



ТЕХНОСФЕРА

Лекция 6 Архитектуры глубинных сетей

Полыковский Даниил
Храбров Кузьма

11 марта 2019 г.

Архитектуры CNN

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

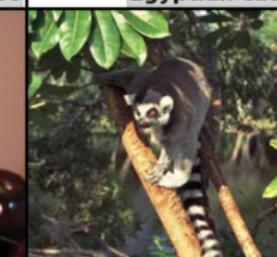
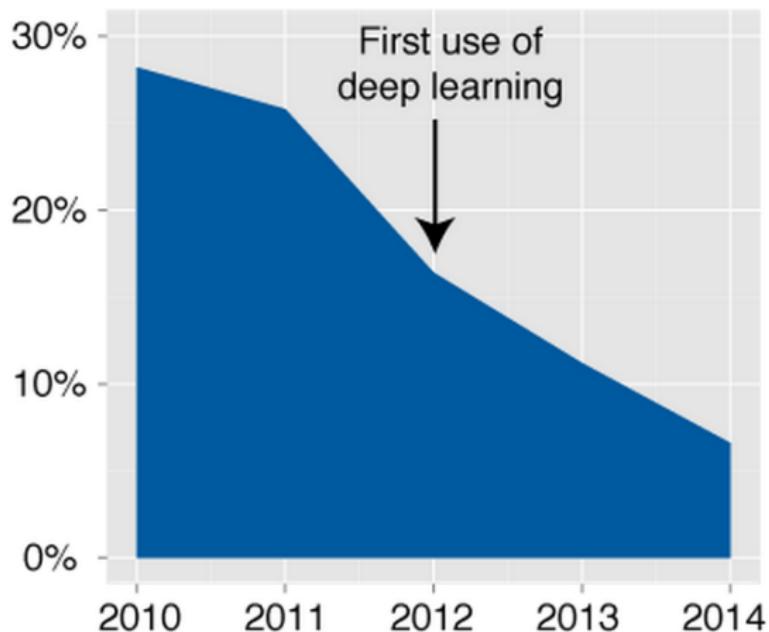
			
mite black widow cockroach tick starfish	container ship lifeboat amphibian fireboat drilling platform	motor scooter go-kart moped bumper car golfcart	leopard jaguar cheetah snow leopard Egyptian cat
			
grille convertible grille pickup beach wagon fire engine	mushroom agaric mushroom jelly fungus gill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

