



ТЕХНОСФЕРА

Лекция n5 Сверточные нейронные сети

Нестеров Павел

15 апреля 2016 г.

План лекции

Биологические предпосылки

Формализация теории

Современные глубокие сверточные нейронные сети

Transfer learning

Применение transfer learning

Другие применения сверточных сетей

Карикатура на мозг человека¹

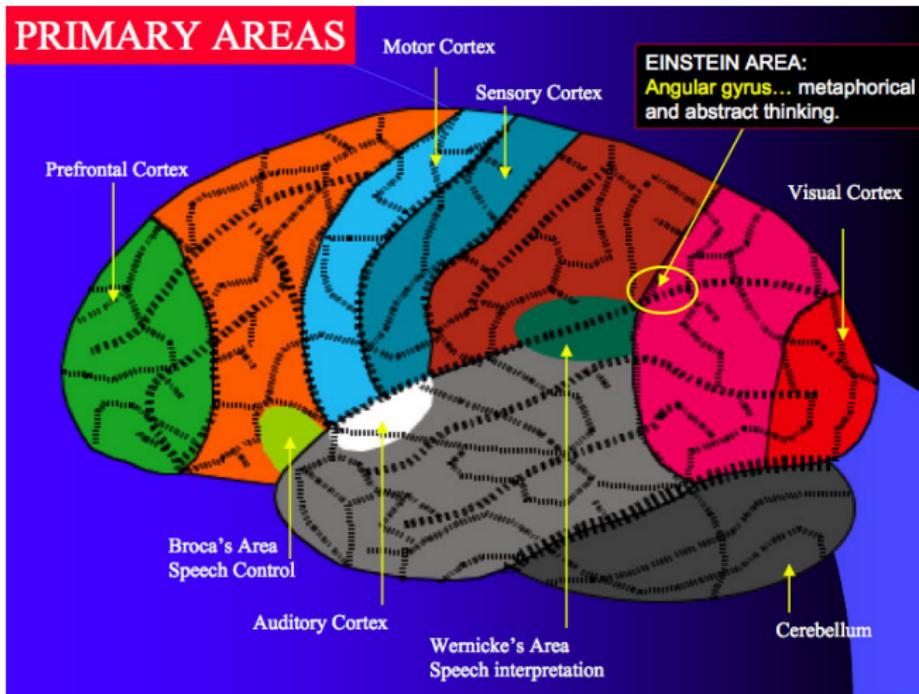
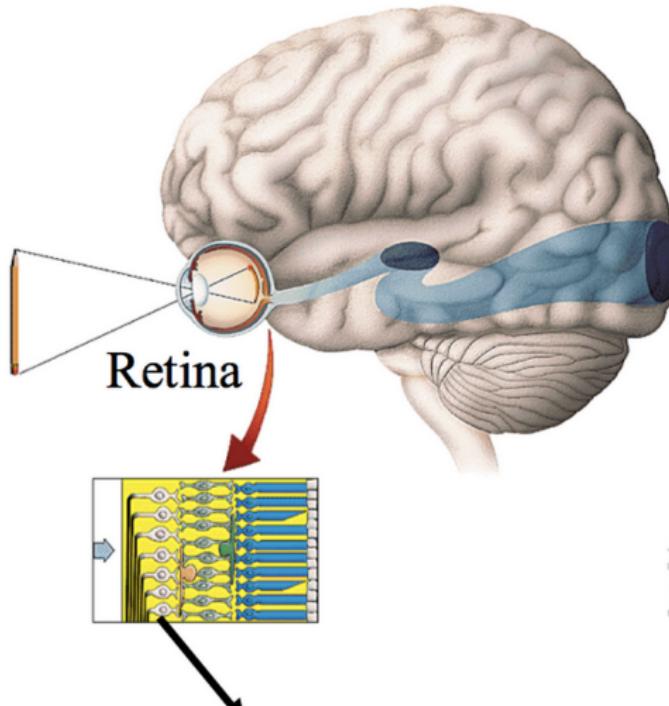


Рис.: Модель верхней коры головного мозга человека

¹Michael I. Jordan, <http://habrahabr.ru/company/1cloud/blog/258219/>

Глаз человека, #1



Retinal Ganglion Cells

Рис.: Модель зрительной системы

George Eads

Глаз человека, #2

Linear receptive field model

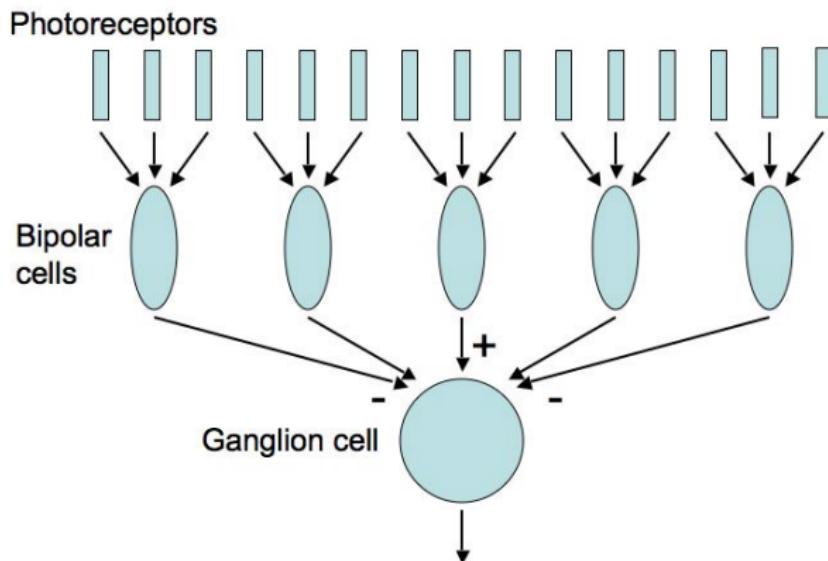
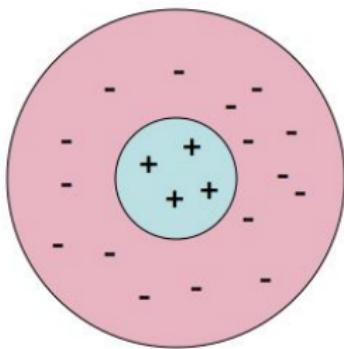


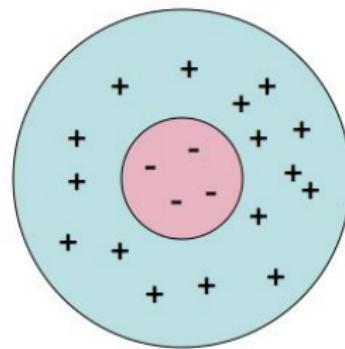
Рис.: Рецептивное поле сенсорного нейрона — участок с рецепторами, которые при воздействии на них определённого стимула приводят к изменению возбуждения этого нейрона.

ON- and OFF-center ganglion cell, #1

Receptive Fields



On-center, Off-surround

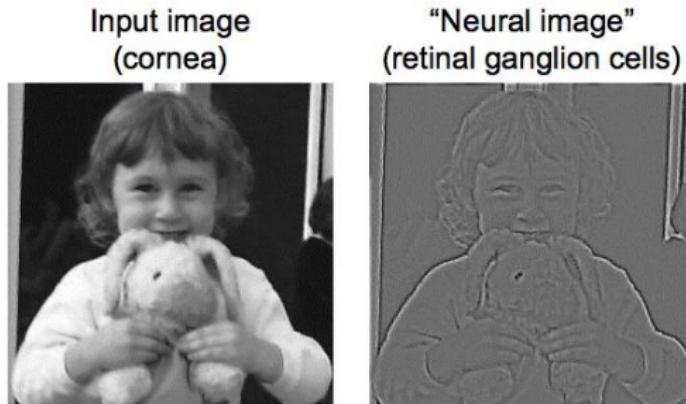


Off-center, On-surround

Рис.: Один из типов биполярной клетки

ON- and OFF-center ganglion cell, #2

Retinal ganglion cells respond to edges



Center-surround receptive fields: emphasize edges.

Рис.: Результат - edge detection²

²<http://www.cns.nyu.edu/~david/courses/perception/lecturenotes/ganglion/ganglion.html>

Модель Hubel & Weisel

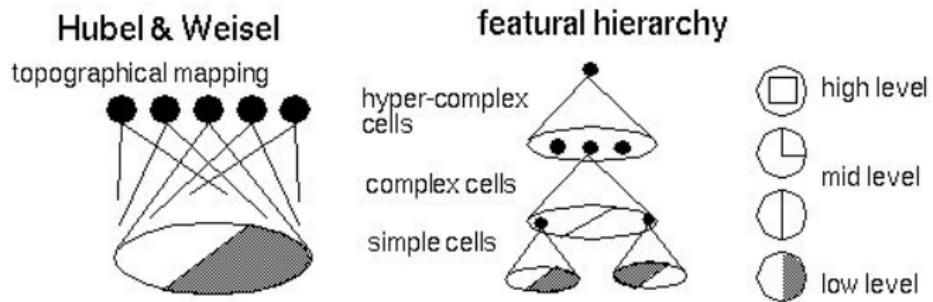


Рис.: Модель Hubel & Weisel³

- ▶ предложена нейрофизиологами Дэвидом Хьюбелом и Торстеном Визельем в 1959 году
- ▶ в 1981 году награждены Нобелевской премией «за открытия, касающиеся принципов переработки информации в нейронных структурах»
- ▶ фундаментальные работы Визеля и Хьюбела по нейрофизиологии зрения открыли основы организации и развития нейронных цепей, ответственных за зрительное распознавание объектов

