

Компьютерное Зрение II: Архитектуры

Сергей Овчаренко

Старший разработчик-исследователь, Яндекс

28 июля 2018 г.

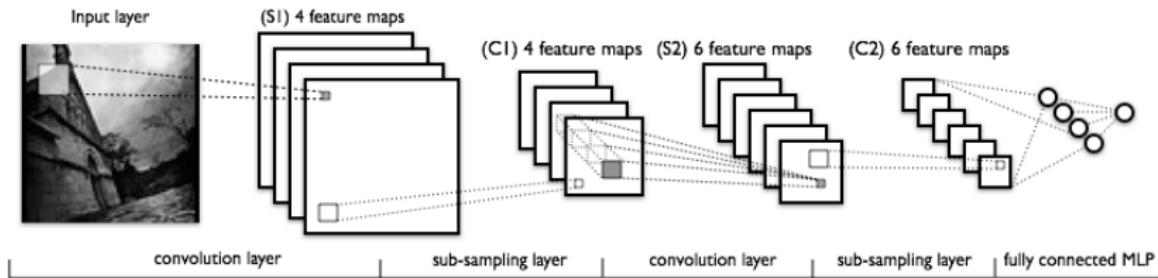
Содержание

1 Архитектуры

- LeNet
- AlexNet
- VGG
- Network in Network
- Inception
- ResNet
- Глубина
- Автоматизация
- Encoder-Decoder
 - Причины
- Transfer learning

2 Почему

LeNet

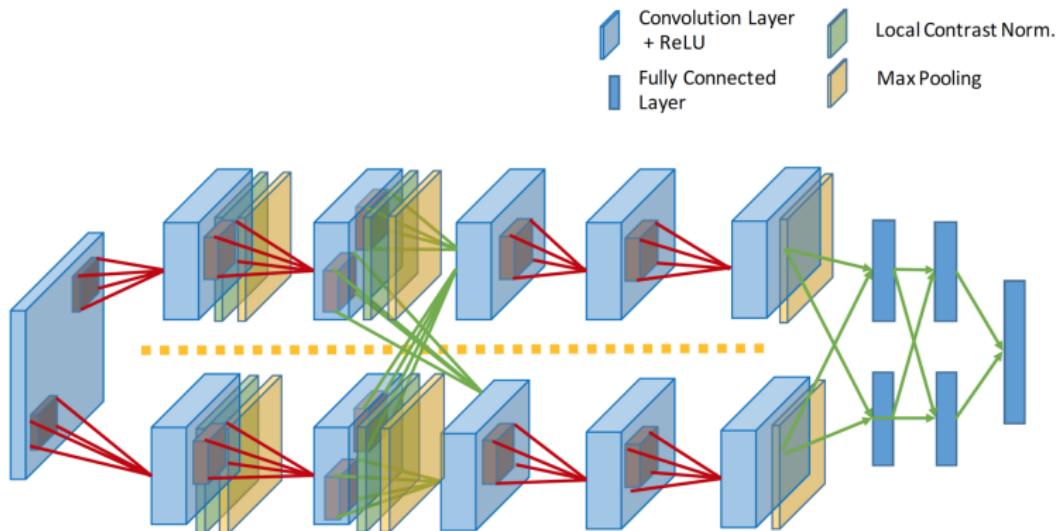


⁰Lecun и др., "Gradient-based learning applied to document recognition".

Архитектуры Почему

LeNet
AlexNet
VGG
Network in Network
Inception
ResNet
Глубина
Автоматизация
Encoder-Decoder

AlexNet

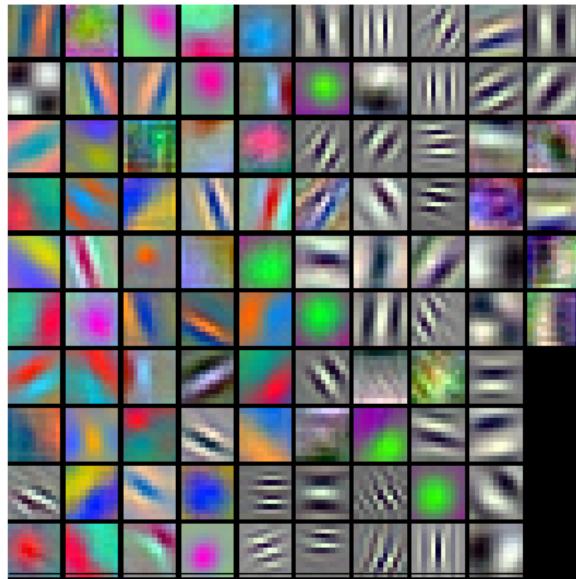


⁰Krizhevsky, Sutskever и Hinton, "ImageNet Classification with Deep Convolutional Neural Networks".

