

cv and ml

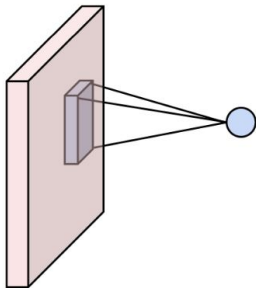
architecture ideas and segmentation

Владимир Глазачев

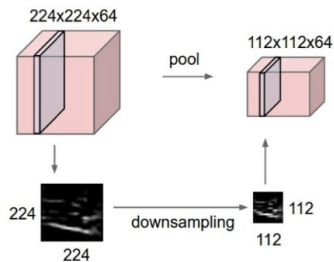
cv в rosebud.ai

Компоненты Conv Network

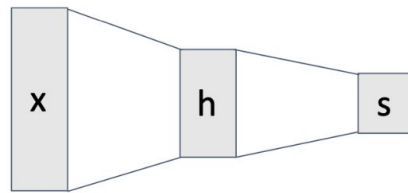
Convolution Layers



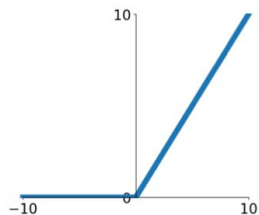
Pooling Layers



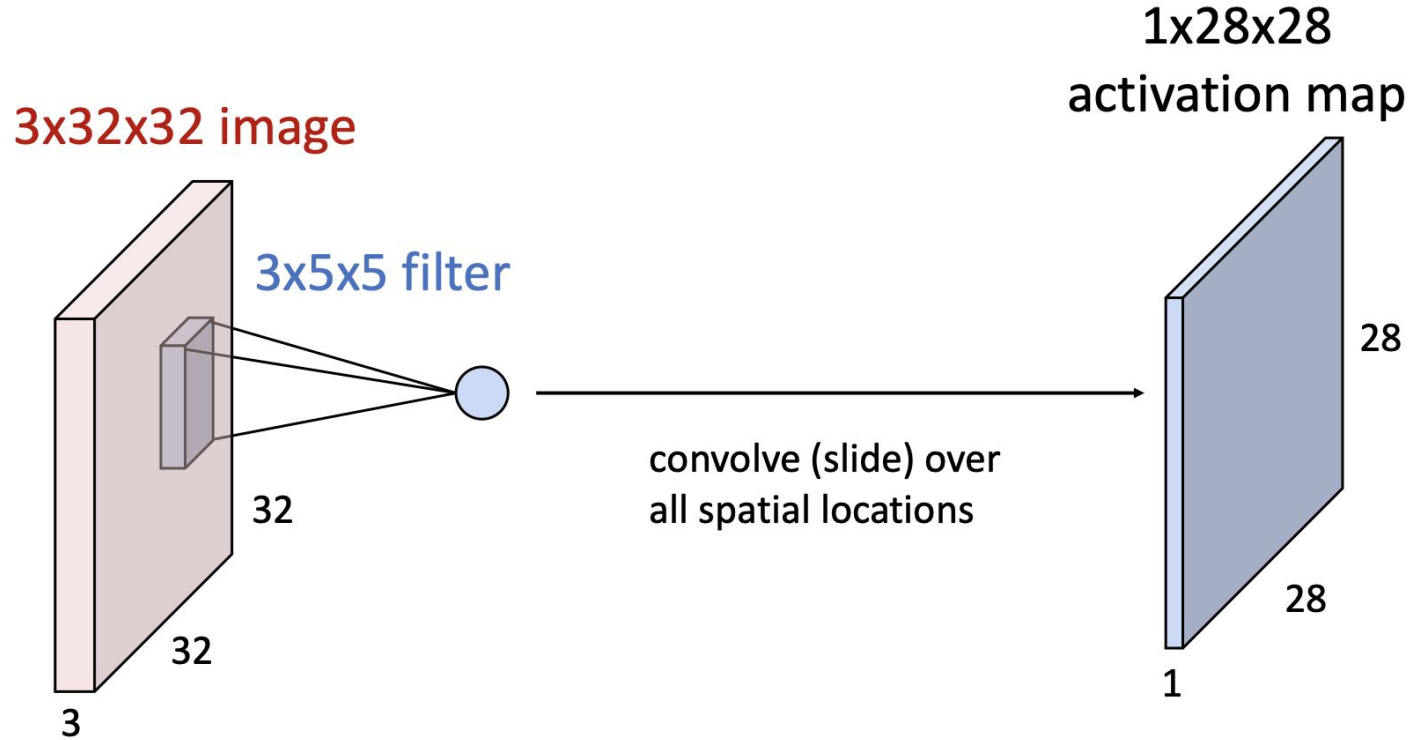
Fully-Connected Layers



Activation Function

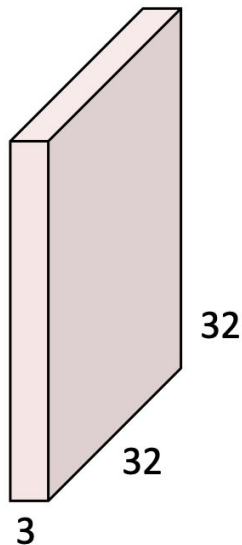


Conv Layer

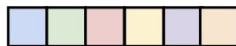


Conv Layer

3x32x32 image

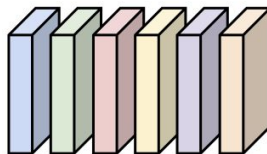


Also 6-dim bias vector:

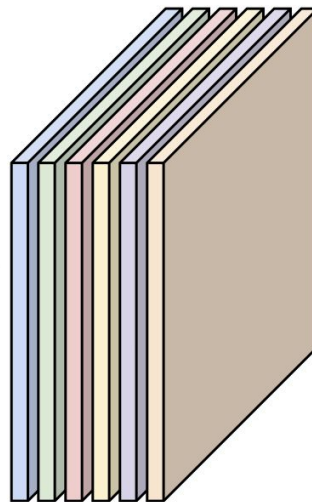


Convolution
Layer

6x3x5x5
filters

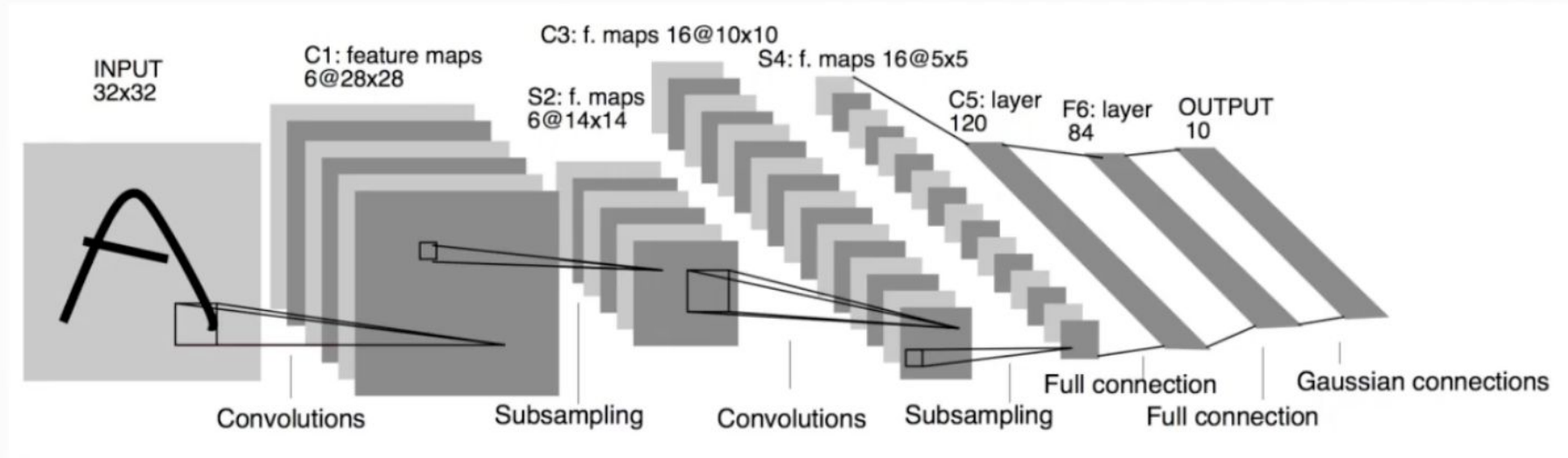


28x28 grid, at each
point a 6-dim vector



Stack activations to get a
6x28x28 output image!

