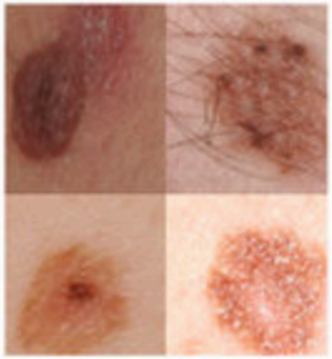
1 Image Classification

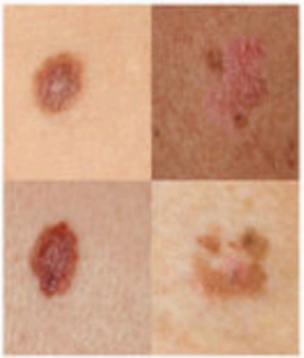
BVM 2018 Tutorial: Advanced Deep Learning Methods

Jakob Wasserthal, Division of Medical Image Computing

Classification of skin cancer

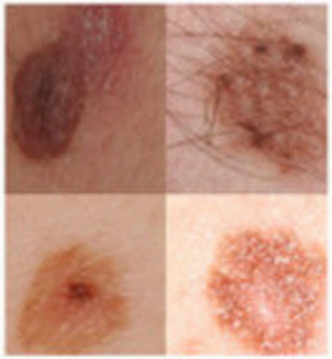


VS



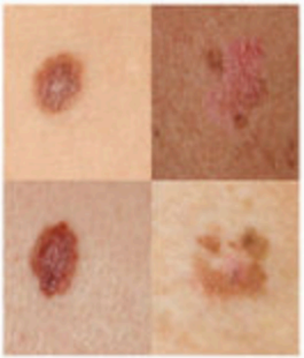
Esteva et al., Dermatologist-level classification of skin cancer with deep neural networks, Nature, 2017

Classification of skin cancer



benign

vs



malignant

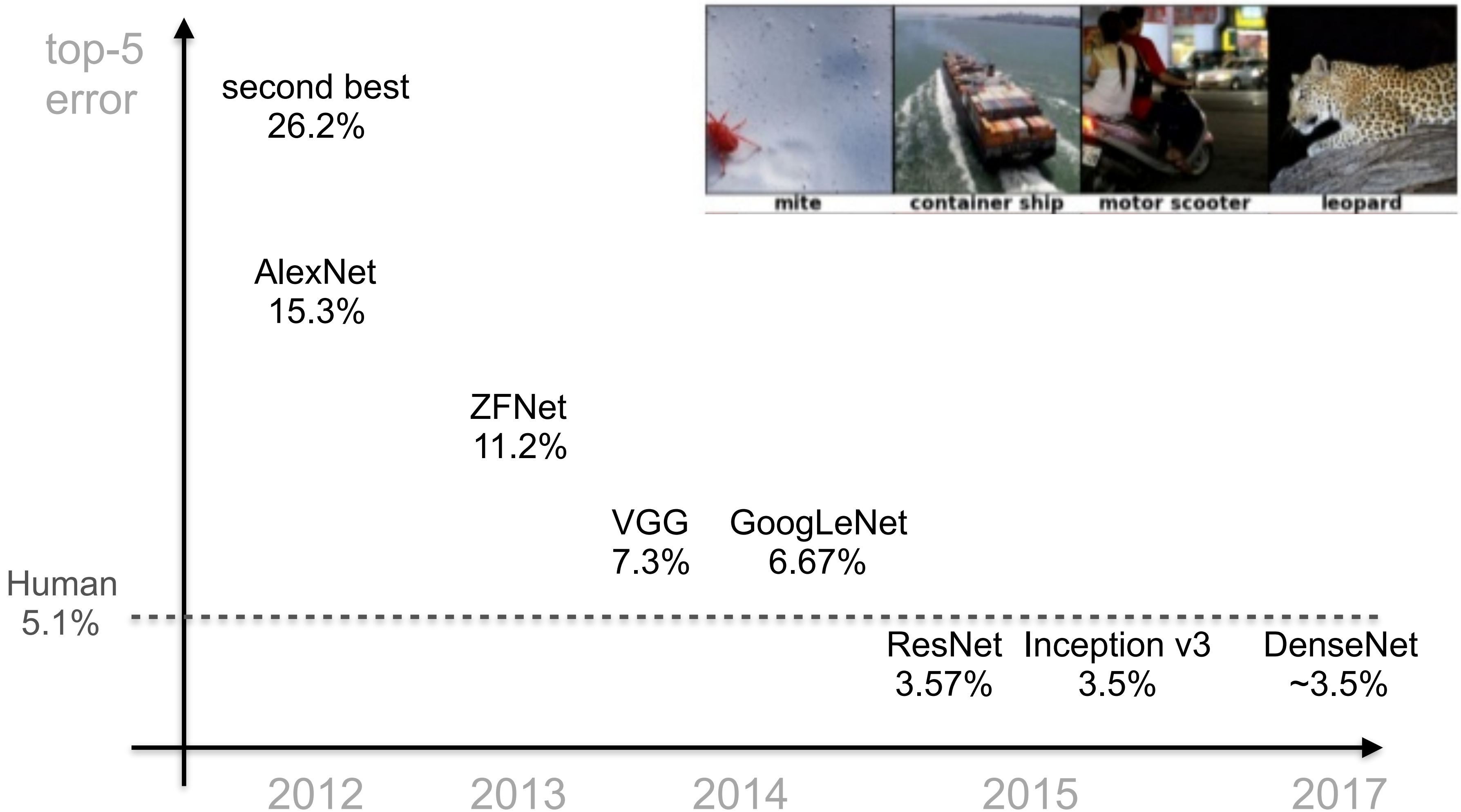
Esteva et al., Dermatologist-level classification of skin cancer with deep neural networks, Nature, 2017

Classification

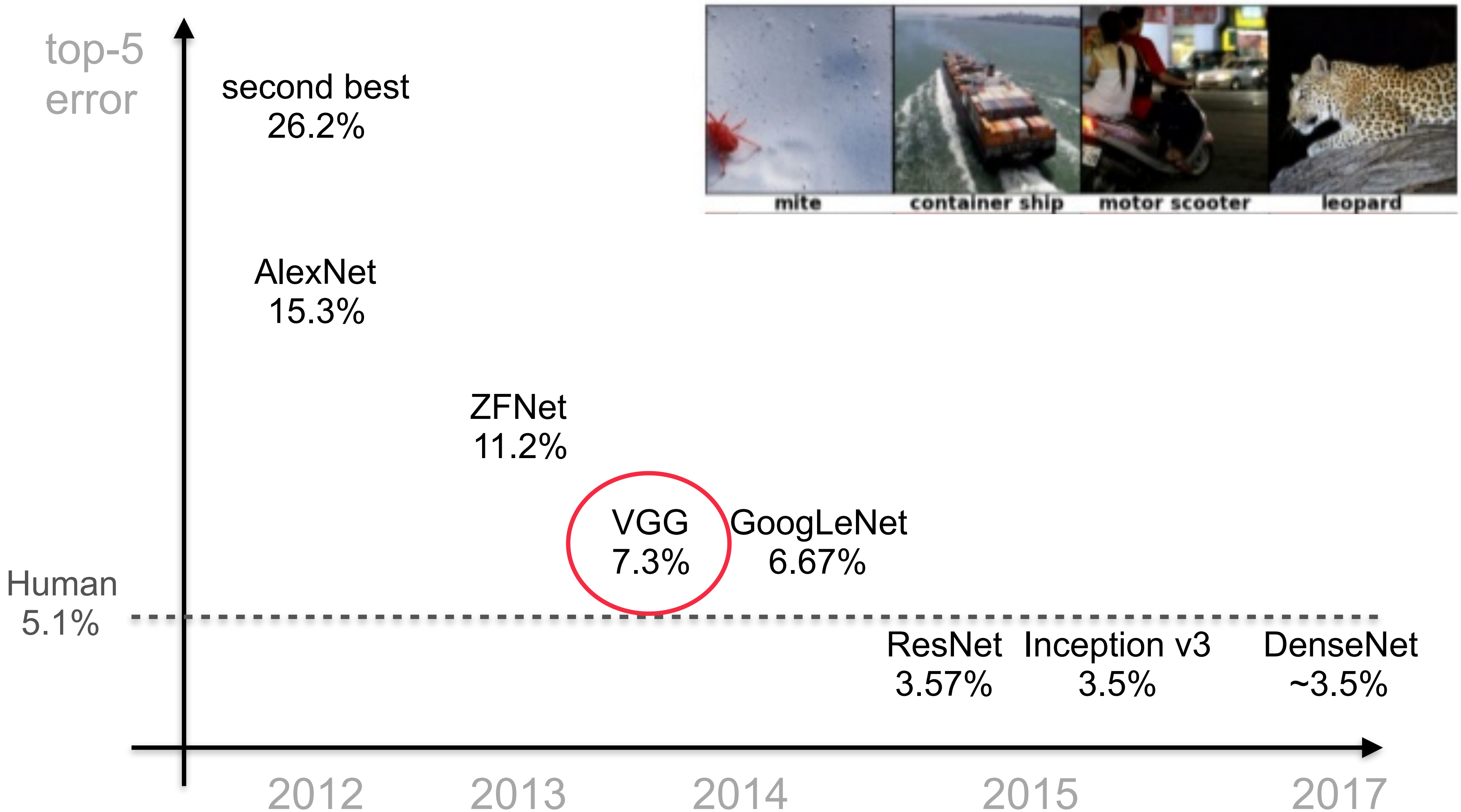
0.98 **p(benign)**

0.02 p(malignant)