³<http://cns-alumni.bu.edu/~slehar/webstuff/pcave/hubel.html>

Персепtron Розенблатта

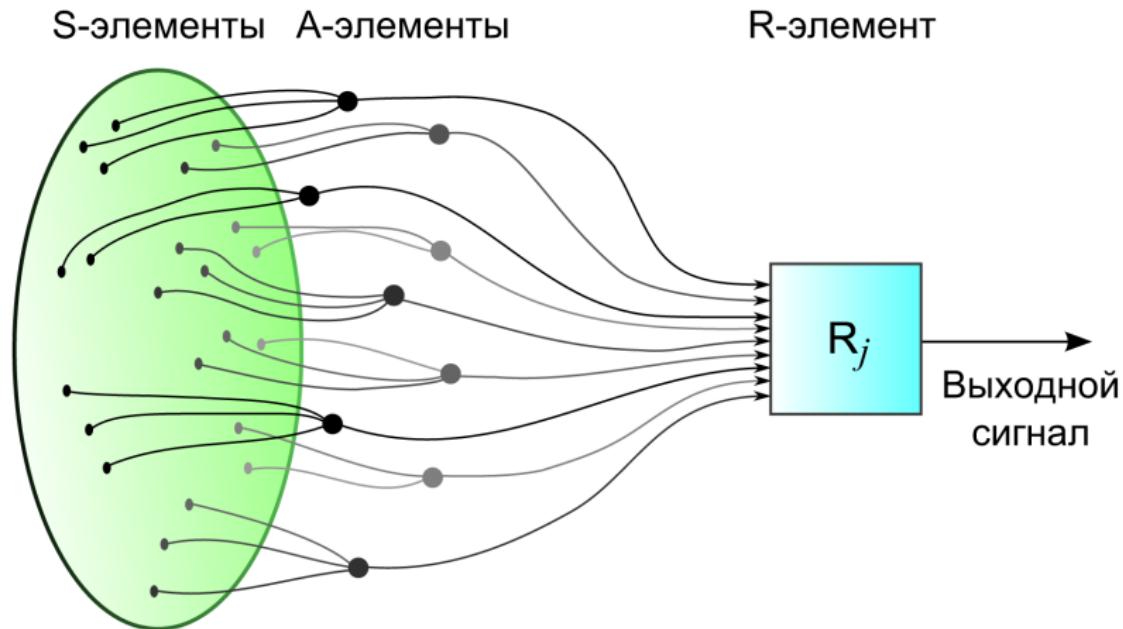


Рис.: Модель персептрана Розенблатта

Когнитрон⁵, #1

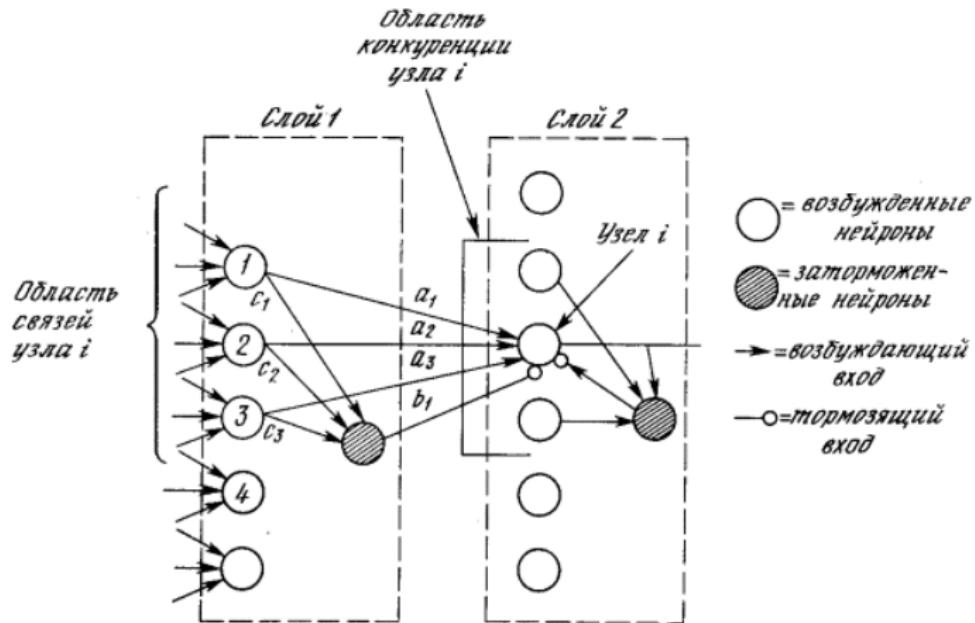


Рис.: Модель когнитрона⁴

⁴<http://uran.donetsk.ua/~masters/2011/fknt/sova/library/Wosserman.htm>

⁵Fukushima K. 1975. Cognitron: A self-organizing multilayered neural network.

Когнитрон, #2

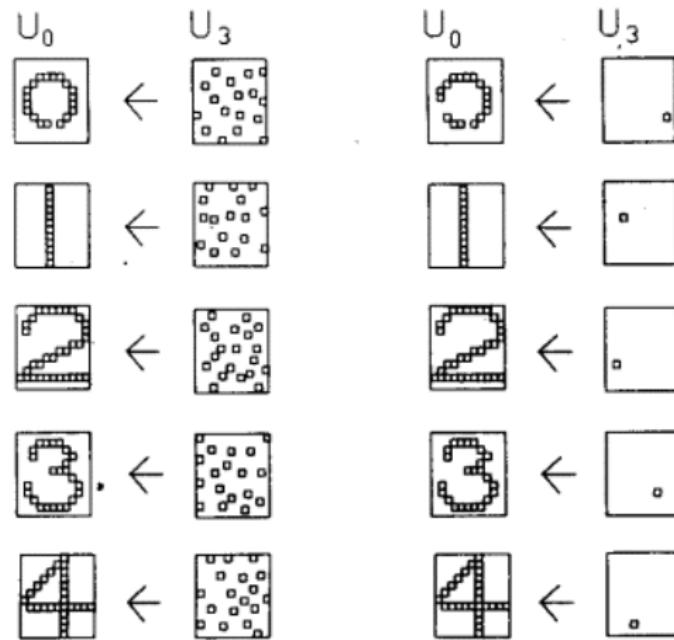


Рис.: Активация на скрытых слоях когнитрона в результате самоорганизации⁶

⁶<http://uran.donetsk.ua/~masters/2011/fknt/sova/library/Wosserman.htm>

Неокогнитрон⁸, #1

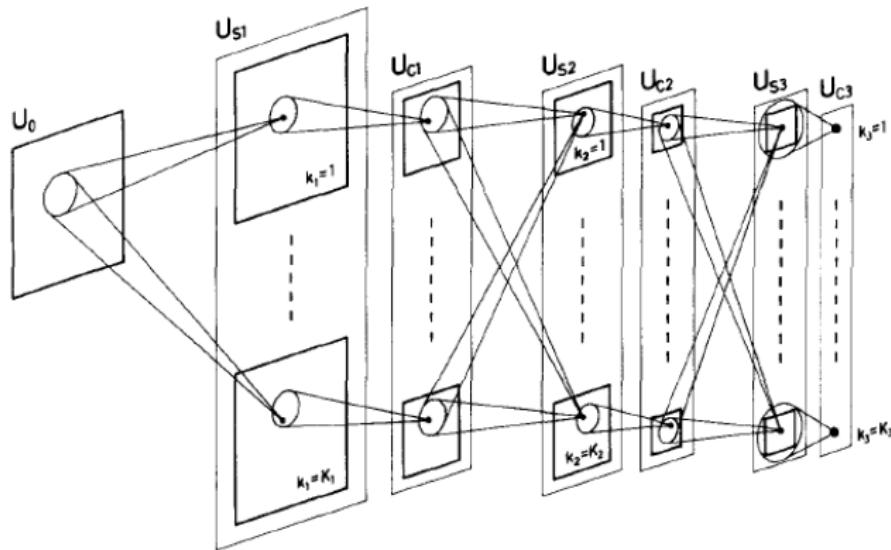


Рис.: Модель неокогнитрона: добавились комплексные узлы, задачей которых является уменьшение зависимости реакции системы от позиции образов во входном поле.⁷

⁷ <http://www.cs.princeton.edu/courses/archive/spr08/cos598B/Readings/Fukushima1980.pdf>

⁸ Fukushima K. 1980. Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position

Неокогнитрон, #2

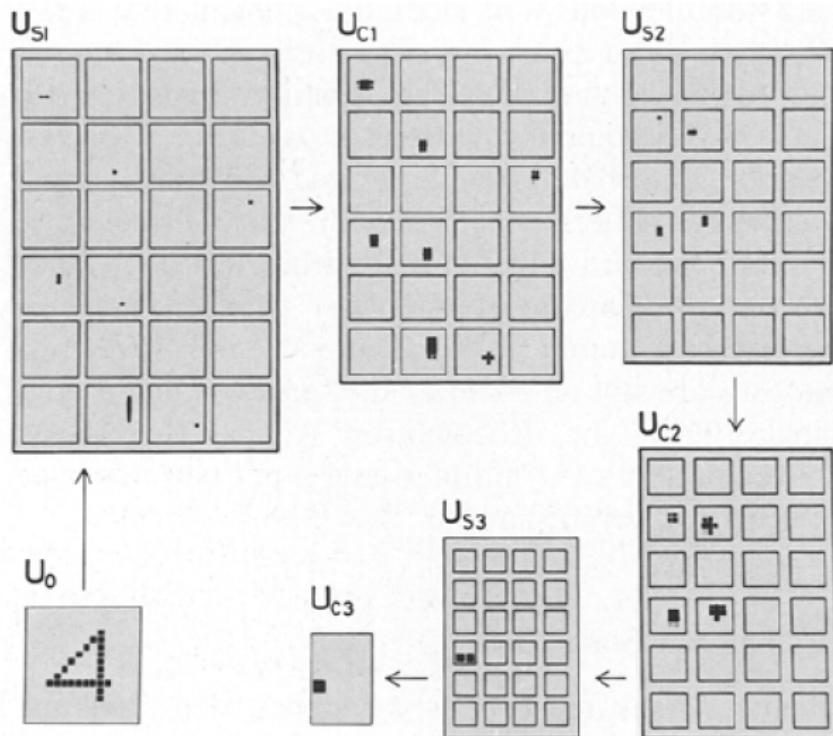


Рис.: Активация на скрытых слоях неокогнитрона в результате самоорганизации⁹

Результаты

Исследование Фукусимы заложило фундамент сверточных нейронных сетей:

- ▶ local feature extraction: каждый нерон имеет свою рецептивную область, что позволяет собирать локальную, а не глобальную статистику, как например в RBM;
- ▶ feature mapping: каждый слой состоит из набора нейронов (filter bank), в результате свертки каждого с входным образом (фильтрация) получается образ меньшей размерности; это дает некоторую инвариантность к положению признака, а так же снижает количество свободных параметров;
- ▶ subsampling/pooling: после каждого сверточного слоя следует пулинг слой (average, max, etc), который еще раз уменьшает размерность, и создает эффект инвариантности к положению и другим видам шума.