Conv networks



Conv Network

- + Мало весов в конволюциях
 - + Хорошая структурная регуляризация и переиспользование данных
 - + Как то работает, не надо руками крафтить фичи
-
- Много весов в dense слоях
 - Работаем с фиксированным размером входа
 - В таком виде - плохо обучалось
 - Исторически до гри было оч медленно :)

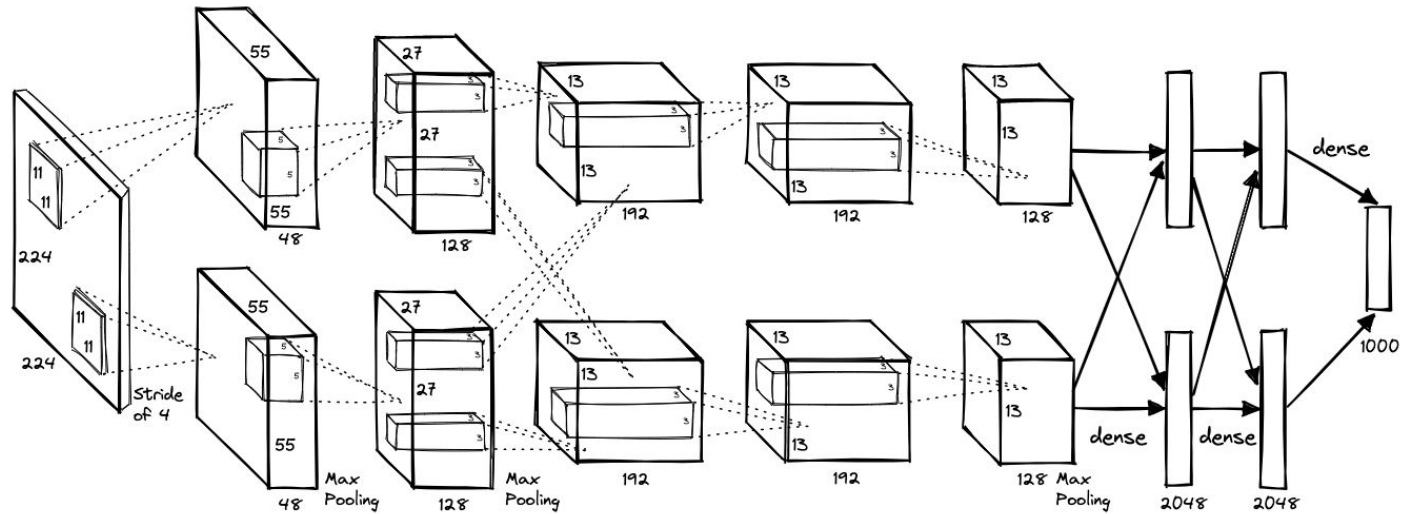
Проблемы

- Нужно много размеченных данных, даже простая архитектура - вычислительно довольно тяжелая на старом железе
 - cuda вычисления двинули это все в массы
- все еще инпут картинки фиксированного размера
 - это можно решить какими то трюками :)
- плохо учится - частично обсудим
 - инициализация
 - нормализация
 - затухание градиентов

Напоминание про градиенты

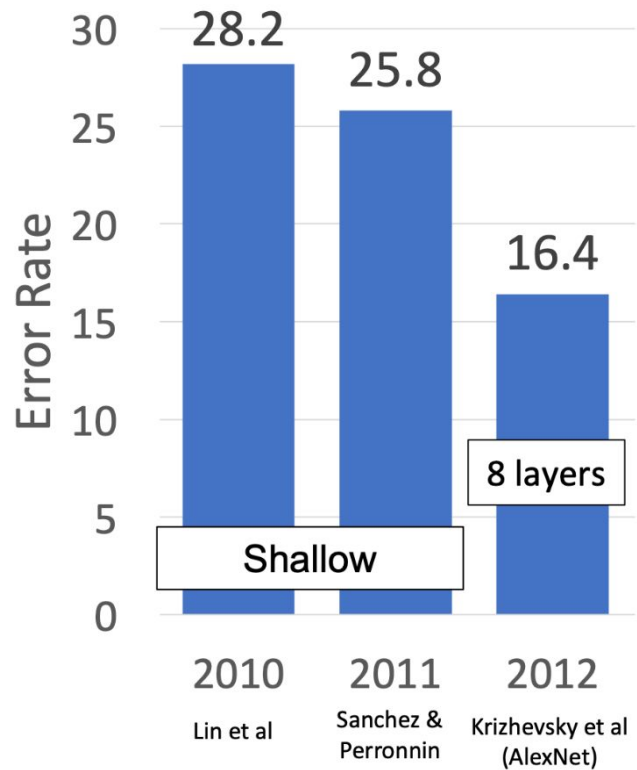
AlexNet

- ImageNet: > 15 миллионов картинок, ≈ 22000 категорий.
- Размечены руками (Amazon Mechanical Turk).
- Классическая сеть AlexNet (Krizhevsky, Sutskever, Hinton, 2012).



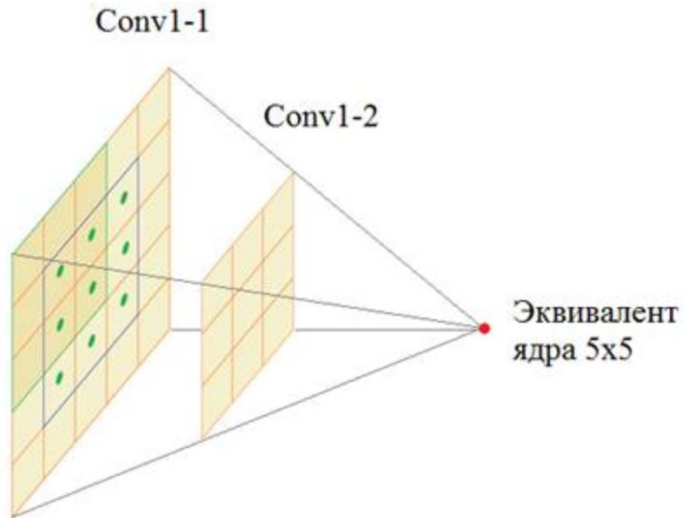
На большом массиве данных получилось обучить convnet, до этого - игрушечные задачи вроде mnist

AlexNet



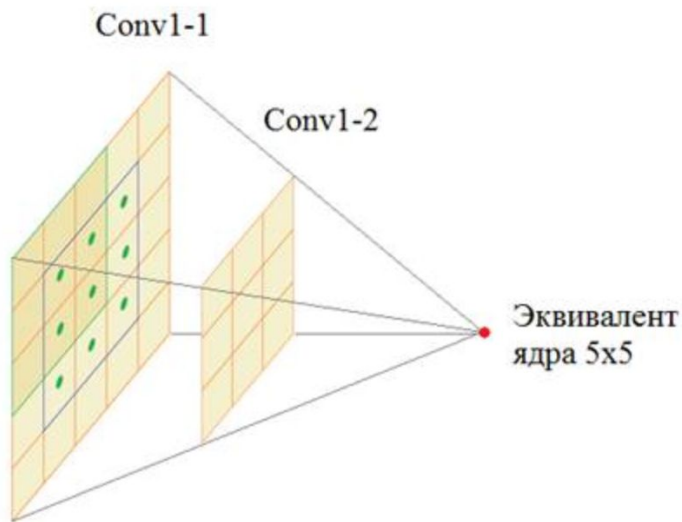
VGG

- В прошлый раз кратко обсудили - какой нужен размер свертки
- В AlexNet свертки 11x11, 5x5, 3x3



VGG

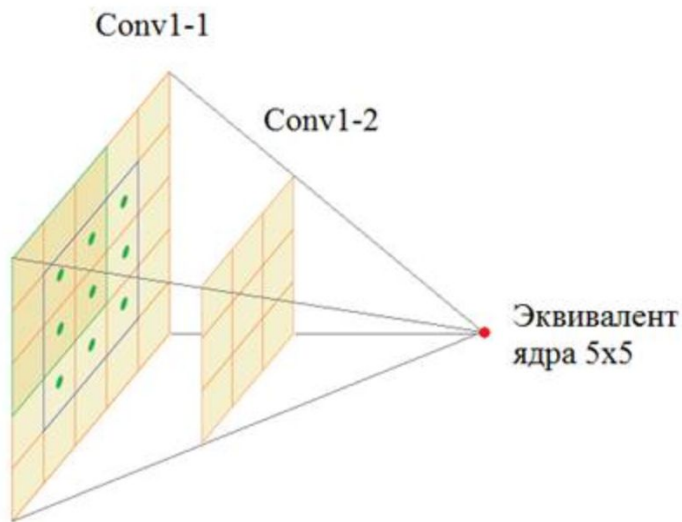
- В прошлый раз кратко обсудили - какой нужен размер свертки
- В AlexNet свертки 11×11 , 5×5 , 3×3



В свертке 5×5 25 параметров
в 2 свертках 3×3 - 18 параметров

VGG

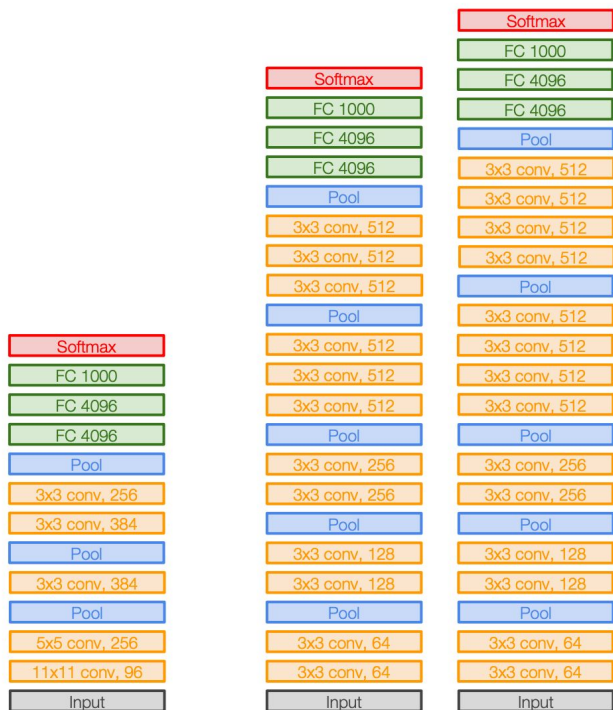
- В прошлый раз кратко обсудили - какой нужен размер свертки
- В AlexNet свертки 11x11, 5x5, 3x3



