RNP-VC

Aula 5: Transfer Learning (Transferência de aprendizado)

Lucas Pereira, Rafael Teixeira, Lucas Assis, Anderson Soares
Instituto de Informática
Universidade Federal de Goiás (UFG)

Sumário

- No último episódio...
- O que é transfer learning?
- Model zoo
- No próximo episódio...

No último episódio...

- Hiperparâmetros

No último episódio...

- Parâmetros:

$$W^{[1]}, b^{[1]}, W^{[2]}, b^{[2]}, W^{[3]}, b^{[3]} \; \ldots$$

- Hiperparâmetros:
  - Pré-processamento
  - Arquitetura
  - Função de ativação
  - Inicialização de parâmetros
  - Algoritmo de otimização
  - Taxa de aprendizado (learning rate)
  - Técnicas de regularização
  - [...]

O que é Transfer Learning?

- Tarefa 1: cachorros vs gatos

Dados

Modelo

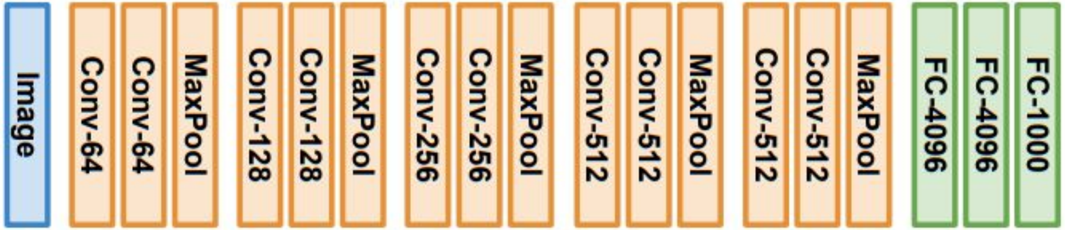
Custo

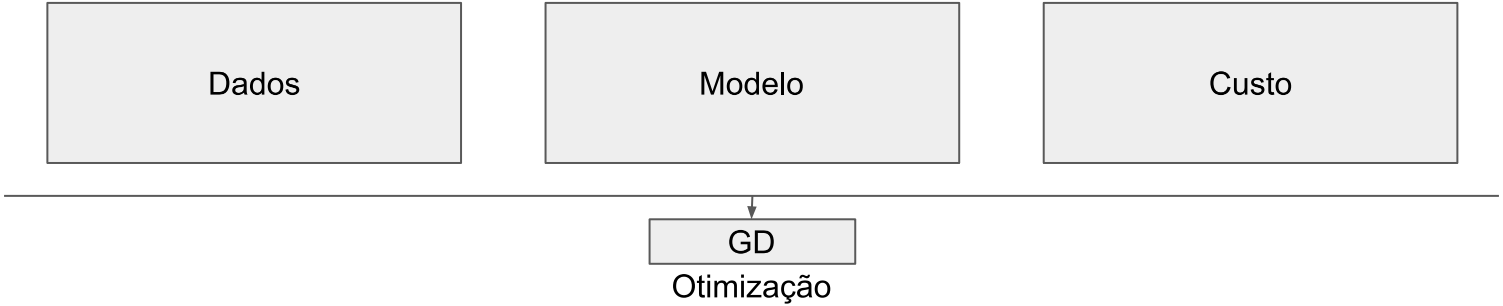
GD

Otimização

O que é Transfer Learning?

- Tarefa 1: cachorros vs gatos





O que é Transfer Learning?

- Tarefa 1: cachorros vs gatos

$$CE = -\sum_{i}^{C} t_i log(s_i)$$

Dados

Modelo

Custo

GD

Otimização

O que é Transfer Learning?

- Tarefa 1: cachorros vs gatos

$$CE = -\sum_{i}^{C} t_i log(s_i)$$

Dados

Modelo

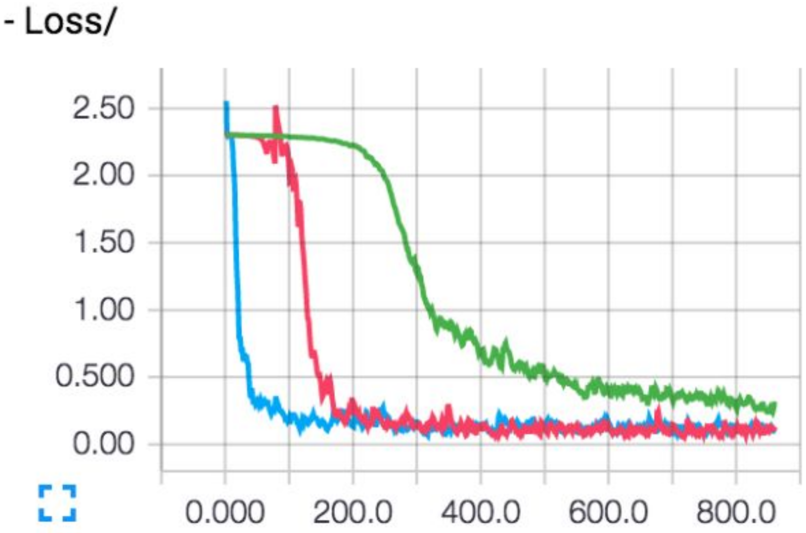
Custo

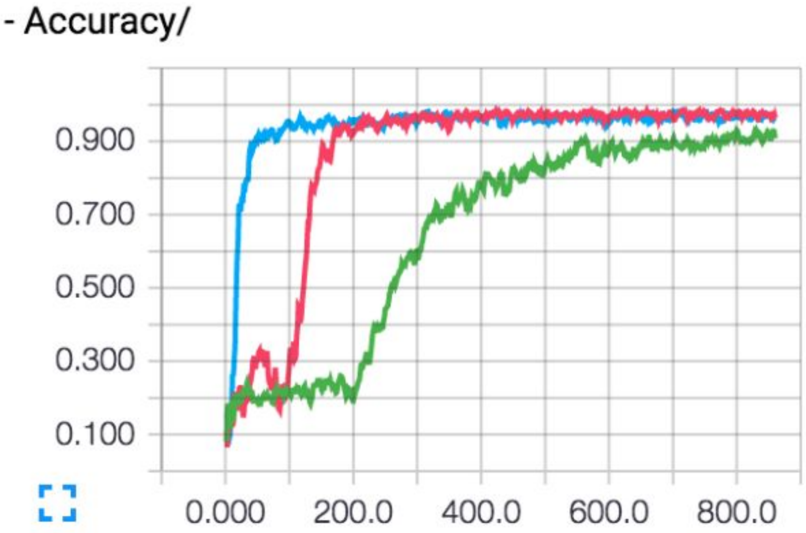
GD

Otimização

O que é Transfer Learning?

- Tarefa 1: cachorros vs gatos





O que é Transfer Learning?

- Tarefa 2: lobos vs onças

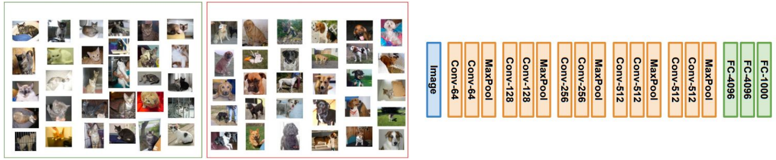
O que é Transfer Learning?

- Tarefa 2: lobos vs onças

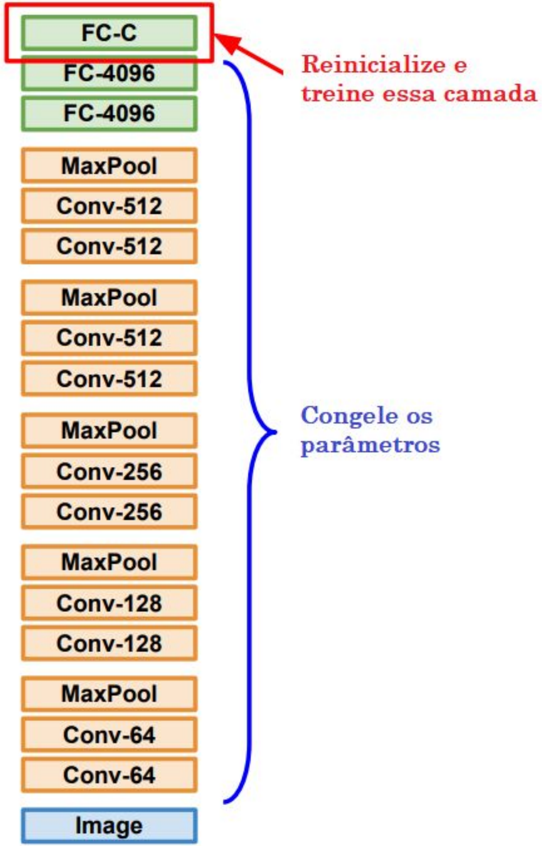
?

O que é Transfer Learning?

- Tarefa 2: lobos vs onças



!



O que é Transfer Learning?