LeNet¹²

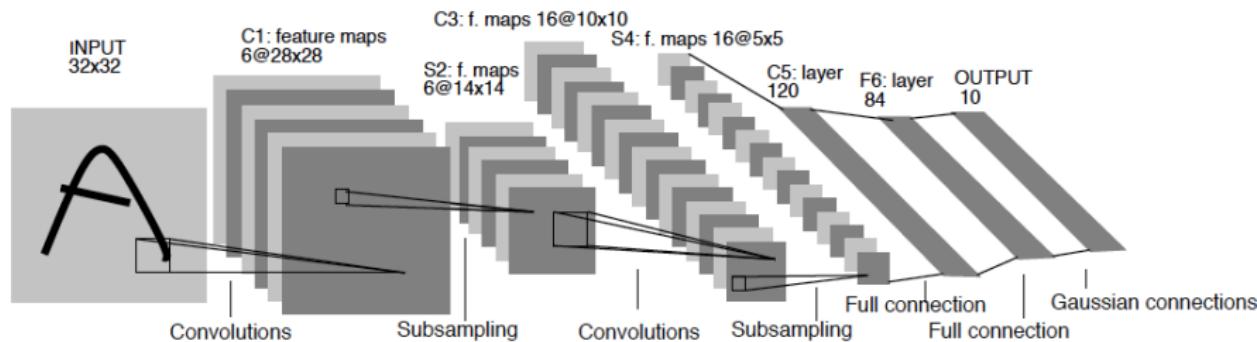


Рис.: Архитектура LeNet-5¹⁰¹¹

¹⁰<http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf>

¹¹<http://yann.lecun.com/exdb/lenet/index.html>

¹²LeCun, Yann; Léon Bottou; Yoshua Bengio; Patrick Haffner (1998). "Gradient-based learning applied to document recognition"

Kernel convolution

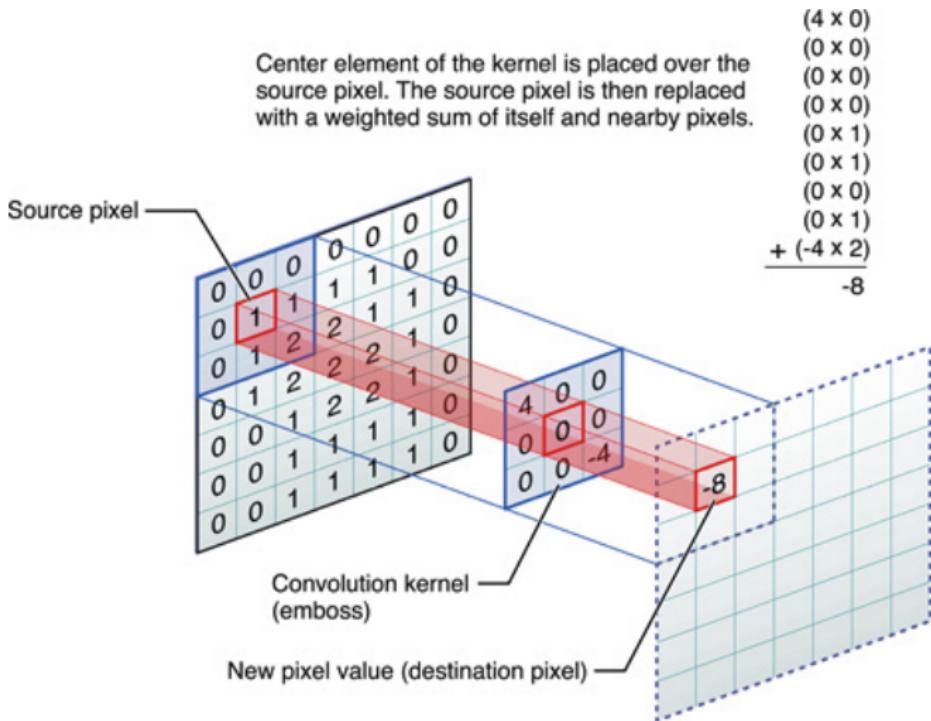
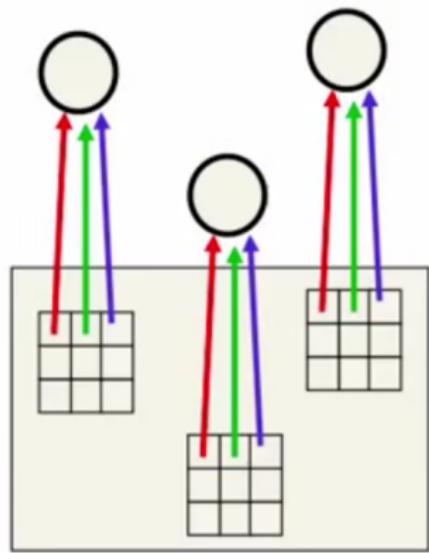


Рис.: Свертка изображения¹³

¹³<https://developer.apple.com/library/ios/documentation/Performance/Conceptual/vImage/ConvolutionOperations/ConvolutionOperations.html>

Replicated features



- ▶ нелинейная функция активации от результата свертки
- ▶ множество копий одного фильтра \Rightarrow снижается количество свободных параметров
- ▶ одновременно обучаются несколько фильтров \Rightarrow представление картинки с различных "точек зрения"

Рис.: Связи одного цвета
имеют одинаковое значение
веса

Weights sharing, #1

Алгоритм backprop легко модифицируется так, что бы можно было наложить любые линейные ограничения на веса. Допустим мы хотим, что бы $w_1 = w_2$:

- ▶ как модифицировать алгоритм?

Weights sharing, #2

Алгоритм backprop легко модифицируется так, что бы можно было наложить любые линейные ограничения на веса. Допустим мы хотим, что бы $w_1 = w_2$:

- ▶ $w_1 = w_2 \Rightarrow \Delta w_1 = \Delta w_2 \Rightarrow \frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial w_2}$
- ▶ $\Delta w_1^{\text{new}} = \Delta w_2^{\text{new}} = \frac{\partial E}{\partial w_1} + \frac{\partial E}{\partial w_2}$

3D weight tensor for filter bank

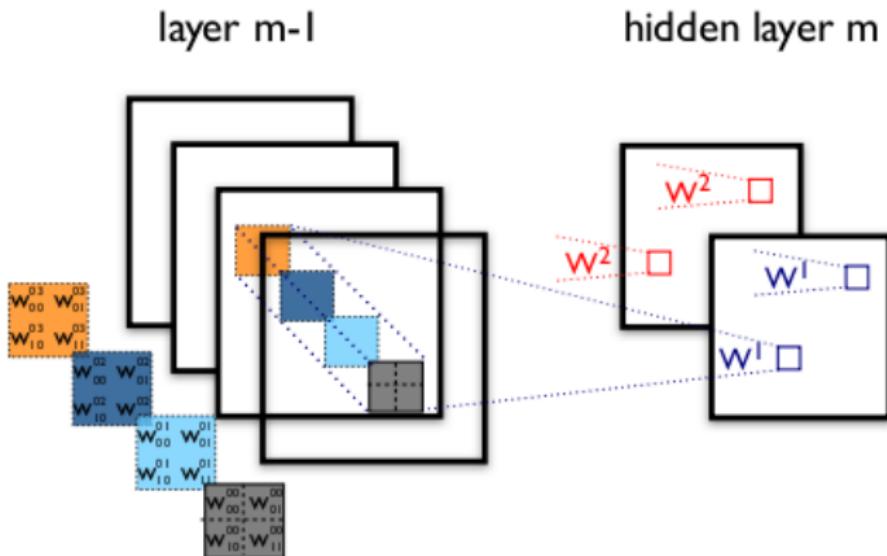


Рис.: Пример сверточного слоя¹⁵

¹⁵<http://deeplearning.net/tutorial/lenet.html>

Изображения внутри filter bank

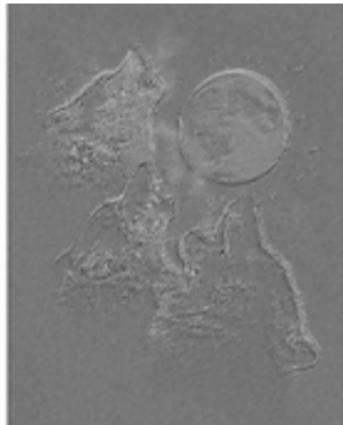


Рис.: Случайная инициализация очень похожа на детектор границ¹⁶

¹⁶<http://deeplearning.net/tutorial/lenet.html>

Pooling, #1

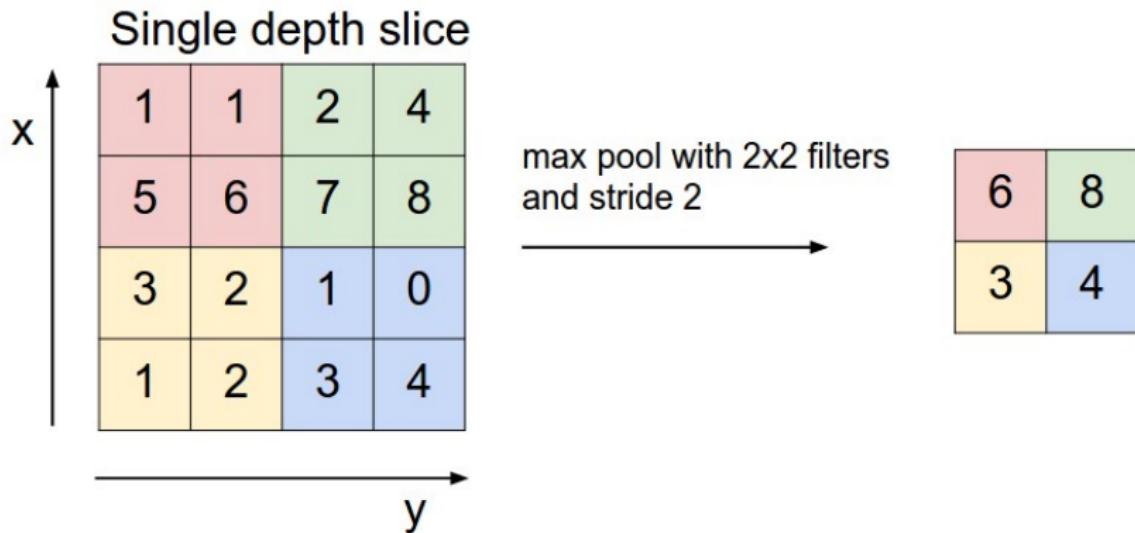


Рис.: Max pooling¹⁷

¹⁷<http://cs231n.github.io/convolutional-networks/>

Pooling, #2

Виды пулинга:

- ▶ sum
- ▶ max
- ▶ average
- ▶ weighted and etc

Функции:

- ▶ некоторая инвариантность к положению \Rightarrow потеря информации об относительном положении частей
- ▶ Большие рецептивные области \Rightarrow больший охват
- ▶ в результате получается нечто похожее на bag-of-word: список признаков на изображении

Результаты LeNet-5 на MNIST



Рис.: Несколько ошибок сети

LeNet-5 <http://yann.lecun.com/exdb/mnist/>

MNIST

MNIST¹⁸ - это набор рукописных букв: 60 000 для тренировки и 10 000 для тестирования

Итогом работ Лекуна и позже других исследователей стала типичная сверточная сеть:

- ▶ сверточный слой
- ▶ пулинг слой
- ▶ ReLu слой
- ▶ dropout слой
- ▶ аугментация данных (смотрим ссылку)
- ▶ любая функция ошибки в зависимости от задачи

¹⁸<http://yann.lecun.com/exdb/mnist/>

ImageNet, описание



- ▶ 1000 классов
- ▶ около 1000 изображений в каждом классе
- ▶ около 1 000 000 изображений всего
- ▶ несколько номинаций: таких как распознавание и детектирование/локализация

ImageNet, правила

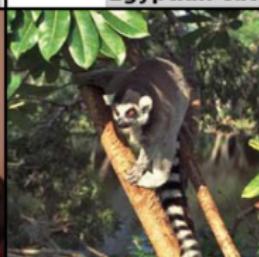
			
mite mite black widow cockroach tick starfish	container ship container ship lifeboat amphibian fireboat drilling platform	motor scooter motor scooter go-kart moped bumper car golfcart	leopard leopard jaguar cheetah snow leopard Egyptian cat
			
grille convertible grille pickup beach wagon fire engine	mushroom agaric mushroom jelly fungus gill fungus dead-man's-fingers	cherry dalmatian grape elderberry ffordshire bullterrier currant	Madagascar cat squirrel monkey spider monkey titi indri howler monkey

Рис.: Примеры прогнозов

ImageNet, победители

	Classification		
	1st	2nd	3rd
2014	GoogLeNet	VGG	MSRA ¹⁹
2013	Clarifai ²⁰	NUS ²¹	ZF ²²
2012	SuperVision ²³	ISI	VGG

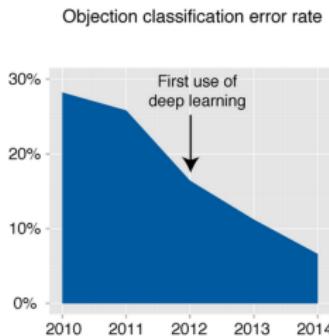


Рис.: Прогресс в конкурсе по распознаванию

¹⁹Spatial Pyramid Pooling in Deep Convolutional Networks

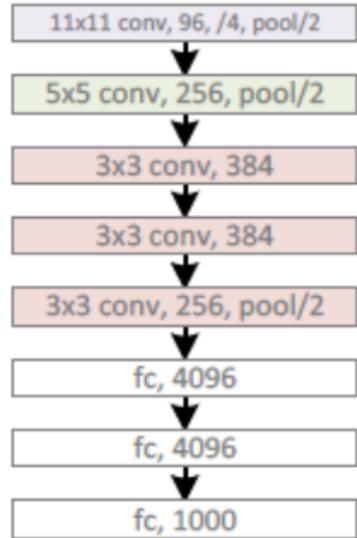
²⁰large deep convolutional network

²¹SVM+DCNN (AlexNet based)

²²ensemble of large CNN

²³AlexNet

AlexNet²⁴



- ▶ 5 сверточных и 3 полносвязных слоя
- ▶ 60M параметров, 650к нейронов
- ▶ эффективное распараллеливание на 2 GPU/CUDA
- ▶ свёртки 11x11, 5x5, 3x3

²⁴Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton: ImageNet Classification with Deep Convolutional Neural Networks, 2012
<http://www.cs.toronto.edu/~fritz/absps/imagenet.pdf>

VGG²⁵

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

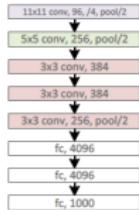
- ▶ VGG-19 (E):
144M
параметров
- ▶ Very Deep
Convolutional
Networks for
Large-Scale
Image
Recognition^a

^aK. Simonyan, A. Zisserman

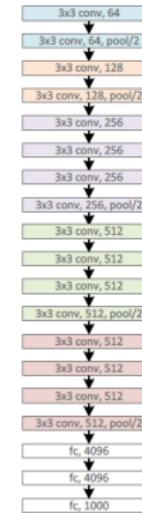
²⁵<http://arxiv.org/pdf/1409.1556.pdf>

Revolution of Depth²⁶, #1

AlexNet, 8 layers
(ILSVRC 2012)



VGG, 19 layers
(ILSVRC 2014)



GoogleNet, 22 layers
(ILSVRC 2014)



²⁶Deep Residual Learning for Image Recognition,
<http://arxiv.org/pdf/1512.03385v1.pdf>

Revolution of Depth²⁷, #2

AlexNet, 8 layers
(ILSVRC 2012)



VGG, 19 layers
(ILSVRC 2014)



ResNet, 152 layers
(ILSVRC 2015)



²⁷Deep Residual Learning for Image Recognition,
<http://arxiv.org/pdf/1512.03385v1.pdf>

Revolution of Depth, #3

- ▶ [http://josephpcohen.com/w/
visualizing-cnn-architectures-side-by-side-with-mxnet/](http://josephpcohen.com/w/visualizing-cnn-architectures-side-by-side-with-mxnet/)

National Data Science Bowl²⁹

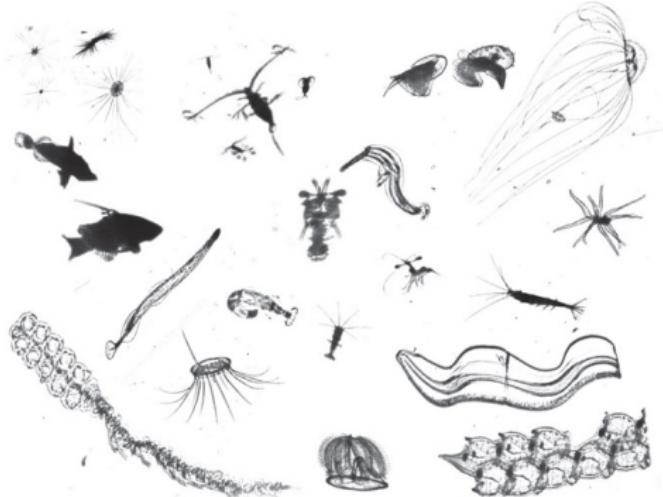


Рис.: Примеры изображений из набора данных: всего 121 несбалансированный класс, 30 000 изображений

Победители ²⁸ использовали архитектуру VGG-16, а так же рад приемов.

²⁸<http://benanne.github.io/2015/03/17/plankton.html>

²⁹<https://www.kaggle.com/c/datasciencebowl>

Data augmentation, #1

Data augmentation:

- ▶ rotation: random with angle between 0 and 360 degree (uniform)
- ▶ translation: random with shift between -10 and 10 pixels (uniform)
- ▶ rescaling: random with scale factor between 1/1.6 and 1.6 (log-uniform)
- ▶ flipping: yes or no (bernoulli)
- ▶ shearing: random with angle between -20 and 20 degree (uniform)
- ▶ stretching: random with stretch factor between 1/1.3 and 1.3 (log-uniform)

Данный аугментировались в реальном времени, т.к. это значительно увеличивает размер датасета.

Data augmentation, #2

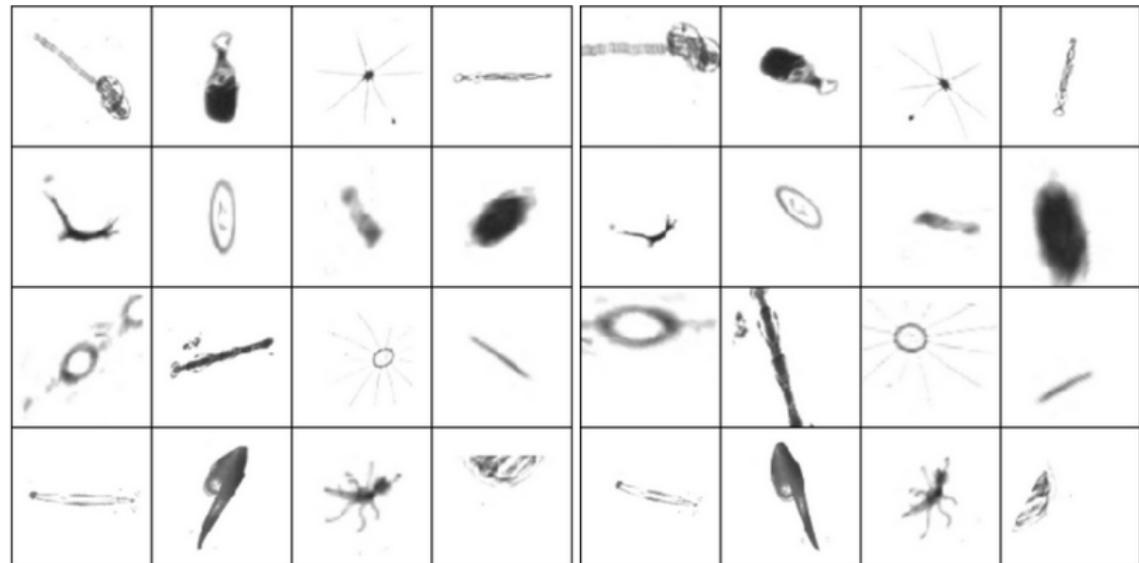


Рис.: Оригиналы и аугментированные образы

Cyclic pooling

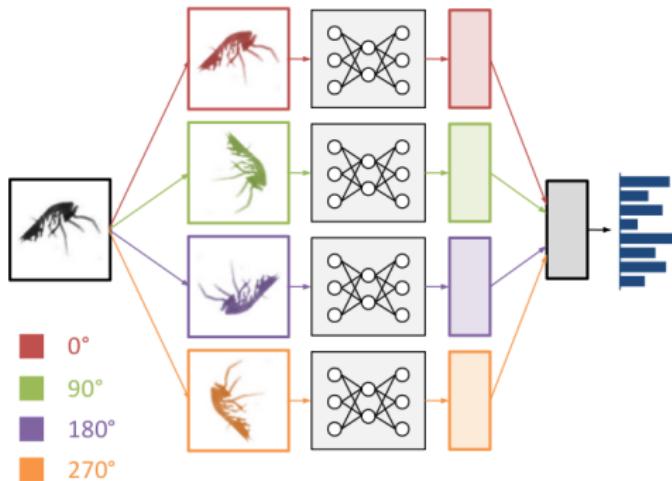


Рис.: Schematic representation of a convnet with cyclic pooling

- ▶ объединение результатов извлечения признаков после прохода по нескольким копиям одной сети параллельно
- ▶ это позволило сократить размер батча в 4 раза (со 128 до 32)
- ▶ root-mean-square pooling оказался эффективнее других

