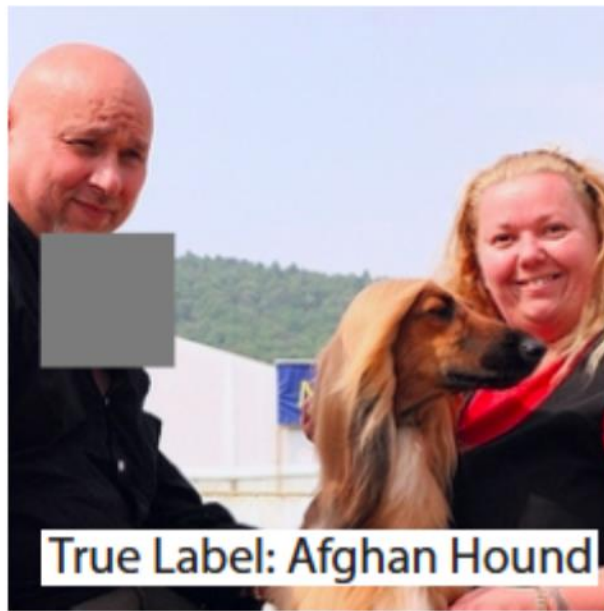


Total activation in most active 5th layer feature map

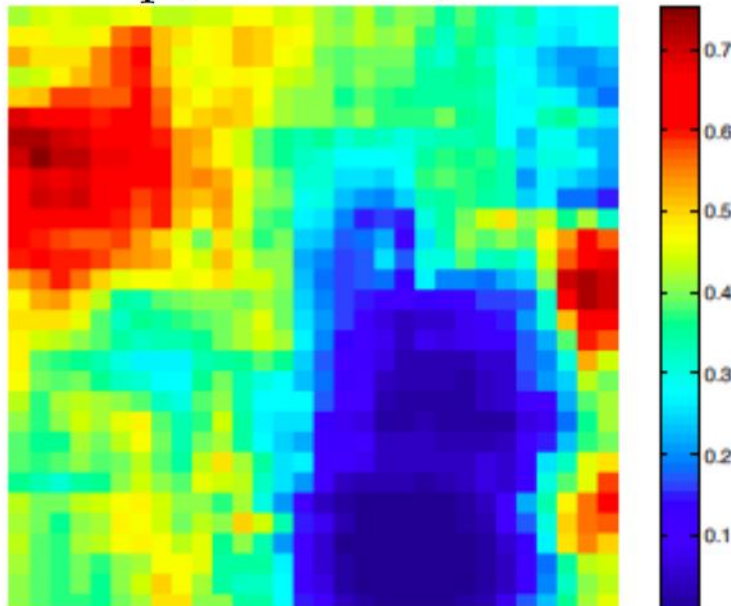


Other activations from same feature map

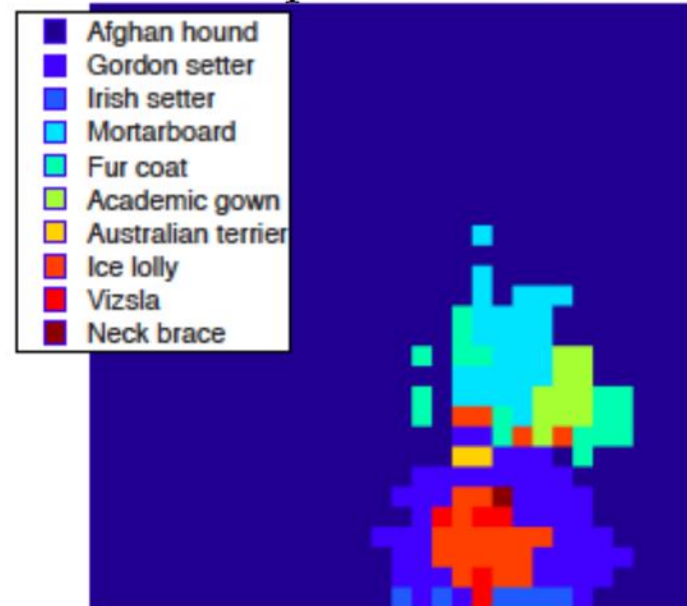