- Tarefa 3: elefantes vs ursos

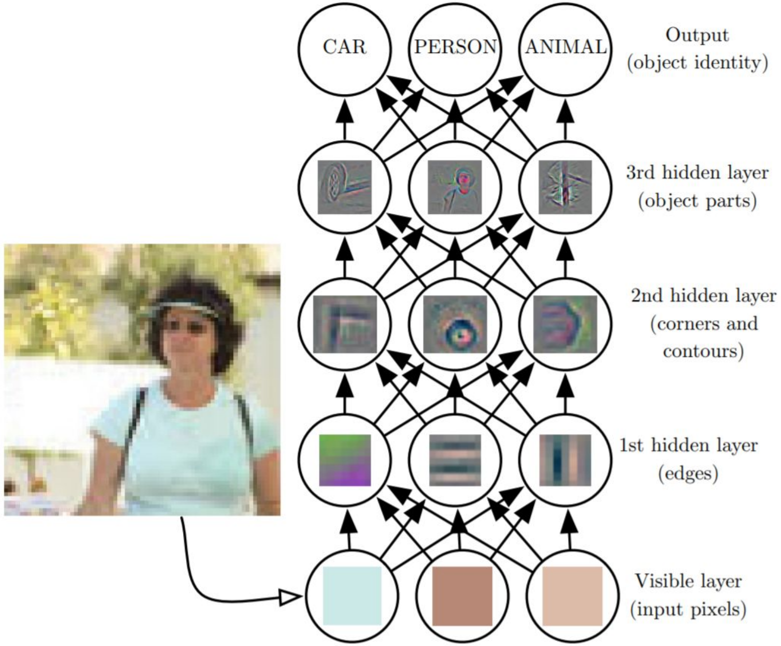
O que é Transfer Learning?

- Tarefa 3: elefantes vs ursos

?

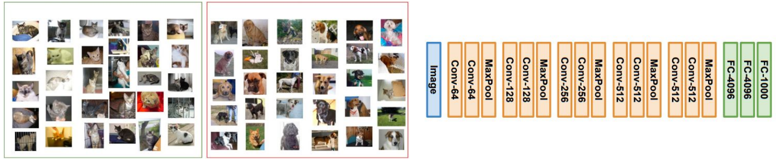
O que é Transfer Learning?

- Composição de representações

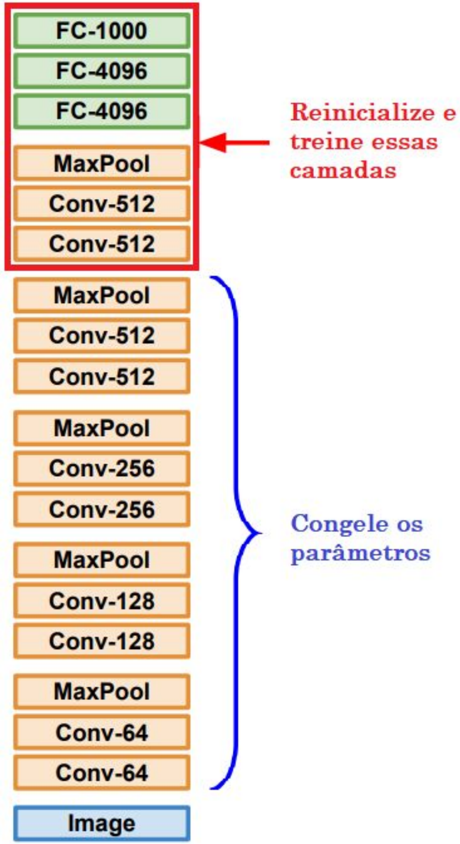


O que é Transfer Learning?

- Tarefa 3: elefantes vs ursos

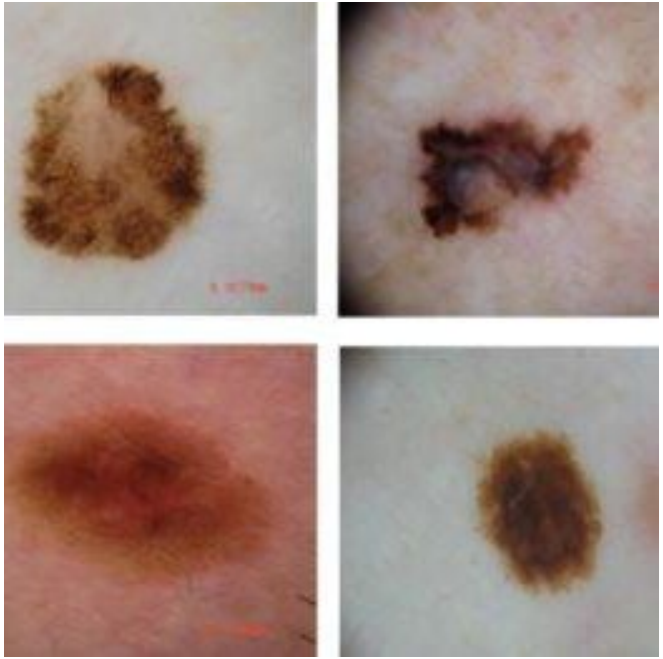


!



O que é Transfer Learning?

- Tarefa 4: tipos de câncer de pele



O que é Transfer Learning?

- Tarefa 4: tipos de câncer de pele

?

O que é Transfer Learning?

- Tarefa 4: tipos de câncer de pele

!

O que é Transfer Learning?

- Tarefa 4: tipos de câncer de pele

!?

O que é Transfer Learning?

- Quando usar TL

	Datasets Similares	Datasets Distintos
Muitos dados disponíveis	Treine algumas camadas do modelo base e o classificador de saída	Treine um número maior de camadas (ou todas elas)
Poucos dados disponíveis	Treine apenas o classificador de saída	É... Houston, we have a problem

Model Zoo

- ImageNet
  - ~1.2M de imagens

IMAGENET

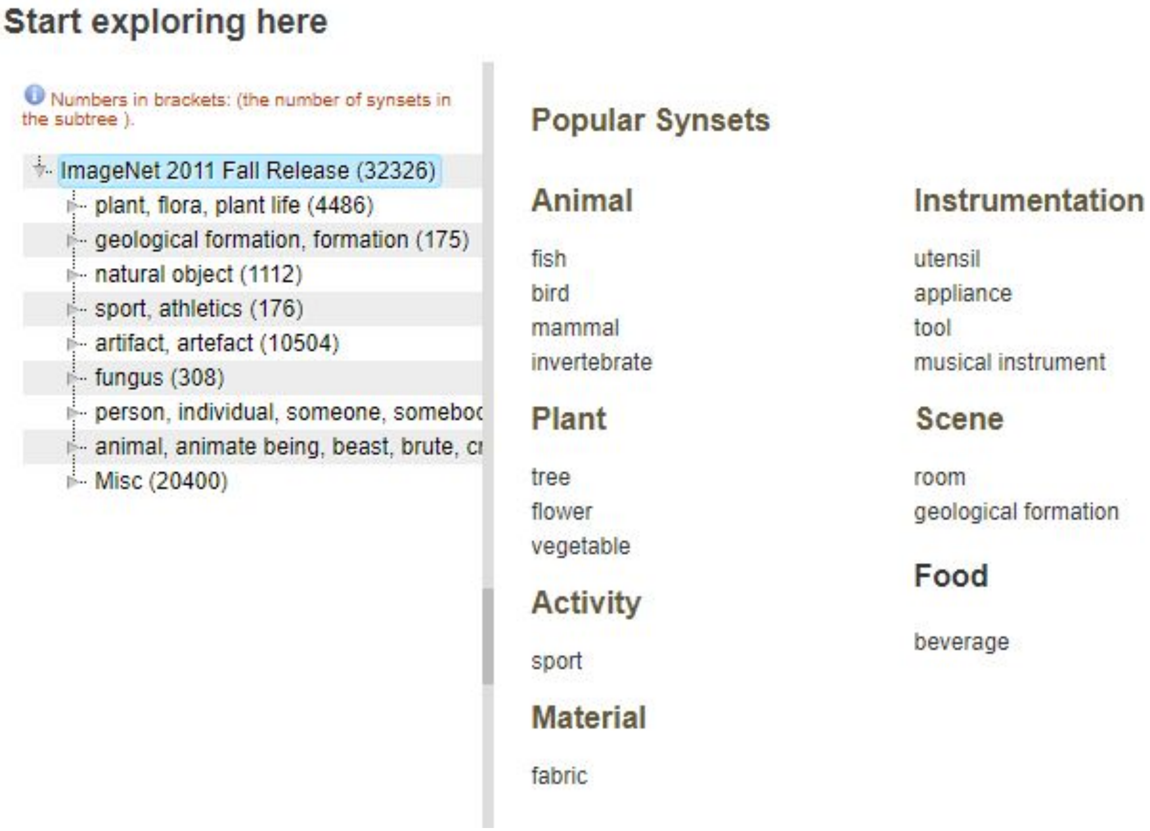
1000 objetos/seres

Model Zoo

- ImageNet
  - ~1.2M de imagens

IMAGENET

1000 objetos/seres



Model Zoo

- AlexNet, 2012

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

Abstract

We trained a large, deep convolutional neural network to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art. The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully-connected layers with a final 1000-way softmax. To make training faster, we used non-saturating neurons and a very efficient GPU implementation of the convolution operation. To reduce overfitting in the fully-connected layers we employed a recently-developed regularization method called "dropout" that proved to be very effective. We also entered a variant of this model in the ILSVRC-2012 competition and achieved a winning top-5 test error rate of 15.3%, compared to 26.2% achieved by the second-best entry.