ILSVRC challenge / ImageNet

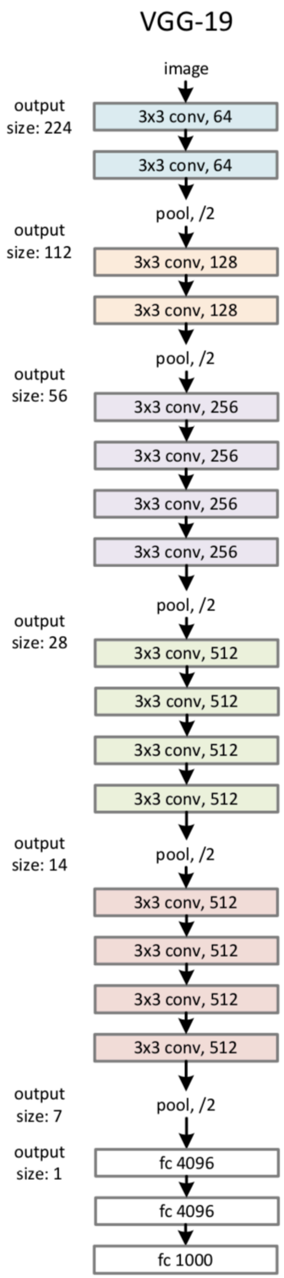


ILSVRC challenge / ImageNet



VGG

- simple structure
- 160M parameters



Simonyan et al.,Very deep convolutional networks for large-scale image recognition, arXiv, *2014*

He et al., Deep Residual Learning for Image Recognition, arXiv, 2015

ILSVRC challenge / ImageNet

top-5 error

second best
26.2%

AlexNet
15.3%

ZFNet
11.2%

VGG
7.3%

GoogLeNet
6.67%

Human
5.1%

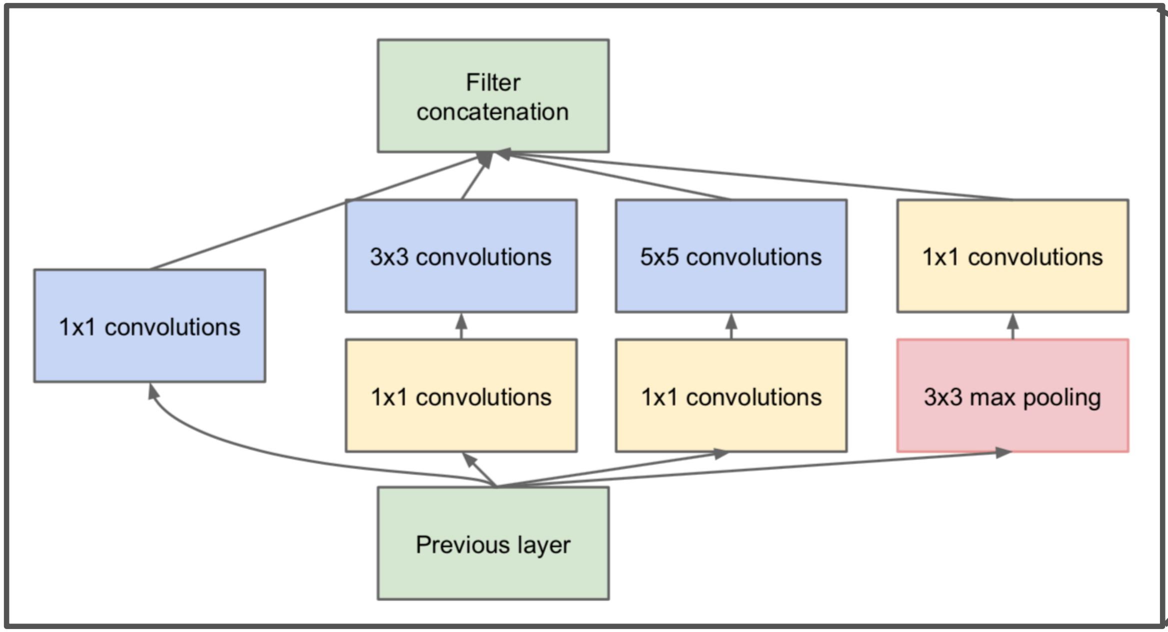
ResNet
3.57%

Inception v3
3.5%

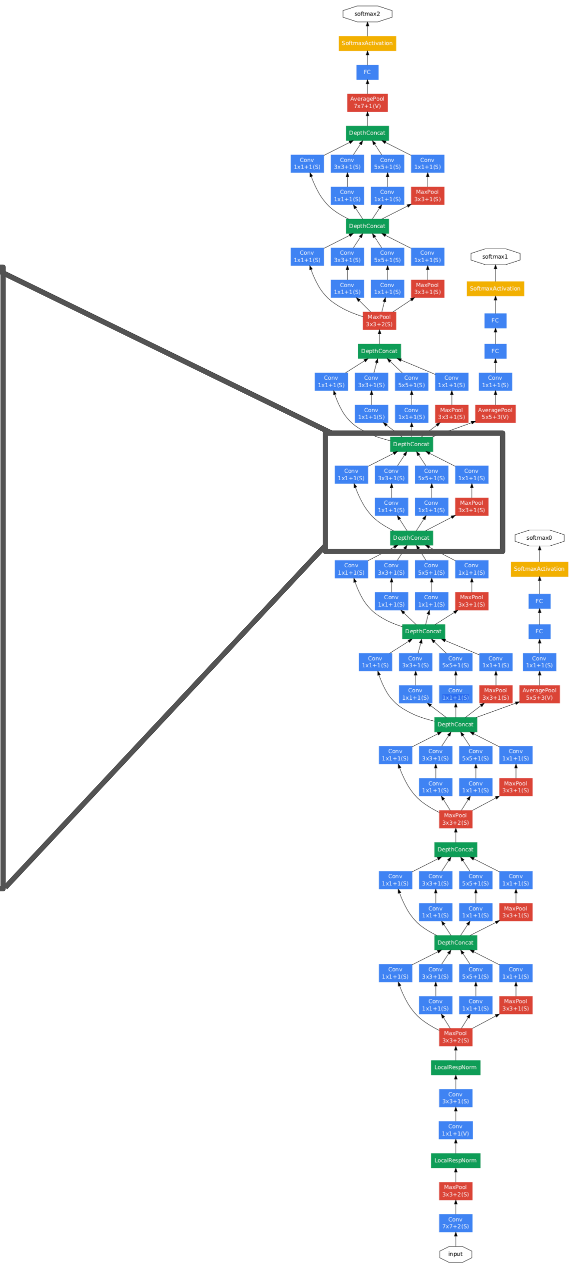
DenseNet
~3.5%

2012 2013 2014 2015 2017

GoogLeNet

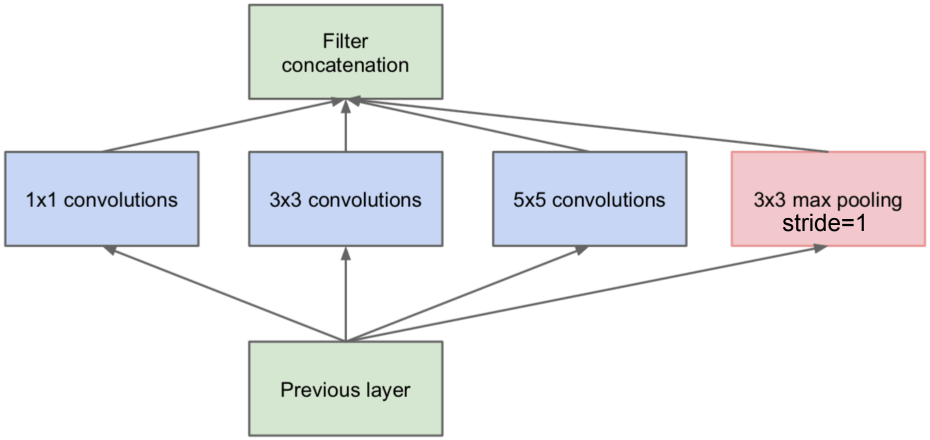


Inception module



Szegedy et al., Going Deeper with Convolutions, arXiv, 2014

GoogLeNet

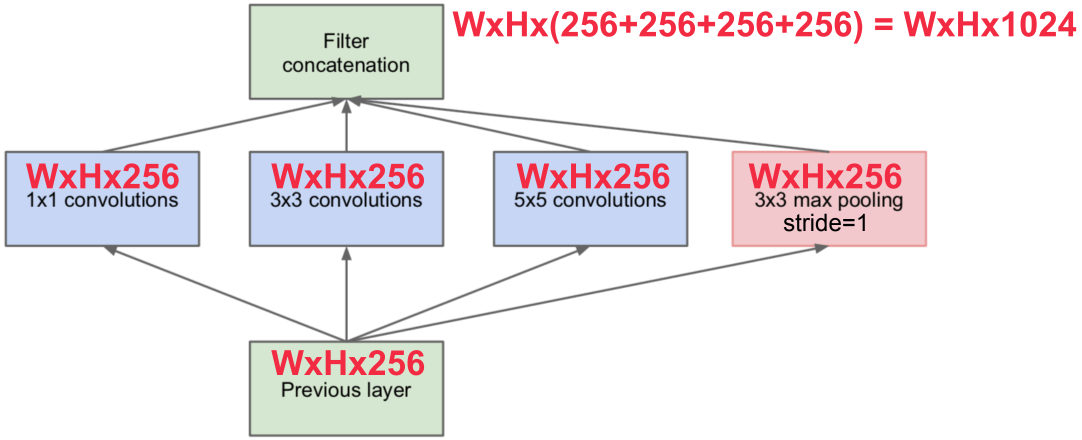


(a) Inception module, naïve version

Szegedy et al., Going Deeper with Convolutions, arXiv, 2014

[Width x Height x Nr of Filters]