Архитектуры
Почему

LeNet
AlexNet
VGG
Network in Network
Inception
ResNet
Глубина
Автоматизация
Encoder-Decoder

AlexNet Интепретация

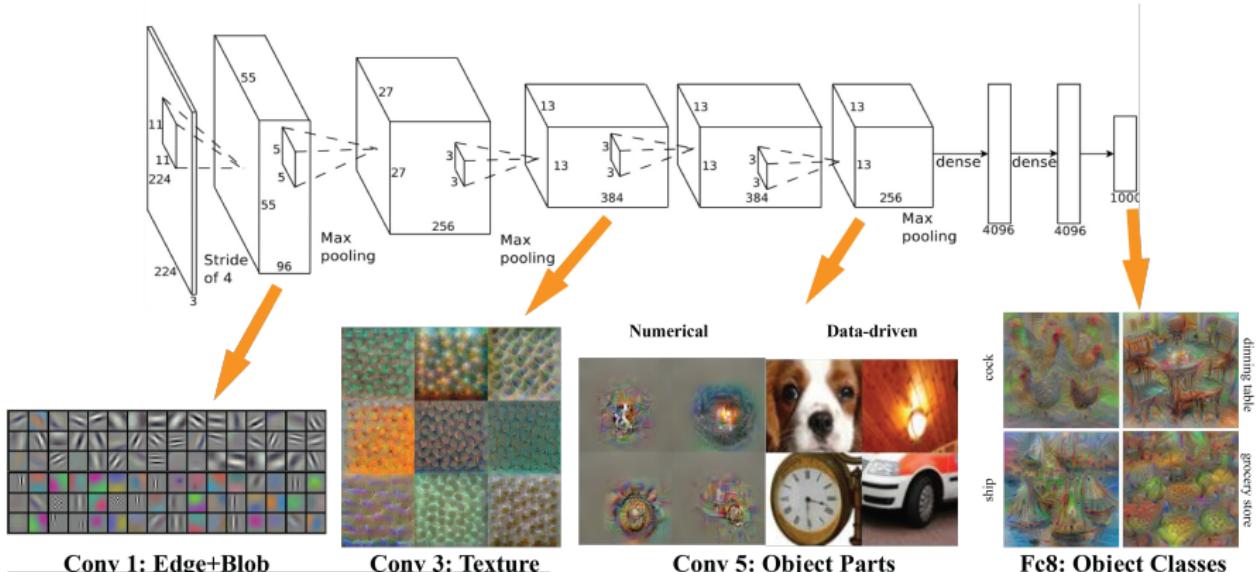


⁰Krizhevsky, Sutskever и Hinton, "ImageNet Classification with Deep Convolutional Neural Networks".

Архитектуры
Почему

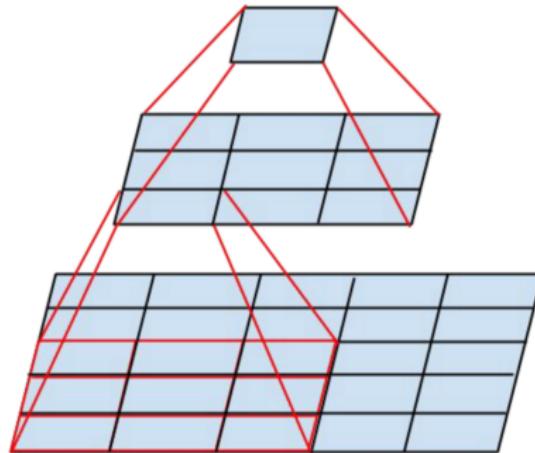
LeNet
AlexNet
VGG
Network in Network
Inception
ResNet
Глубина
Автоматизация
Encoder-Decoder

AlexNet Интепретация II



⁰Krizhevsky, Sutskever и Hinton, "ImageNet Classification with Deep Convolutional Neural Networks".