В свертке 5x5 25 параметров
в 2 свертках 3x3 - 18 параметров

работает как регуляризация, качество не теряем, большие свертки почти не используются

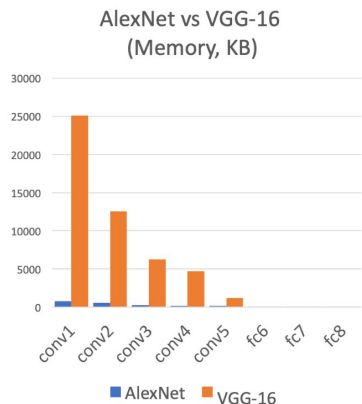
VGG



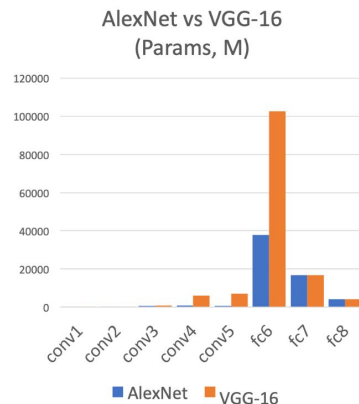
AlexNet

VGG16

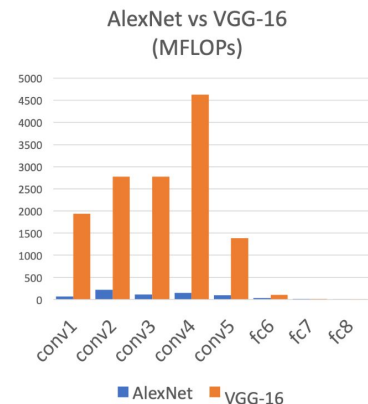
VGG19



AlexNet total: 1.9 MB
VGG-16 total: 48.6 MB (25x)

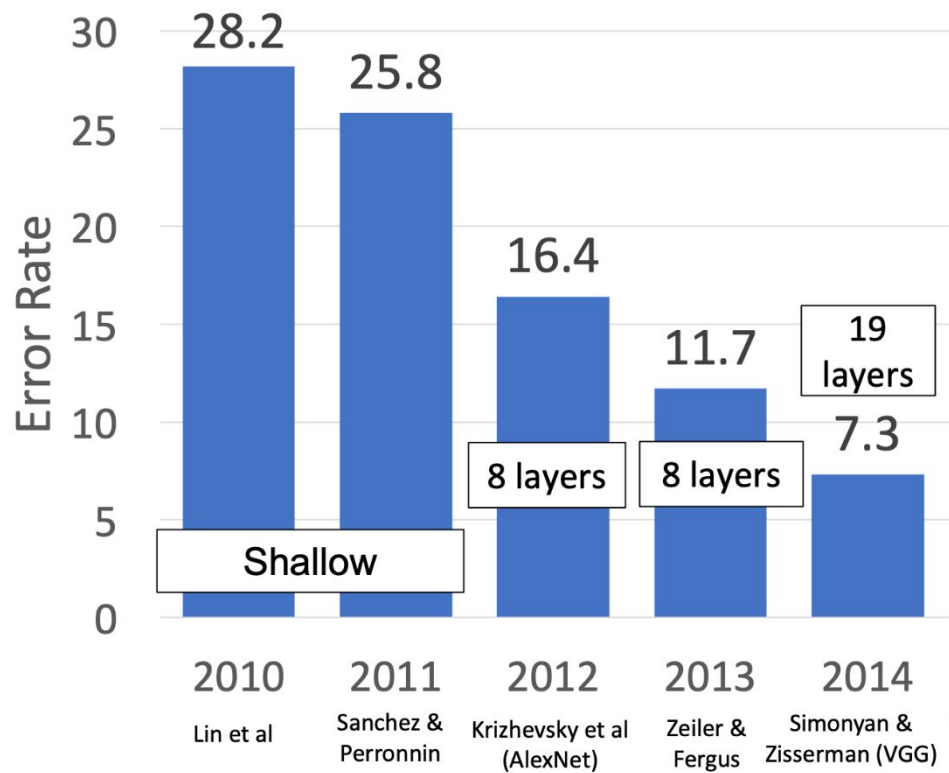


AlexNet total: 61M
VGG-16 total: 138M (2.3x)



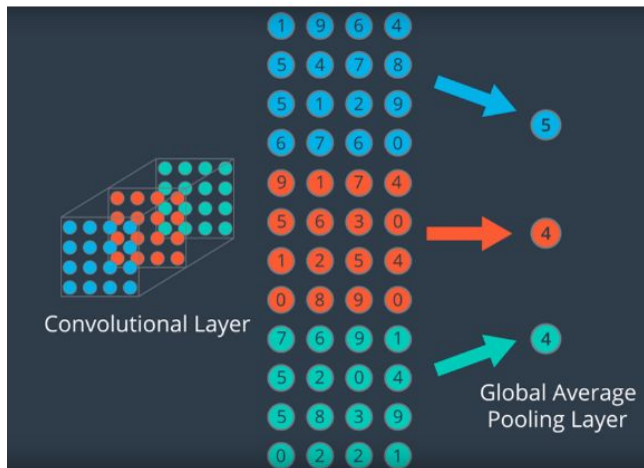
AlexNet total: 0.7 GFLOP
VGG-16 total: 13.6 GFLOP (19.4x)

VGG



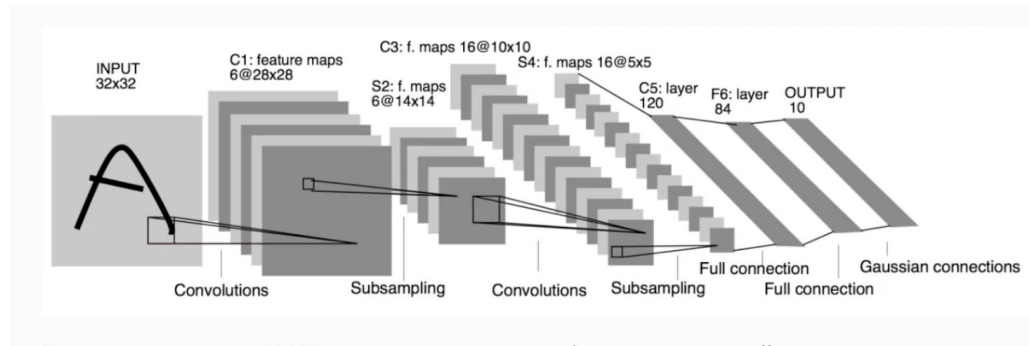
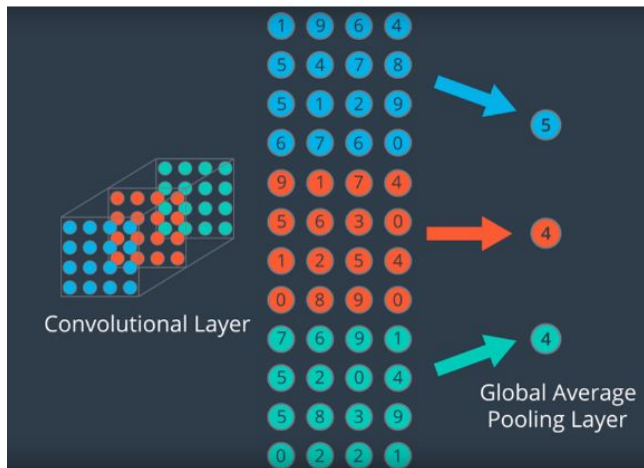
GoogLeNet

- Хотим избавиться от dense слоев в конце - большинство параметров все еще там
- Global Average Pooling



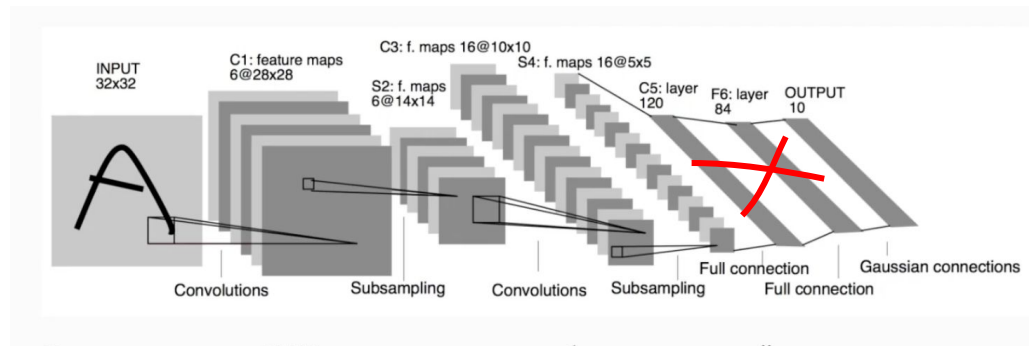
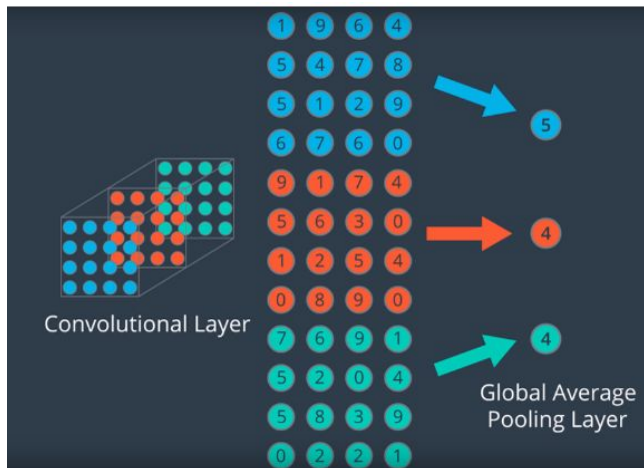
GoogLeNet