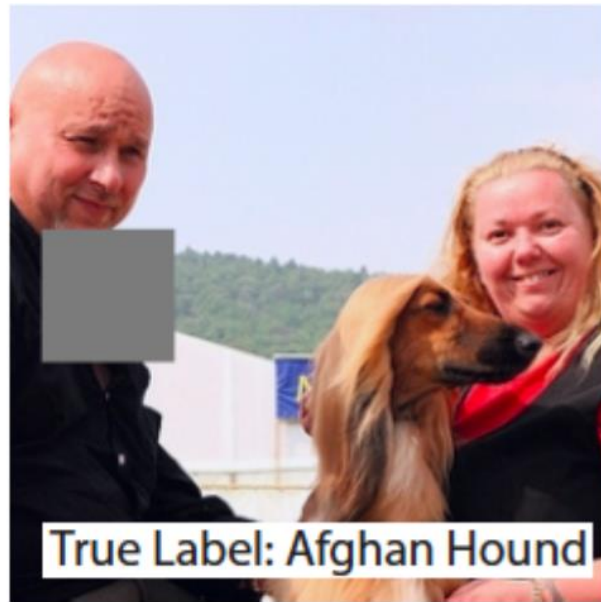


$p(\text{True class})$

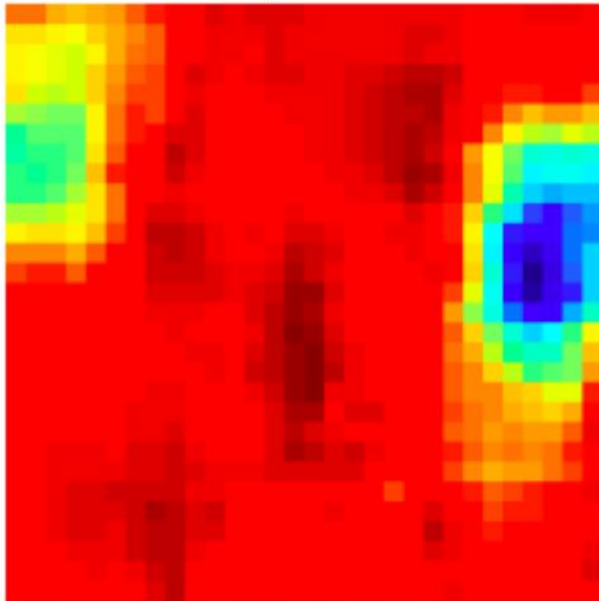


Most probable class

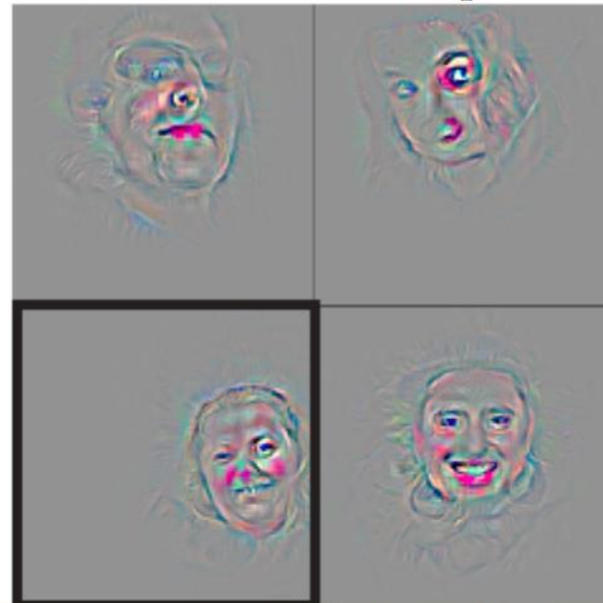




Total activation in most active 5th layer feature map