VGG I



⁰Simonyan и Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition".

Архитектуры Почему

LeNet
AlexNet
VGG
Network in Network
Inception
ResNet
Глубина
Автоматизация
Encoder-Decoder

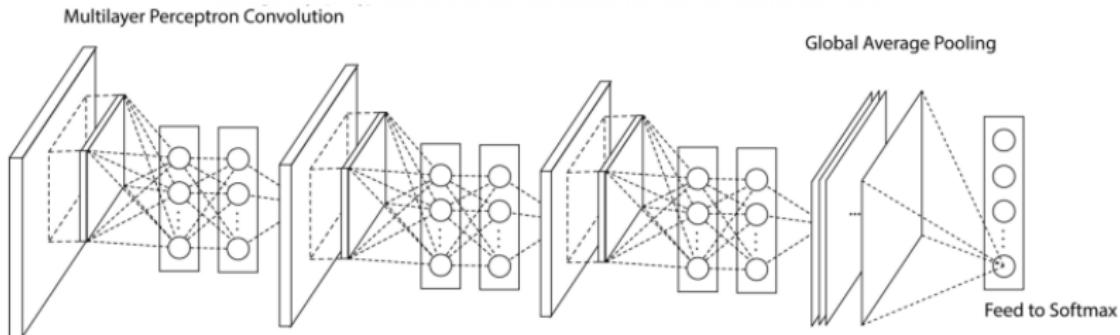
VGG II

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					

Table 2: **Number of parameters** (in millions).

Network	A,A-LRN	B	C	D	E
Number of parameters	133	133	134	138	144

Network in Network

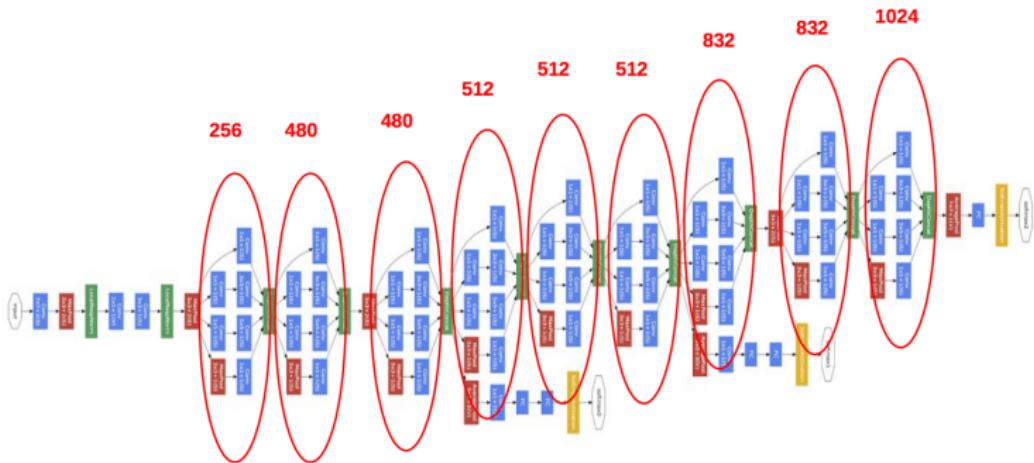


⁰Lin, Chen и Yan, "Network In Network".

Архитектуры
Почему

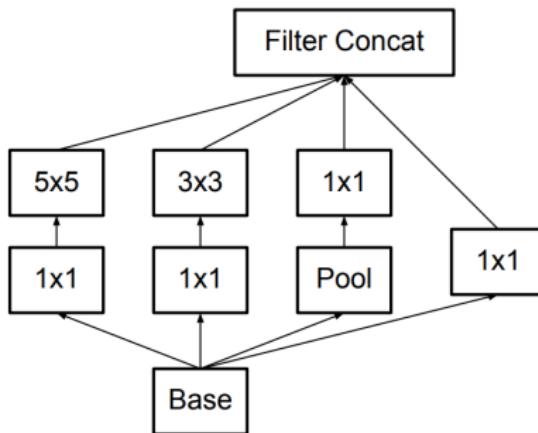
LeNet
AlexNet
VGG
Network in Network
Inception
ResNet
Глубина
Автоматизация
Encoder-Decoder

InceptionV1 aka GoogleNet



⁰Szegedy и др., "Going Deeper with Convolutions".