Rolling feature maps

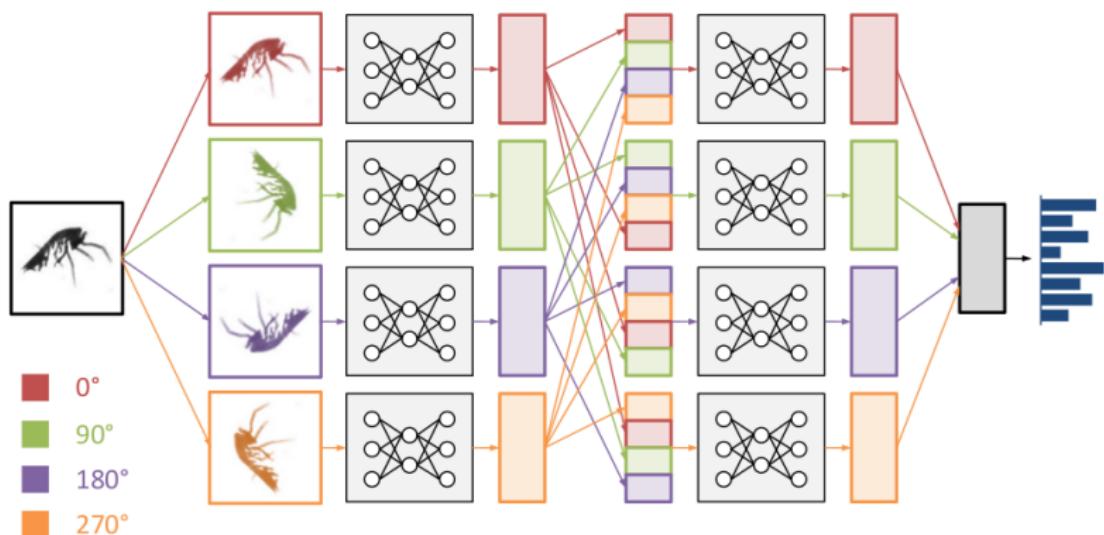


Рис.: Schematic representation of a roll operation inside a convnet with cyclic pooling.

Другие моды

- ▶ leaky ReLU: $f(x) = \max(x, a \cdot x)$, оказалось что для глубокой сети (10+ слоев) значения $a > 0.5$ улучшают качество на данном датасете
- ▶ добавление к сверточным признакам других признаков перед полносвязным слоем (Hu moments, Zernike moments, atc)
- ▶ инициализация весов: Exact solutions to the nonlinear dynamics of learning in deep linear neural networks³⁰
- ▶ предобучение: Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction³¹
- ▶ Pseudo-labeling: Distilling the Knowledge in a Neural Network³²
- ▶ исходники: <https://github.com/benanne/kaggle-ndsb>

³⁰<http://arxiv.org/pdf/1312.6120v3.pdf>

³¹<http://people.idsia.ch/~ciresan/data/icann2011.pdf>

³²<http://arxiv.org/abs/1503.02531>

Извлечение признаков, история

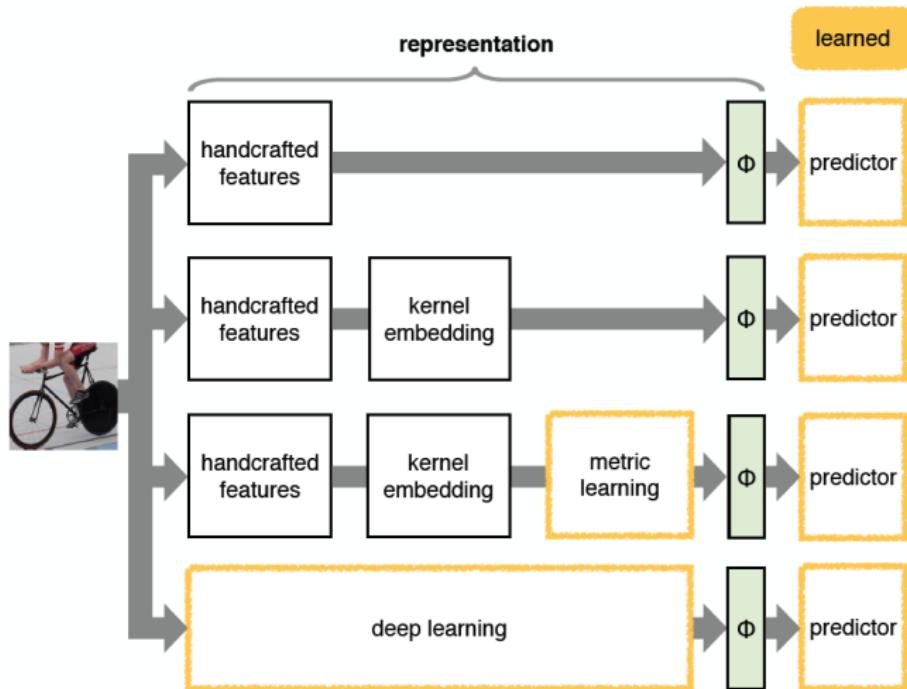


Рис.: Глубокое обучение³³

³³Learning visual representations (Andrea Vedaldi)