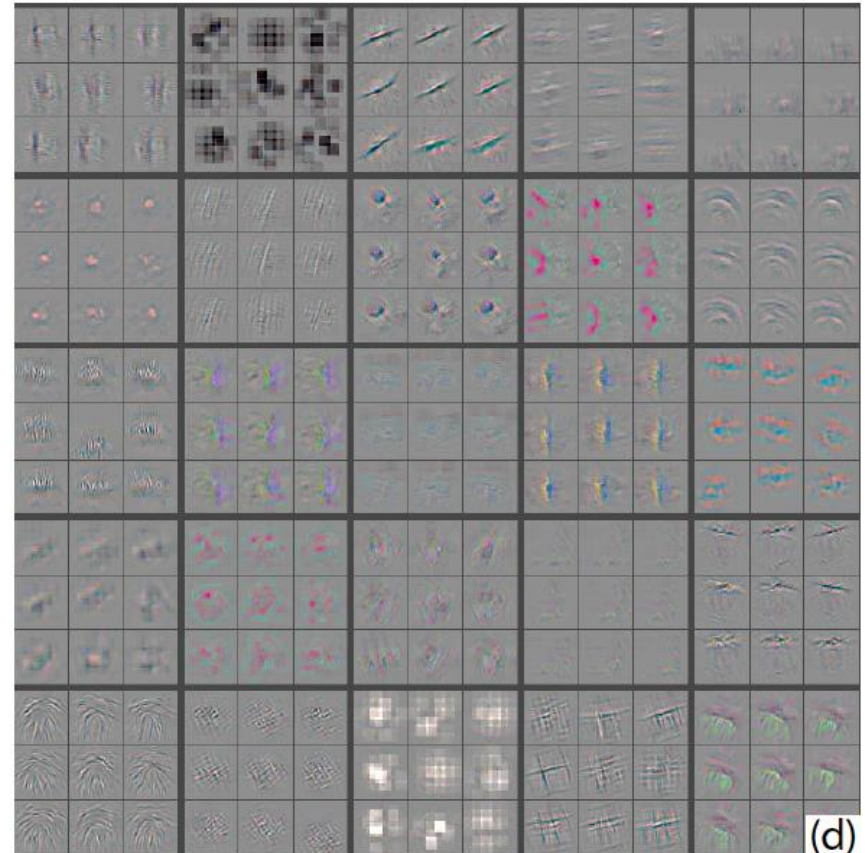
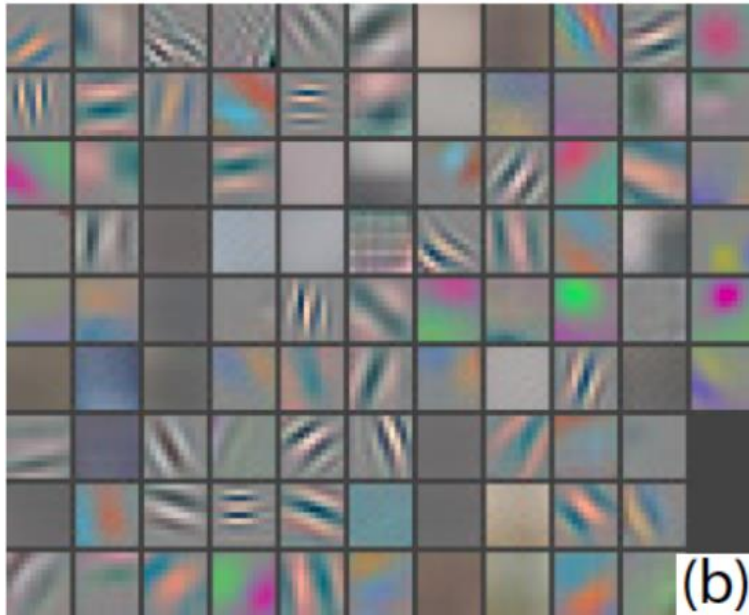


Other activations from same feature map



Анализ фильтров первых слоёв

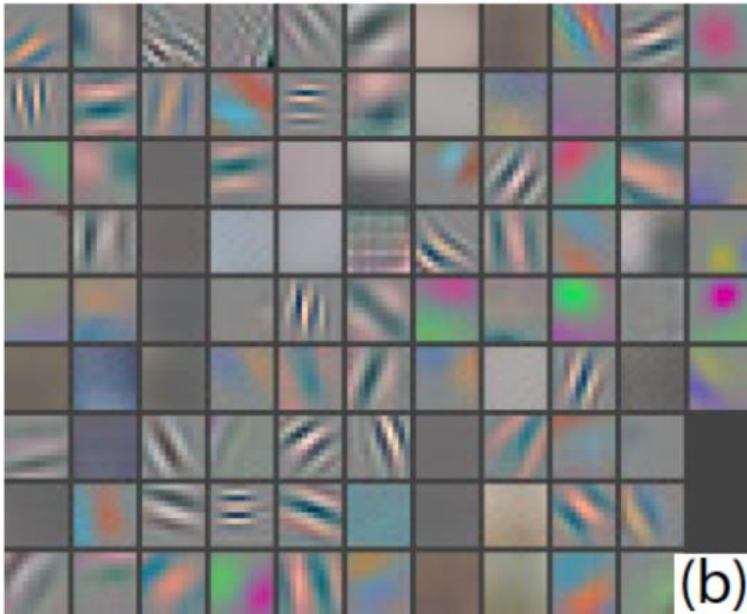
Krizhevsky et al.