Inception Module



InceptionV2: Принципы

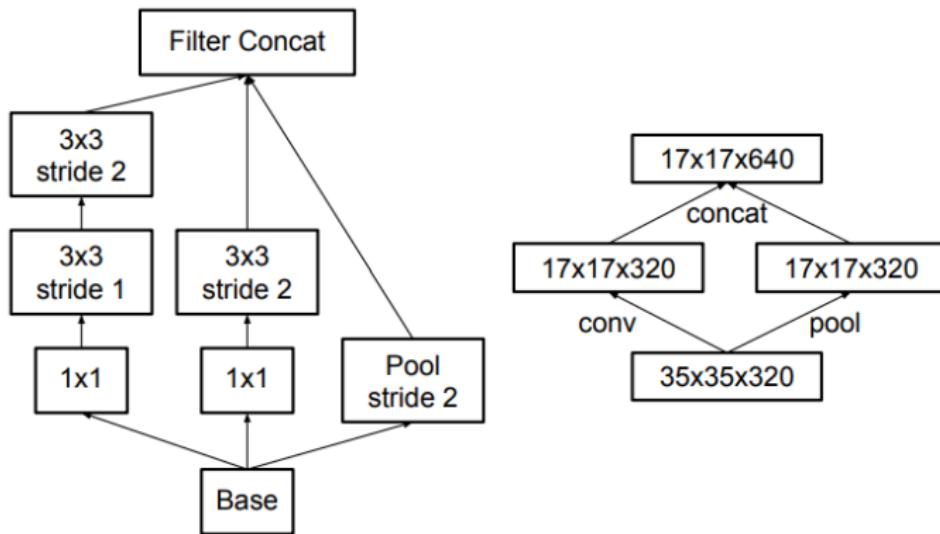
- ❶ Избегайте representational bottlenecks
- ❷ Представления высокой размерности проще обрабатывать локально
- ❸ Пространственная агрегация может происходить в пространстве пониженной размерности без значительный потерь в выразительности
- ❹ Следует искать баланс глубины и ширины сети

⁰Szegedy и др., “Rethinking the Inception Architecture for Computer Vision”.

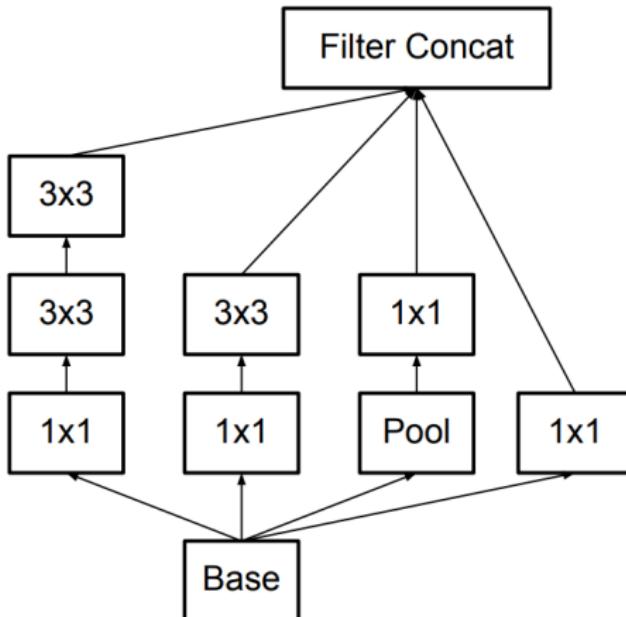
Архитектуры
Почему

LeNet
AlexNet
VGG
Network in Network
Inception
ResNet
Глубина
Автоматизация
Encoder-Decoder