GoogLeNet

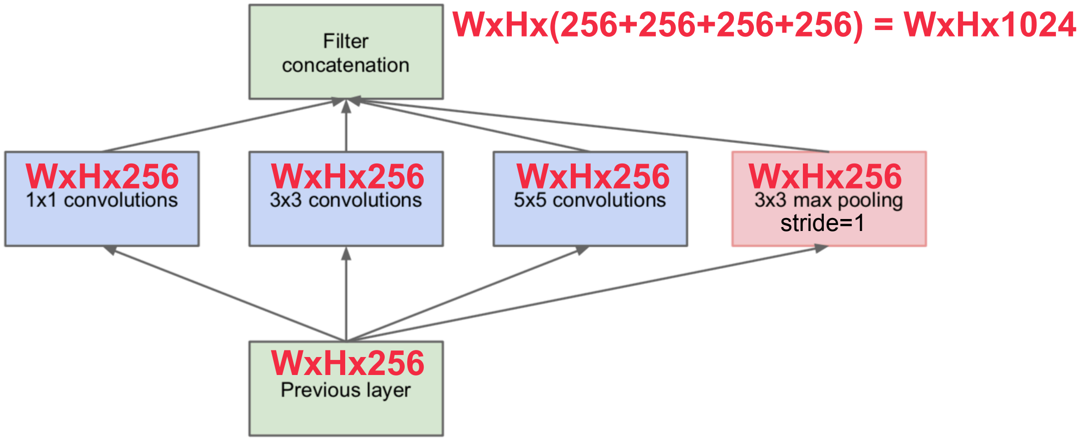


(a) Inception module, naïve version

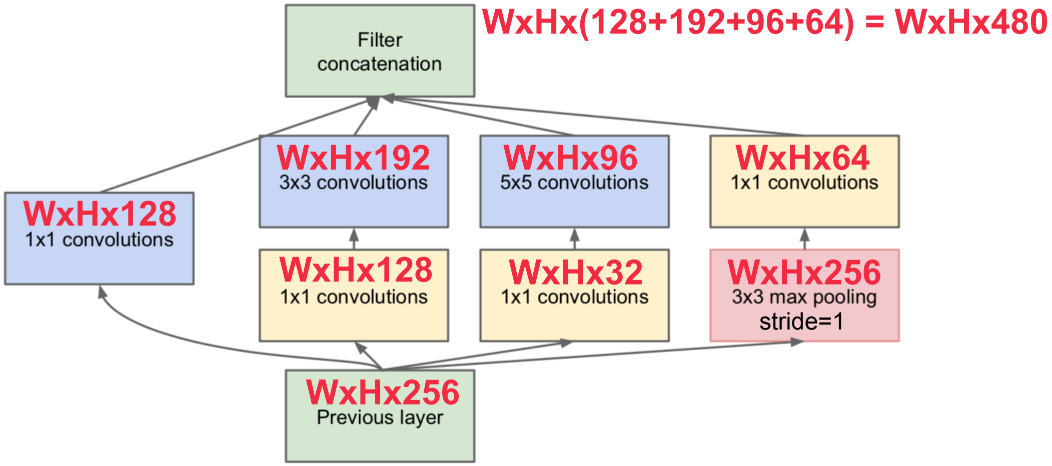
Szegedy et al., Going Deeper with Convolutions, arXiv, 2014

[Width x Height x Nr of Filters]