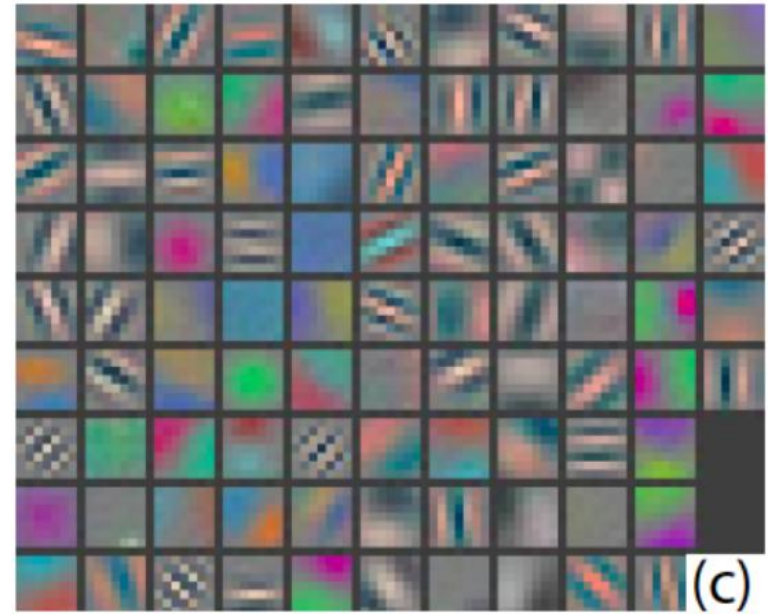
- 1) Много «мёртвых» фильтров
- 2) Есть высоко-частотные и низко-частотные фильтры. Средние частоты не представлены
- 3) Blocking artifacts на втором слое

Фильтры первого слоя

Krizhevsky et al.