InceptionV2: Принцип 1



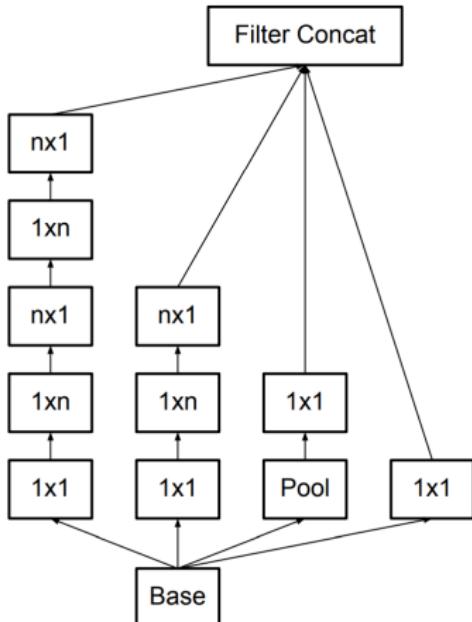
InceptionV2: Принцип 3



Архитектуры
Почему

LeNet
AlexNet
VGG
Network in Network
Inception
ResNet
Глубина
Автоматизация
Encoder-Decoder

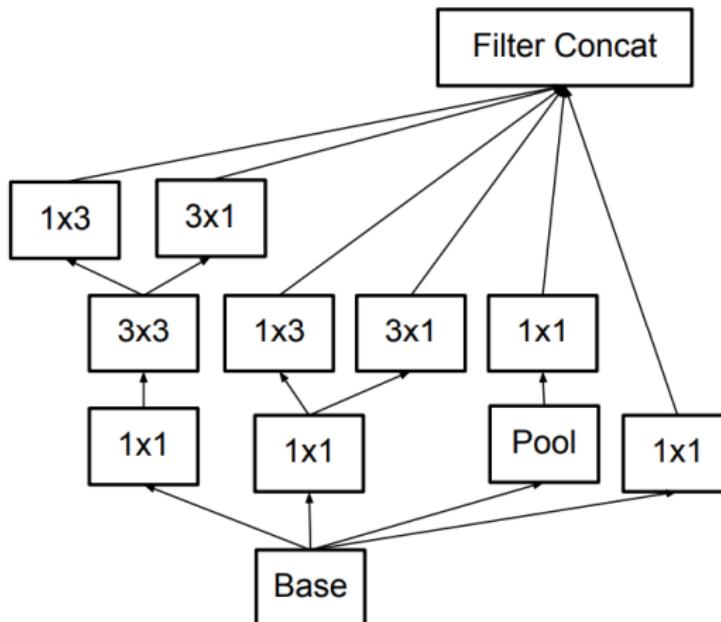
InceptionV2



Архитектуры
Почему

LeNet
AlexNet
VGG
Network in Network
Inception
ResNet
Глубина
Автоматизация
Encoder-Decoder

InceptionV2: Принцип 2



ResNet: Идея

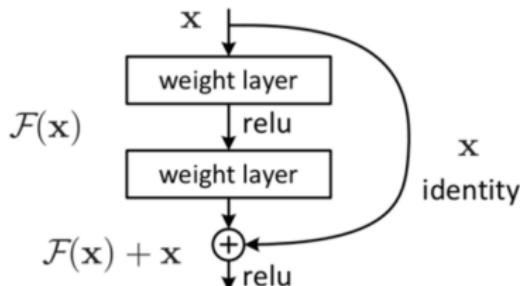


Figure 2. Residual learning: a building block.

$\mathcal{H}(x)$ is the true function we want to learn

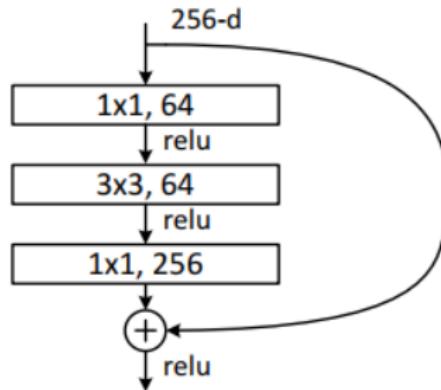
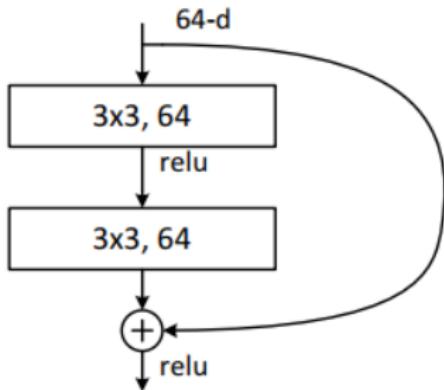
Let's pretend we want to learn
 $\mathcal{F}(x) := \mathcal{H}(x) - x$
 instead.