GoogLeNet



(a) Inception module, naïve version



(b) Inception module with dimensionality reduction

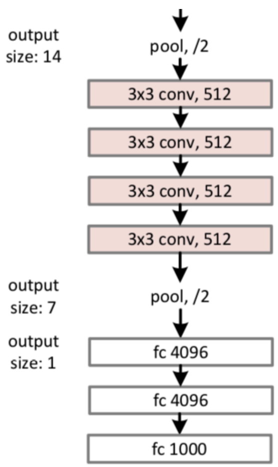
Szegedy et al., 2014

GoogLeNet

VGG

Data dimensions

#Parameters



14x14x512

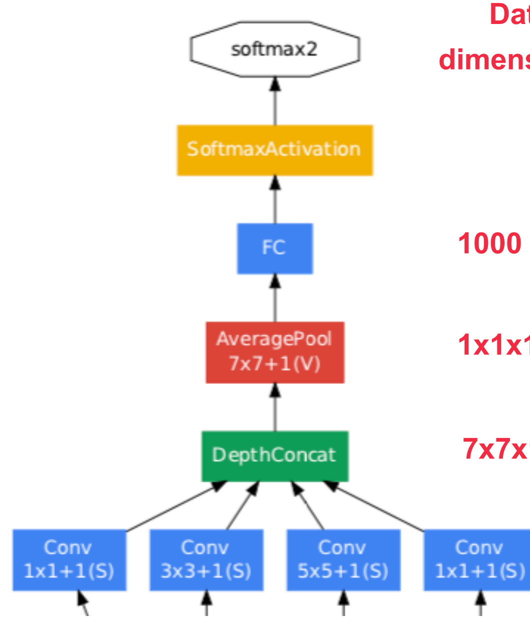
7x 7x512=25.088

4094 **25088*4094=102M**

GoogLeNet

Data dimensions

#Parameters



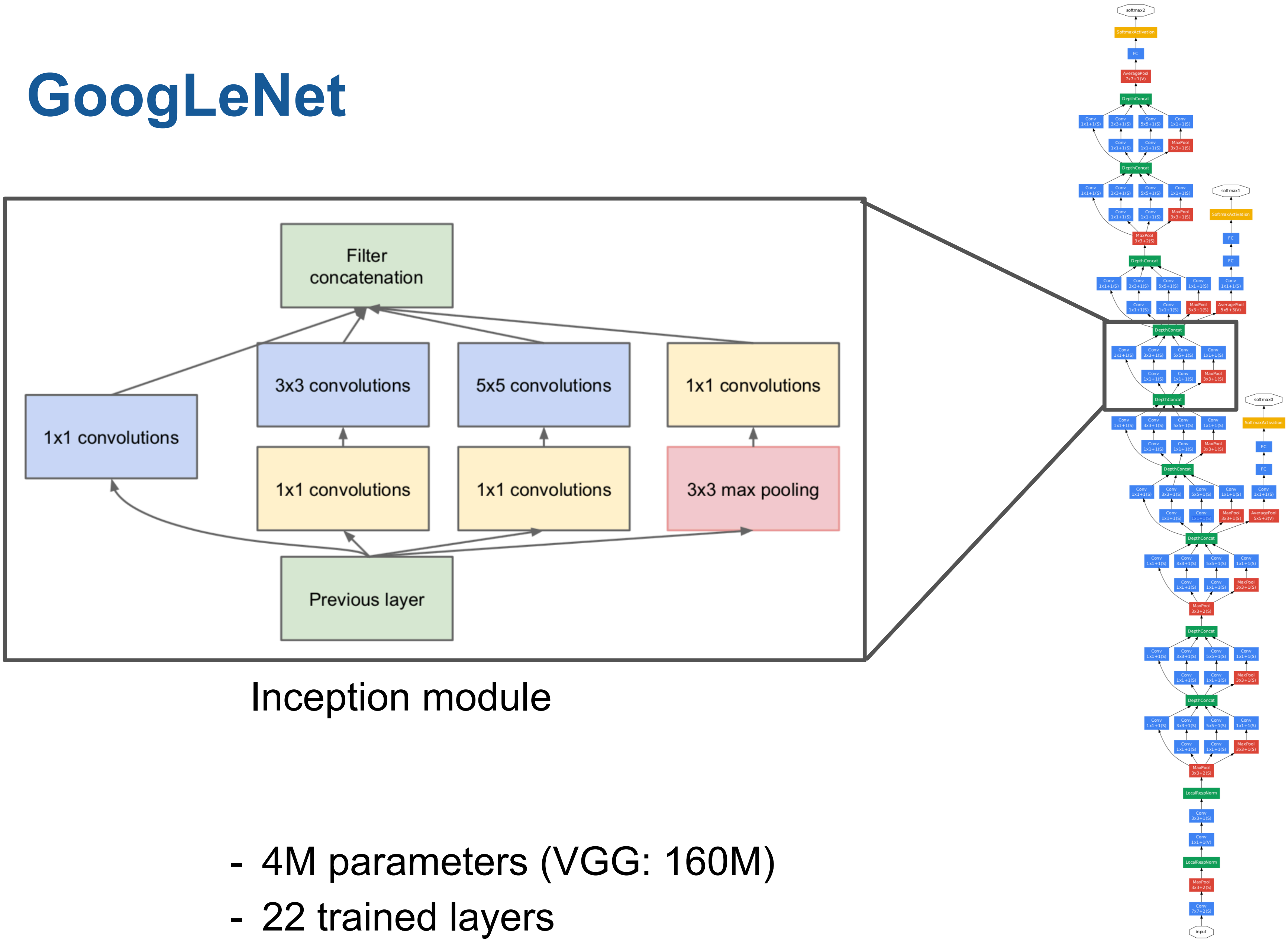
1000 **1000*1024=1M**

1x1x1024

7x7x1024

Szegedy et al., Going Deeper with Convolutions, arXiv, 2014

GoogLeNet

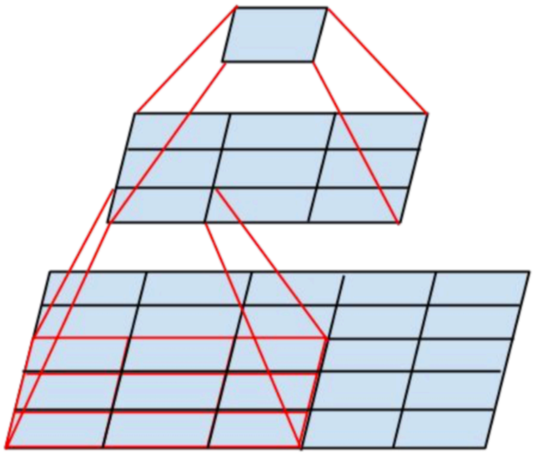


Inception module

- 4M parameters (VGG: 160M)
- 22 trained layers

Szegedy et al., Going Deeper with Convolutions, arXiv, 2014

Inception v3 - Improvement 1



Parameters:

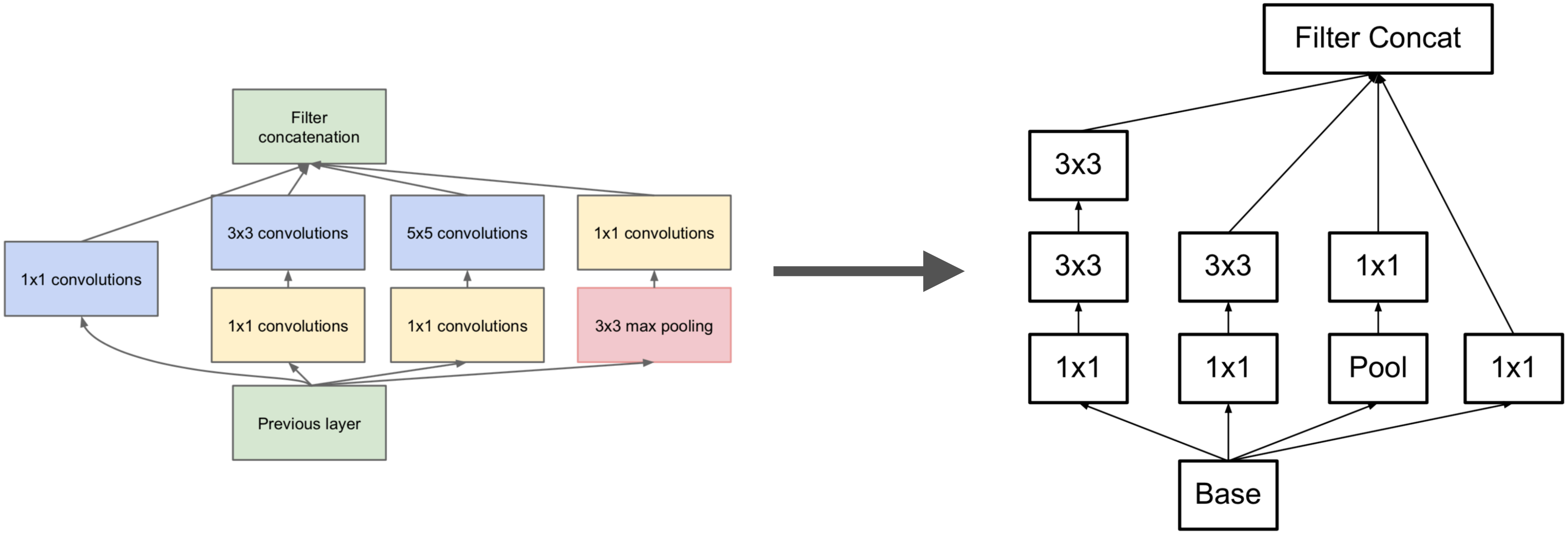
5x5-convolution: 5*5=25

2* 3x3-convolution: 2* (3*3)=18

=> ~30% less parameters and computations

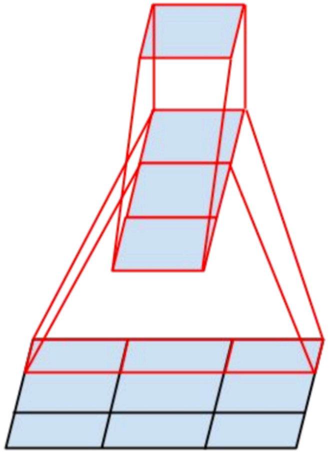
Szegedy et al., Rethinking the Inception Architecture for Computer Vision, arXiv, 2015

Inception v3 - Improvement 1



Szegedy et al., Rethinking the Inception Architecture for Computer Vision, arXiv, 2015

Inception v3 - Improvement 2



Parameters:

3x3-convolution: 3*3=9

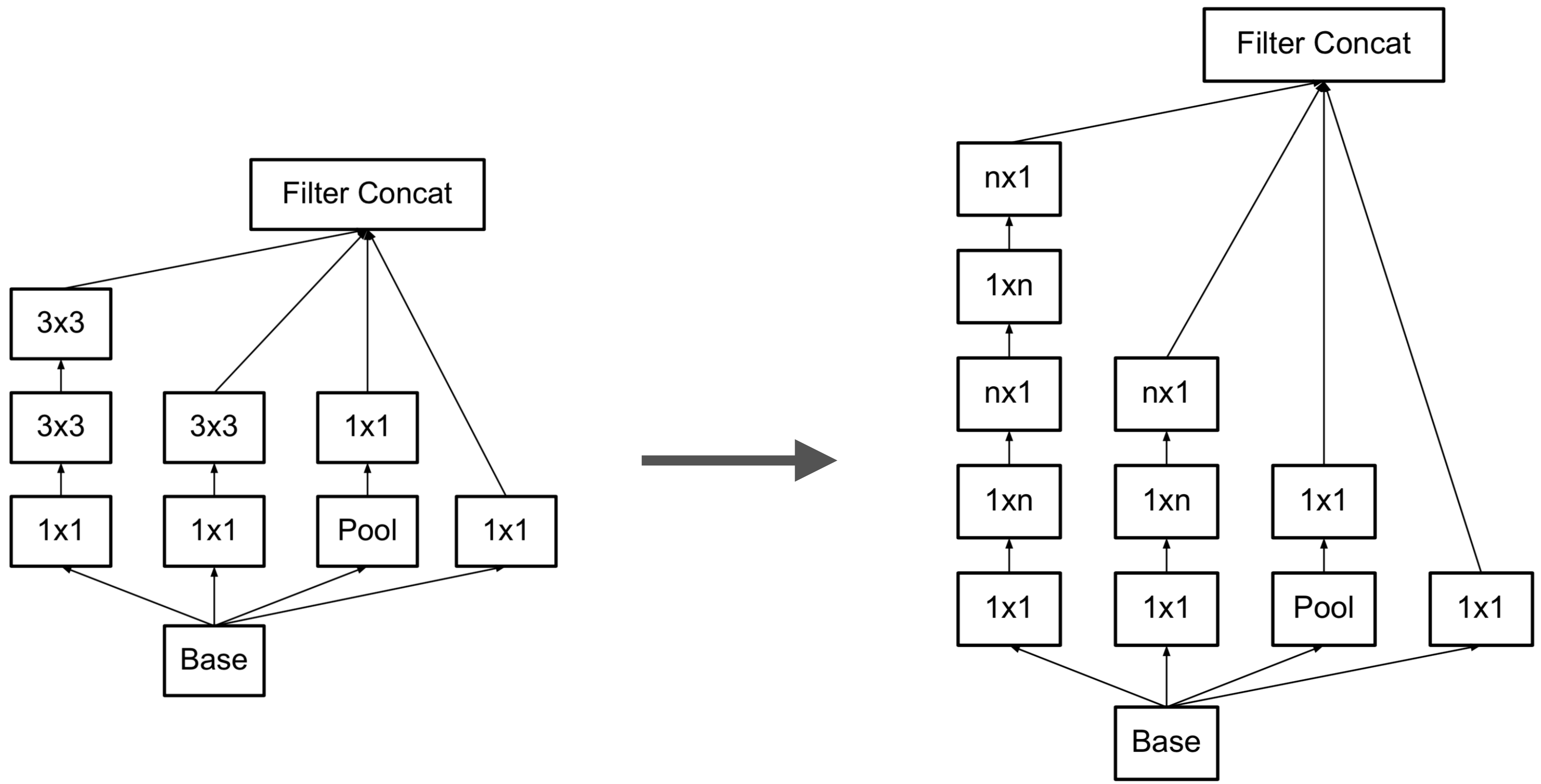
2* 1x3-convolution: 2* (1*3)=6

=> ~33% less parameters and computations

Szegedy et al., Rethinking the Inception Architecture for Computer Vision, arXiv, 2015

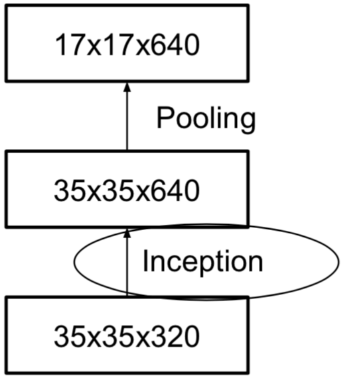
dkfz.