- Хотим избавиться от dense слоев в конце - большинство параметров все еще там
- Global Average Pooling



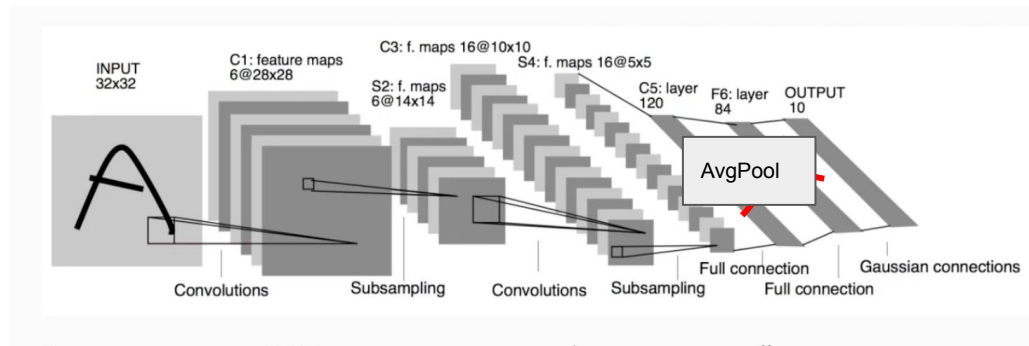
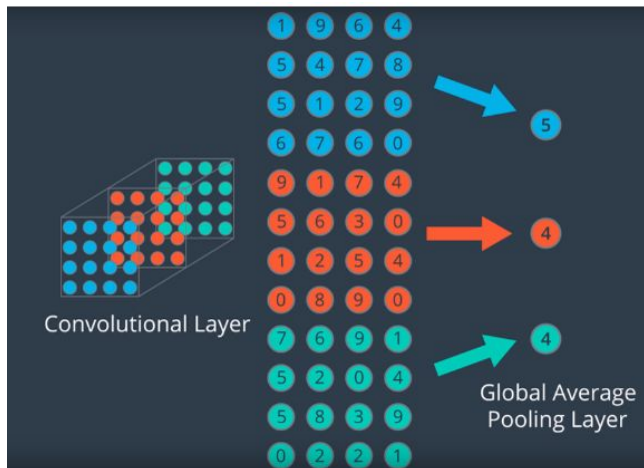
GoogLeNet

- Хотим избавиться от dense слоев в конце - большинство параметров все еще там
- Global Average Pooling



GoogLeNet

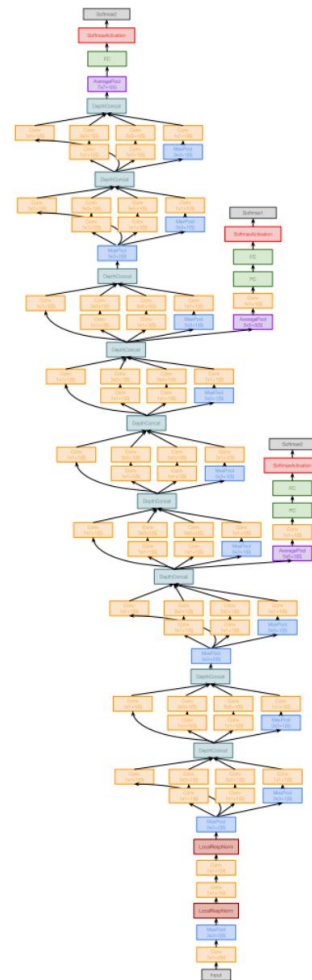
- Хотим избавиться от dense слоев в конце - большинство параметров все еще там
- Global Average Pooling



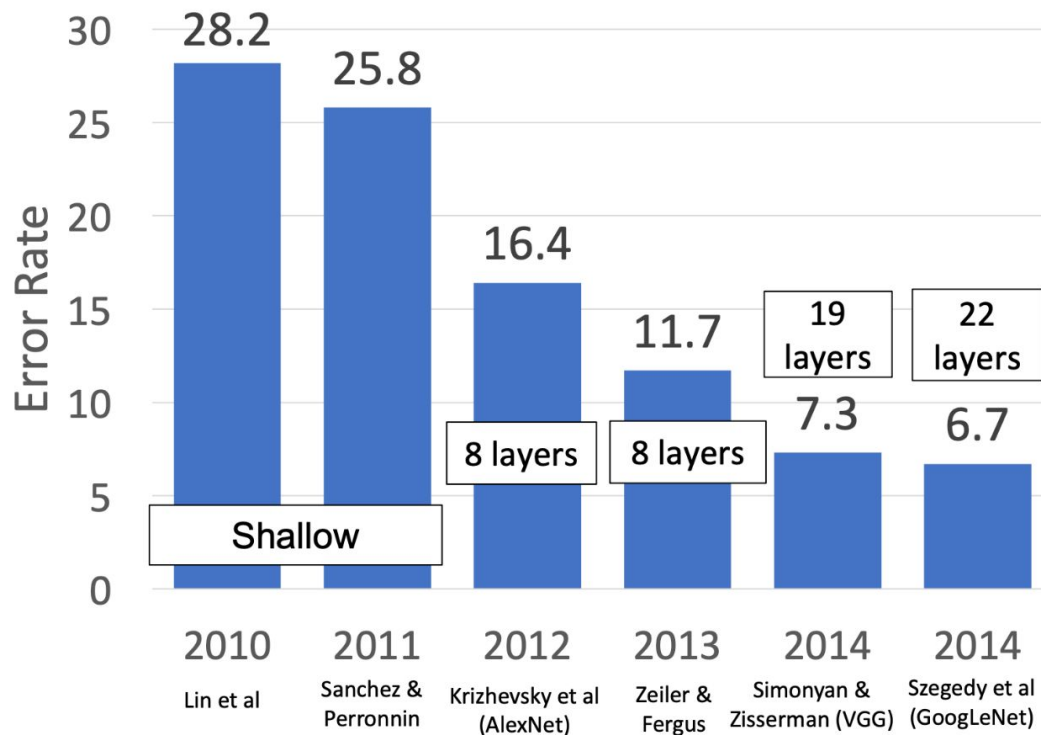
S4 -> avgpool -> вектор размера 16
Соединяем его с выходом dense слоя

GoogLeNet

- Глубокие модели плохо обучаются
- Почему? Насыщение (затухание) градиентов

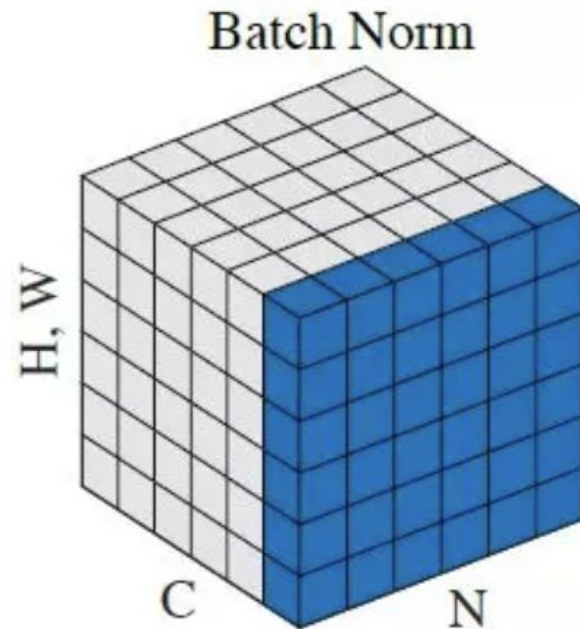


GoogLeNet



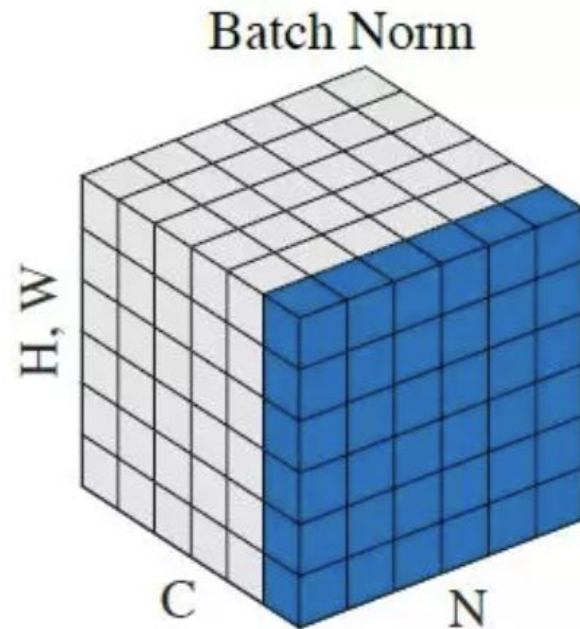
BatchNorm

- Где то тут появился батчнорм
- Доп головы не нужны, можем обучать глубокие сети (10+ слоев) особо не думая