The original function is then

$$\mathcal{F}(x) + x$$

⁰He и др., "Deep Residual Learning for Image Recognition".

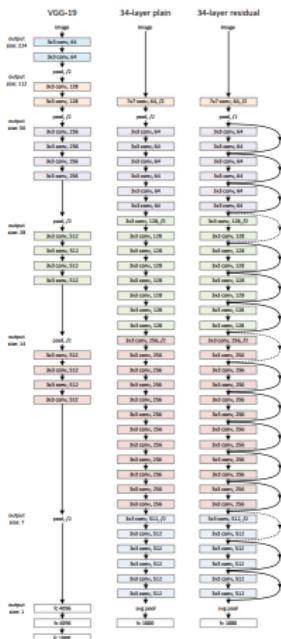
ResNet: Реализация



Архитектуры Почему

LeNet
AlexNet
VGG
Network in Network
Inception
ResNet
Глубина
Автоматизация
Encoder-Decoder

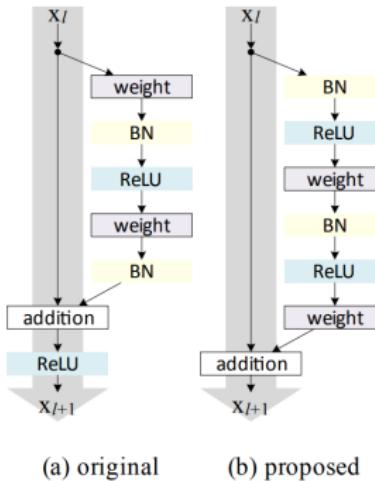
ResNet: Результаты



Архитектуры
Почему

LeNet
AlexNet
VGG
Network in Network
Inception
ResNet
Глубина
Автоматизация
Encoder-Decoder

ResNet: Identity mappings



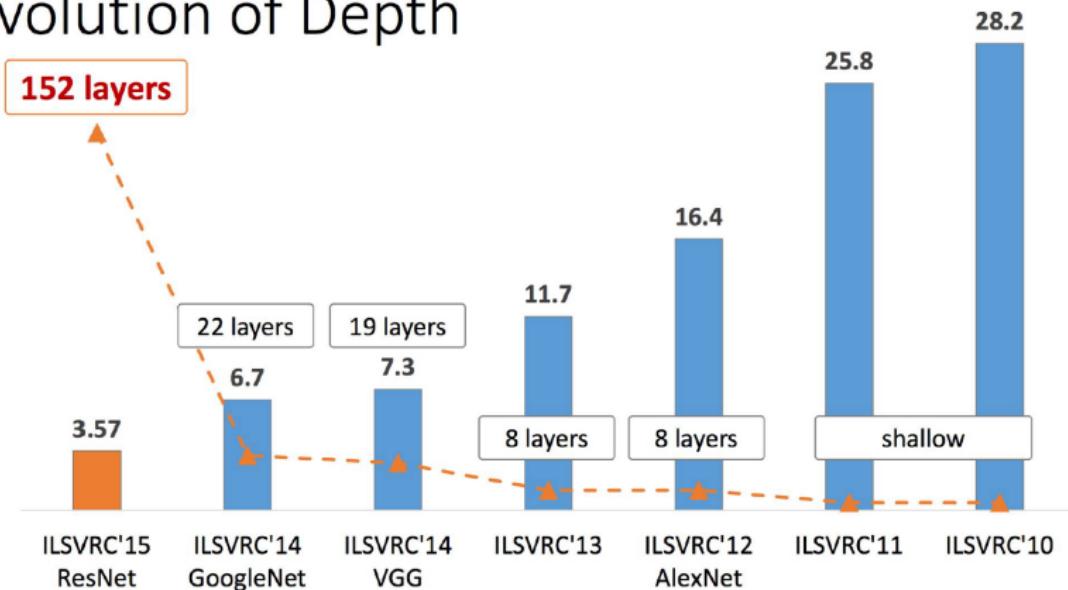
⁰Не и др., “Identity Mappings in Deep Residual Networks”.