Извлечение признаков, примеры

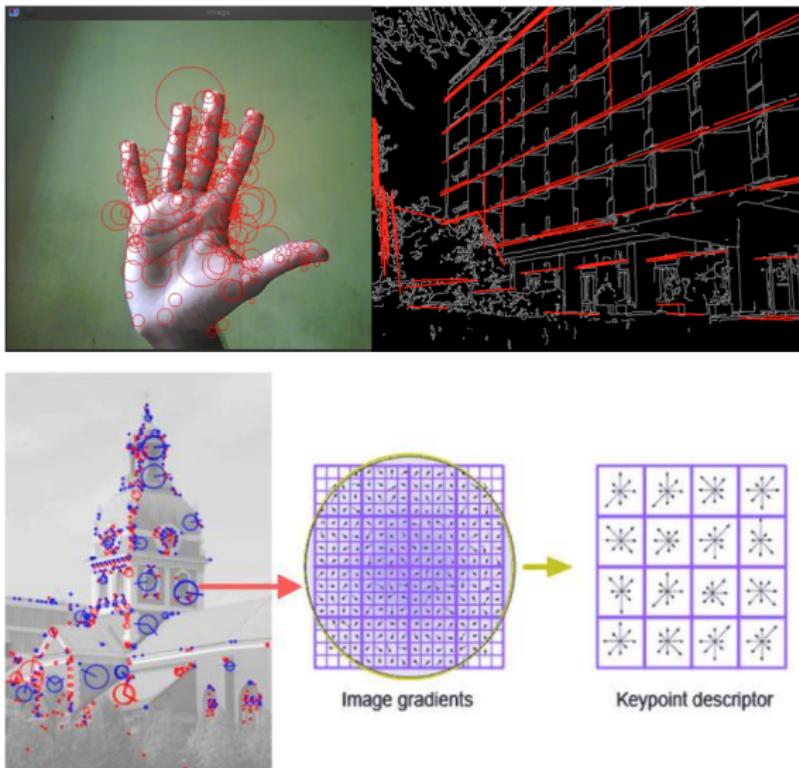


Рис.: Классический подход к извлечению признаков

Извлечение признаков, фильтр Габора

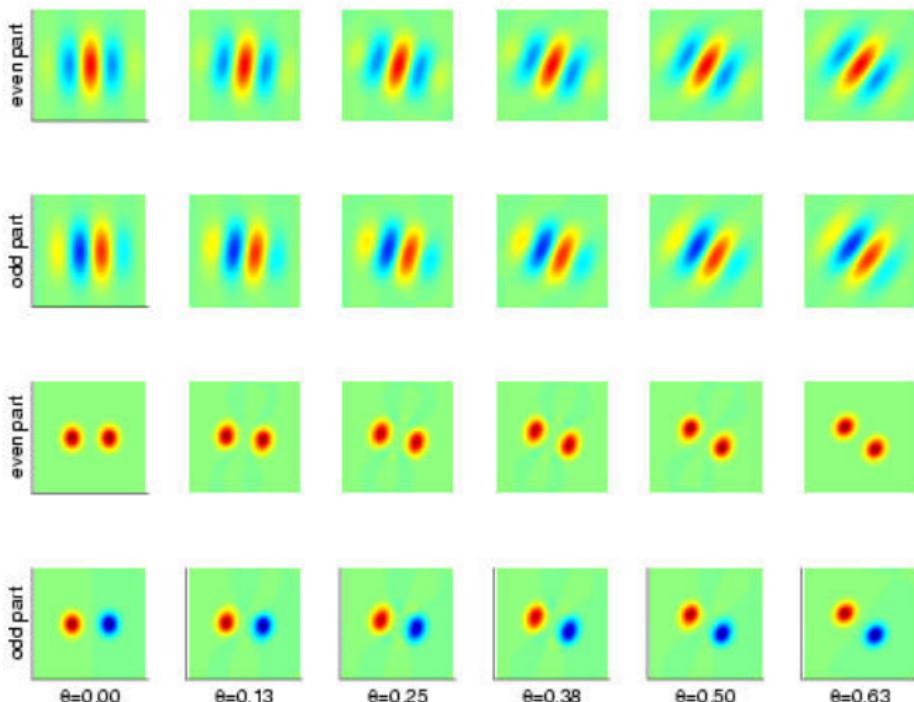


Рис.: Визуализация 2D фильтров Габора с различными параметрами

Выученные признаки, #1

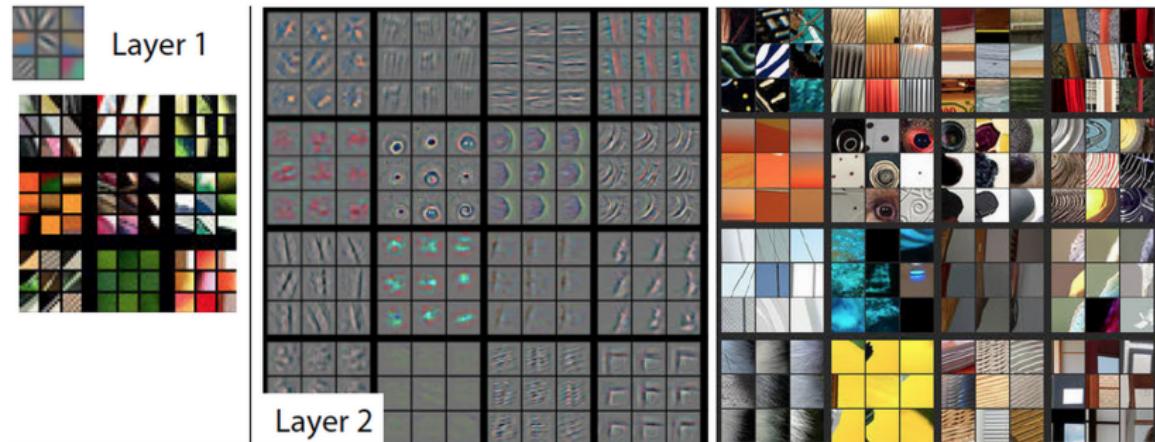


Рис.: Visualizing and Understanding Convolutional Networks³⁴

³⁴Matthew D. Zeiler and Rob Fergus

Выученные признаки, #2

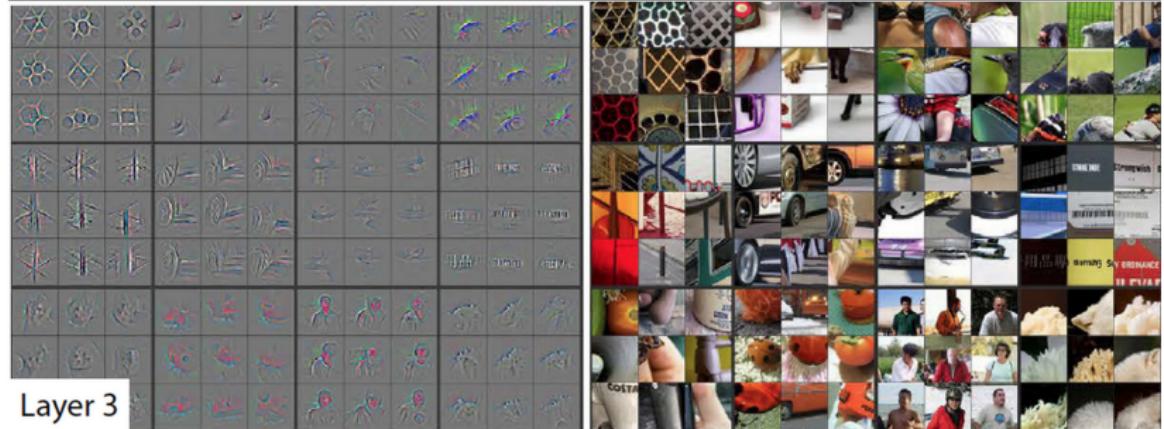


Рис.: Visualizing and Understanding Convolutional Networks³⁵

³⁵Matthew D. Zeiler and Rob Fergus

Выученные признаки, #3

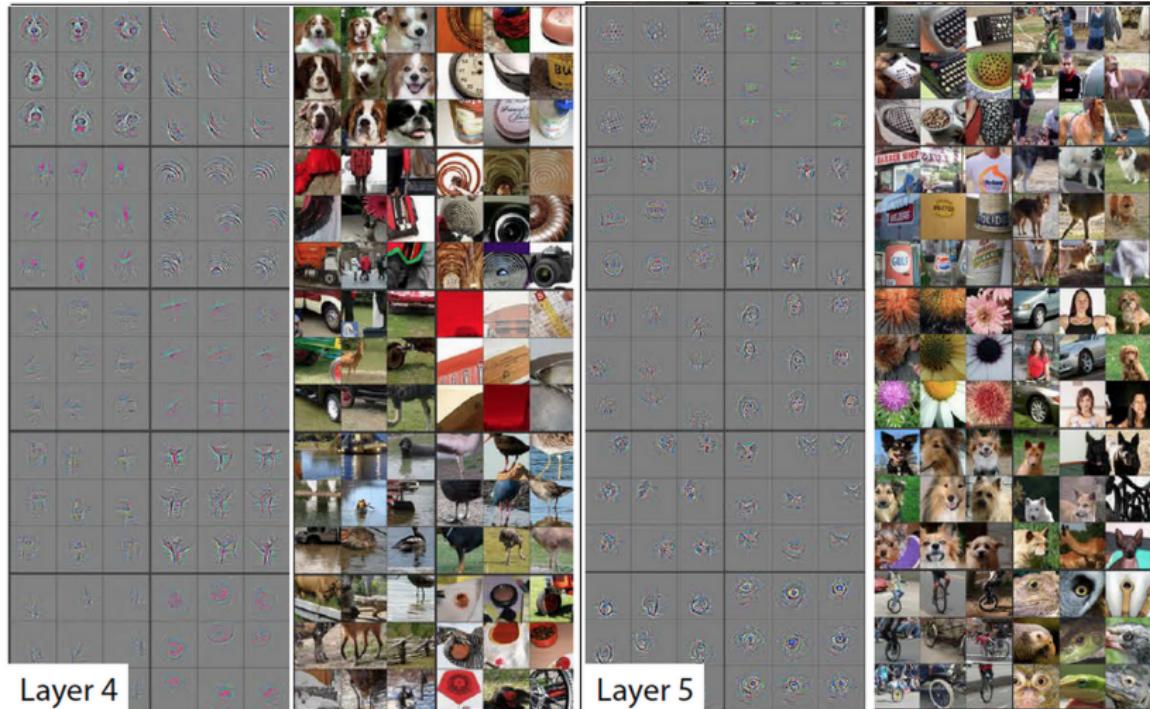


Рис.: Visualizing and Understanding Convolutional Networks³⁶

³⁶Matthew D. Zeiler and Rob Fergus

Синаптическая нейропластичность

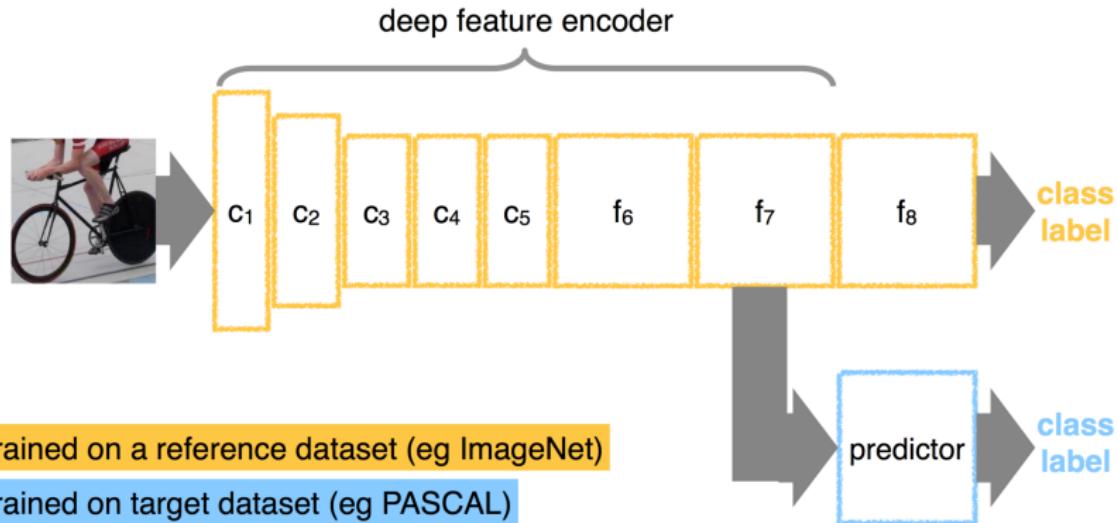


Рис.: Трансфер между двумя глубокими сетями³⁷

³⁷Learning visual representations (Andrea Vedaldi)

Transfer learning³⁸, #1

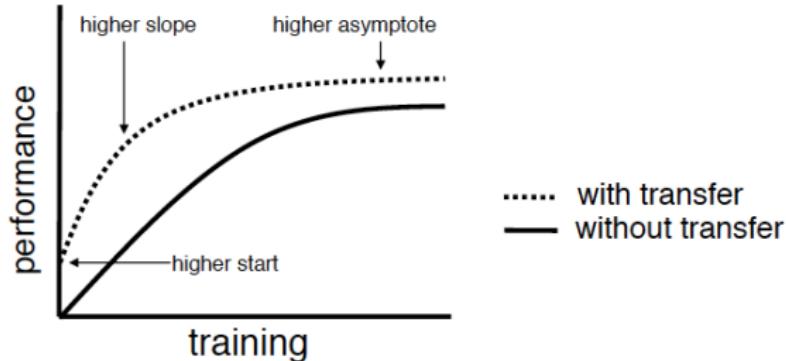


Рис.: Цели трансфера знаний

- ▶ higher start — улучшение качества обучения уже на начальных итерациях за счет более тщательной подборки начальных параметров модели или какой-либо другой априорной информации;
- ▶ higher slope — ускорение сходимости алгоритма обучения;
- ▶ higher asymptote — улучшение верхней достижимой границы качества.

³⁸ftp:

//ftp.cs.wisc.edu/machine-learning/shavlik-group/torrey.handbook09.pdf

Transfer learning, #2

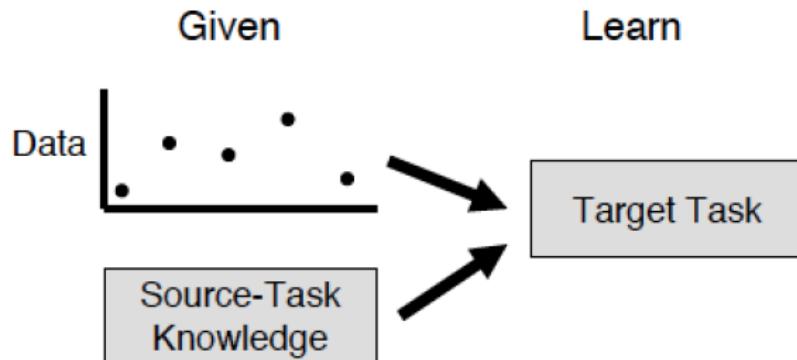


Рис.: Модель решения задачи в рамках парадигмы трансфера знаний

Transfer learning, #3

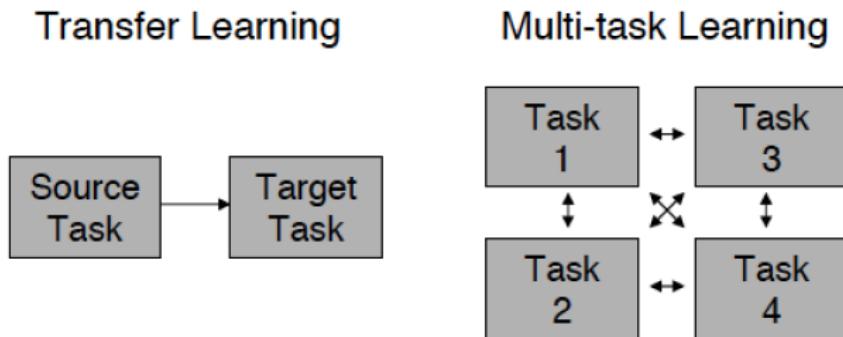


Рис.: Предобучение глубокой сети с помощью RBM - это multitask learning, не путать с transfer learning