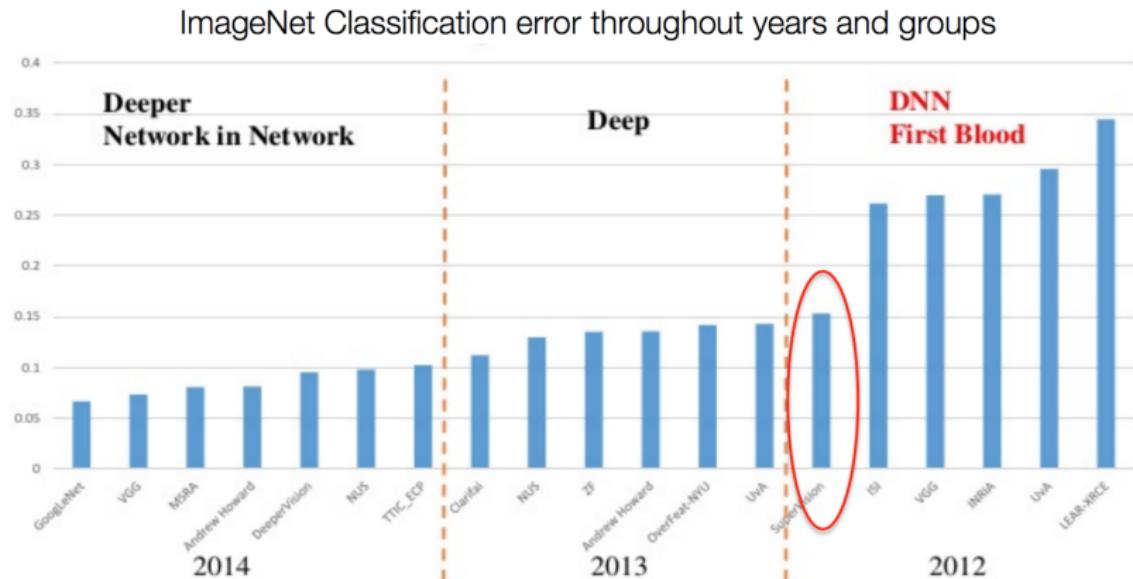
Рис.: Примеры прогнозов

ImageNet, прогресс

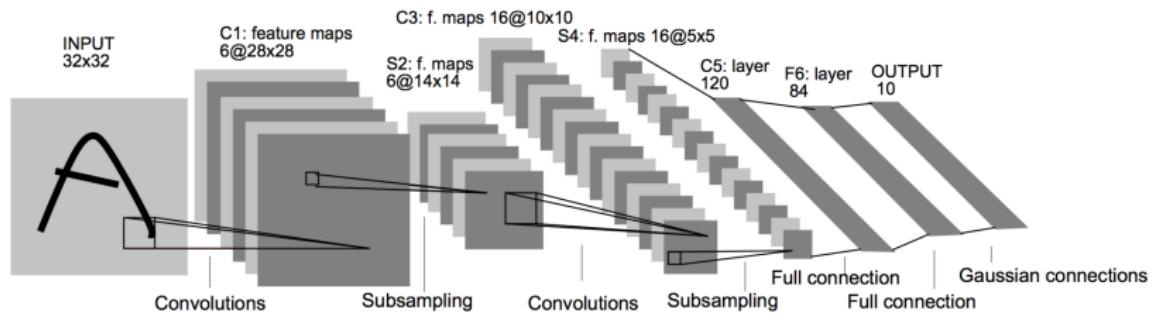
Objection classification error rate



ImageNet, прогресс

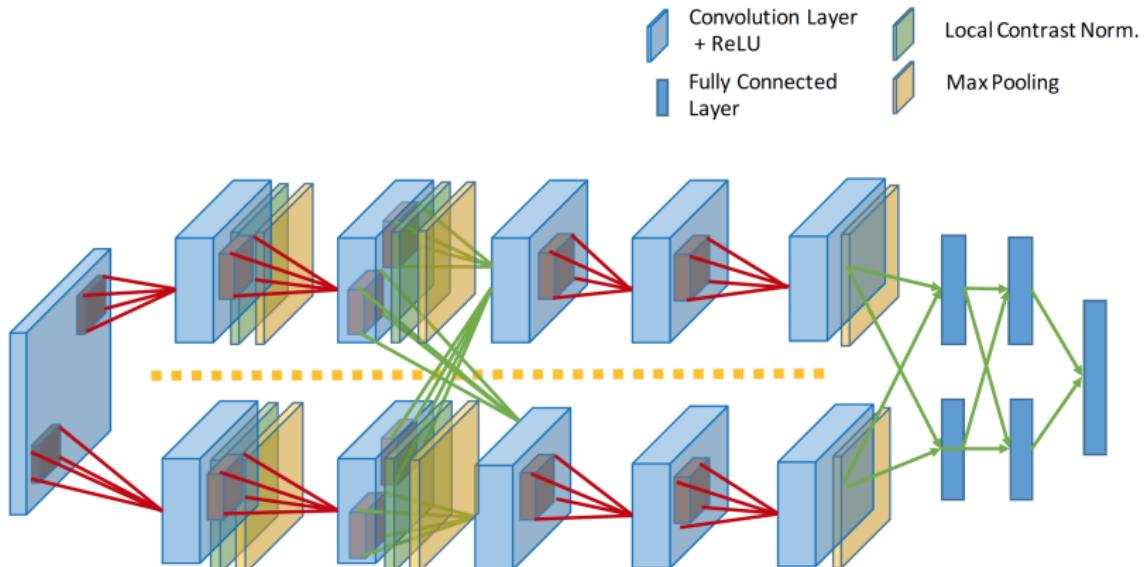


Lenet¹



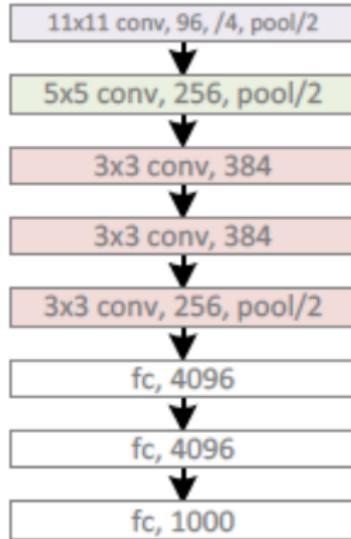
¹<http://yann.lecun.com/exdb/lenet>

AlexNet²



²Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton: ImageNet Classification with Deep Convolutional Neural Networks, 2012

<http://www.cs.toronto.edu/~fritz/absps/imagenet.pdf>



- ▶ 5 сверточных и 3 полносвязных слоя
- ▶ 60M параметров, 650к нейронов
- ▶ эффективное распараллеливание на 2 GPU/CUDA
- ▶ свёртки 11x11, 5x5, 3x3
- ▶ ReLU, т.к. не надо вычислять exp

VGG³

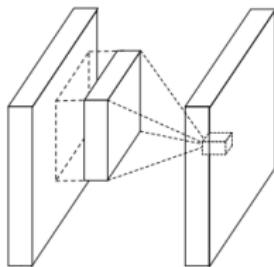
ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	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
- ▶ Только
свертки 3x3

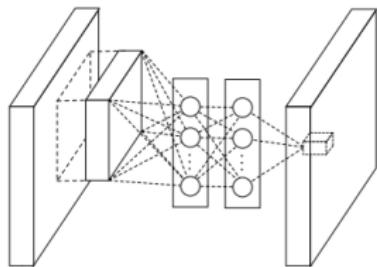
^aK. Simonyan, A. Zisserman

³<http://arxiv.org/pdf/1409.1556.pdf>

NIN⁴

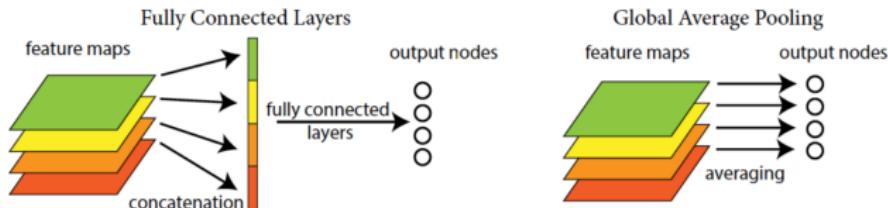


(a) Linear convolution layer



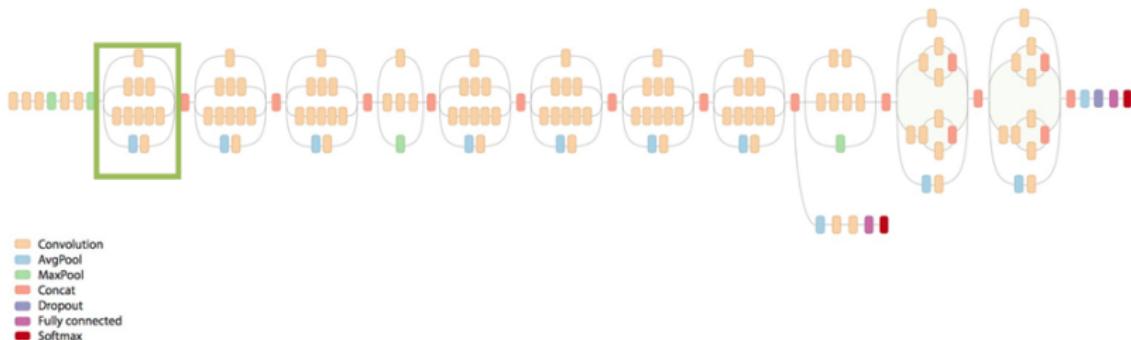
(b) Mlpconv layer

- ▶ “Полносвязные слои” (свертки 1×1) внутри свертки
- ▶ Глобальный пулинг для вытягивания изображения (+ доп. регуляризация)

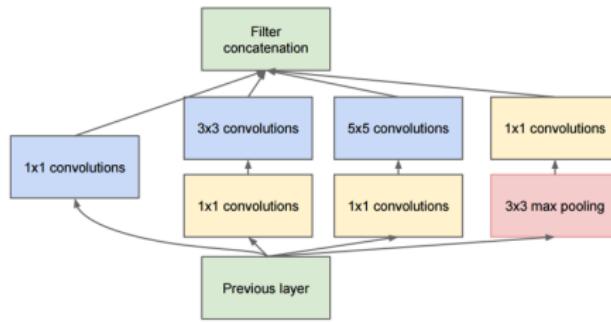


⁴<https://arxiv.org/pdf/1312.4400v3.pdf>

GoogLeNet⁵



Green box shows parallel region of GoogLeNet



Full Inception module

⁵http://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Szegedy_Going_Deeper_With_2015_CVPR_paper.pdf

Revolution of Depth

AlexNet, 8 layers
(ILSVRC 2012)



VGG, 19 layers
(ILSVRC 2014)



ResNet, 152 layers
(ILSVRC 2015)