Архитектуры
Почему

LeNet
AlexNet
VGG
Network in Network
Inception
ResNet
Глубина
Автоматизация
Encoder-Decoder

ResNet: ImageNet

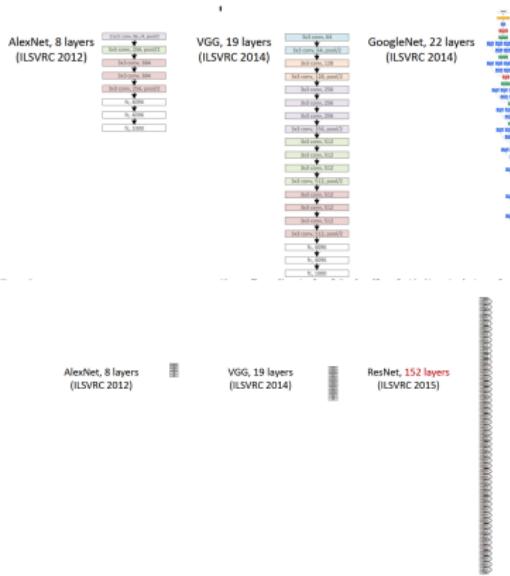
Revolution of Depth



Архитектуры Почему

LeNet
AlexNet
VGG
Network in Network
Inception
ResNet
Глубина
Автоматизация
Encoder-Decoder

ResNet: All your layers are belong to us



Архитектуры Почему

LeNet
AlexNet
VGG
Network in Network
Inception
ResNet
Глубина
Автоматизация
Encoder-Decoder

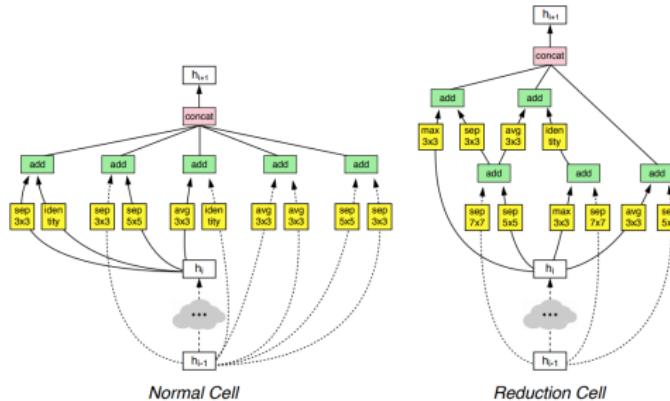
Stack more layers



Архитектуры
Почему

LeNet
AlexNet
VGG
Network in Network
Inception
ResNet
Глубина
Автоматизация
Encoder-Decoder

Пусть нейронная сеть подбирает архитектуру нейронной сети

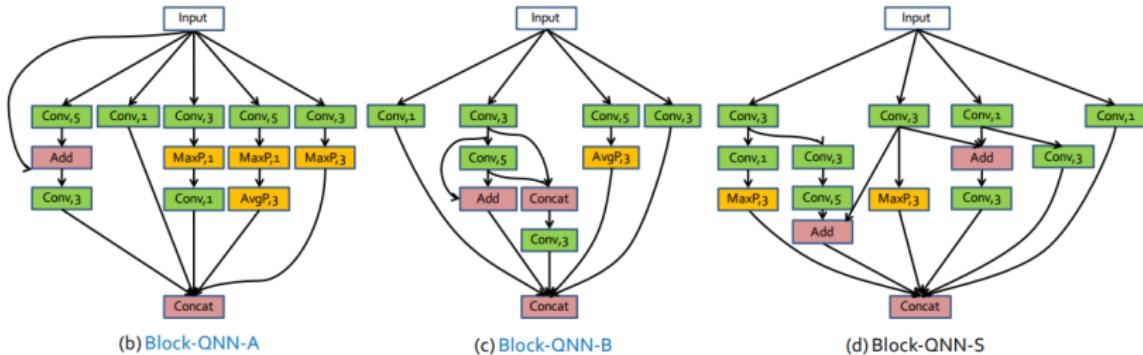


⁰Zoph и др., “Learning Transferable Architectures for Scalable Image Recognition”.

Архитектуры Почему

LeNet
AlexNet
VGG
Network in Network
Inception
ResNet
Глубина
Автоматизация
Encoder-Decoder

Пусть нейронная сеть подбирает архитектуру нейронной сети



⁰Zhong, Yan и Liu, "Practical Network Blocks Design with Q-Learning".