Transfer learning, #4

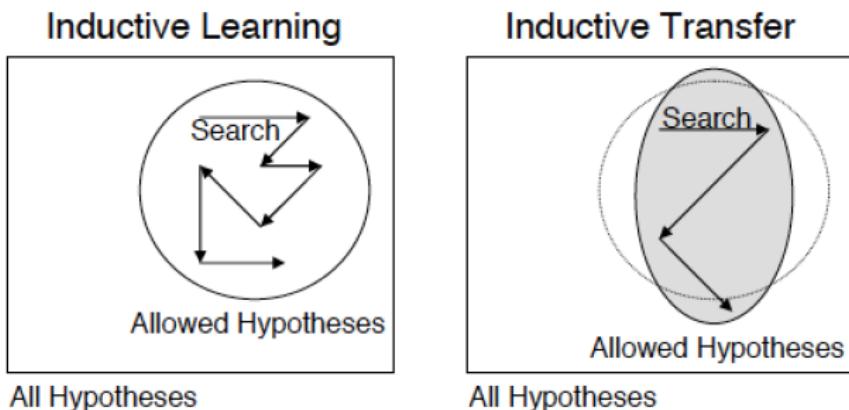


Рис.: Трансфер знаний можно также рассматривать как некоторую регуляризацию, которая ограничивает пространство поиска до определенного набора допустимых и хороших гипотез

Семантический код

Семантический кодирование - это такое кодирование, при котором для каждого объекта из исходного пространства (высокой размерности), ставится в соответствие объект (код) из нового пространства (меньшей размерности), такой, что семантически похожие объекты имеют похожие коды (в смысле некоторой меры).

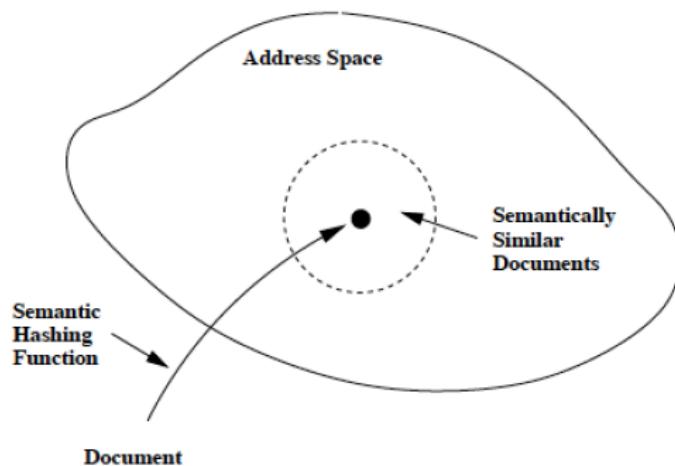


Рис.: Схематическое представление семантического кодирования³⁹

³⁹Semantic Hashing (Salakhutdinov, Hinton)

DBN (Deep RBM) и Deep AE

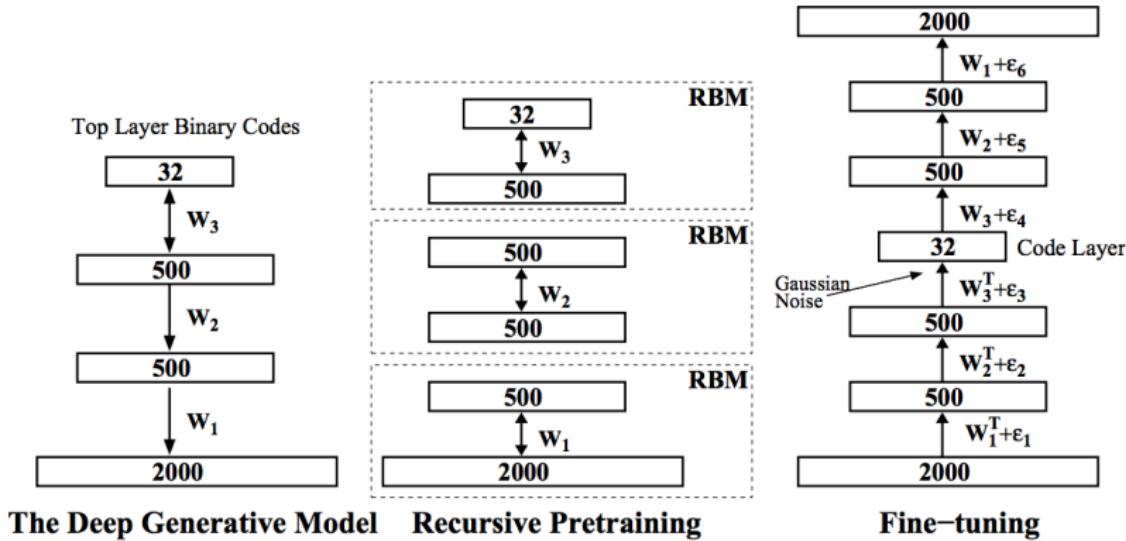
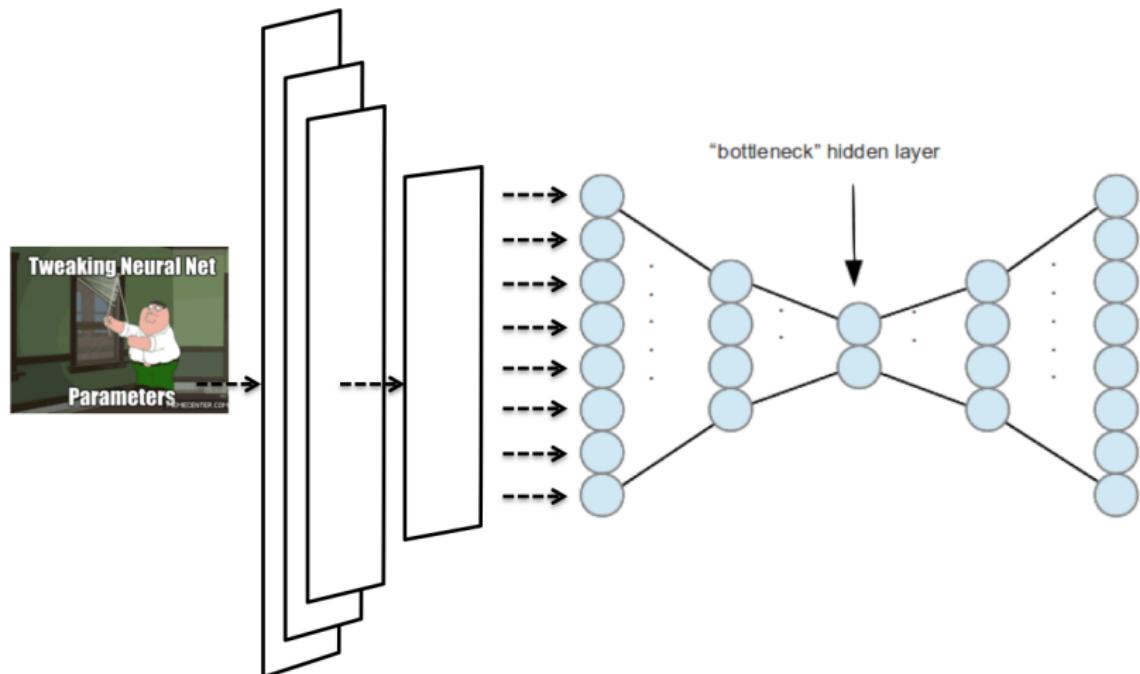


Рис.: Глубокий автоенкодер⁴⁰

⁴⁰Semantic Hashing (Salakhutdinov, Hinton)

Трансфер из VGG



Визуализация, #1



1550825.jpg



1620216.jpg



1700151.jpg



1700178.jpg



1701017.jpg



1764506.jpg



1764518.jpg



1877486.jpg



2026084.jpg



2094066.jpg



2109108.jpg



2109153.jpg

Визуализация, #2



0goal.jpg



93186.png



171581.png



364466.jpg



629451.jpg



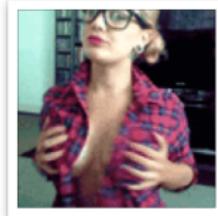
633182.jpg



712078.jpg



808167.png



814004.png



1483664.jpg



1622336.jpg

Визуализация, #3



0goal.jpg



346763.jpg



362624.jpg



444028.jpg



513224.jpg



534278.jpg



571146.jpg



768177.png



847001.jpg



867038.jpg



1604230.jpg

Визуализация, #4



0goal.jpg



378829.jpg



611505.jpg



689011.jpg



899429.jpg



930662.jpg



1020816.jpg



1149542.jpg



1272985.jpg



1282751.jpg



2148107.jpg

Визуализация, #5

2276683425790278213



1115802.jpg



1115868.jpg



1791723.jpg



1897651.jpg



1897678.jpg



2006720.jpg



2006725.jpg



2006727.jpg



2078335.jpg



2084644.jpg



2084941.jpg



2102445.jpg

Визуализация, #6



1734142.jpg



1734143.jpg



1734153.jpg



1734156.jpg



1734164.jpg



1734165.jpg



1961146.jpg



2014413.jpg



2014426.jpg



2065165.jpg



2069372.jpg



2069437.jpg



2082477.jpg



2100696.jpg



2105538.jpg

Визуализация, #7



0goal.jpg



686049.jpg



704389.jpg



704390.jpg



795077.jpg



950742.jpg



1054508.png



1179832.jpg



1367326.jpg



1429950.jpg



1628913.jpg

Распознавание речи⁴¹

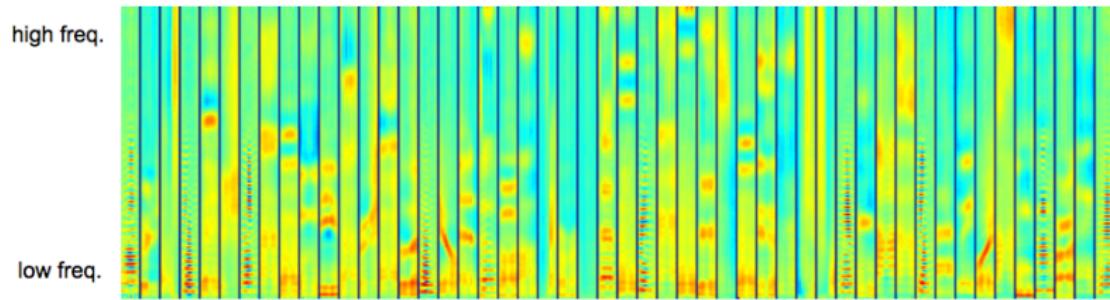


Рис.: Спектрограмма голосового сигнала, каждая колонка - это рецептивное поле нейрона