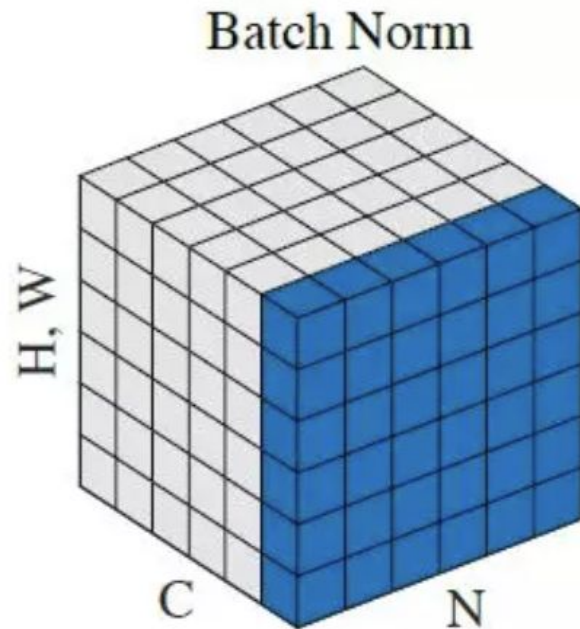
BatchNorm

- Где то тут появился батчнорм
- Доп головы не нужны, можем обучать глубокие сети (10+ слоев) особо не думая
- Почему работает? Пространство становится более гладким



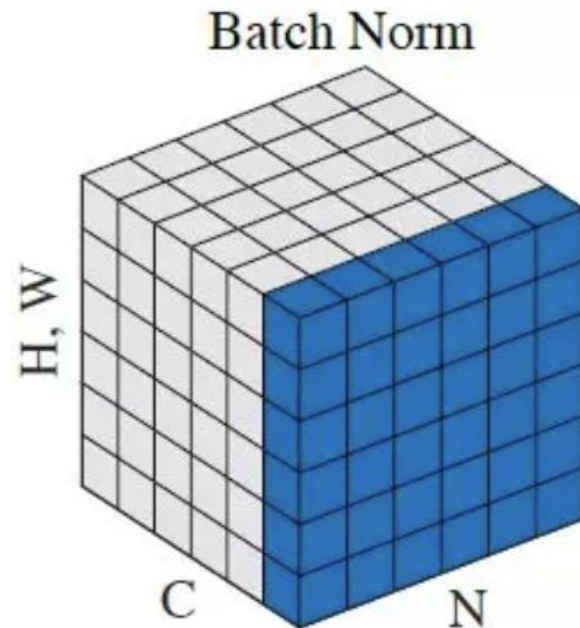
BatchNorm

- Где то тут появился батчнорм
- Доп головы не нужны, можем обучать глубокие сети (10+ слоев) особо не думая
- Почему работает? Пространство становится более гладким
- internal covariate shift , но никто не знает что это такое



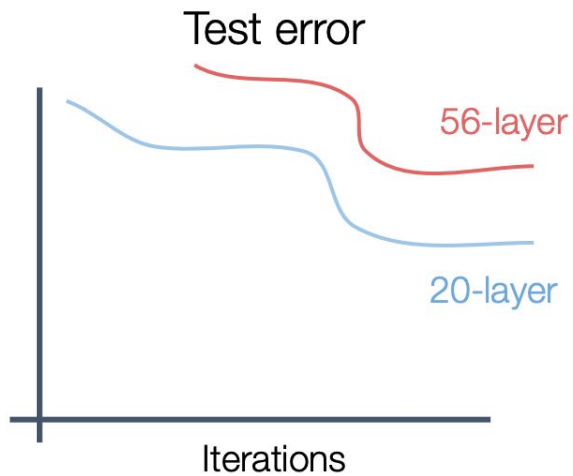
BatchNorm

- Где то тут появился батчнорм
- Доп головы не нужны, можем обучать глубокие сети (10+ слоев) особо не думая
- Почему работает? Пространство становится более гладким
- internal covariate shift , но никто не знает что это такое
- математически - nn - функция, $f(x)$;
Липшицева константа меньше

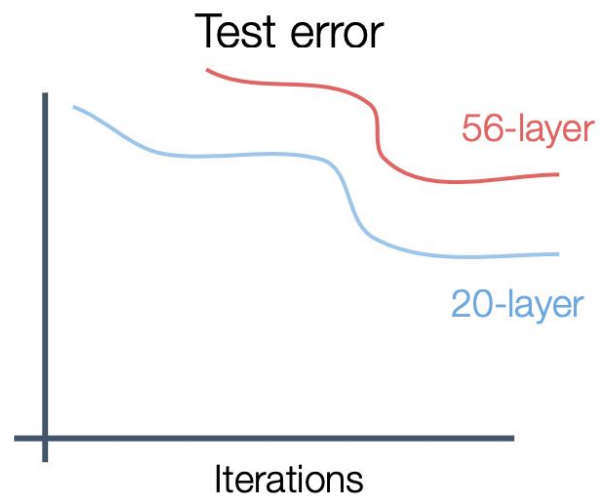
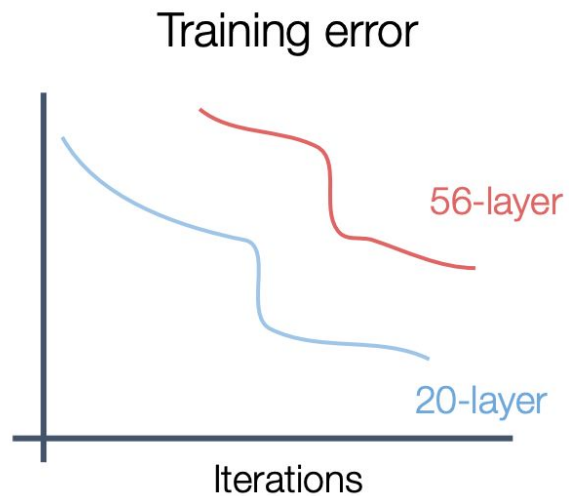


1x1 КОНВОЛЮЦИИ

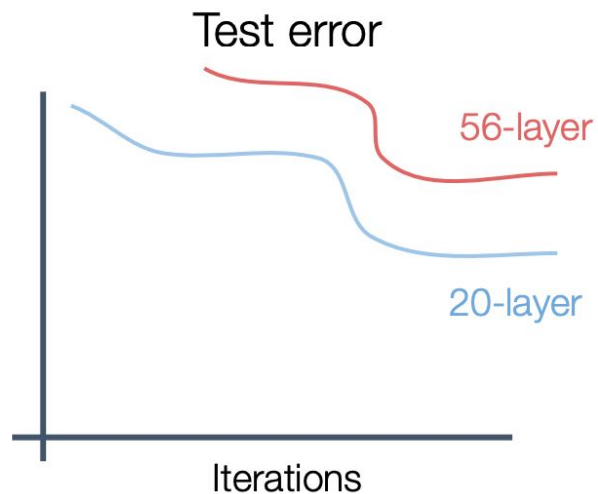
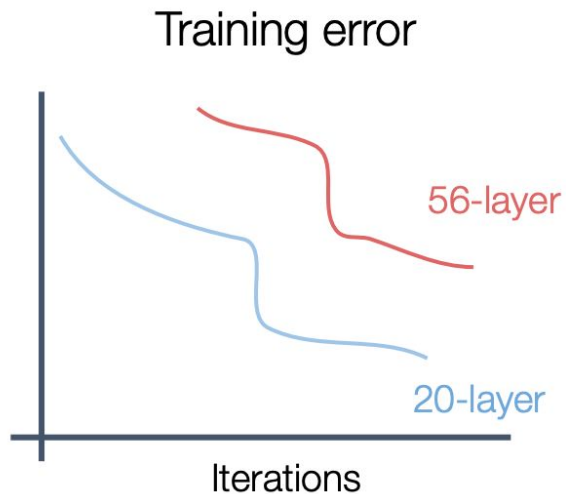
Проблемы глубоких сетей



Проблемы глубоких сетей

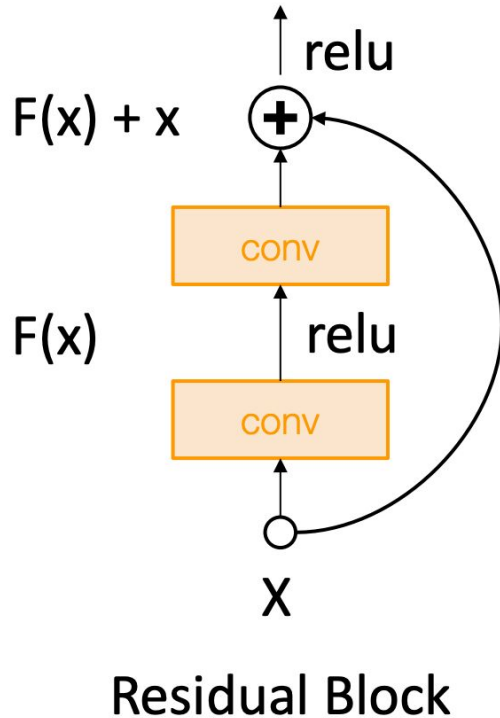
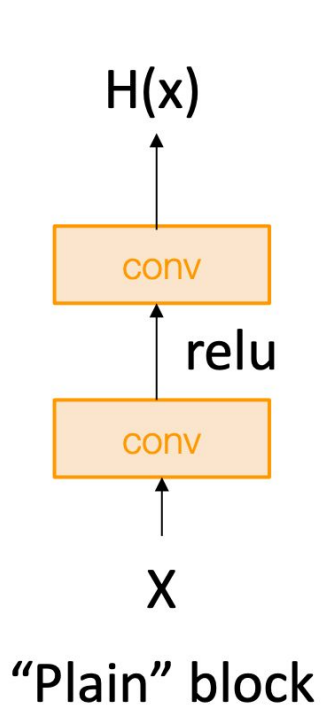