Архитектуры
Почему

LeNet
AlexNet
VGG
Network in Network
Inception
ResNet
Глубина
Автоматизация
Encoder-Decoder

У них неплохо получается!

Method	Input Size	Depth	Top-1	Top-5
VGG [25]	224x224	16	28.5	9.90
Inception V1 [30]	224x224	22	27.8	10.10
Inception V2 [14]	224x224	22	25.2	7.80
ResNet-50 [11]	224x224	50	24.7	7.80
ResNet-152 [11]	224x224	152	23.0	6.70
Xception(our test) [4]	224x224	50	23.6	7.10
ResNext-101(64x4d) [35]	224x224	101	20.4	5.30
Block-QNN-B, N=3	224x224	38	24.3	7.40
Block-QNN-S, N=3	224x224	38	22.6	6.46

⁰Zhong, Yan и Liu, "Practical Network Blocks Design with Q-Learning".

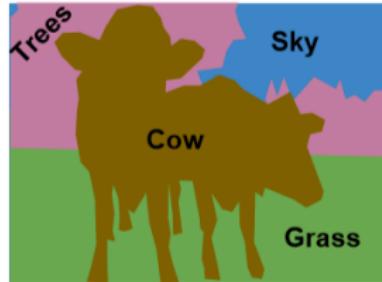
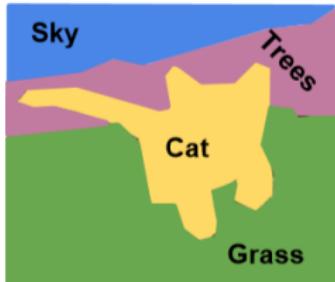
Архитектуры Почему

LeNet
AlexNet
VGG
Network in Network
Inception
ResNet
Глубина
Автоматизация
Encoder-Decoder

Мотивация



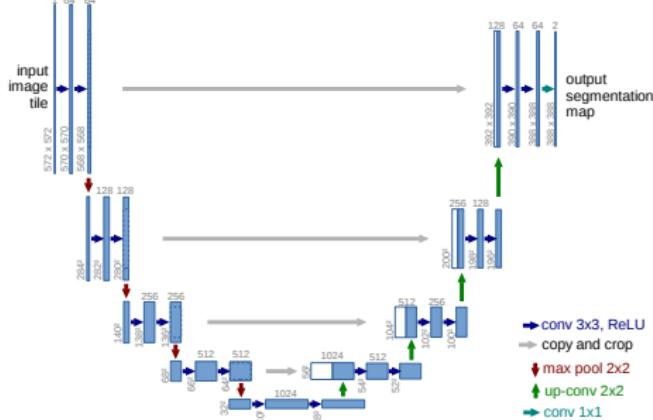
This image is CC0 public domain



Архитектуры Почему

LeNet
AlexNet
VGG
Network in Network
Inception
ResNet
Глубина
Автоматизация
Encoder-Decoder

U-Net

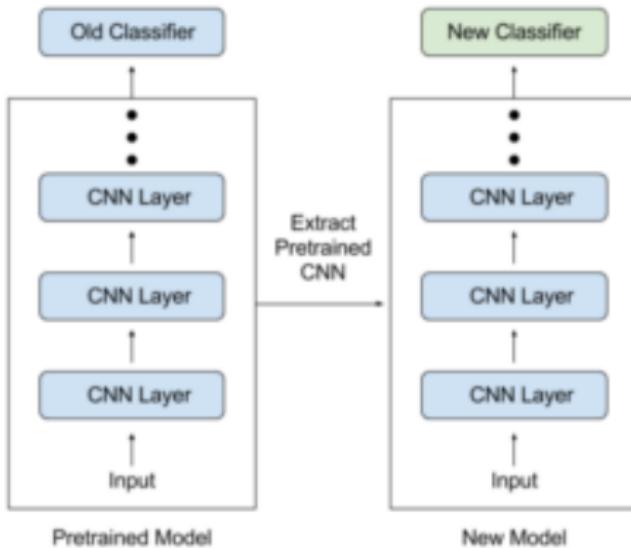


⁰Ronneberger, Fischer и Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”.

Зачем нам готовые архитектуры?

- Разработка архитектуры – трудоёмкий процесс
- Понимание существующих архитектур помогает строить новые
- Transfer learning

Transfer learning I



Transfer learning II