Model Zoo

- AlexNet, 2012

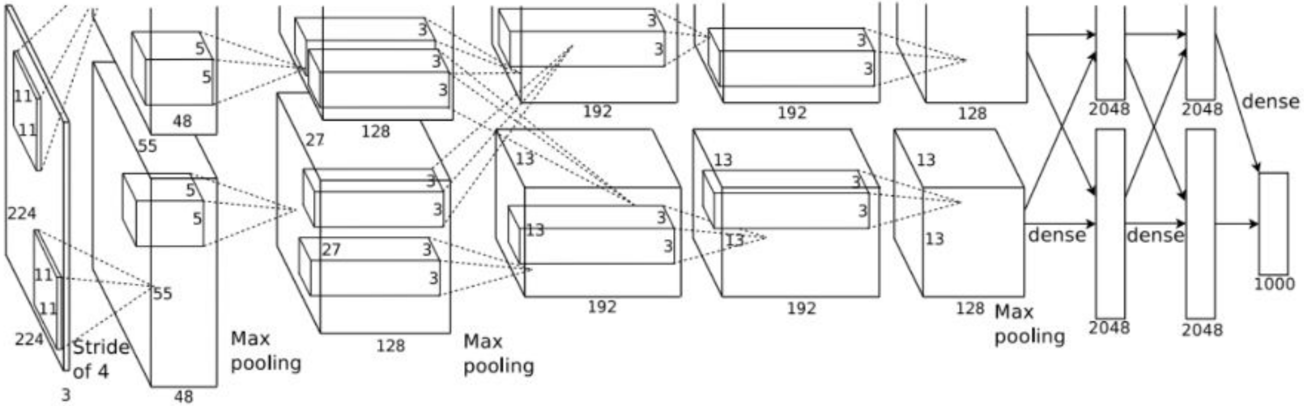


Figure 2: An illustration of the architecture of our CNN, explicitly showing the delineation of responsibilities between the two GPUs. One GPU runs the layer-parts at the top of the figure while the other runs the layer-parts at the bottom. The GPUs communicate only at certain layers. The network's input is 150,528-dimensional, and the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

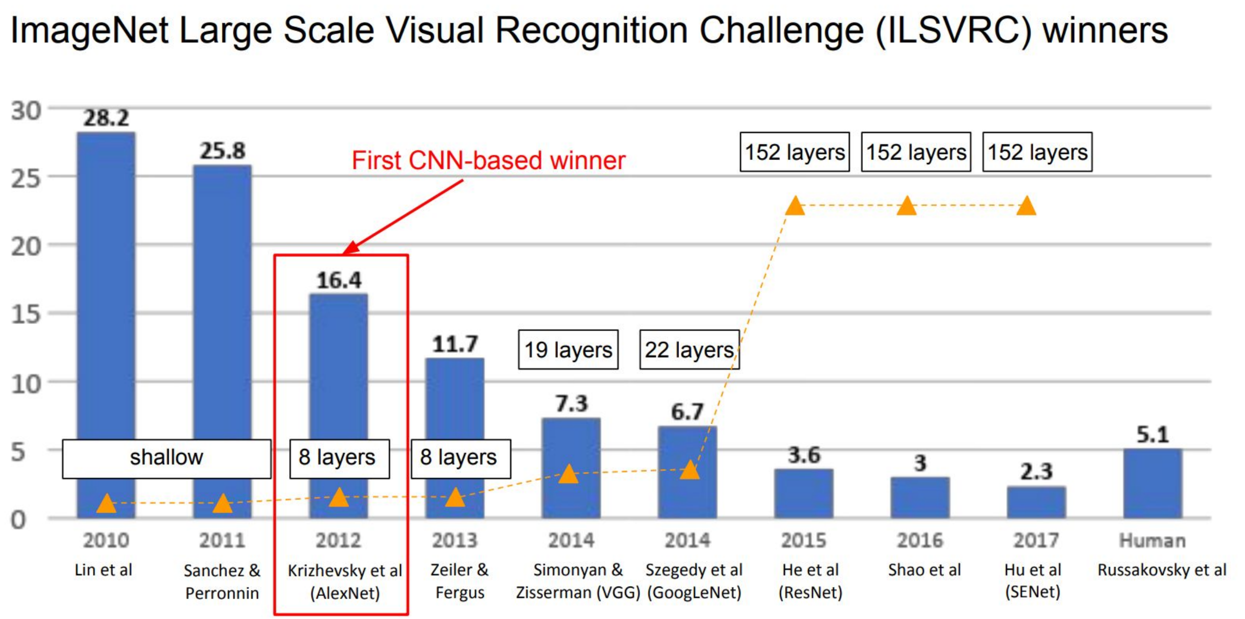
Model Zoo

- AlexNet, 2012

Details/Retrospectives:
- first use of ReLU
- used Norm layers (not common anymore)
- heavy data augmentation
- dropout 0.5
- batch size 128
- SGD Momentum 0.9
- Learning rate 1e-2, reduced by 10 manually when val accuracy plateaus
- L2 weight decay 5e-4
- 7 CNN ensemble: 18.2% -> 15.4%

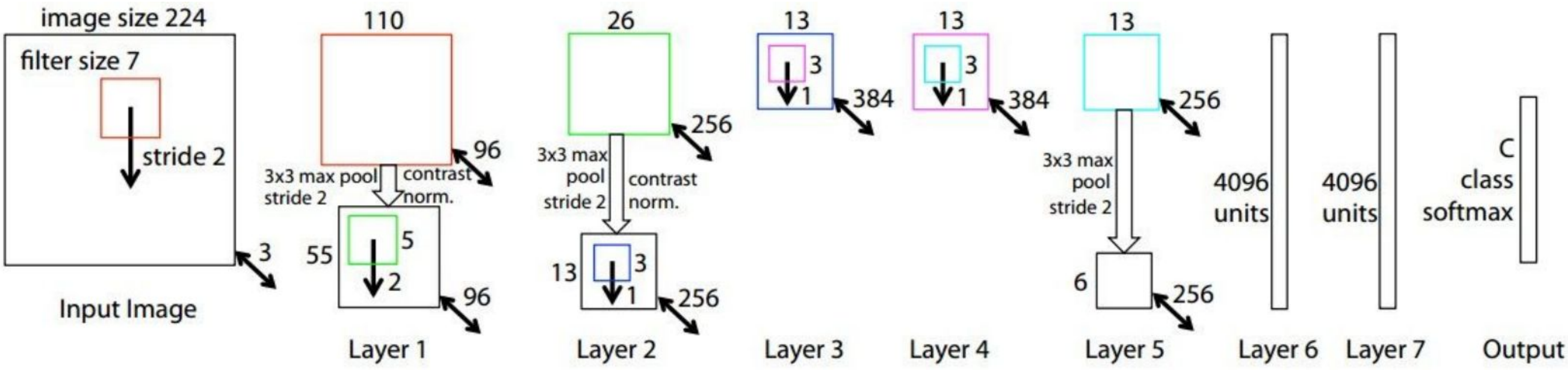
Model Zoo

- AlexNet, 2012



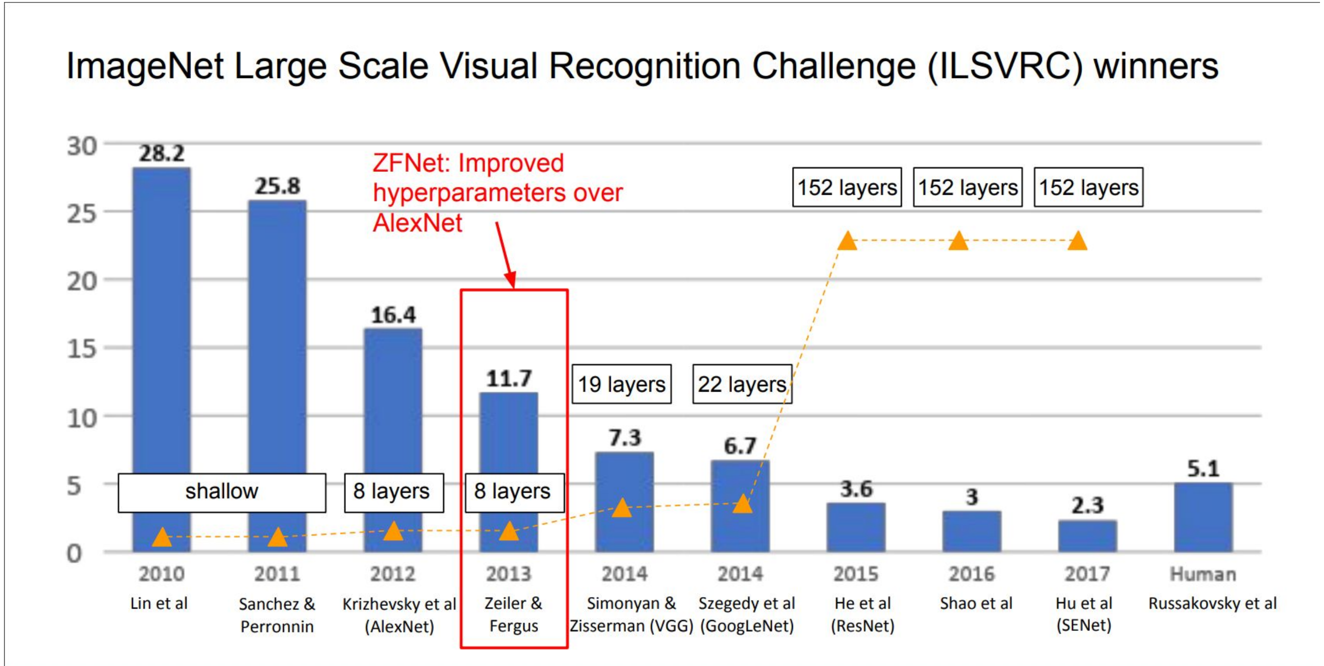
Model Zoo

- ZFNet, 2013
    - AlexNet com pequenas diferenças de hiperparâmetros e camadas mais largas