Inception v3 - Improvement 2

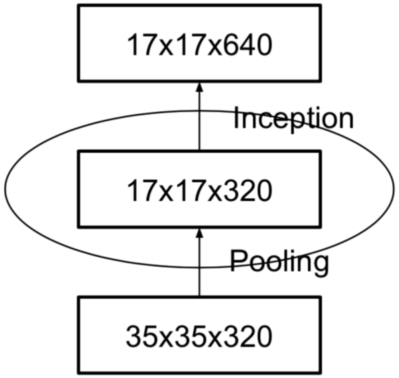


Szegedy et al., Rethinking the Inception Architecture for Computer Vision, arXiv, 2015

Inception v3 - Improvement 3



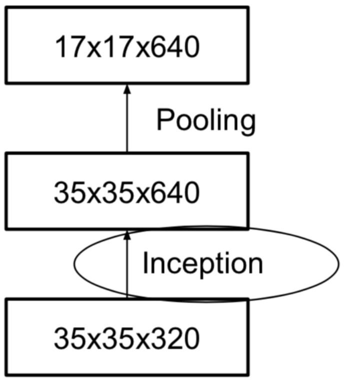
3x more computations



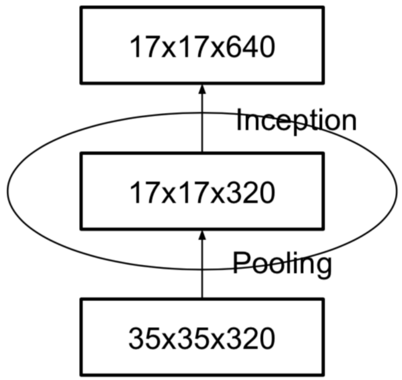
Representational
bottleneck

Szegedy et al., Rethinking the Inception Architecture for Computer Vision, arXiv, 2015

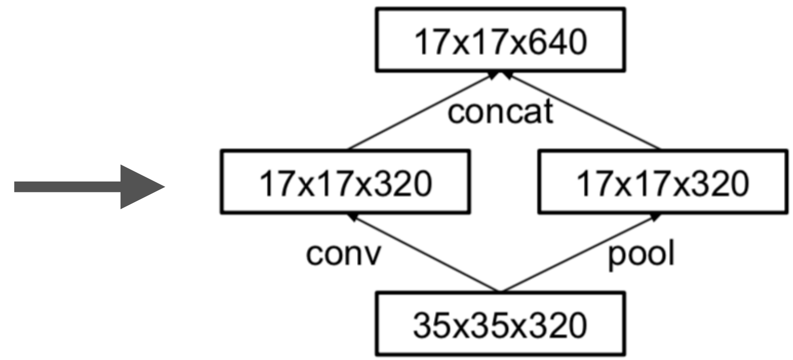
Inception v3 - Improvement 3



3x more computations



Representational bottleneck

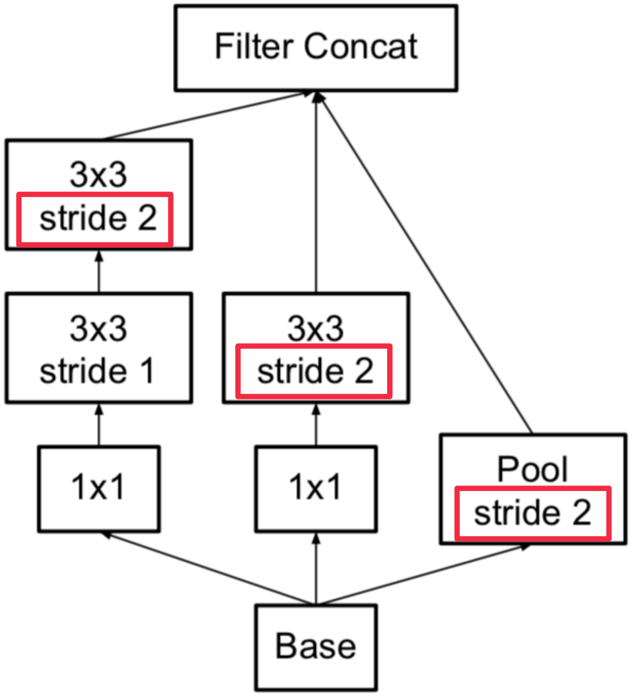


- No bottleneck
- 1x computations

Szegedy et al., Rethinking the Inception Architecture for Computer Vision, arXiv, 2015

dkfz.

Inception v3 - Improvement 3



Optimised Inception module

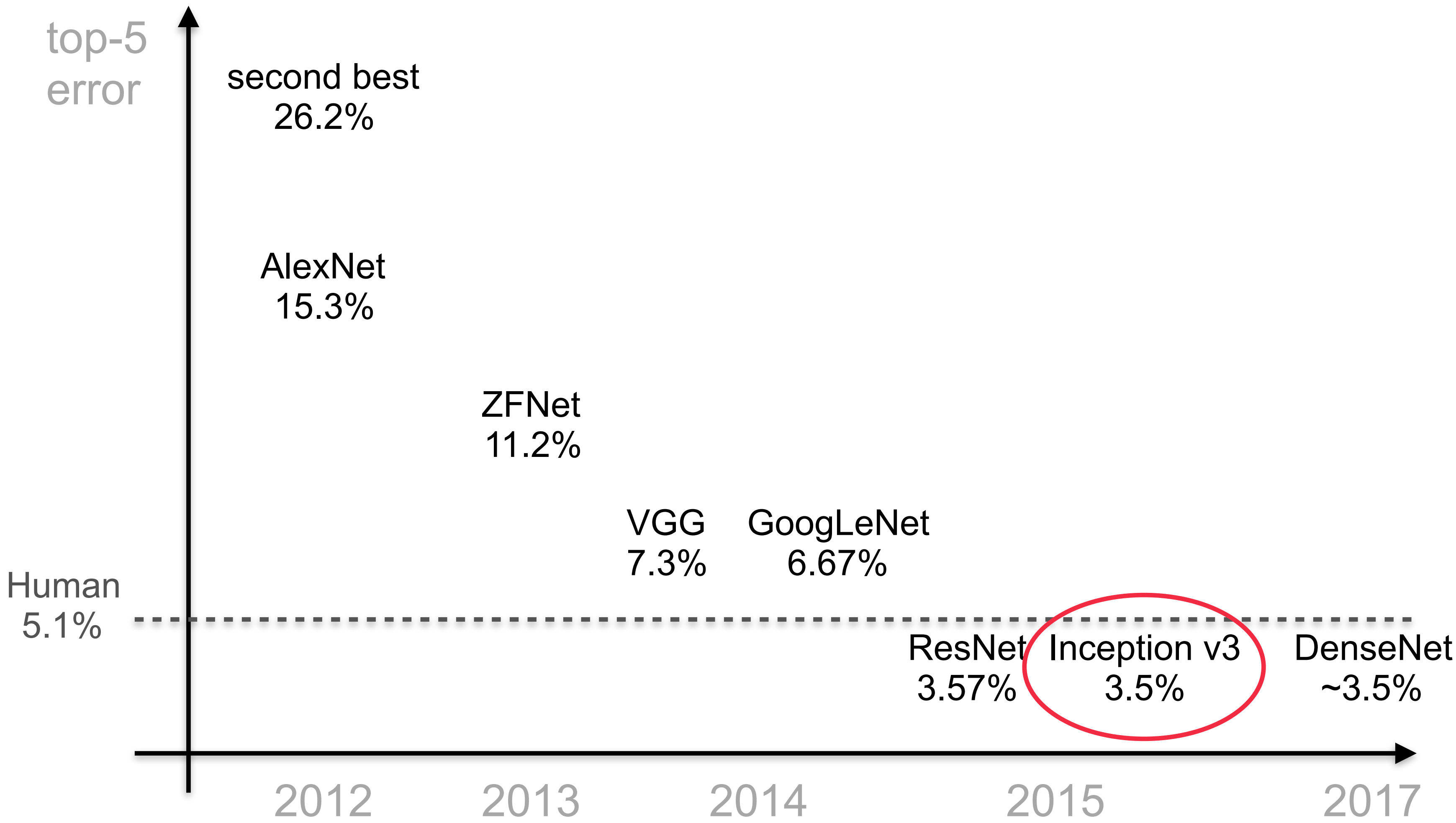
Szegedy et al., Rethinking the Inception Architecture for Computer Vision, arXiv, 2015