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

ResNet⁶

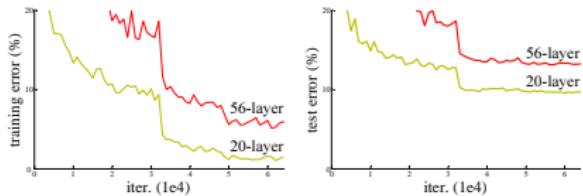


Рис.: Обучение “обычной” сети

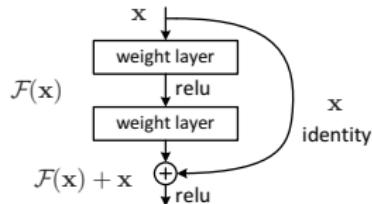


Рис.: Residual block

- ▶ Результат на 56 слоях хуже. Проблема не в переобучении
- ▶ Решение заведомо существует: 20 слоев, затем $\mathcal{F}(x)=x$
- ▶ Выучить $\mathcal{F}(x)=x$ тяжело, а $\mathcal{F}(x)=0$ просто
- ▶ Residual block решает эту проблему

⁶<https://arxiv.org/pdf/1512.03385v1.pdf>

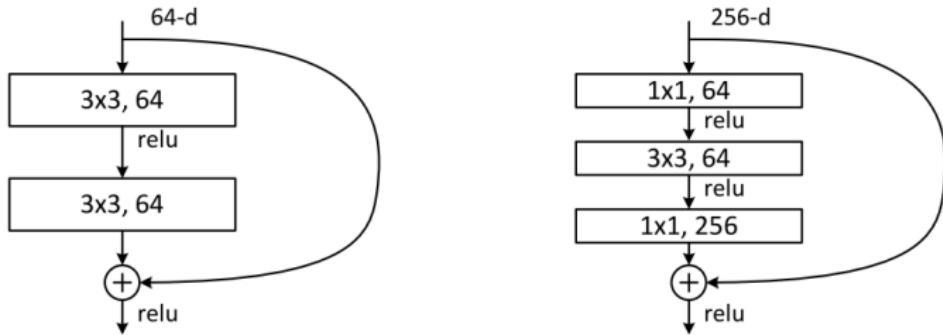


Figure 5. A deeper residual function \mathcal{F} for ImageNet. Left: a building block (on 56×56 feature maps) as in Fig. 3 for ResNet-34. Right: a “bottleneck” building block for ResNet-50/101/152.

Существует множество модификаций ResNet: ResNet in ResNet⁷, DenseNet⁸, ResNext.

⁷ <https://arxiv.org/pdf/1603.08029v1.pdf>

⁸ <https://arxiv.org/pdf/1608.06993v1.pdf>

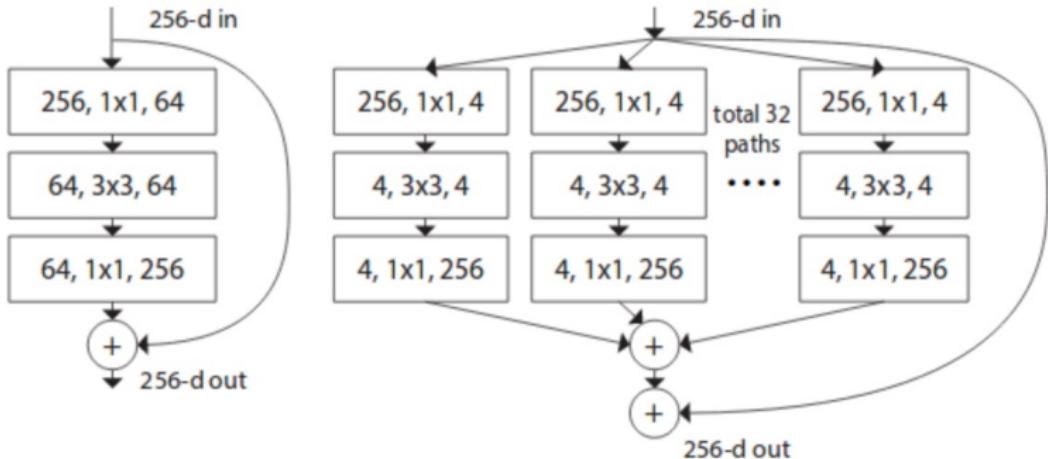


Figure 1. **Left:** A block of ResNet [14]. **Right:** A block of ResNeXt with cardinality = 32, with roughly the same complexity. A layer is shown as (# in channels, filter size, # out channels).

ResNet (left) and ResNeXt (right) Architecture.

Сравнение моделей

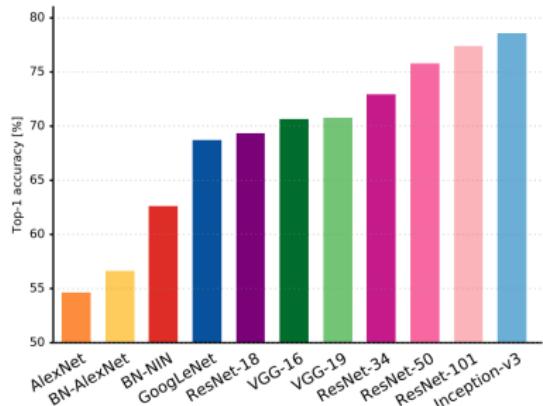


Figure 1: **Top1 vs. network.** Single-crop top-1 validation accuracies for top scoring single-model architectures. We introduce with this chart our choice of colour scheme, which will be used throughout this publication to distinguish effectively different architectures and their correspondent authors. Notice that network of the same group share colour, for example ResNet are all variations of pink.

Рис.: Accuracy

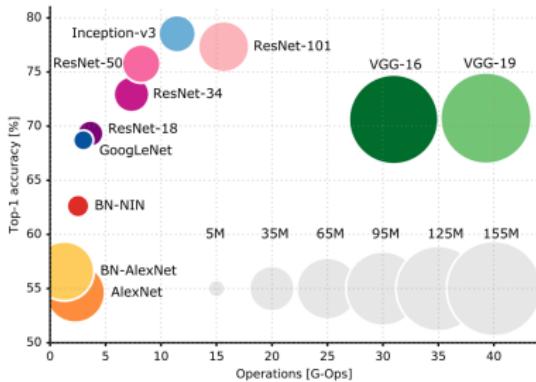


Figure 2: **Top1 vs. operations, size \propto parameters.** Top-1 one-crop accuracy versus amount of operations required for a single forward pass. The size of the blobs is proportional to the number of network parameters; a legend is reported in the bottom right corner, spanning from 5×10^6 to 155×10^6 params.