Model Zoo

- ZFNet, 2013



Model Zoo

- VGG, 2014

VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION

Karen Simonyan[*] **& Andrew Zisserman**[+]
Visual Geometry Group, Department of Engineering Science, University of Oxford
{karen,az}@robots.ox.ac.uk

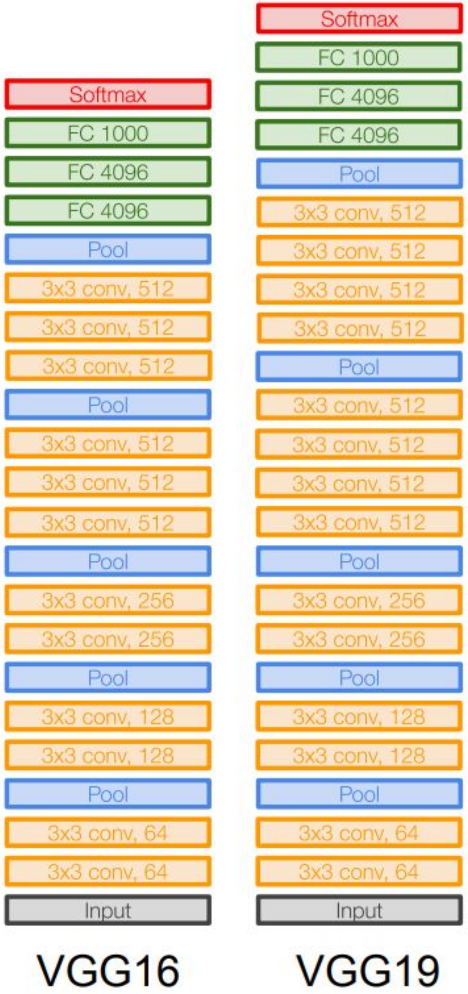
ABSTRACT

In this work we investigate the effect of the convolutional network depth on its accuracy in the large-scale image recognition setting. Our main contribution is a thorough evaluation of networks of increasing depth using an architecture with very small (3×3) convolution filters, which shows that a significant improvement on the prior-art configurations can be achieved by pushing the depth to 16–19 weight layers. These findings were the basis of our ImageNet Challenge 2014 submission, where our team secured the first and the second places in the localisation and classification tracks respectively. We also show that our representations generalise well to other datasets, where they achieve state-of-the-art results. We have made our two best-performing ConvNet models publicly available to facilitate further research on the use of deep visual representations in computer vision.

Model Zoo

- VGG, 2014
  - 16 e 19 camadas
  - só 3x3 convs

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					



Model Zoo

- GoogLeNet, 2014

Going deeper with convolutions

Christian Szegedy
Google Inc.

Wei Liu
University of North Carolina, Chapel Hill

Yangqing Jia
Google Inc.

Pierre Sermanet
Google Inc.

Scott Reed
University of Michigan

Dragomir Anguelov
Google Inc.

Dumitru Erhan
Google Inc.

Vincent Vanhoucke
Google Inc.

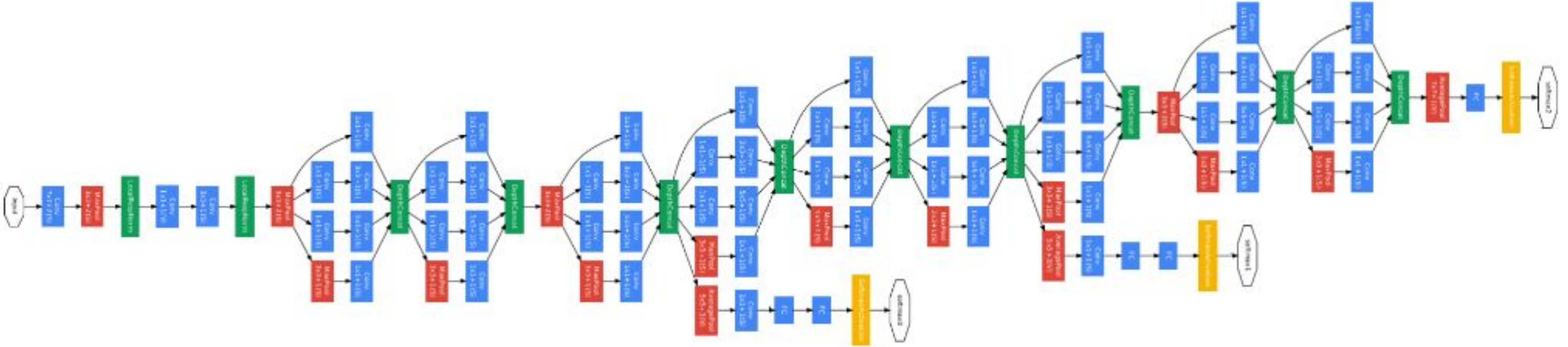
Andrew Rabinovich
Google Inc.

Abstract

We propose a deep convolutional neural network architecture codenamed Inception, which was responsible for setting the new state of the art for classification and detection in the ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC14). The main hallmark of this architecture is the improved utilization of the computing resources inside the network. This was achieved by a carefully crafted design that allows for increasing the depth and width of the network while keeping the computational budget constant. To optimize quality, the architectural decisions were based on the Hebbian principle and the intuition of multi-scale processing. One particular incarnation used in our submission for ILSVRC14 is called GoogLeNet, a 22 layers deep network, the quality of which is assessed in the context of classification and detection

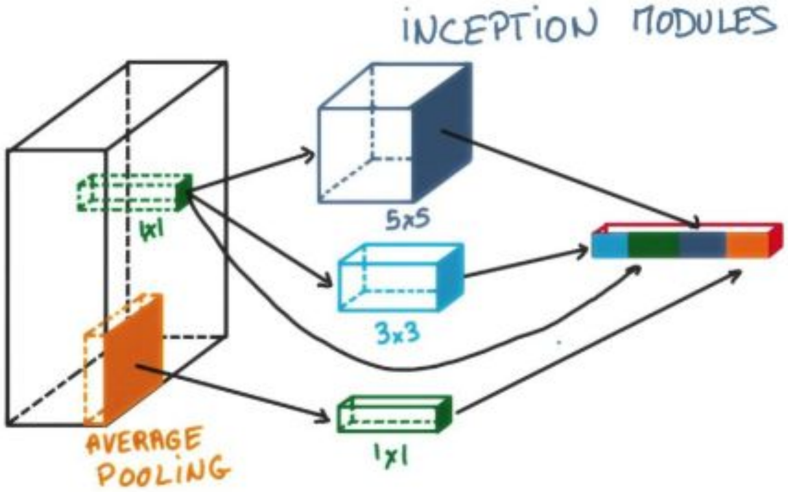
Model Zoo

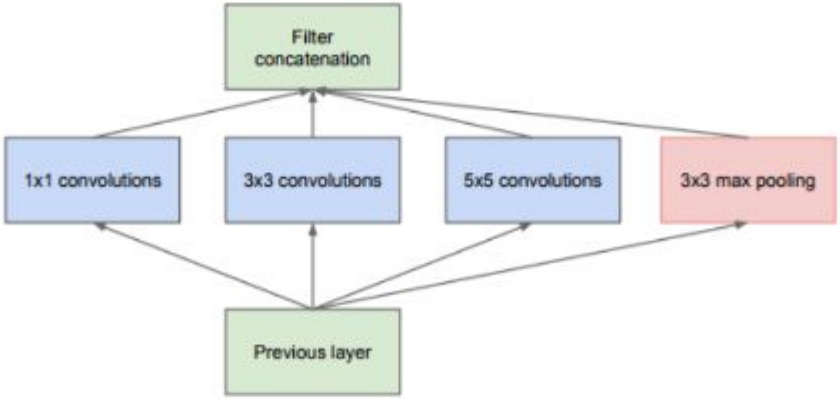
- GoogLeNet, 2014
  - Módulo Inception
  - 22 camadas