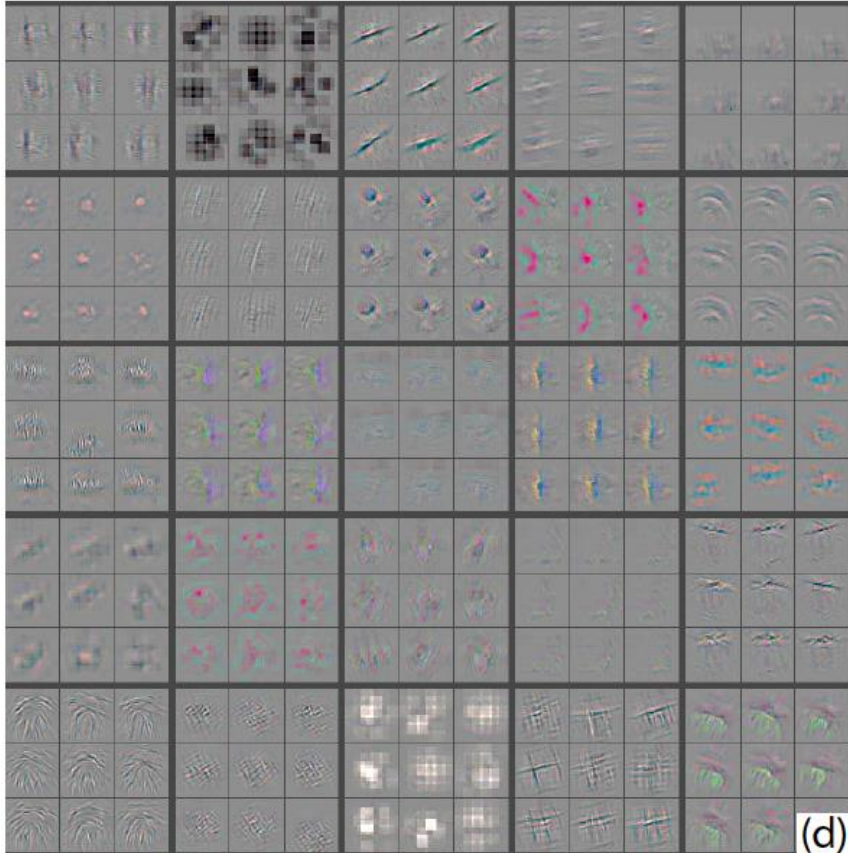
Zeiler and Fergus



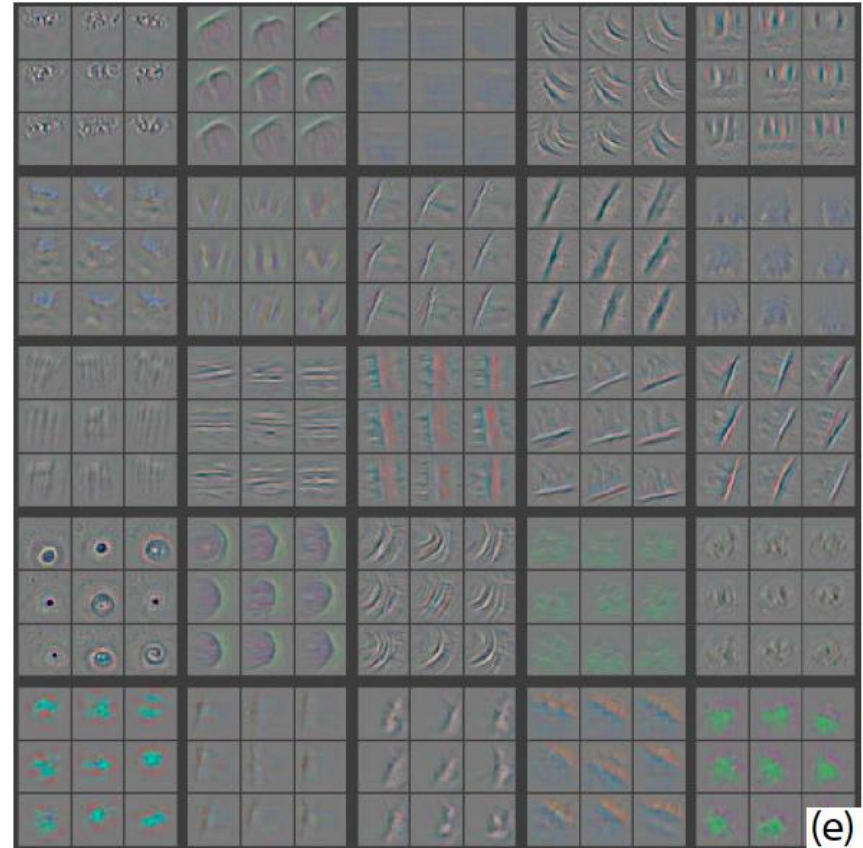
- 1) Уменьшен размер фильтров (с 11x11 до 7x7)
- 2) уменьшено смещение между соседними операциями Pooling.

Фильтры второго слоя

Krizhevsky et al.