Сравнение моделей

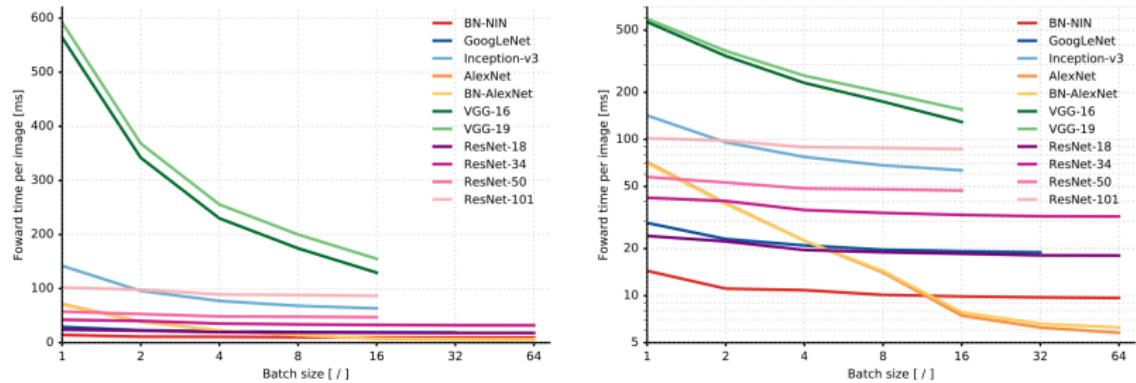


Figure 3: **Inference time vs. batch size.** These two charts show inference time across different batch sizes with a linear and logarithmic ordinate respectively and logarithmic abscissa. Missing data points are due to lack of enough system memory required to process bigger batches.

Рис.: Скорость

Сравнение моделей

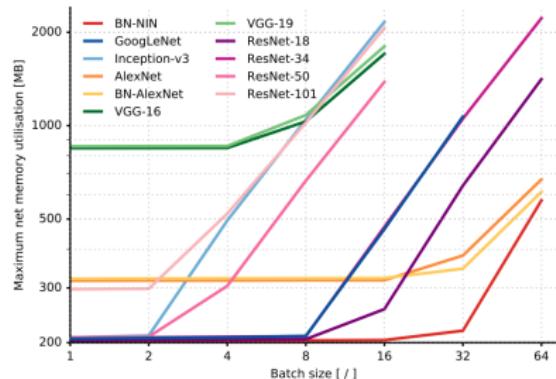


Figure 5: **Memory vs. batch size.** Maximum system memory utilisation for batches of different sizes. Memory usage shows a knee graph, due to the network model memory static allocation and the variable memory used by batch size.

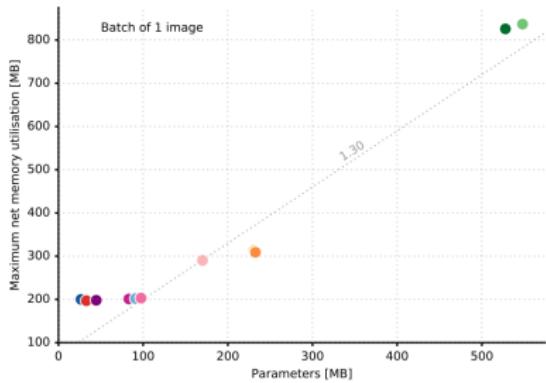


Figure 6: **Memory vs. parameters count.** Detailed view on static parameters allocation and corresponding memory utilisation. Minimum memory of 200 MB, linear afterwards with slope 1.30.

Рис.: Потребление памяти

Squeeze-and-Excitation Networks

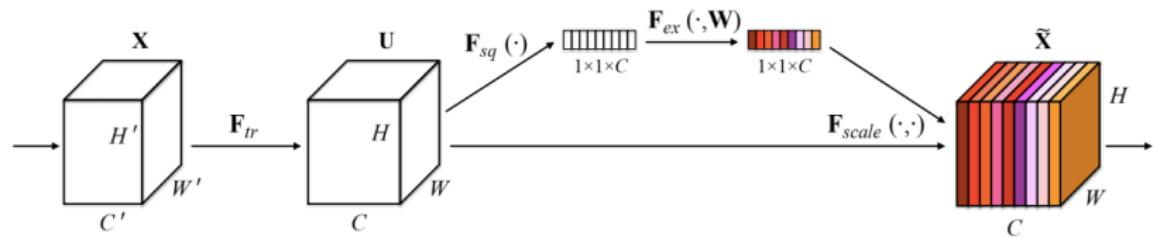
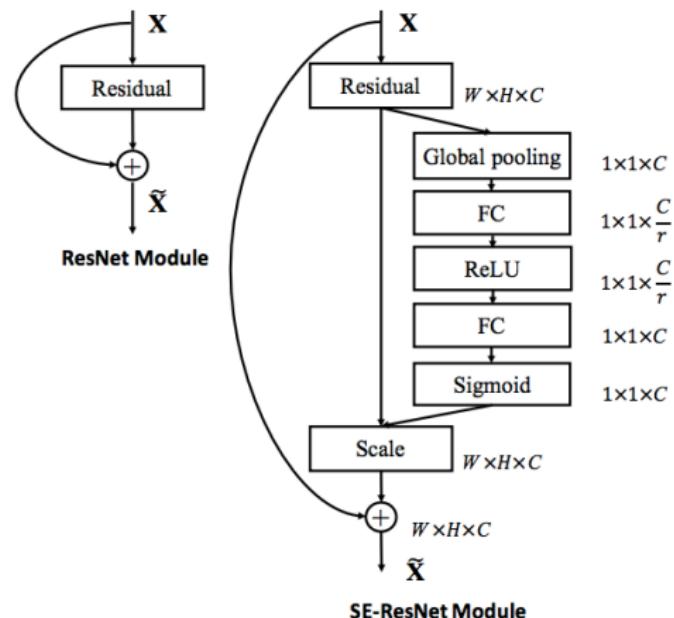
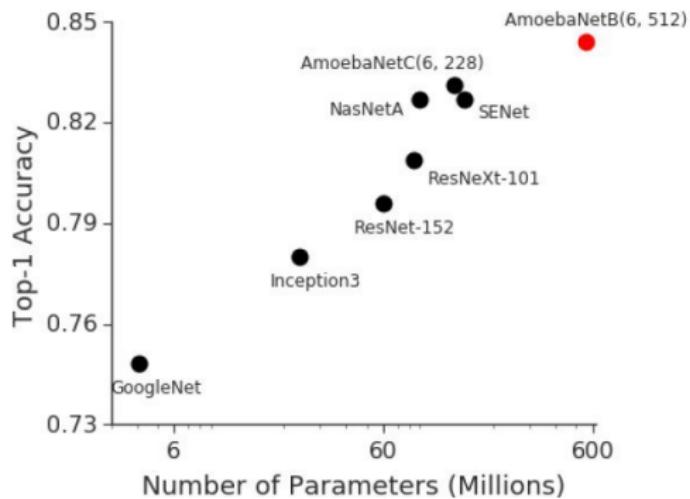


Рис.: <https://arxiv.org/abs/1709.01507>

Squeeze-and-Excitation Networks



Last results



Case study

National Data Science Bowl

National Data Science Bowl¹⁰



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

Победители⁹ использовали архитектуру VGG-16, а так же ряд специальных трюков.