Inception v3

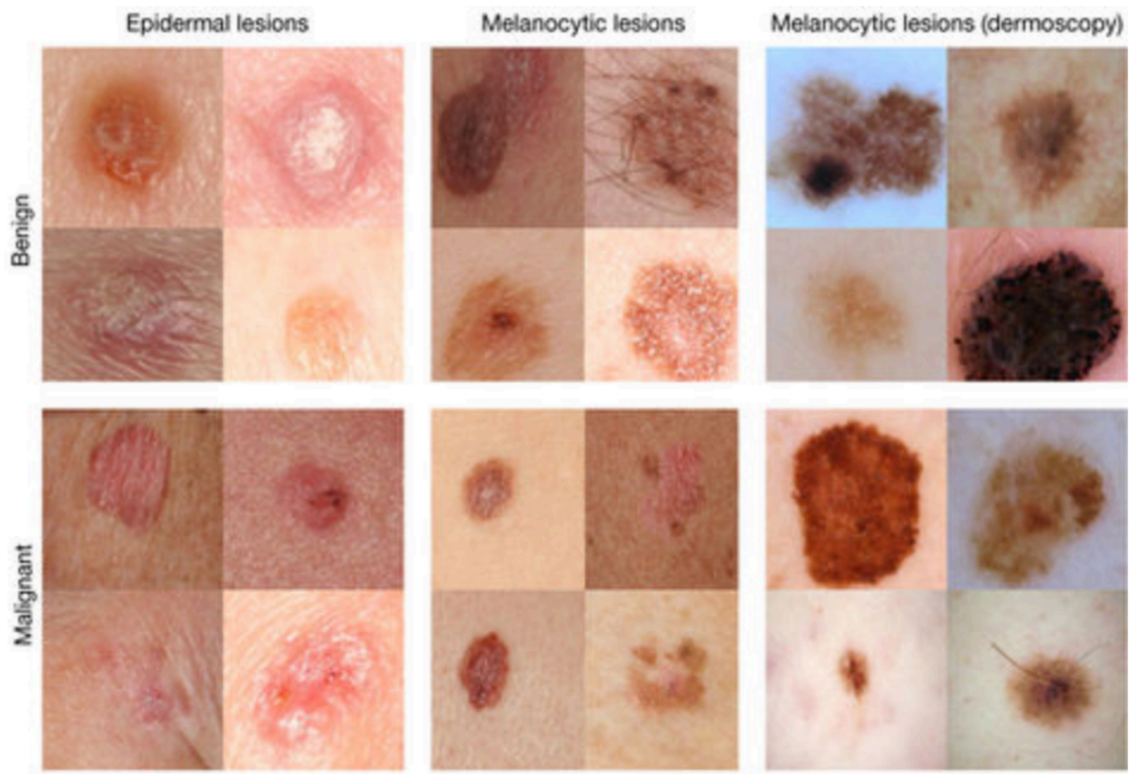
- 3.5% top-5 error
- 42 Layers
- 2.5x number of parameters of GoogLeNet

Szegedy et al., Rethinking the Inception Architecture for Computer Vision, arXiv, 2015

ILSVRC challenge / ImageNet



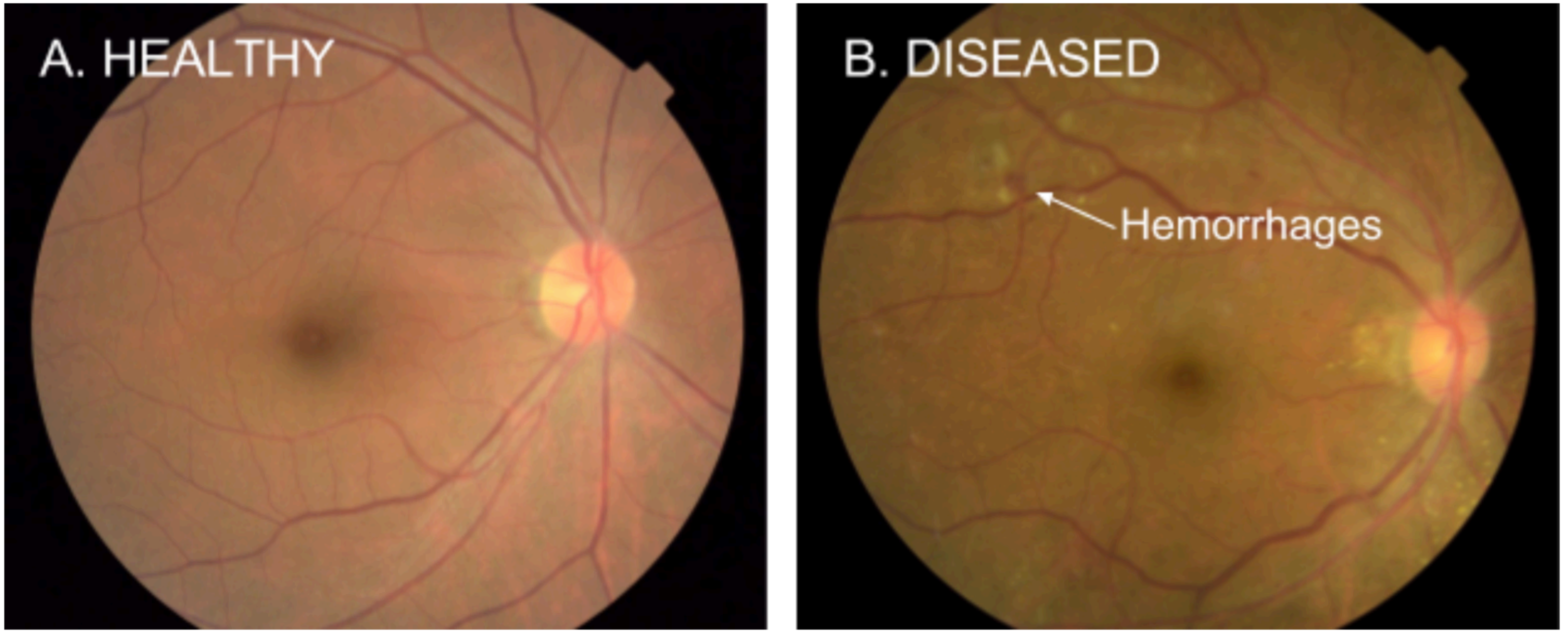
Classification of skin cancer



- Inception v3 pretained on ImageNet
- Dermatologist-level accuracy

Esteva et al., Dermatologist-level classification of skin cancer with deep neural networks, Nature, 2017