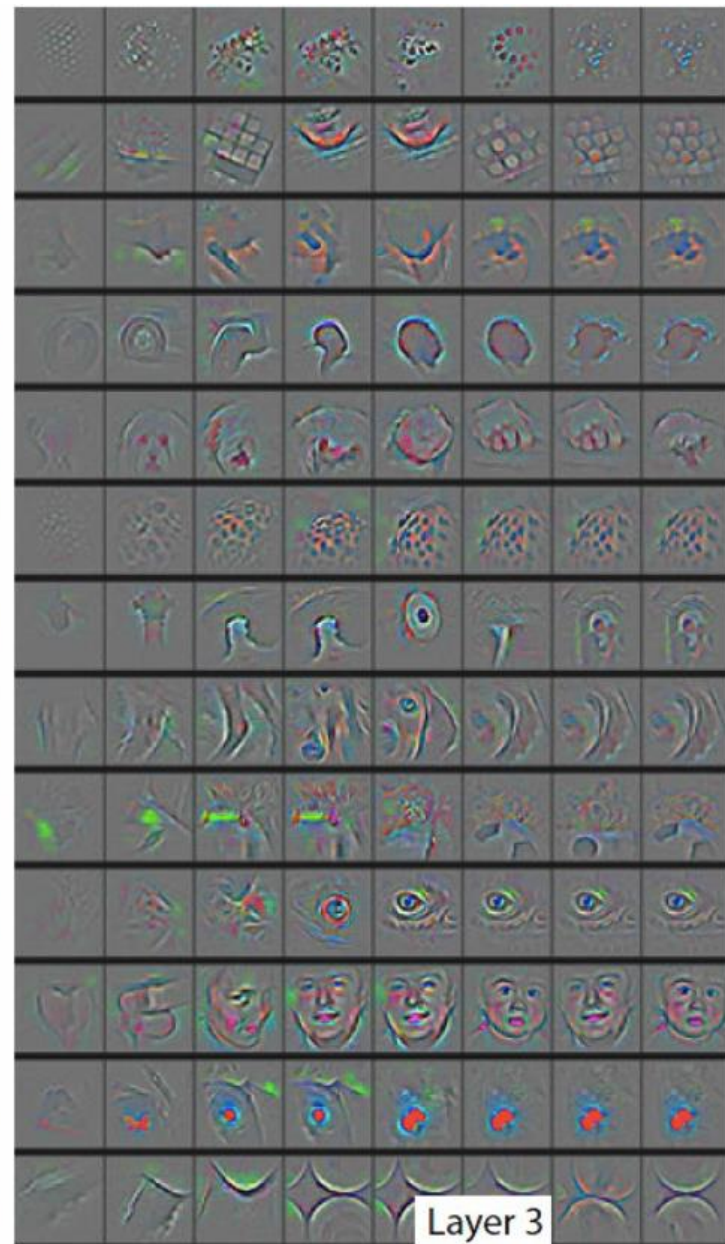
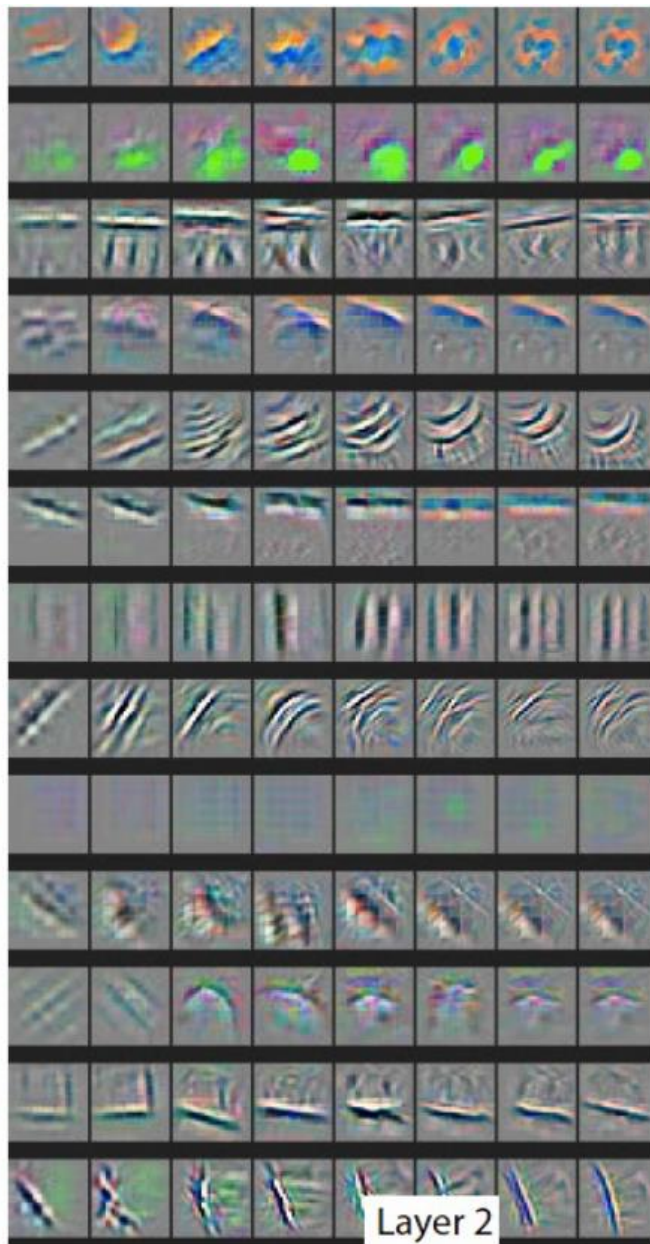