⁹ <http://benanne.github.io/2015/03/17/plankton.html>

¹⁰ <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)

Данные нужно аугментировать в реальном времени (иначе — значительное увеличение размера датасета). Не стоит недооценивать test-time аугментацию.

Data augmentation, #2

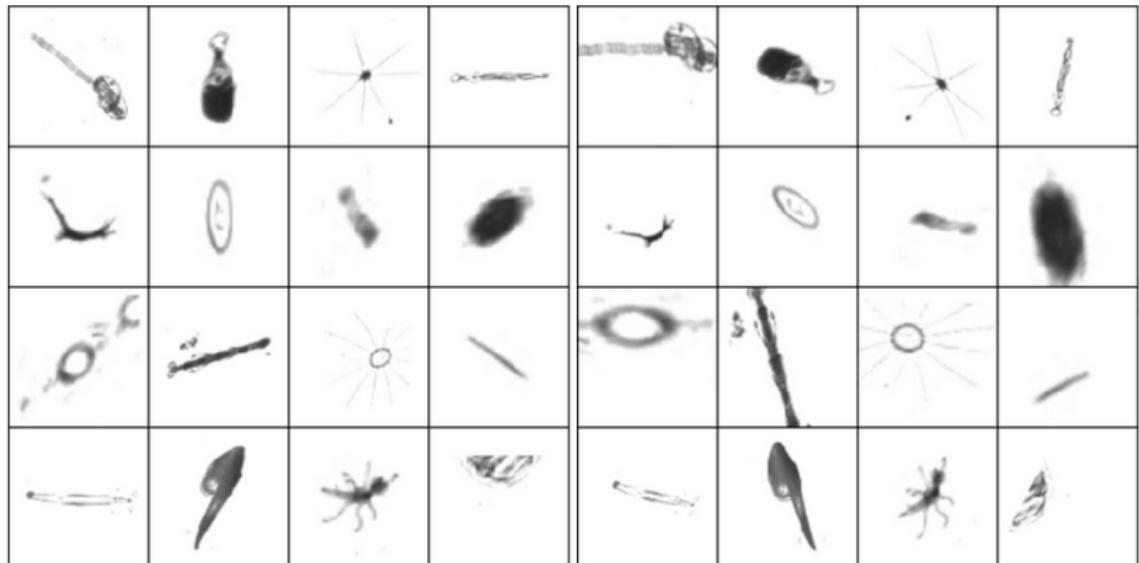


Рис.: Оригиналы и аугментированные образы

Cyclic pooling

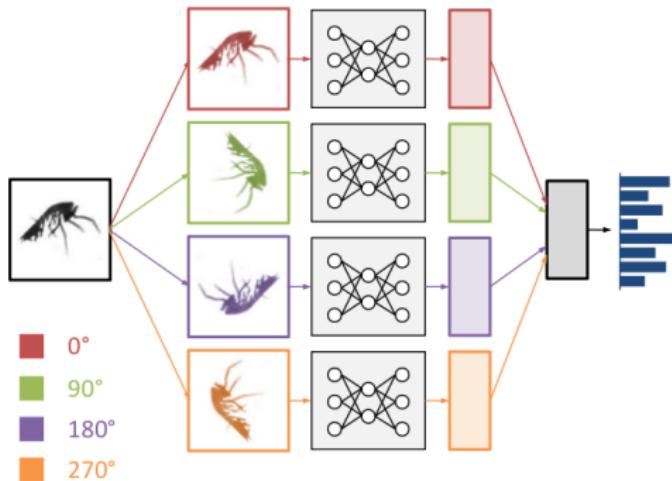


Рис.: Schematic representation of a convnet with cyclic pooling

- ▶ 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

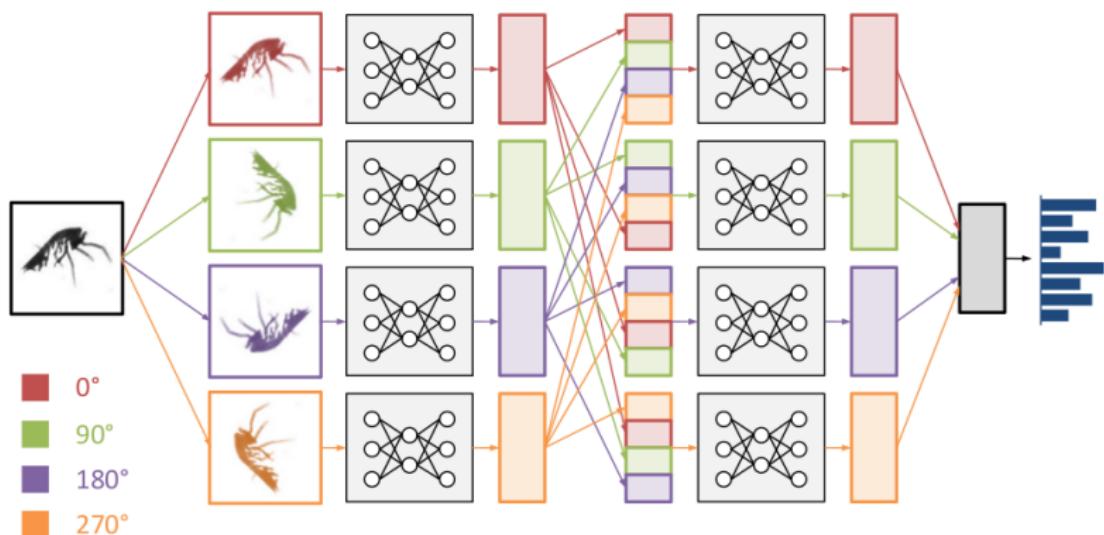


Рис.: Schematic representation of a roll operation inside a convnet with cyclic pooling.

Советы

- ▶ leaky/parameterized ReLU: $f(x) = \max(x, a \cdot x)$, $a \leq 0$
- ▶ добавление к сверточным признакам других признаков перед полносвязным слоем (Hu moments, Zernike moments, atc)
- ▶ использовать предобученные модели
- ▶ self-training с тестовой выборкой
- ▶ аугментировать данные