Проблемы глубоких сетей

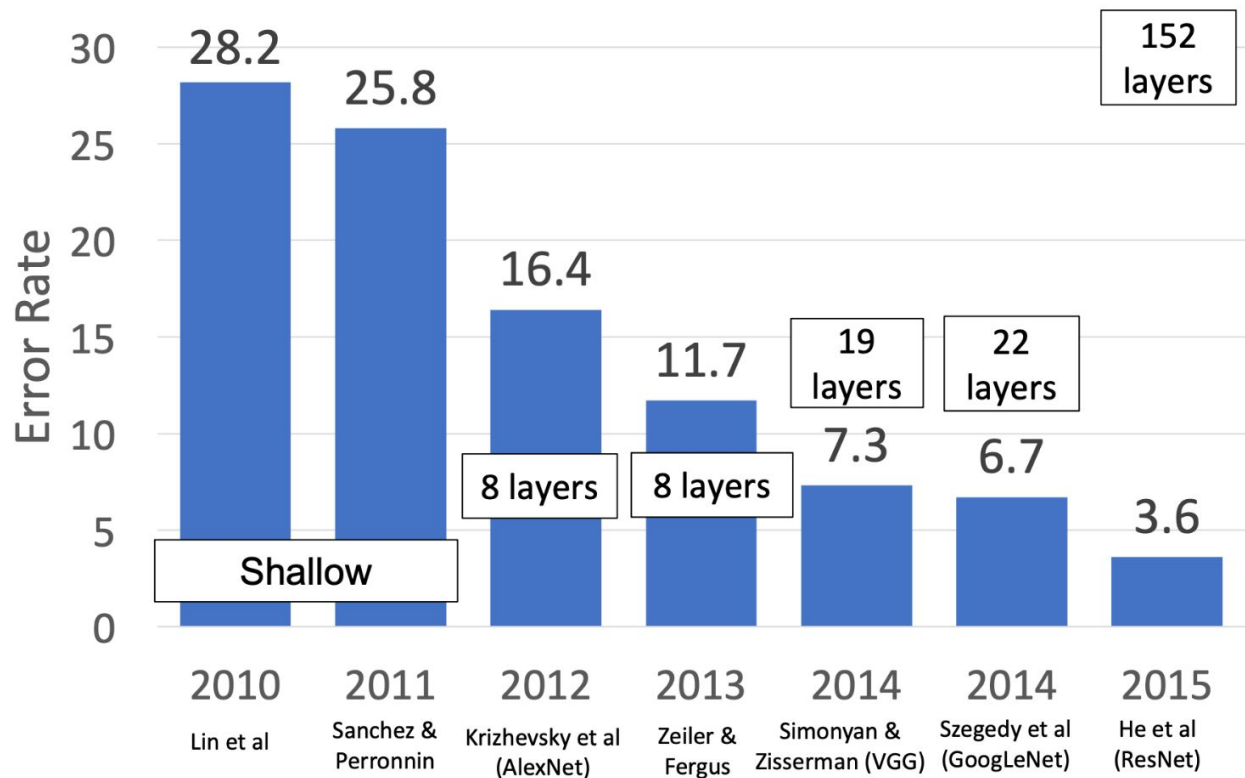


Большая сеть имеет маленькую своей подсетью, так что проблема не в архитектуре а в обучении

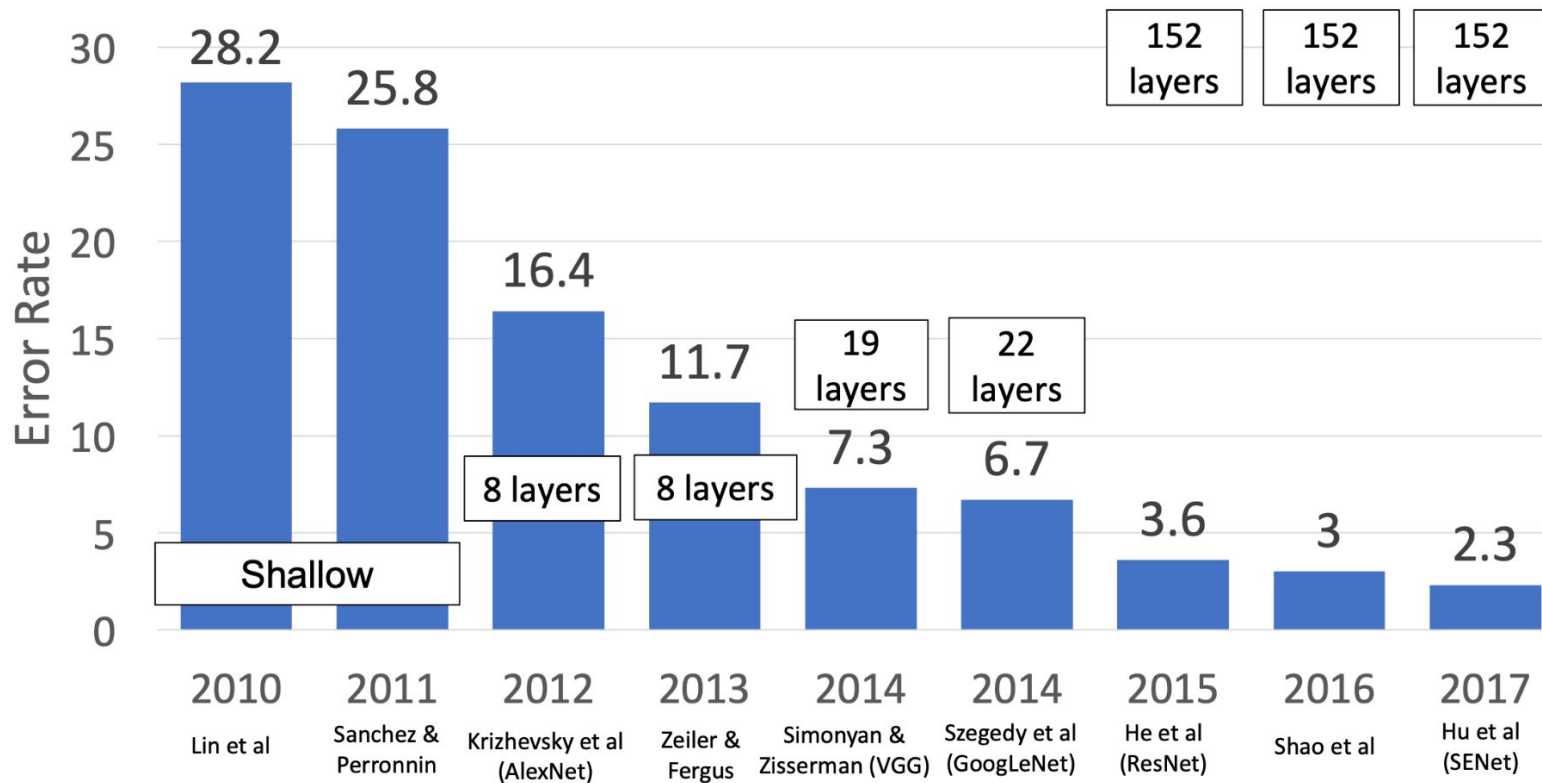
ResNet



ResNet



Конец ImageNet



К чему это все пришло

- делаем качество выше
 - делаем скорость выше
 - количество параметров меньше
 - воюем во все стороны сразу
-
- есть еще подмножество задач про микро модели, когда хочется засунуть в embed устройство или старый телефон (в новый телефон влезет много что)
 - на современном встроенном на iphone чипе вы можете делать тоже самое что на довольно мощной карте, порядка 1080 ; т.е. хороший реалтайм

Аугментации

Original image



augmentation



Horizontal Flip



Crop



Median Blur



Contrast



Hue / Saturation / Value



Gamma



TestTime аугментации



Все еще кот так что можно агрегировать, всегда (когда имеет смысл)
может улучшить результат, но дорого

А нужна ли нам классификация?

- Классификация даст ответ на вопрос есть ли что то на картинке
- Какие бизнес задачи мы через это можем решить?

А нужна ли нам классификация?

- Какие задачи есть
 - self driving
 - распознать окружение и разметку
 - распознать людей
 - понять собственную ориентацию в пространстве
 - ...
 - medical
 - найти аномалию на рентгене
 - посчитать какие нибудь клетки
 - подсветить опухоли
 - manufacturing
 - найти аномалию
 - складская навигация
 - robotics на конвейерах
 - security
 - идентификация
 - retail автоматические магазины без продавцов
 - agro
 - поля размечать
 - предсказывать урожайность
 - искать аномалии и проблемы, болезни, жуков
 - fun
 - фильтр в инстаграме сделать
 - лицо натянуть на дипфейк
 - ... - мало где тут задача является классификацией

Детекция и сегментация

Is this a dog?



Image Classification

What is there in image
and where?



Object Detection

Which pixels belong to
which object?

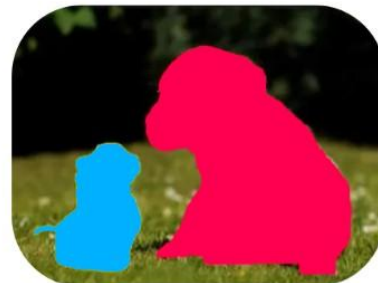


Image Segmentation

Детекция и сегментация

Is this a dog?