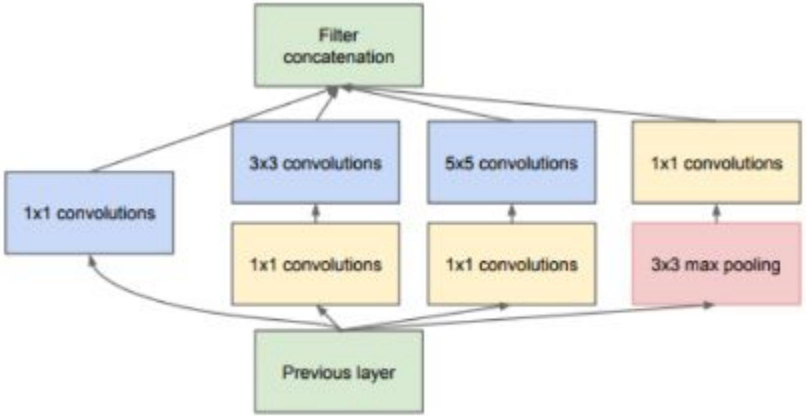
Model Zoo

- GoogLeNet, 2014
    - Módulo Inception





(a) Inception module, naïve version



(b) Inception module with dimension reductions

Model Zoo

- GoogLeNet, 2014
    - Módulo Inception

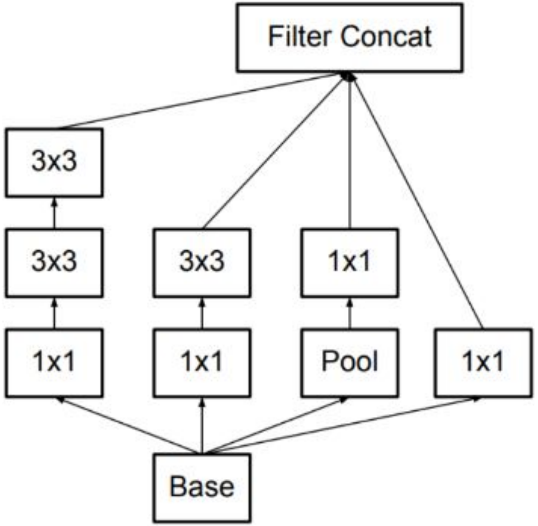


Figure 5. Inception modules where each 5×5 convolution is replaced by two 3×3 convolution, as suggested by principle 3 of Section 2.

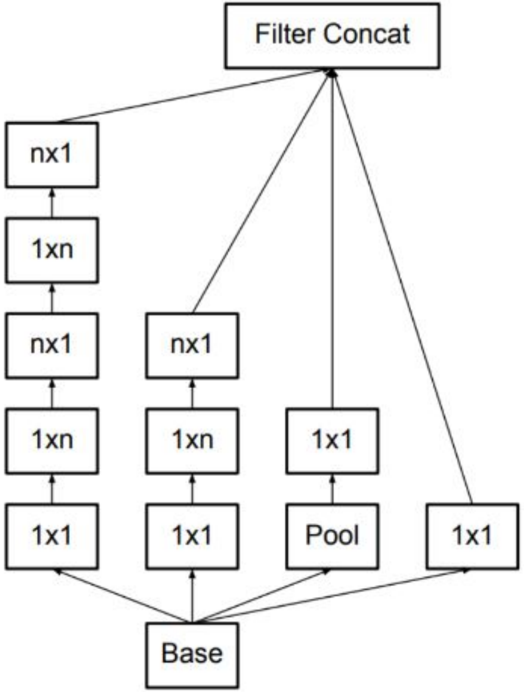
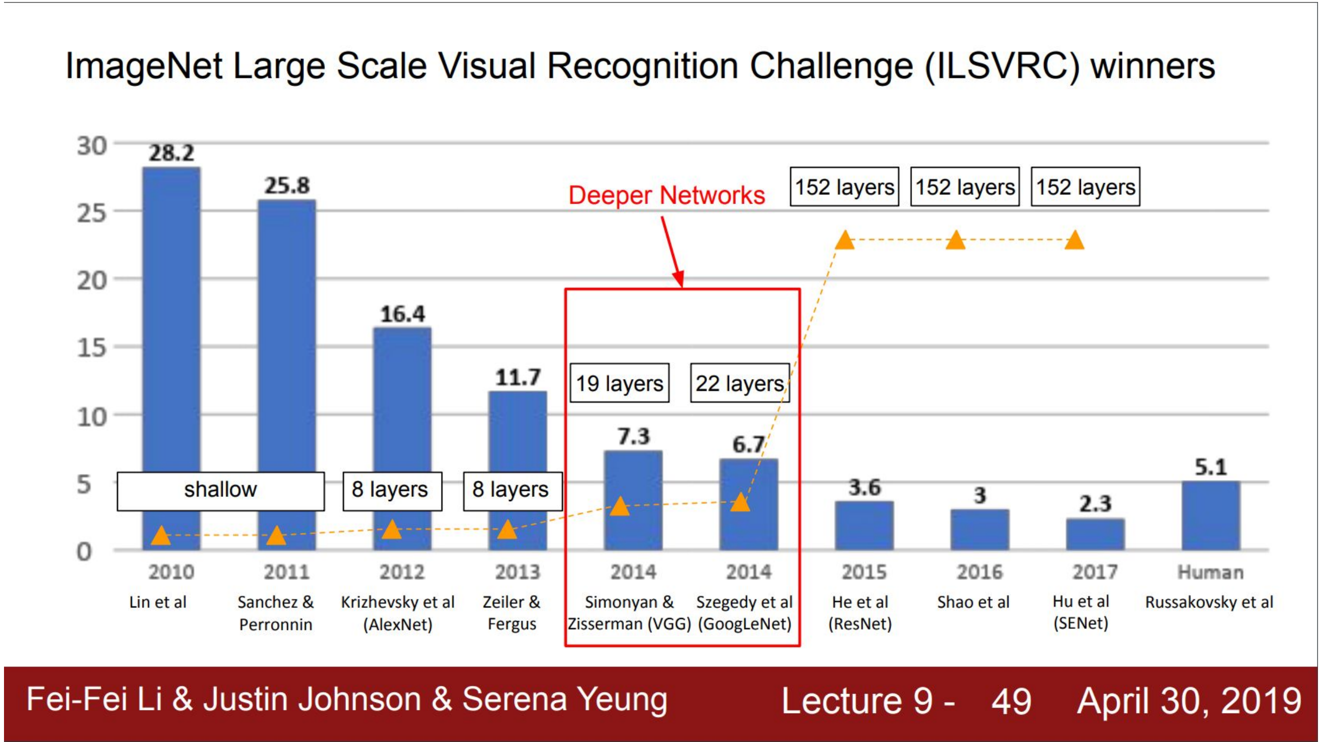


Figure 6. Inception modules after the factorization of the $n \times n$ convolutions. In our proposed architecture, we chose $n = 7$ for the 17×17 grid. (The filter sizes are picked using principle 3)

Model Zoo

- VGG e GoogLeNet, 2014



Model Zoo

- ResNet, 2015

Deep Residual Learning for Image Recognition

Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun
Microsoft Research
{kahe, v-xiangz, v-shren, jiansun}@microsoft.com

Abstract

Deeper neural networks are more difficult to train. We present a residual learning framework to ease the training of networks that are substantially deeper than those used previously. We explicitly reformulate the layers as learning residual functions with reference to the layer inputs, instead of learning unreferenced functions. We provide comprehensive empirical evidence showing that these residual networks are easier to optimize, and can gain accuracy from considerably increased depth. On the ImageNet dataset we evaluate residual nets with a depth of up to 152 layers—8× deeper than VGG nets [41] but still having lower complexity. An ensemble of these residual nets achieves 3.57% error on the ImageNet test *set. This result won the 1st place on the ILSVRC 2015 classification task. We also present analysis on CIFAR-10 with 100 and 1000 layers.*

The depth of representations is of central importance for many visual recognition tasks. Solely due to our extremely deep representations, we obtain a 28% relative improvement on the COCO object detection dataset. Deep residual nets are foundations of our submissions to ILSVRC & COCO 2015 competitions[1], where we also won the 1st places on the tasks of ImageNet detection, ImageNet localization, COCO detection, and COCO segmentation.

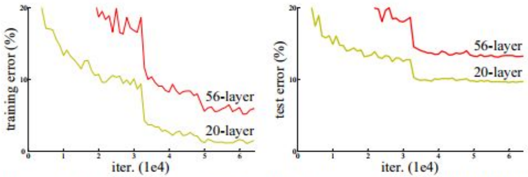


Figure 1. Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer "plain" networks. The deeper network has higher training error, and thus test error. Similar phenomena on ImageNet is presented in Fig. 4.

greatly benefited from very deep models.

Driven by the significance of depth, a question arises: *Is learning better networks as easy as stacking more layers?* An obstacle to answering this question was the notorious problem of vanishing/exploding gradients [1, 9], which hamper convergence from the beginning. This problem, however, has been largely addressed by normalized initialization [23, 9, 37, 13] and intermediate normalization layers [16], which enable networks with tens of layers to start converging for stochastic gradient descent (SGD) with backpropagation [22].