Другие задачи компьютерного зрения

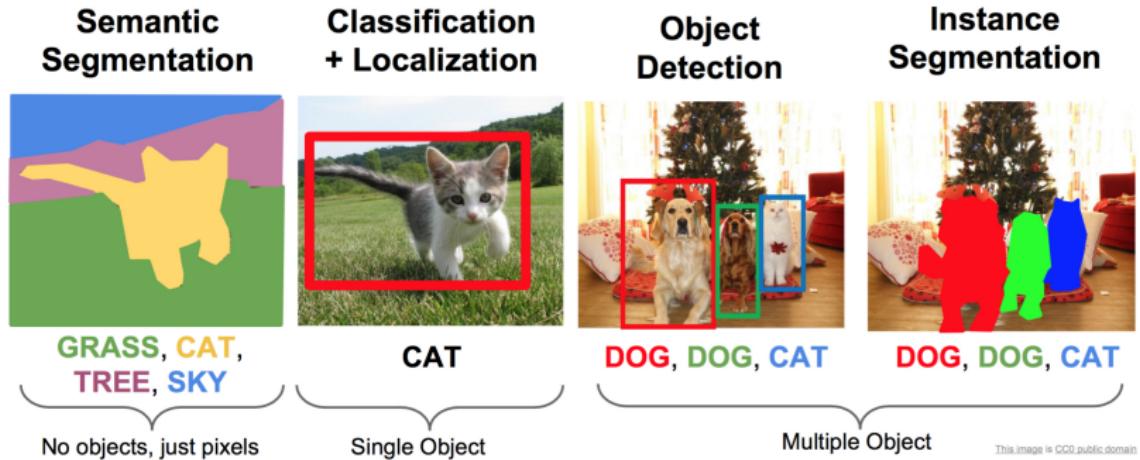


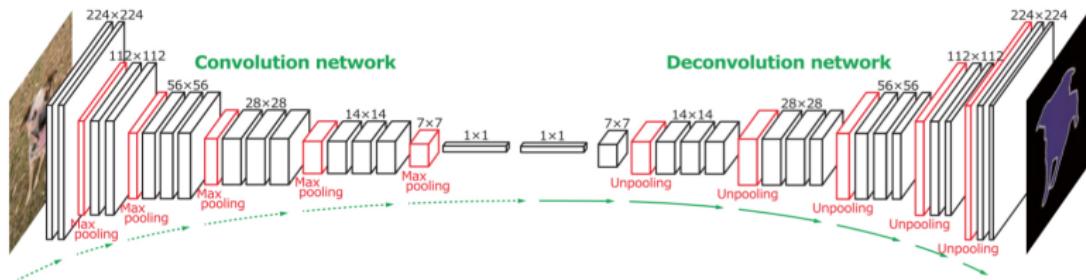
Рис.: Semantic segmentation, Localization, Detection and Instance Segmentation (Source: CS231n (Stanford) lecture slides)

Semantic segmentation



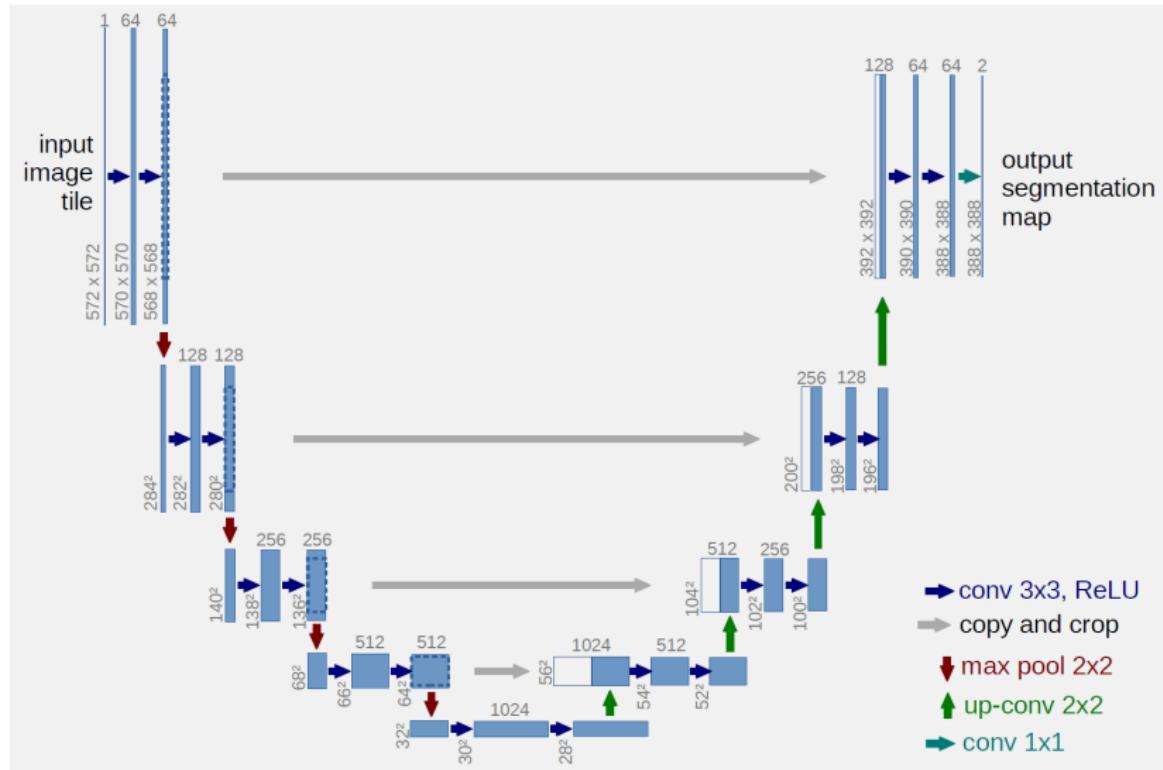
Задача - понять к какой категории относится каждая точка картинки.

Semantic segmentation. Подходы. ¹¹



¹¹<https://arxiv.org/abs/1505.04366v1>

U-NET¹²



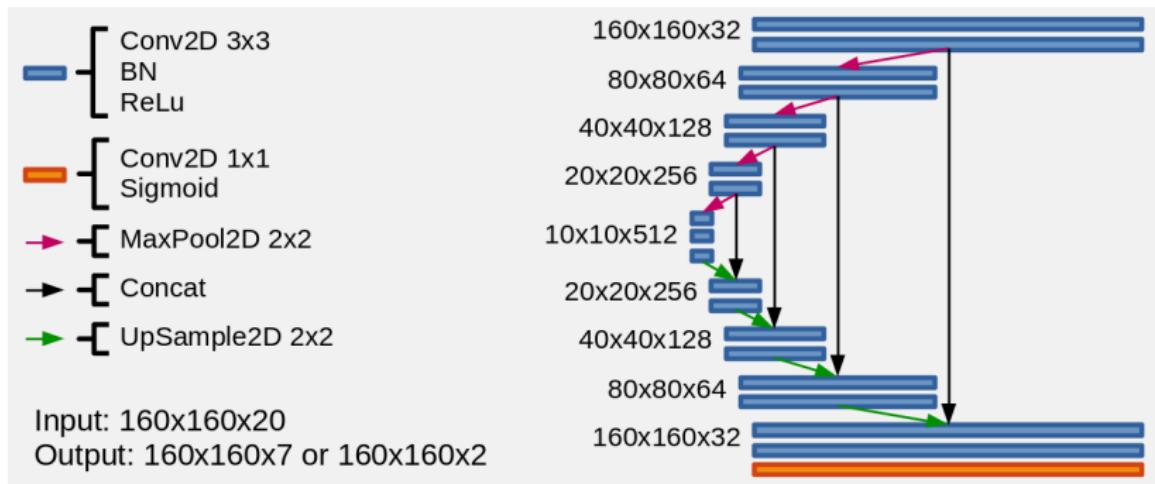
¹²<https://arxiv.org/abs/1505.04597>

U-NET

Задача:

- ▶ 25 спутниковых изображений в Train
- ▶ 425 спутниковых изображений в Test
- ▶ 10 классов объектов
- ▶ Изображения участков поверхности земли 1x1 км

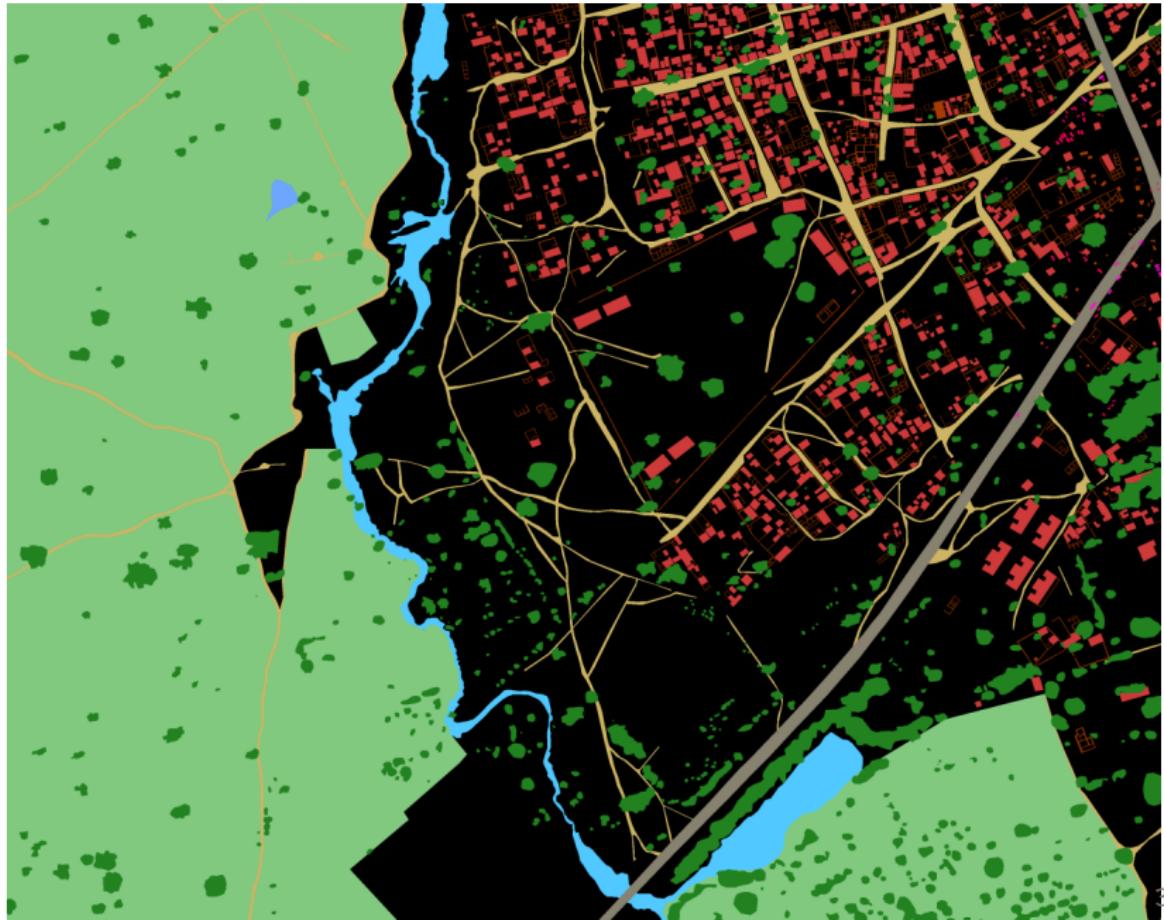
U-NET



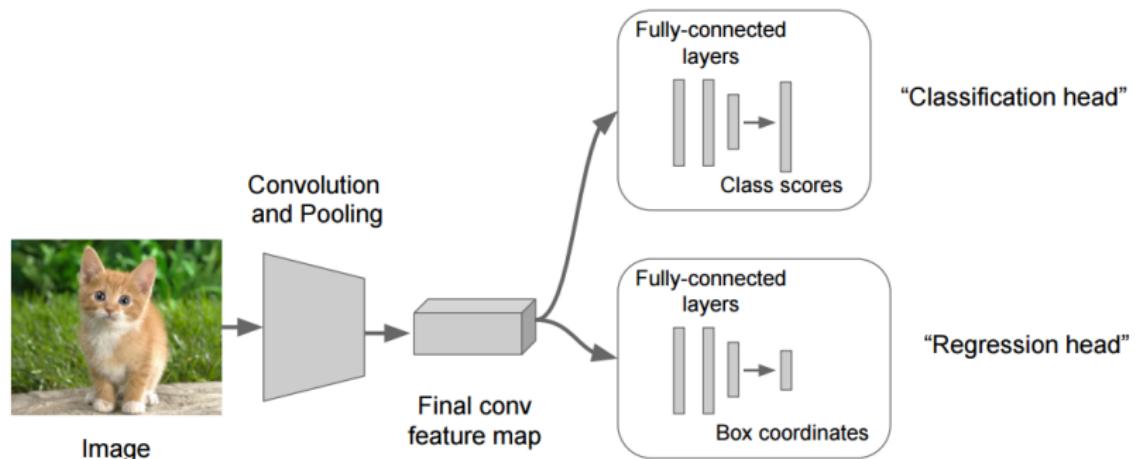
U-NET



U-NET



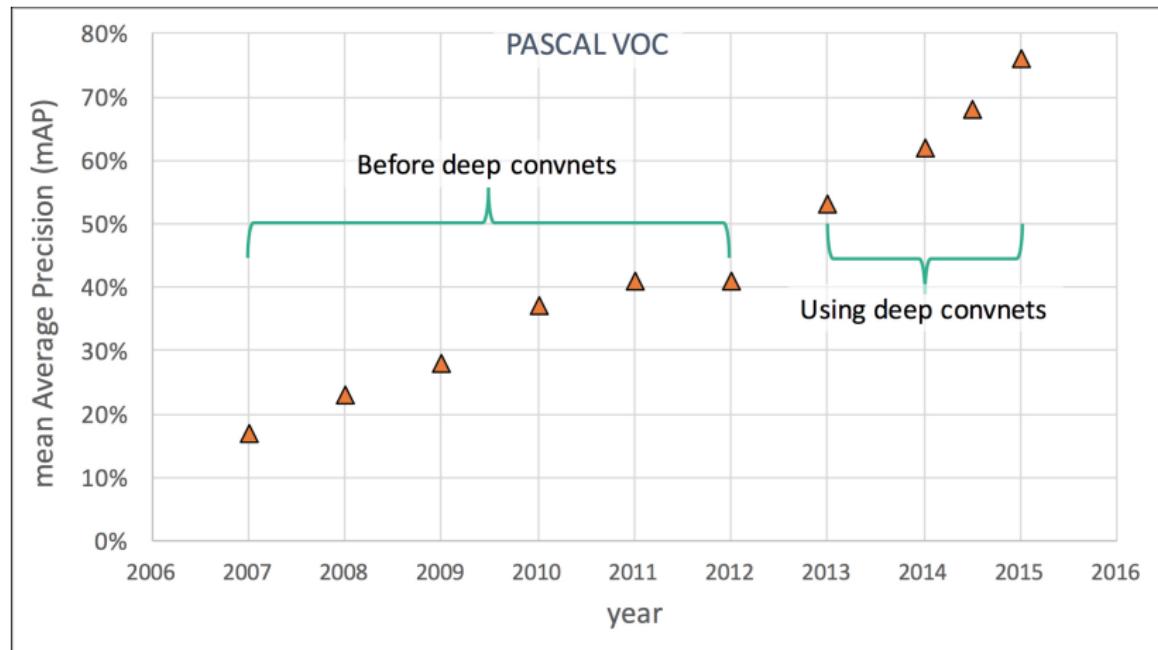
Object Classification + localization.



Object detection. Pascal VOC competition



Object detection. История



Detection. Подходы

- ▶ Регрессия. Так же как и с одним объектом.
Обычно не известно, какие объекты мы ищем и сколько их.
- ▶ Классификация.
Выбираем области, в которых могут быть объекты и применяем классификатор.

Detection. Выбор областей. Скользящее окно



Рис.: Background