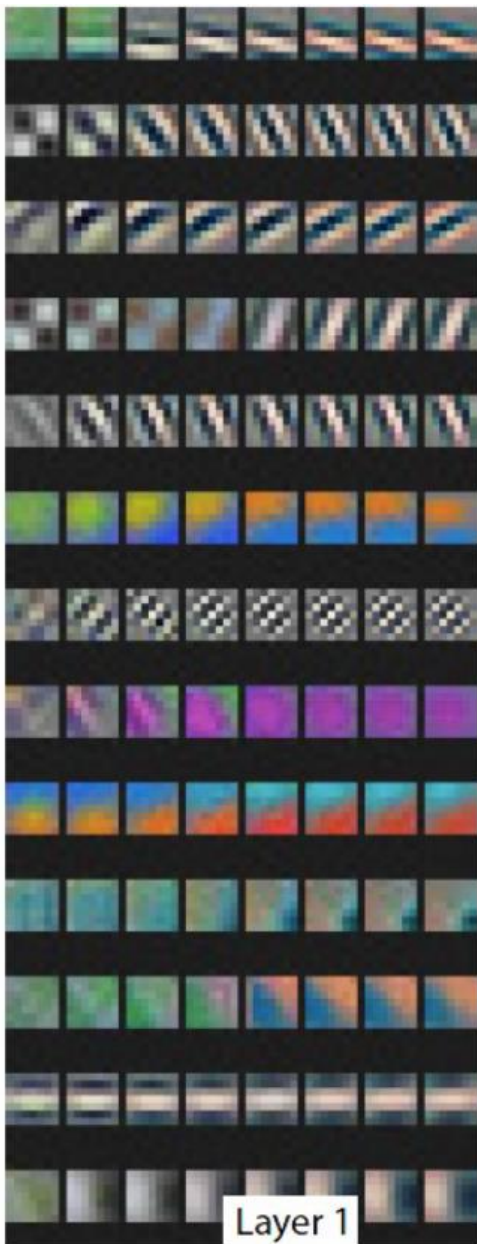