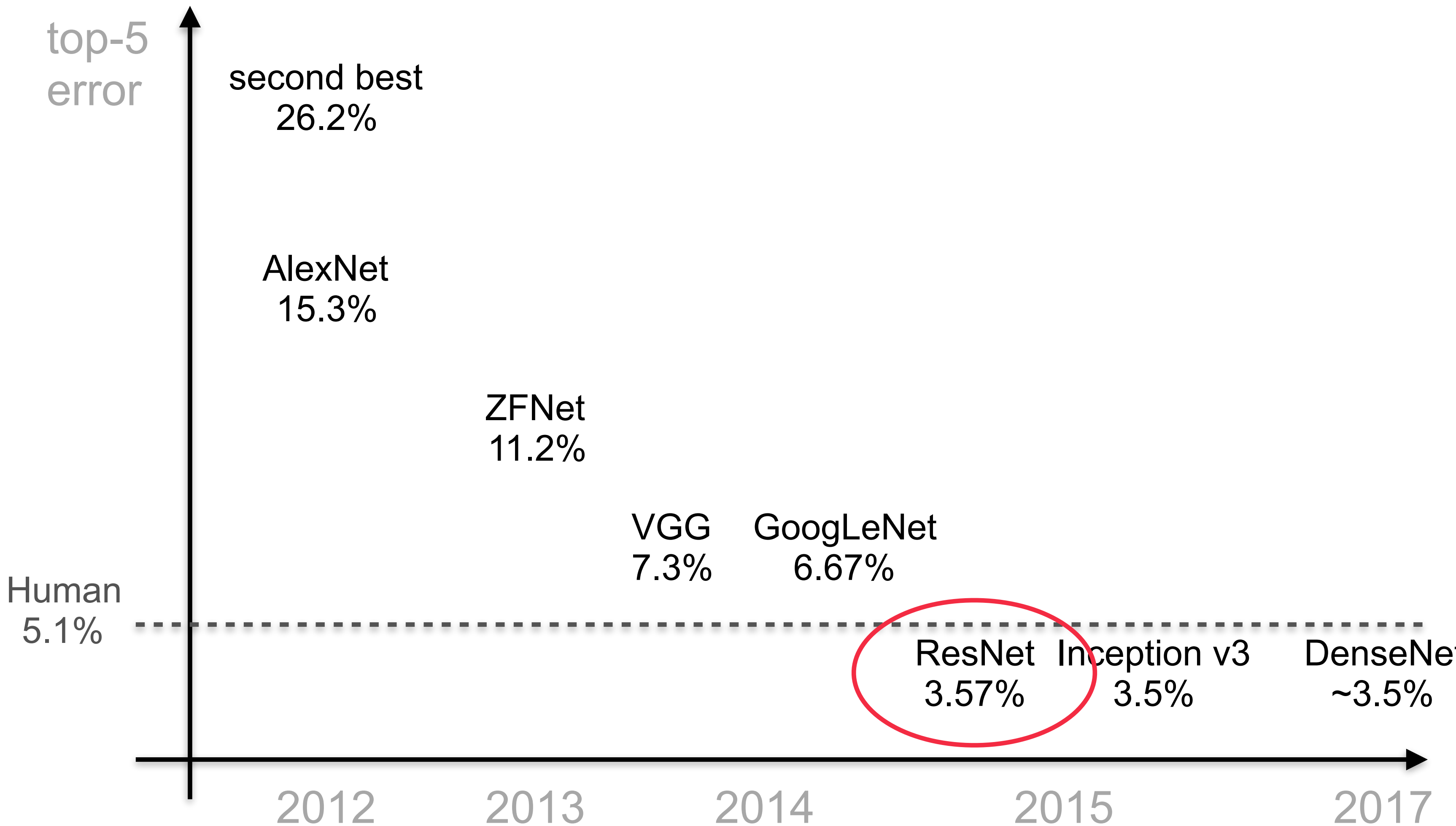
Classification of diabetic retinopathy



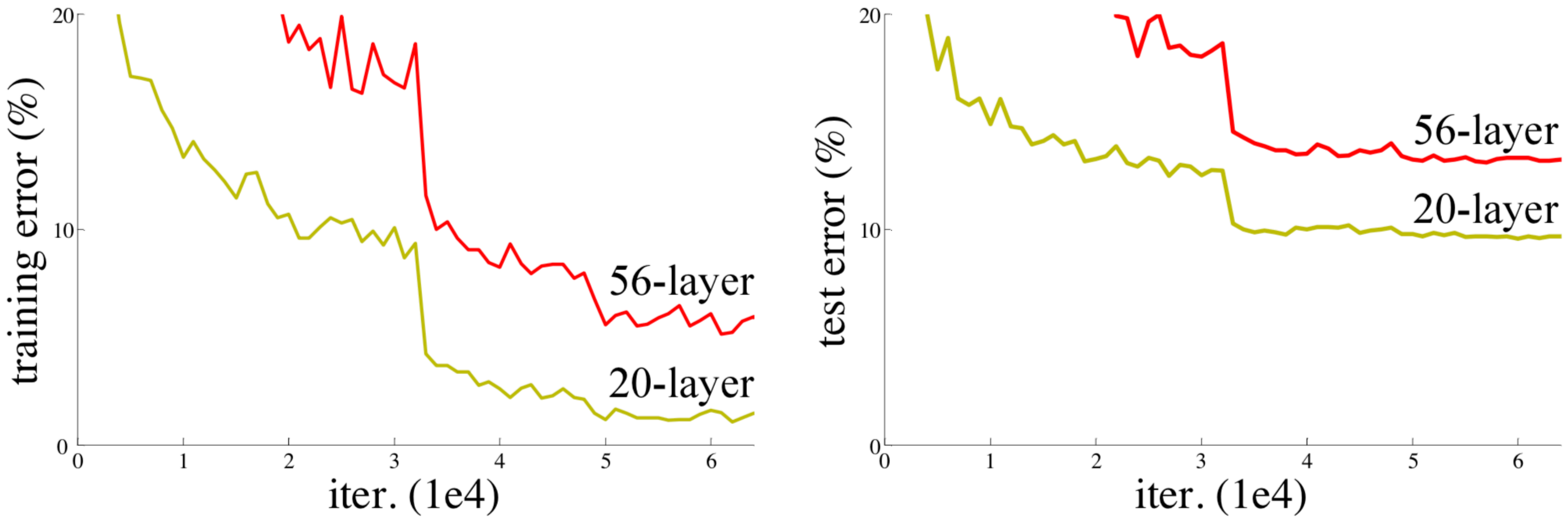
- Inception v3 pretained on ImageNet
- Expert-level accuracy

Gulshan et al., Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs, JAMA, 2016

ILSVRC challenge / ImageNet

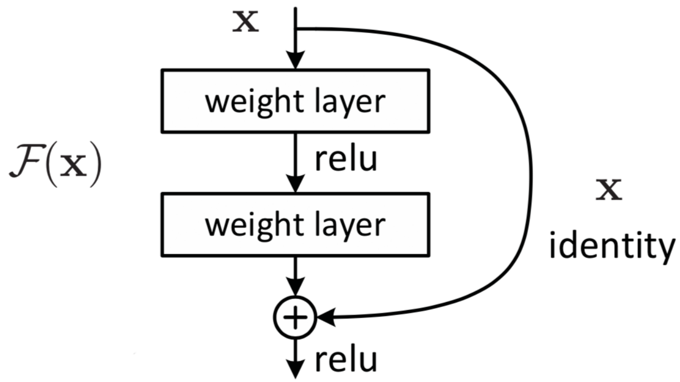


ResNet



He et al., Deep Residual Learning for Image Recognition, arXiv, 2015

ResNet

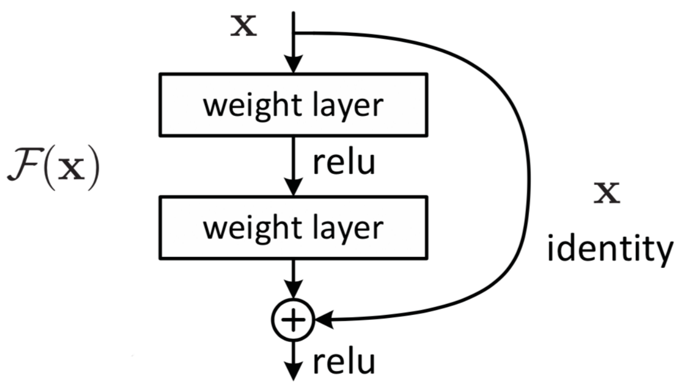


$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x}.$$

He et al., Deep Residual Learning for Image Recognition, arXiv, 2015

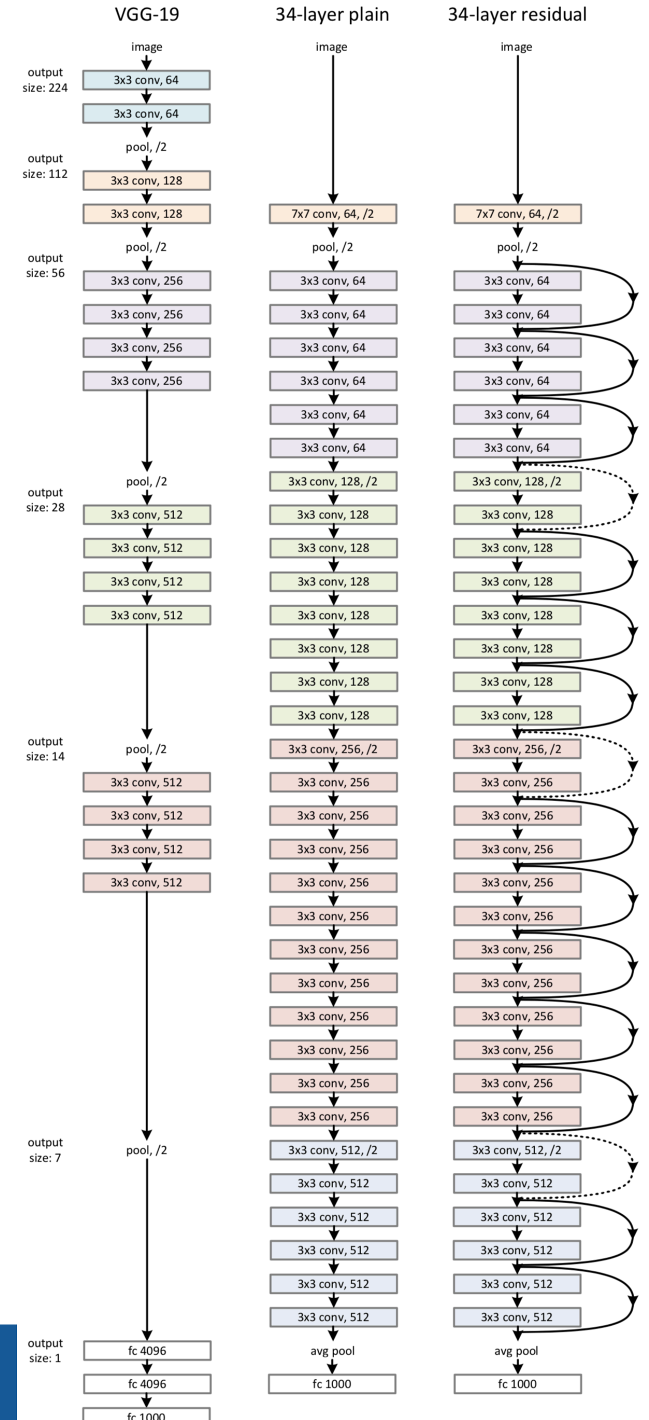
dkfz.