When deeper networks are able to start converging, a *degradation* problem has been exposed: with the network

Model Zoo

- ResNet, 2015

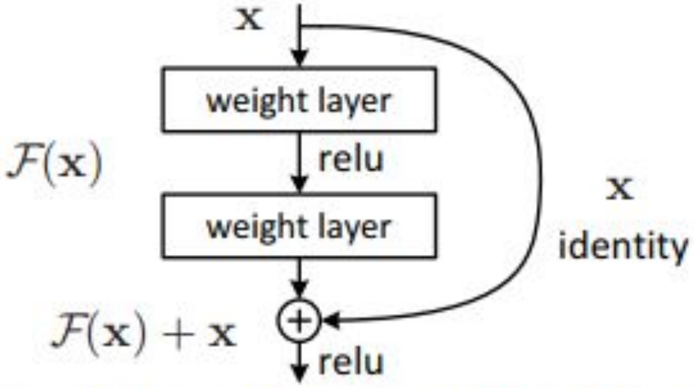
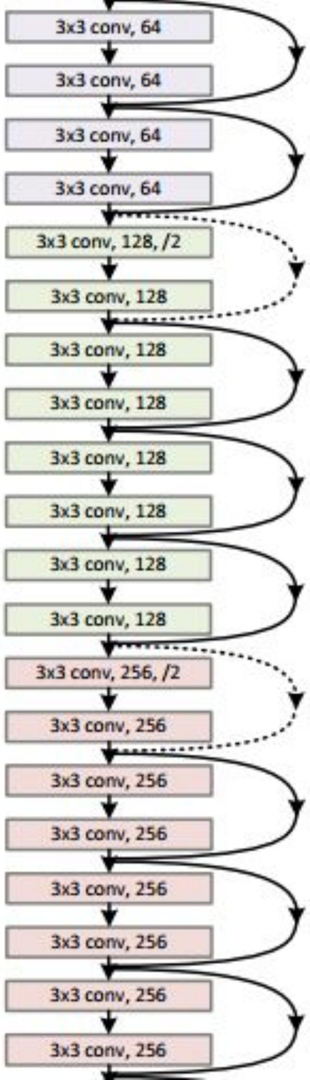
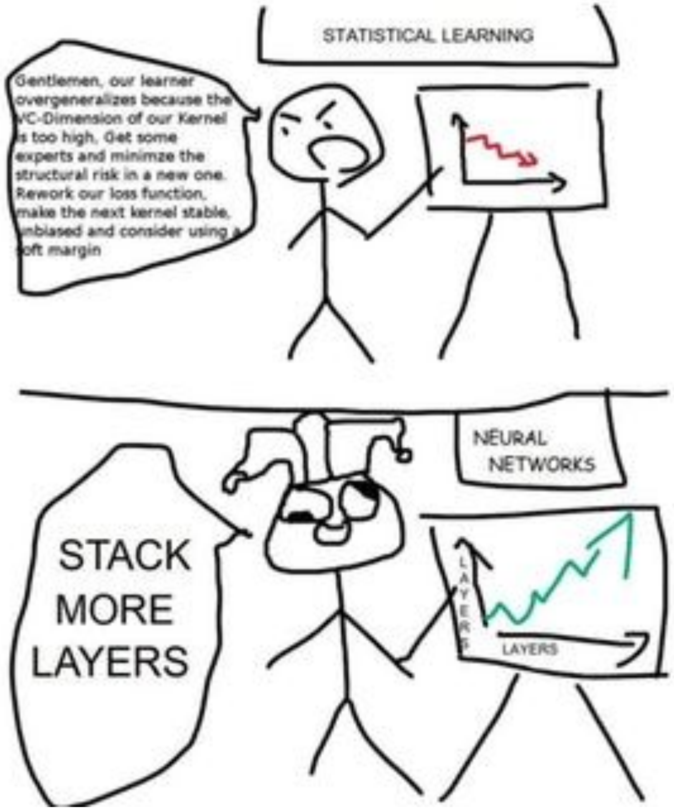


Figure 2. Residual learning: a building block.



Model Zoo

- ResNet, 2015



Model Zoo

- ResNet, 2015

Model Zoo

- ResNet, 2015

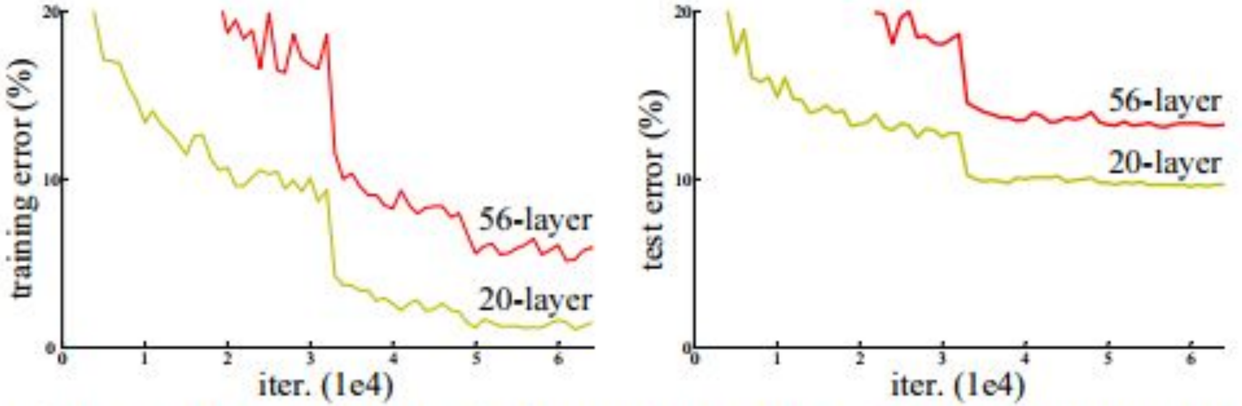
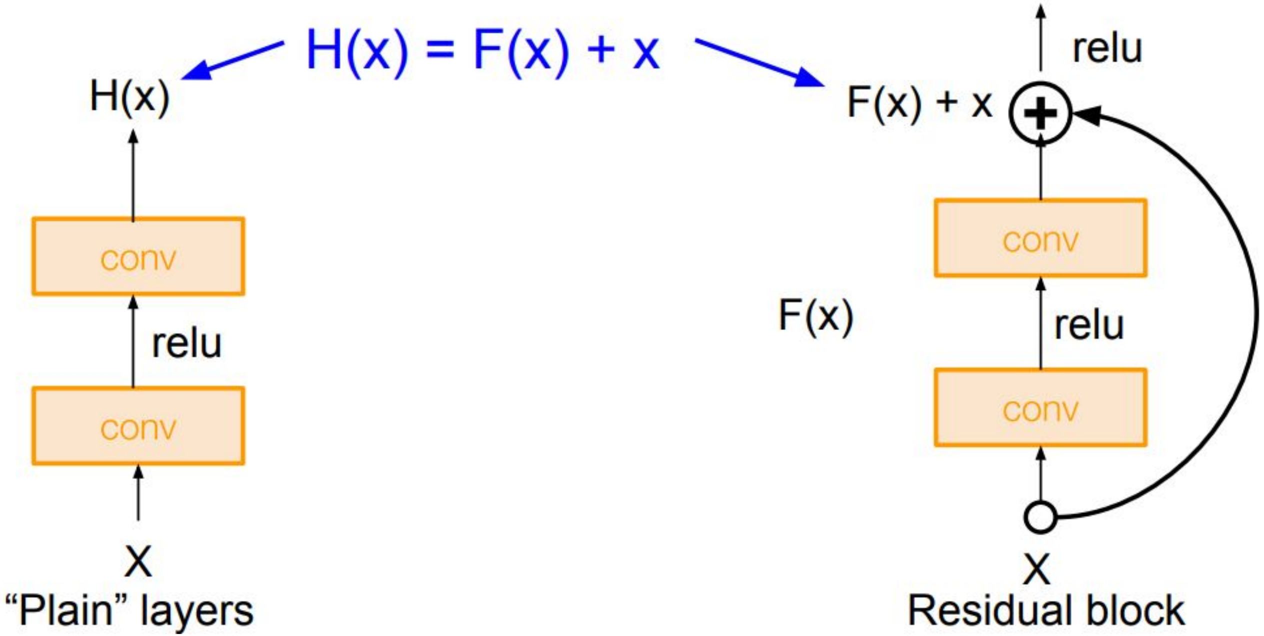


Figure 1. Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer "plain" networks. The deeper network has higher training error, and thus test error. Similar phenomena on ImageNet is presented in Fig. 4.