Zeiler and Fergus

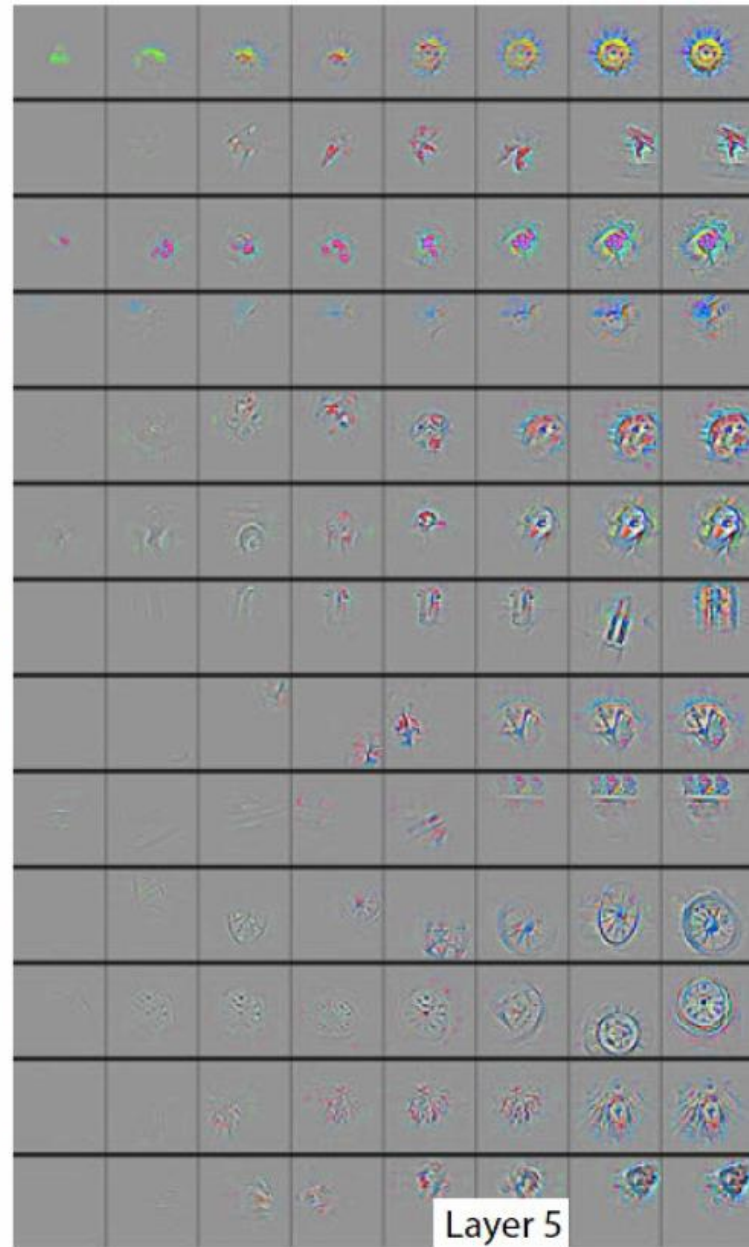
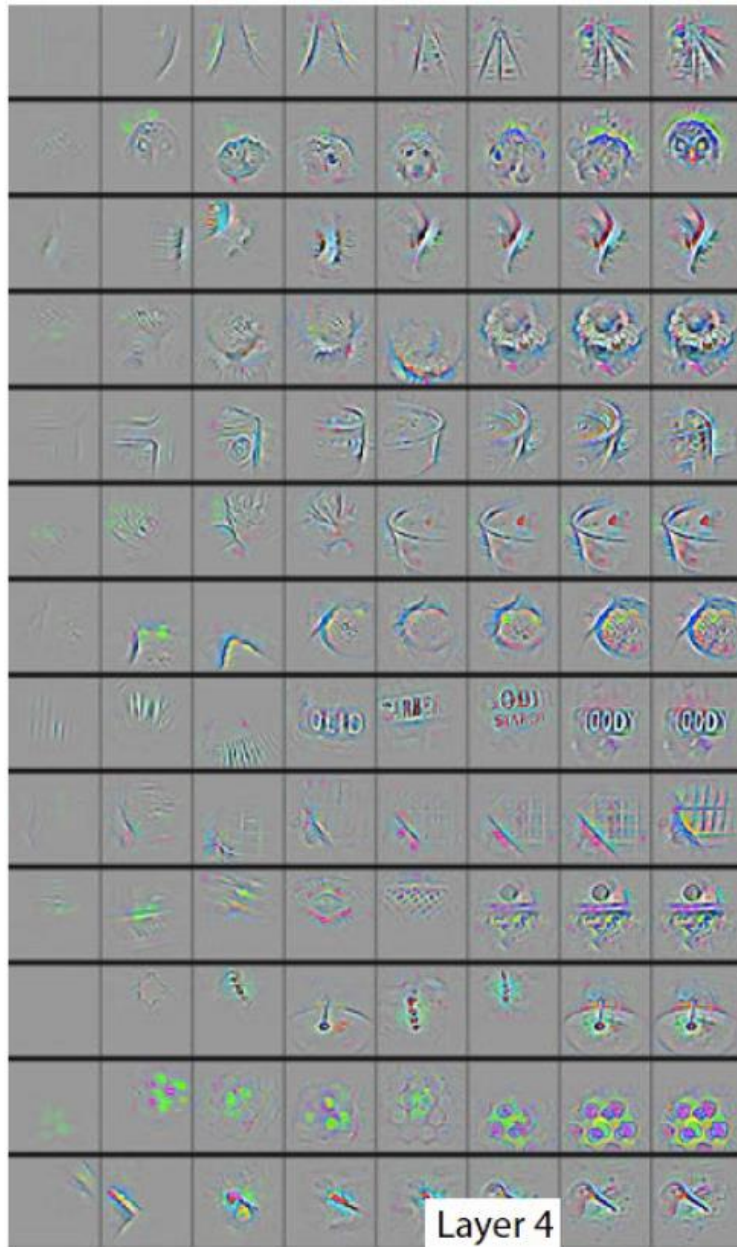


- 1) Уменьшен размер фильтров (с 11x11 до 7x7)
- 2) уменьшено смещение между соседними операциями Pooling.

Изменение откликов в процессе обучения



Изменение откликов в процессе обучения



R. Fergus, Deep Learning for Computer Vision, NIPS
2013 tutorial

Слайды тьюториала в хорошем качестве:

[http://cs.nyu.edu/~fergus/presentations/nips2013_](http://cs.nyu.edu/~fergus/presentations/nips2013_final.pdf)
[final.pdf](http://cs.nyu.edu/~fergus/presentations/nips2013_final.pdf)