Image Classification

What is there in image
and where?



Object Detection

Which pixels belong to
which object?

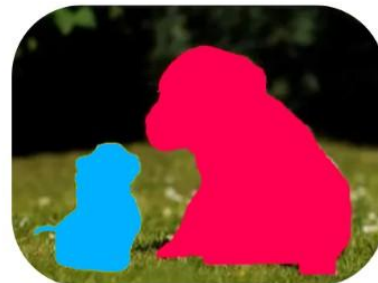


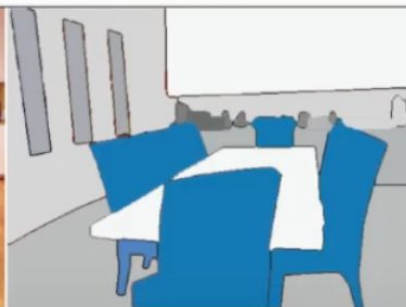
Image Segmentation

А зачем детекция если умеем сегментацию?

Сегментация - задачи



Input Image



Semantic Segmentation

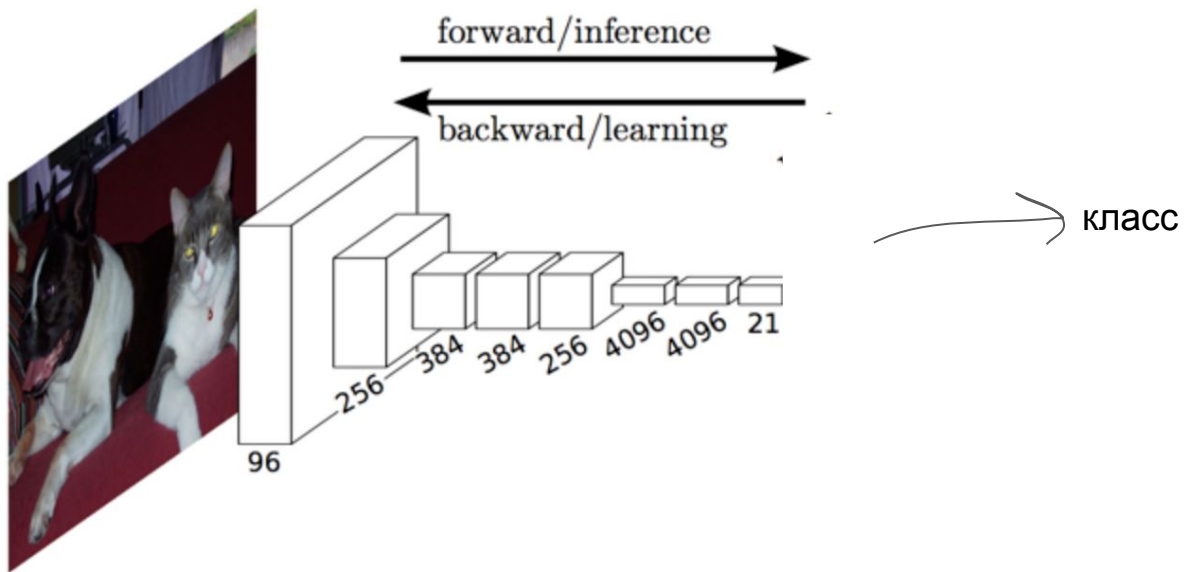


Boundary Segmentation

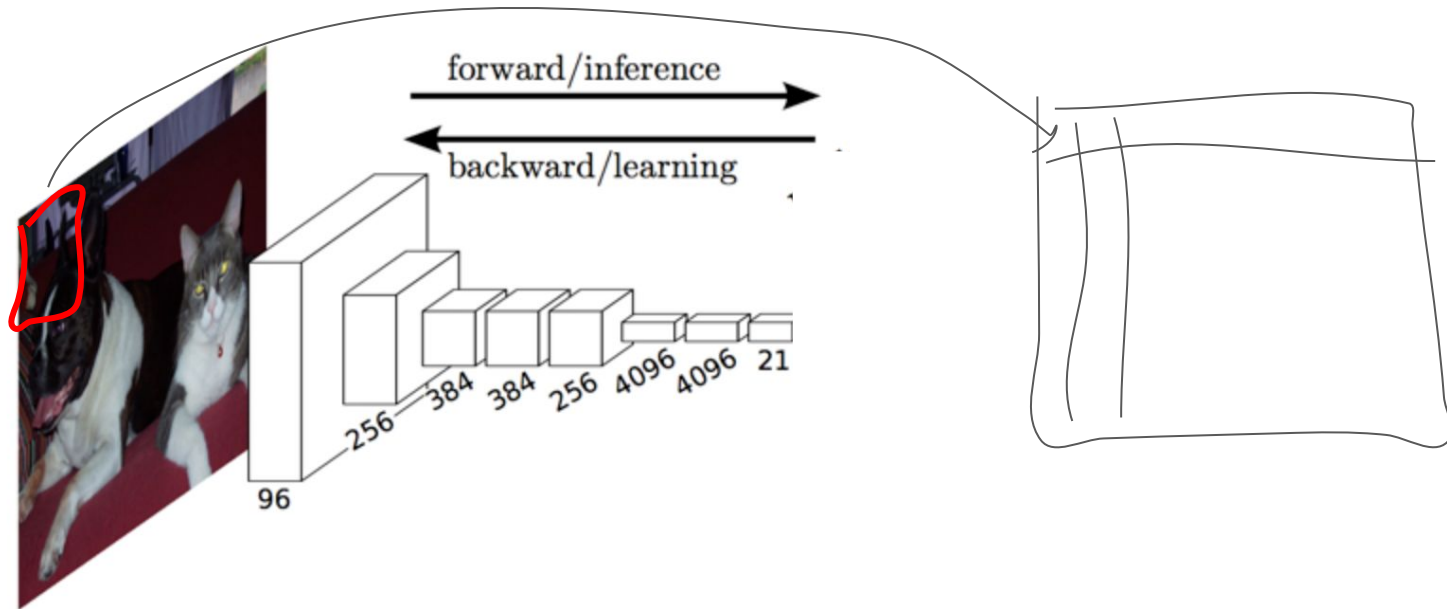


Semantic Instance Segmentation

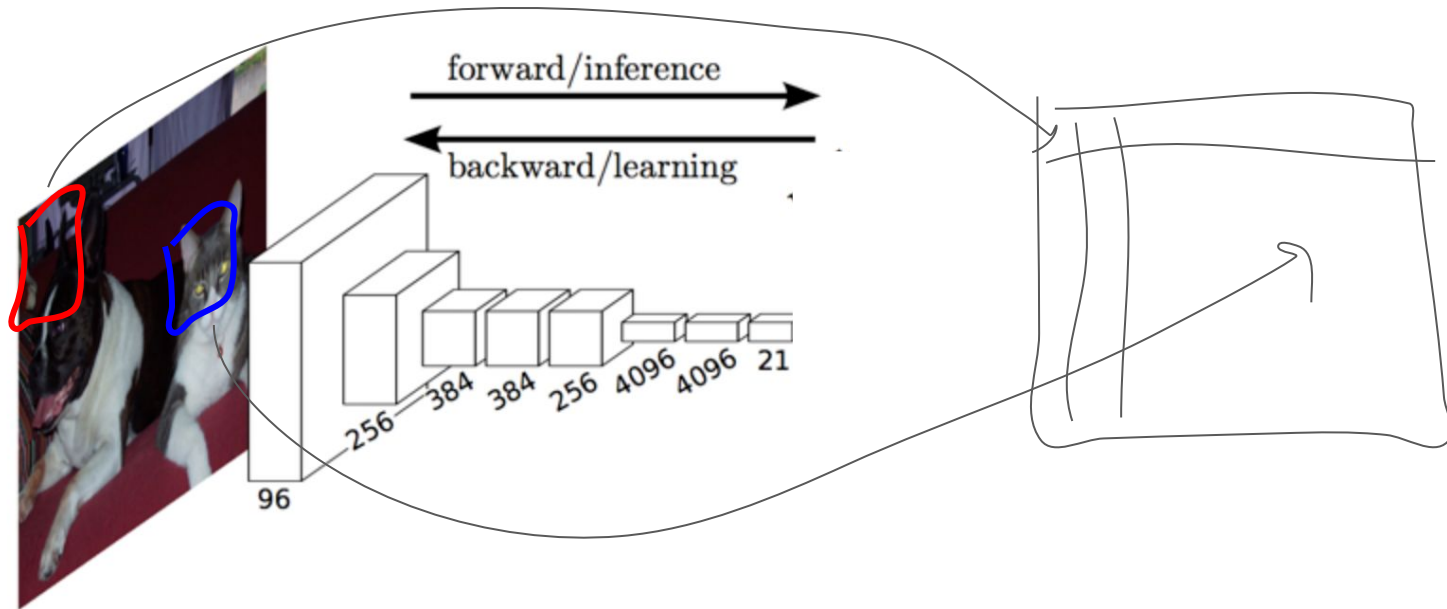
Сегментация - наивно



Сегментация - наивно

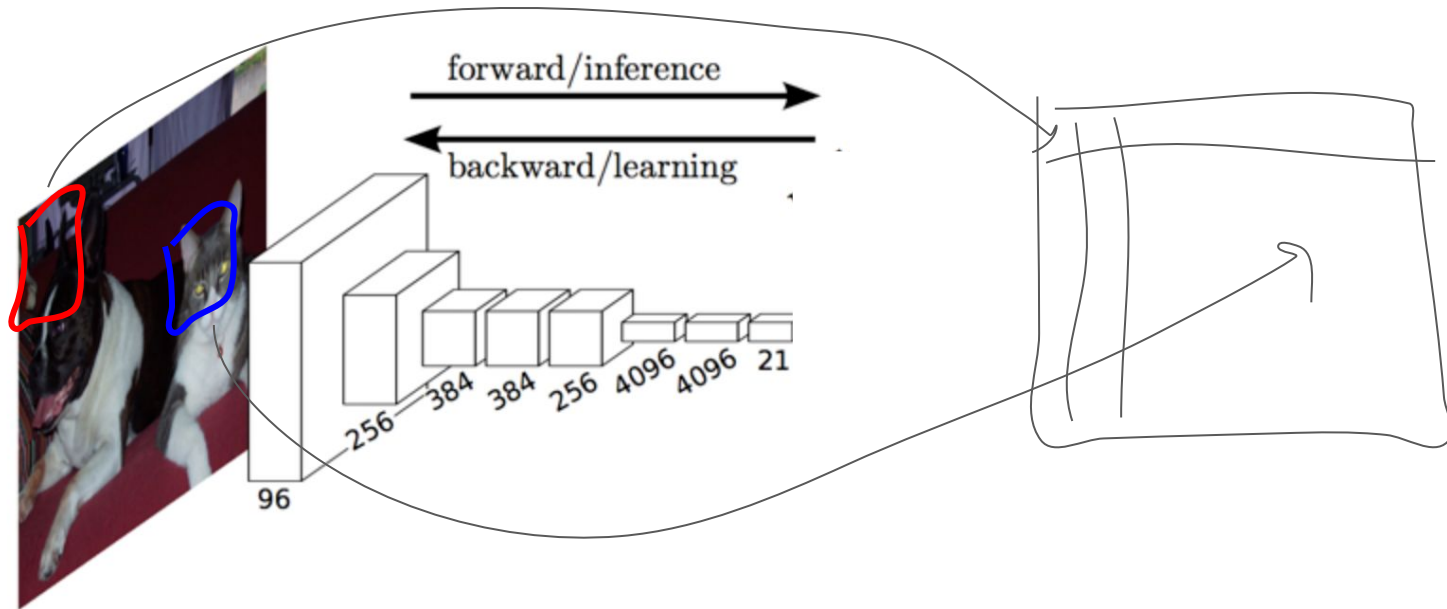


Сегментация - наивно



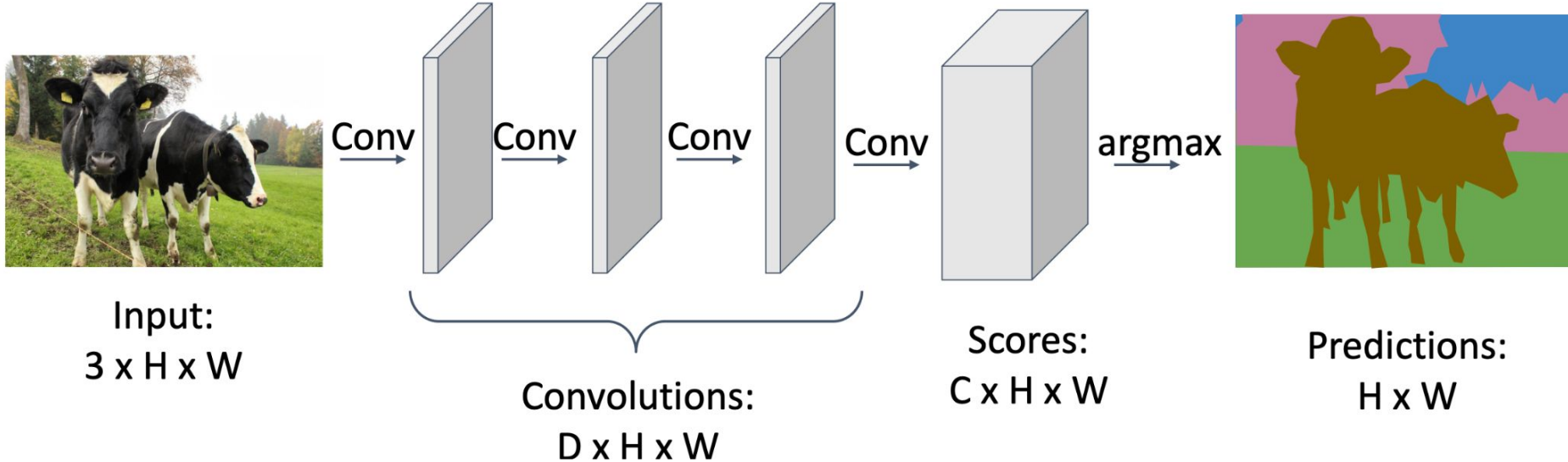
Медленно, не используем маски во время обучения

Сегментация - наивно



Медленно, не используем маски во время обучения

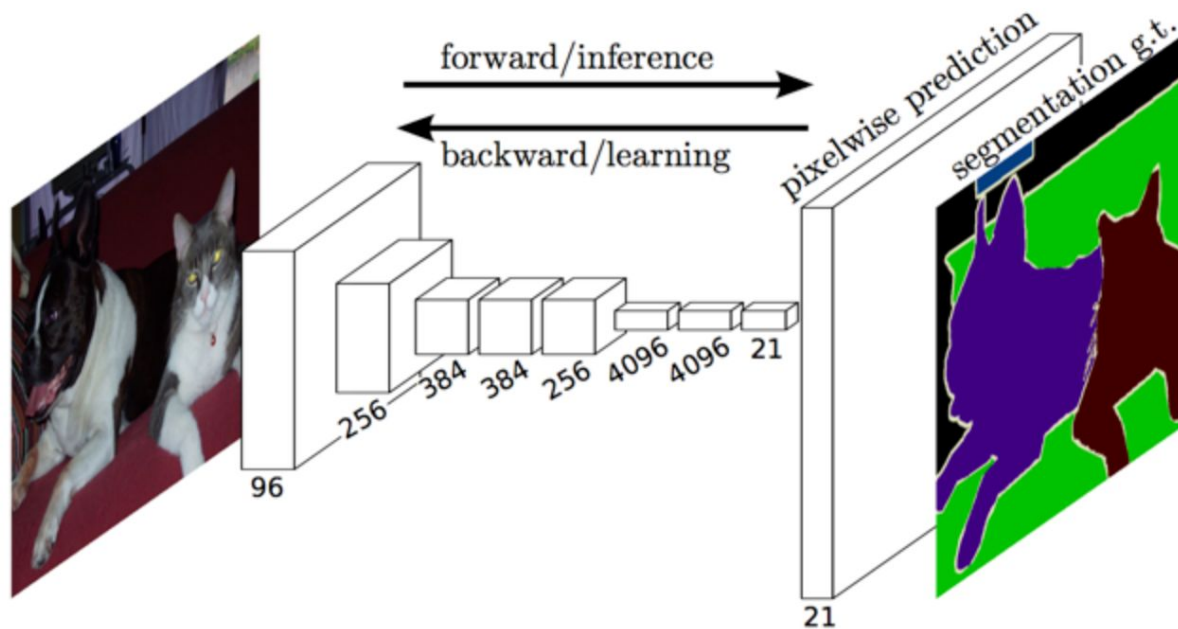
Сегментация - end2end



loss – per pixel cross entropy

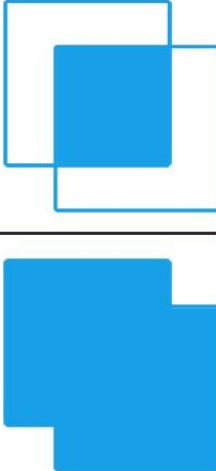
fully conv network - конволюционизация

Сегментация - наивно



сегментация - лосс и проблемы

- кроссентропия (попиксельная класификация)
- может быть довольно unbalanced
- в какой то момент вы умеете хорошо решать задачу так что важны становятся границы - а оно не отбалансено
- iou и дифф версии этого