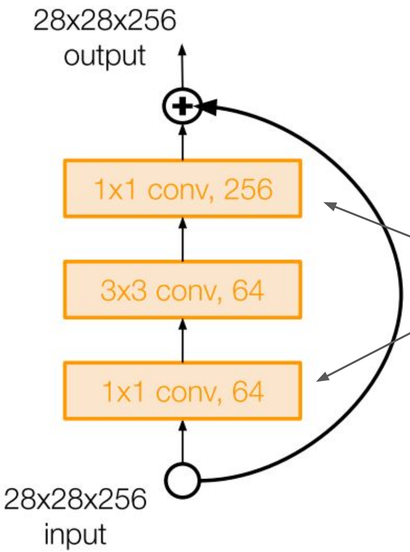
Model Zoo

- ResNet, 2015

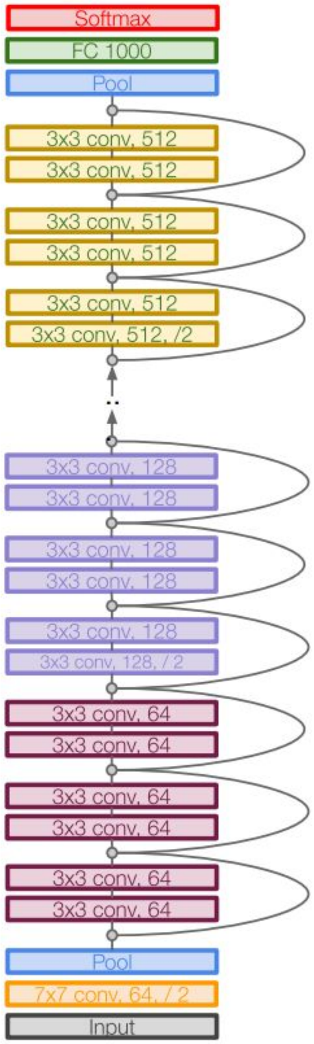


Model Zoo

- ResNet, 2015
  - 34, 50, 101 e 152 camadas



Camadas bottleneck (ResNet-50+)



Model Zoo

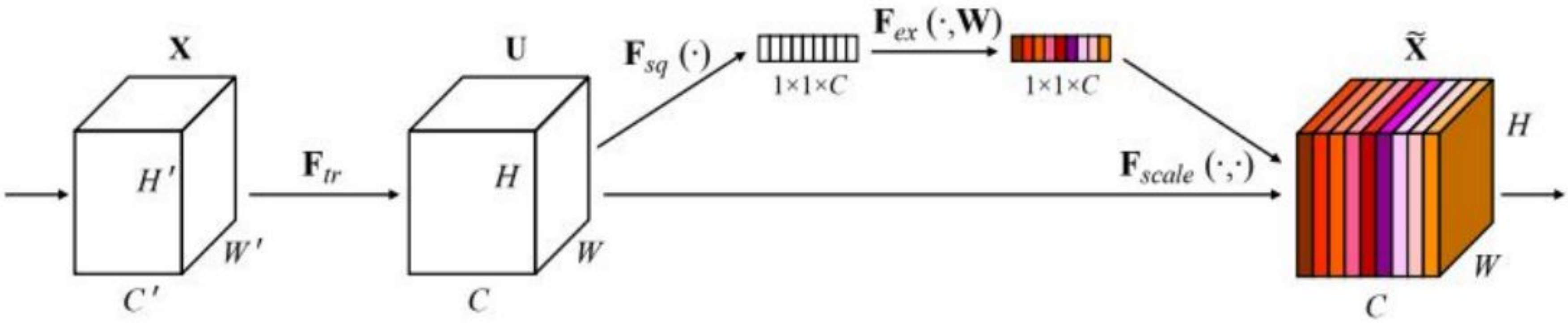
- ResNet, 2015

Training ResNet in practice:

- Batch Normalization after every CONV layer
- Xavier 2/ initialization from He et al.
- SGD + Momentum (0.9)
- Learning rate: 0.1, divided by 10 when validation error plateaus
- Mini-batch size 256
- Weight decay of 1e-5
- No dropout used