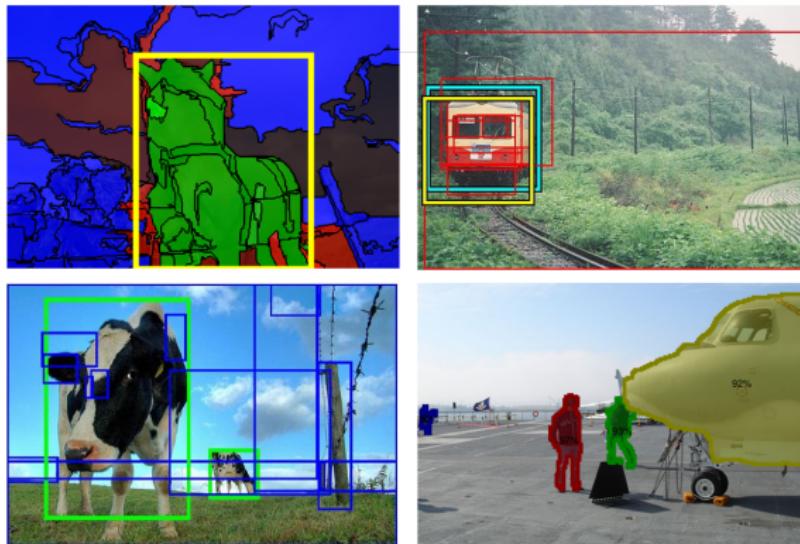


Рис.: Dog



Рис.: Cat

Detection. Автоматический выбор областей.



Shape Sharing for Object Segmentation. J. Kim and K. Grauman. ECCV 2012.

Uijlings et al, "Selective Search for Object Recognition", IJCV 2013

Category-Independent Object Proposals with Diverse Ranking. Ian Endres and Derek Hoiem. PAMI 2014.

Zitnick and Dollar, "Edge boxes: Locating object proposals from edges", ECCV 2014

Detection. R-CNN.

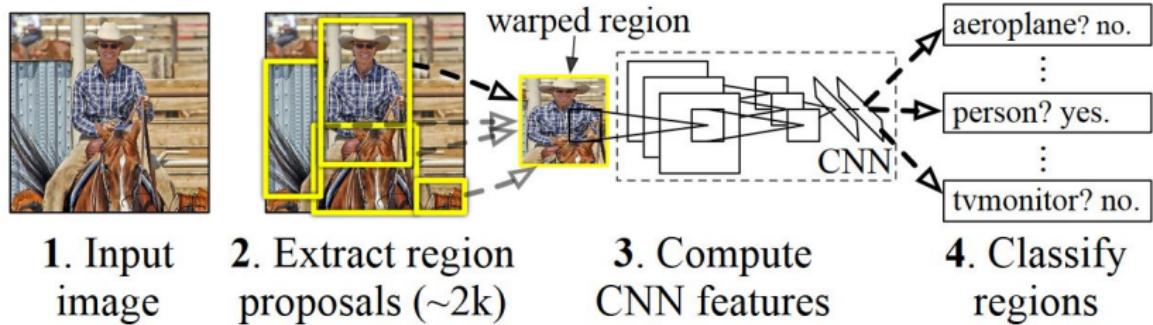


Рис.: R-CNN paper

Detection. Faster.

Fast R-CNN

Faster R-CNN

YOLO

Instance Segmentation. Mask R-CNN

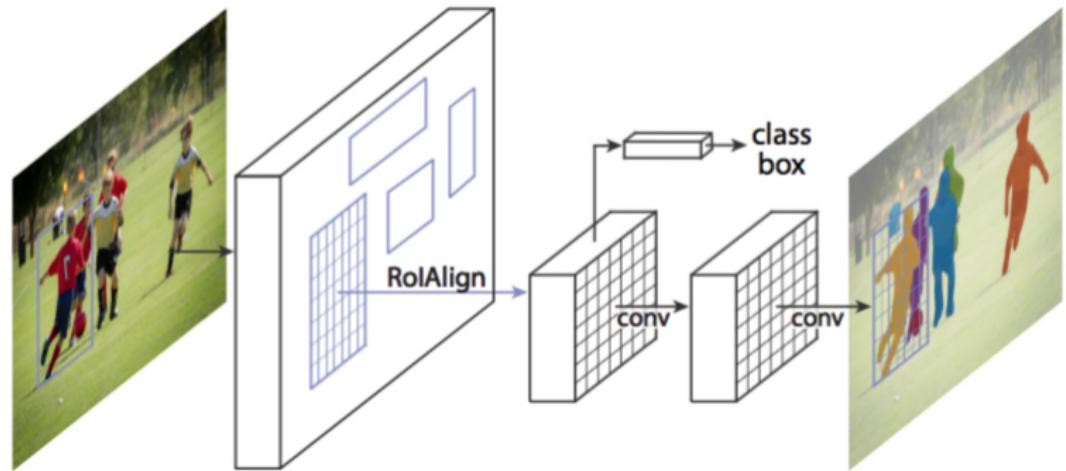


Рис.: Mask R-CNN paper

Вопросы