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$


че делать?

не надо изобретать велосипед

- берете резнет50 или резнект50 ; это даст вам хороший бейзлайн ;
- если на ембединг девайсе - посмотреть не лезут ли пункт выше)) если нет то посмотреть более компактные версии (резнет34 и т д) и мобайл архитектуры
- на сегментации - ищите кто решал похожую задачу ; хорошим бейзлайном будет взять резнет50, накинуть скип слоев и декодера и обучить unet лайк модель ; если ваш датасет ок - то все будет норм

че делать x2?

- сегментация дорогая, че делать?
 - размечаем часть картинок, обучайте модель, сегментируйте с heavy аугментациями, агрегируйте маску - смотрим глазами , мб через это вы получите бесплатно еще кучу примеров + можно выделить то что плохо работало отправить на доразметку
 - обучаем новую модель на новых данных, повторяем процесс
 - через пару итераций датасет будет разумного качества чтоб показать какое то решение
 - !!! обычн это не очень работает в традиционном мл, я пробовал !!! но в дл вы часто можете делать себе данные сами

Ссылки

- в части про архитектуры многие картинки из курса по compvision michigan university <https://www.youtube.com/watch?v=XaZlIVrIO-Q&t=3282s>
- если есть вопросы можно мне написать в телеграмме @vladgl