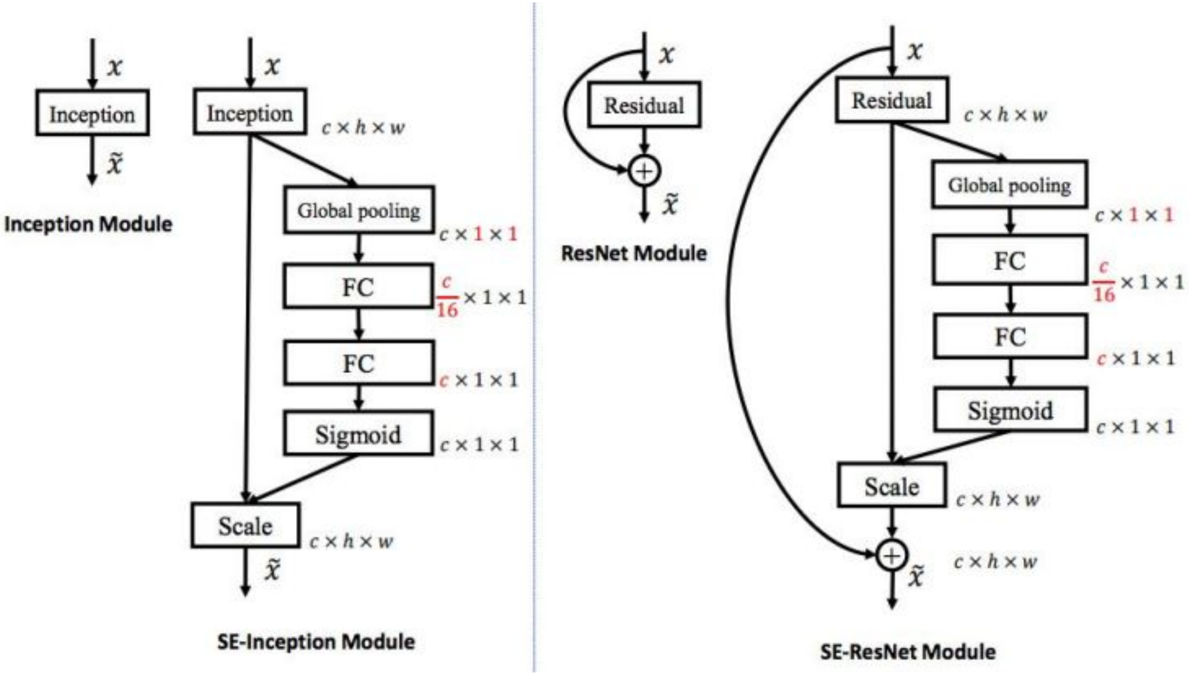
Model Zoo

- Squeeze-and-Excitation Networks (SENet), 2017



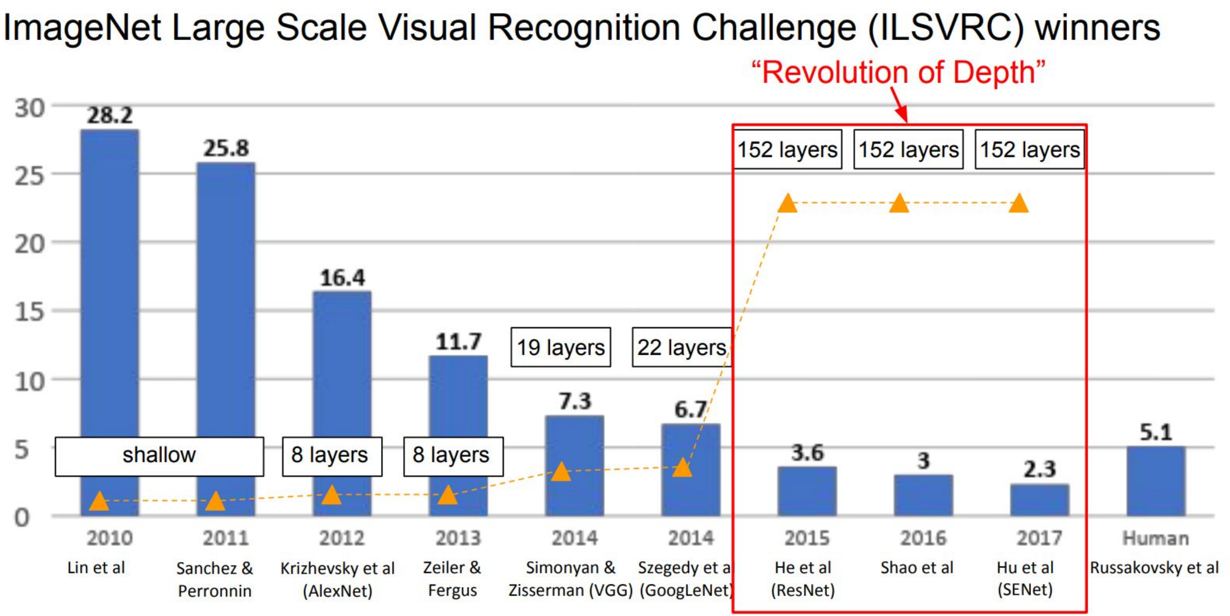
Model Zoo

- Squeeze-and-Excitation Networks (SENet), 2017



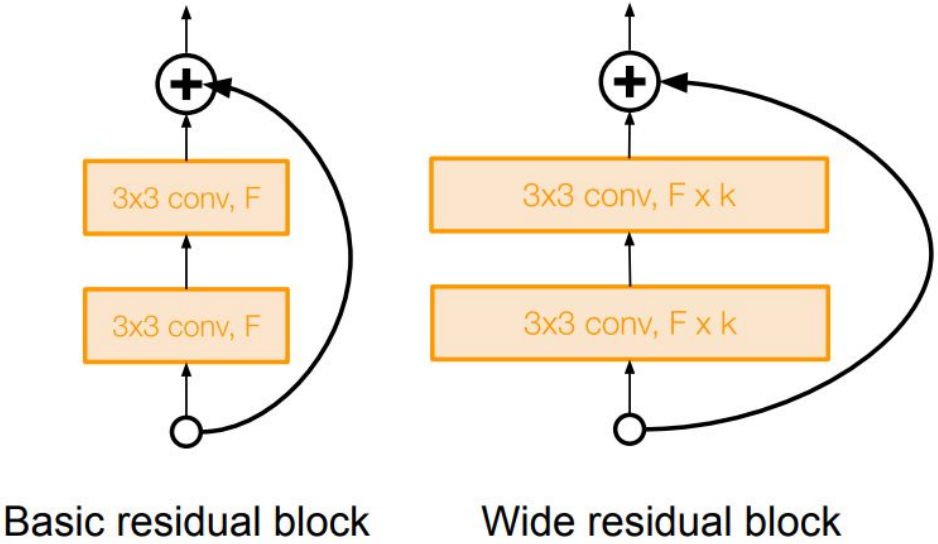
Model Zoo

- ResNet, 2015 e SENet, 2017



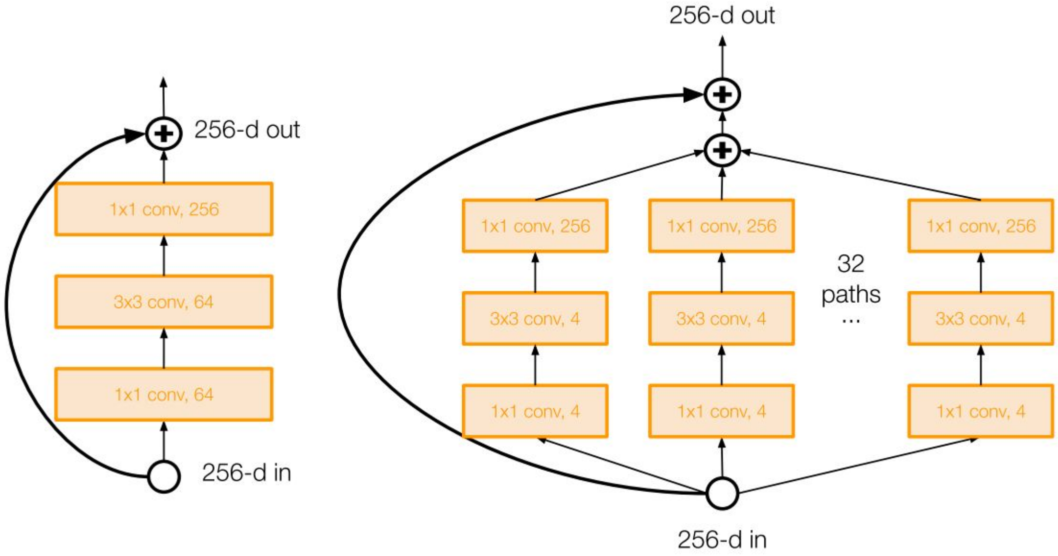
Model Zoo

- Redes menos famosas
    - Wide ResNet, 2016



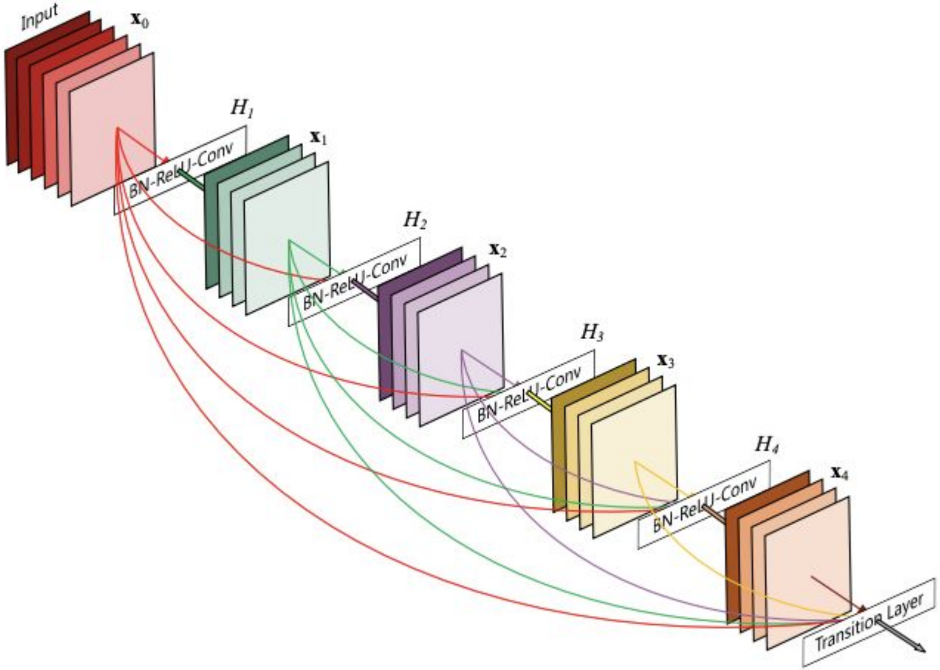
Model Zoo

- Redes menos famosas
  - Wide ResNet, 2016
  - ResNext, 2016



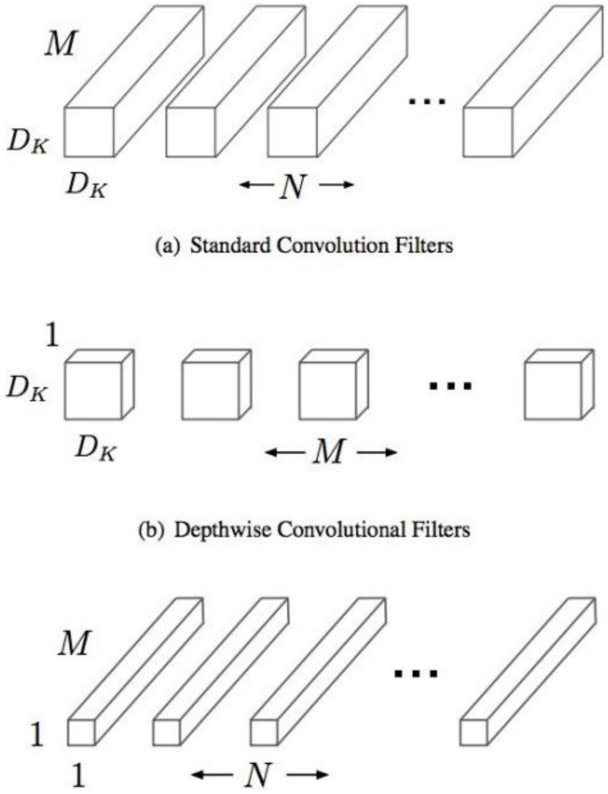
Model Zoo

- Redes menos famosas
    - Wide ResNet, 2016
    - ResNext, 2016
    - DenseNet, 2017



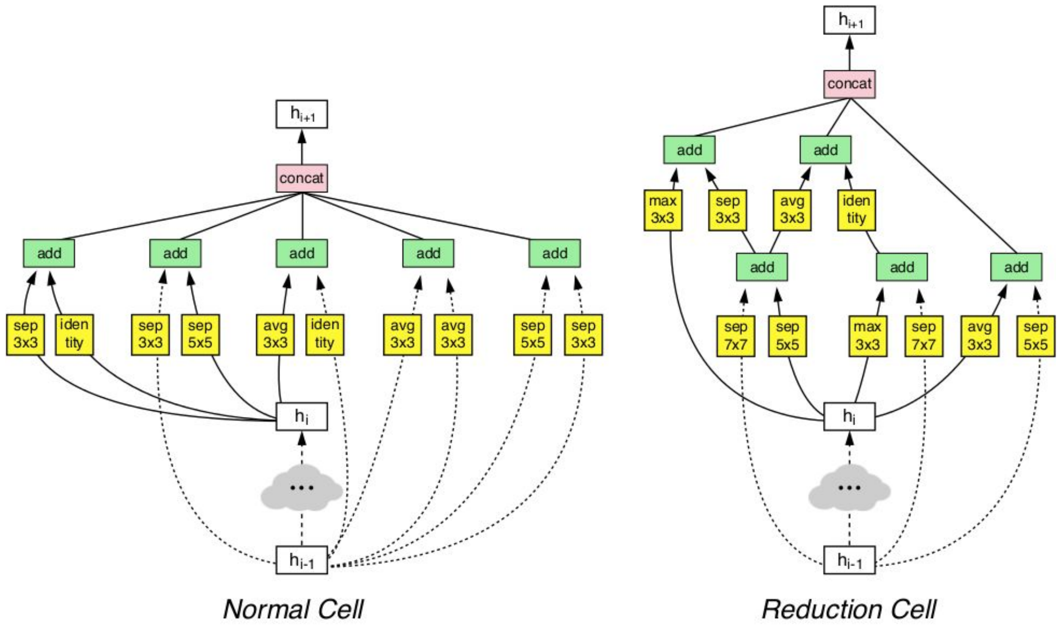
Model Zoo

- Redes menos famosas
    - Wide ResNet, 2016
    - ResNext, 2016
    - DenseNet, 2017
    - MobileNets, 2017



Model Zoo

- Redes menos famosas
  - Wide ResNet, 2016
  - ResNext, 2016
  - DenseNet, 2017
  - MobileNets, 2017
  - NASNet, 2016-17



No próximo episódio...

- Hands-on