⁴¹<http://ai.stanford.edu/~ang/papers/nips09-AudioConvolutionalDBN.pdf>

Текст как изображение⁴²⁴³

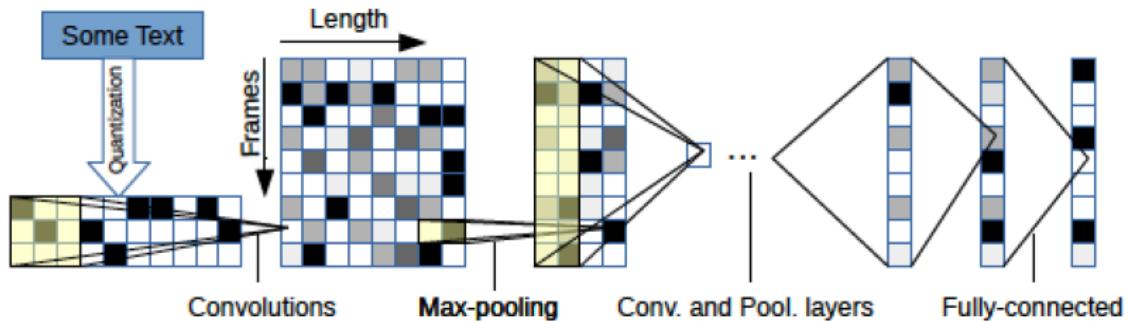
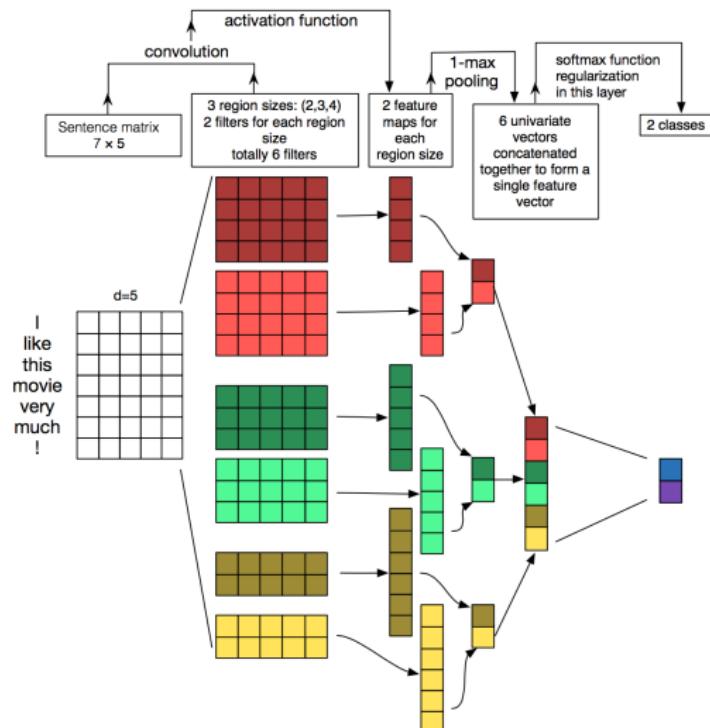


Рис.: Обработка изображения представляющего текст

⁴²<http://arxiv.org/pdf/1502.01710v1.pdf>

⁴³http://nlp.csail.mit.edu/papers/Kalchbrenner_DCNN_ACL14

Векторное представление текста как изображение⁴⁴



⁴⁴[http://www.wildml.com/2015/11/
understanding-convolutional-neural-networks-for-nlp/](http://www.wildml.com/2015/11/understanding-convolutional-neural-networks-for-nlp/)

Reinforcement learning + CNN⁴⁵

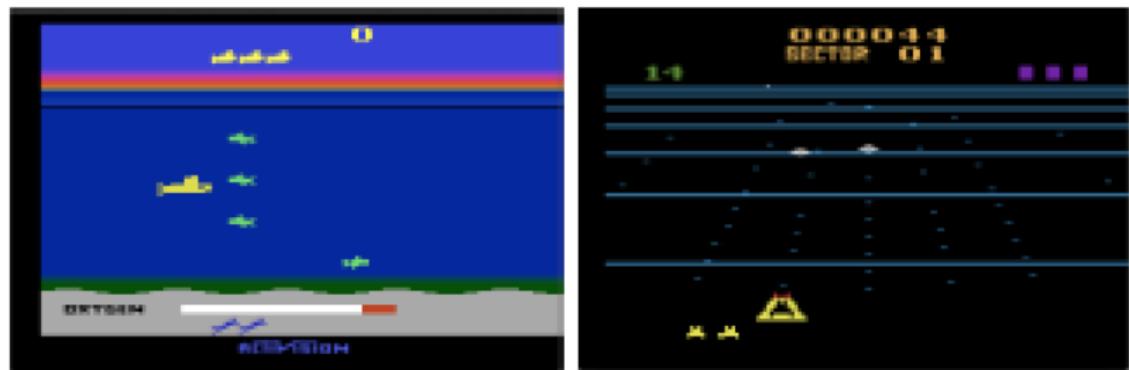


Рис.: Screen shots from five Atari 2600 Games: (Left-to-right) Pong, Breakout, Space Invaders, Seaquest, Beam Rider

CNN for object detection⁴⁶

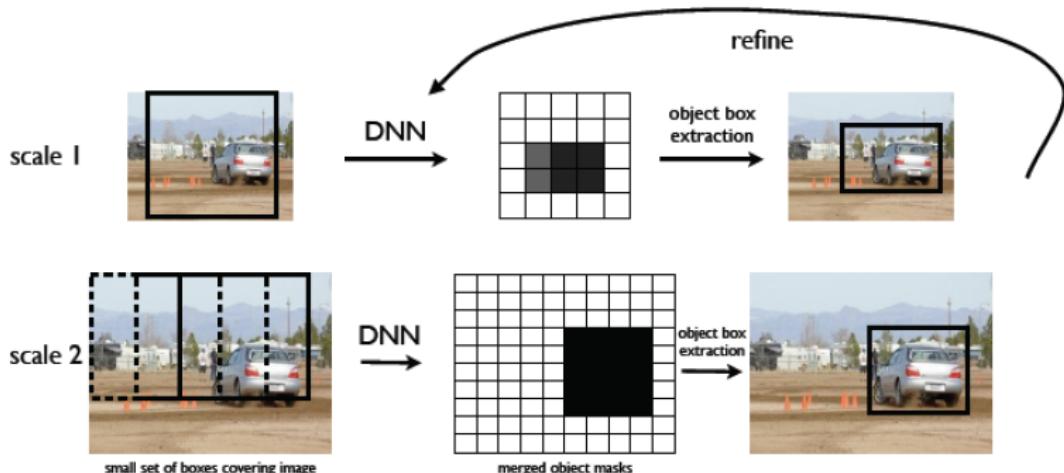
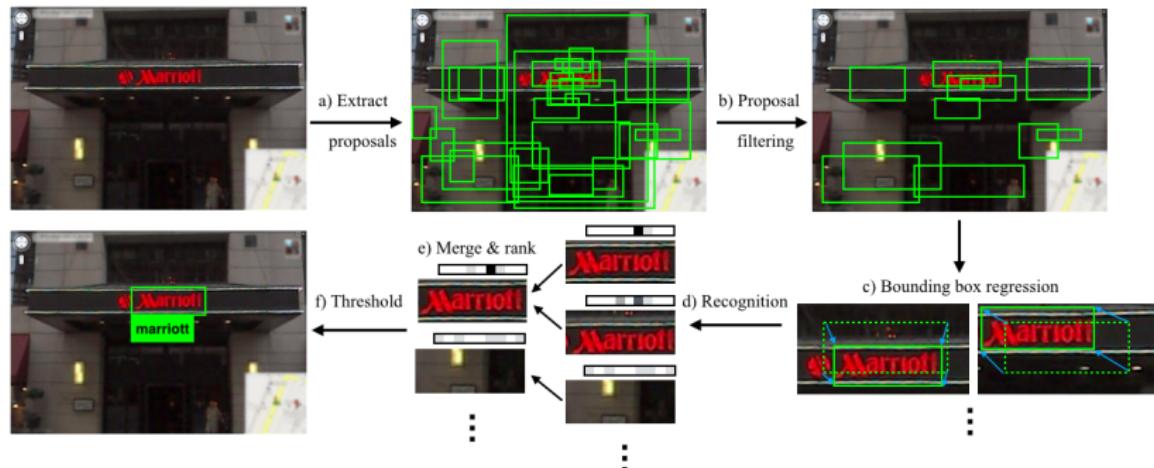


Рис.: Поиск на пирамиде изображения

⁴⁶<http://papers.nips.cc/paper/5207-deep-neural-networks-for-object-detection.pdf>

Hardcore OCR⁴⁷



⁴⁷<http://www.robots.ox.ac.uk/~vgg/research/text/index.html>

Recurrent network + CNN⁴⁸

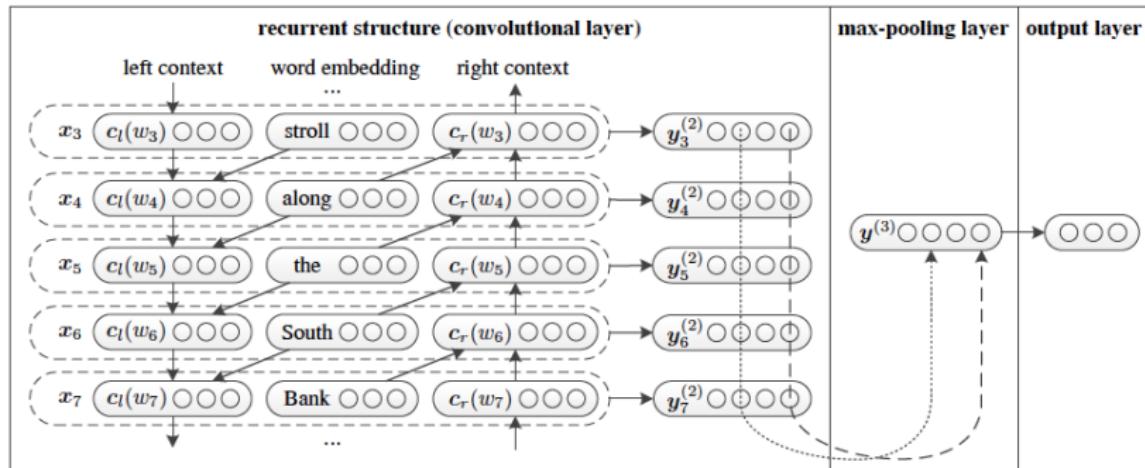


Рис.: The structure of the recurrent convolutional neural network

⁴⁸<http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/viewFile/9745/9552>

Вопросы