ResNet

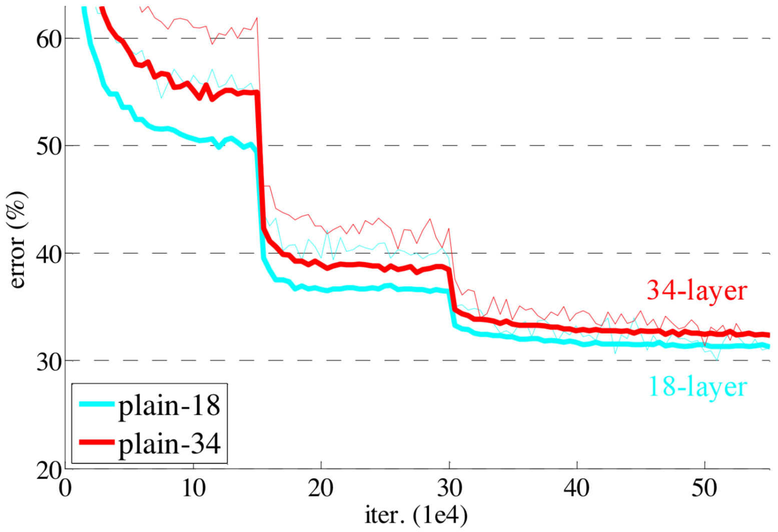


$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x}.$$

- 152 Layers



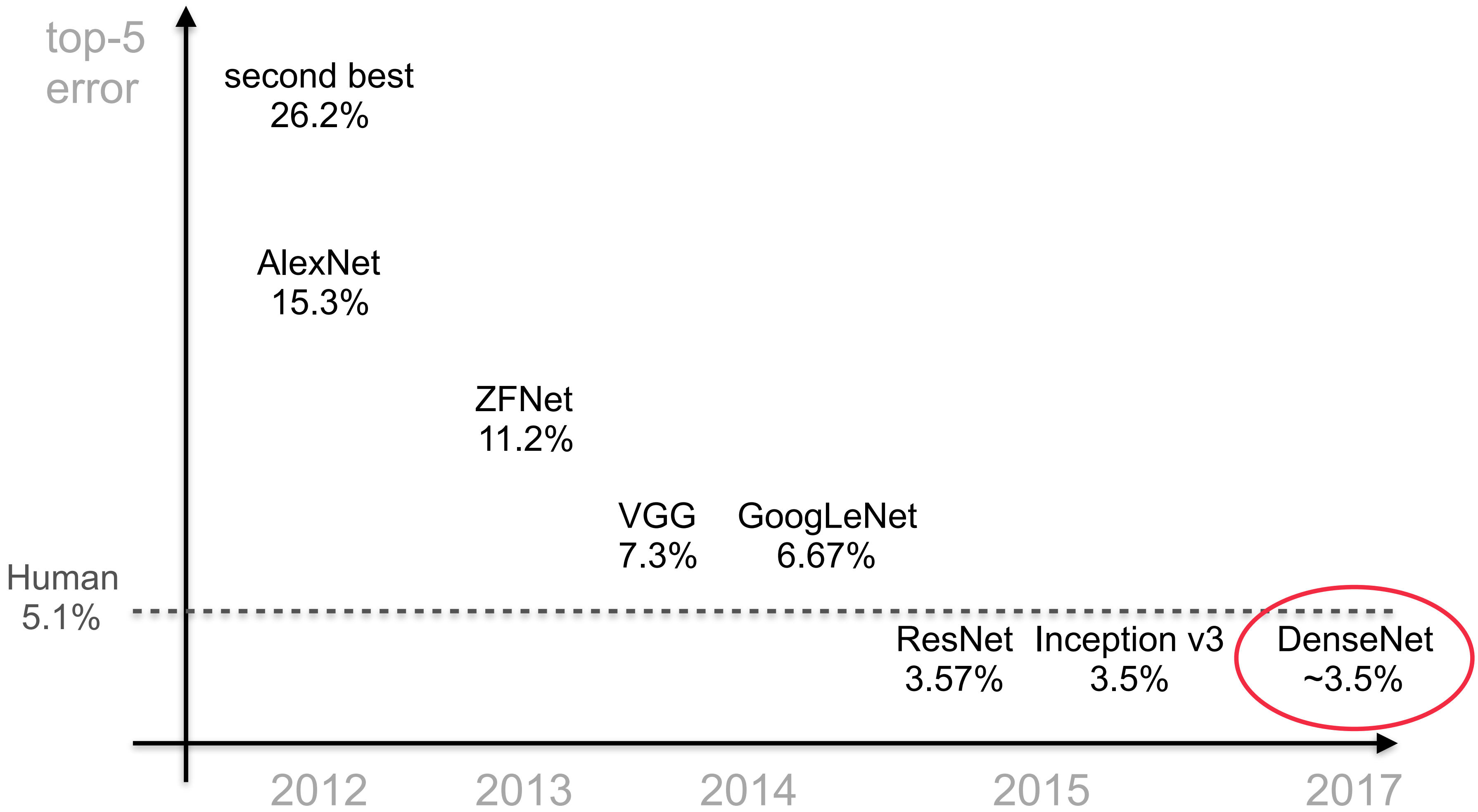
ResNet



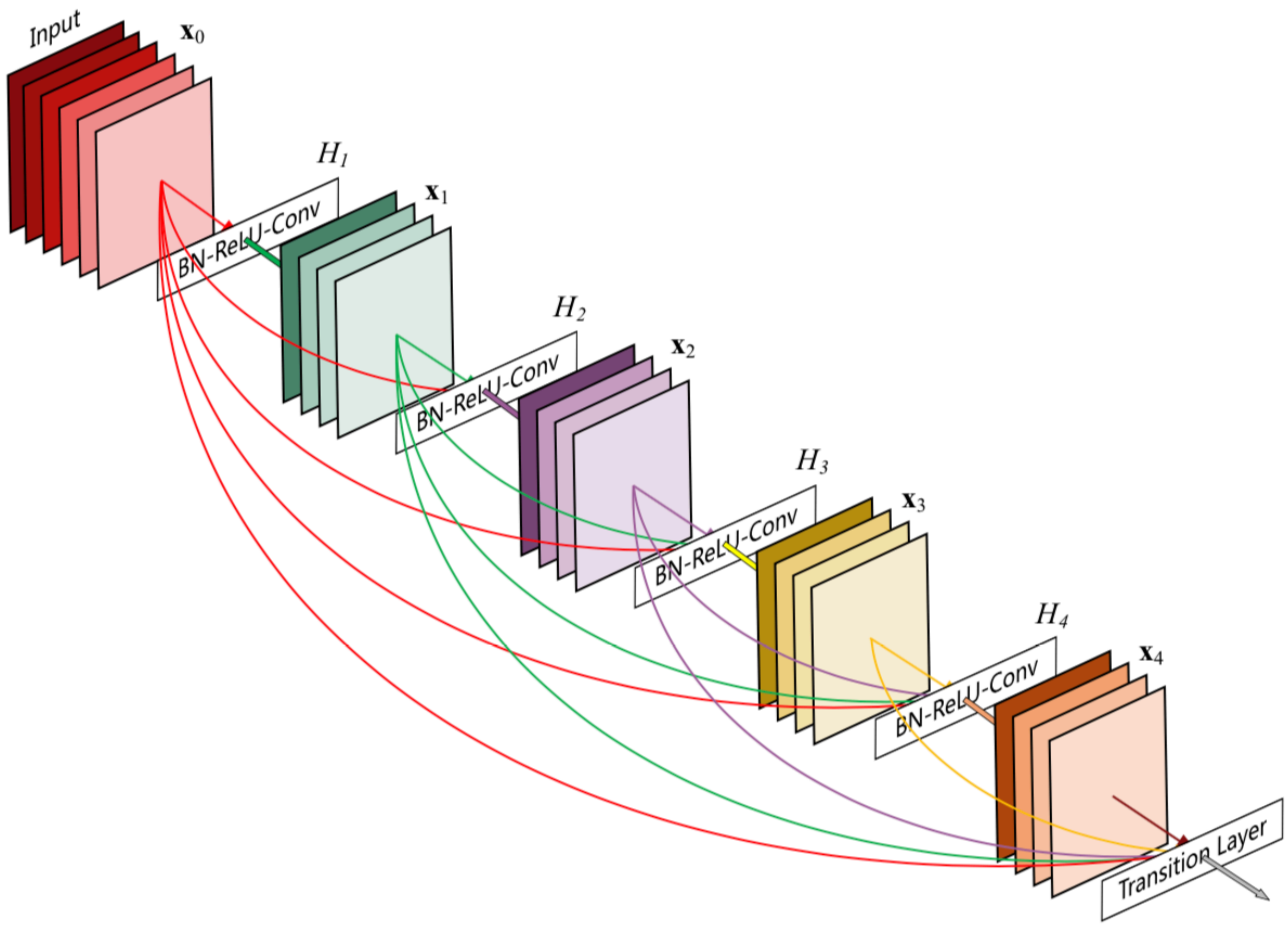
He et al., Deep Residual Learning for Image Recognition, arXiv, 2015

dkfz.

ILSVRC challenge / ImageNet

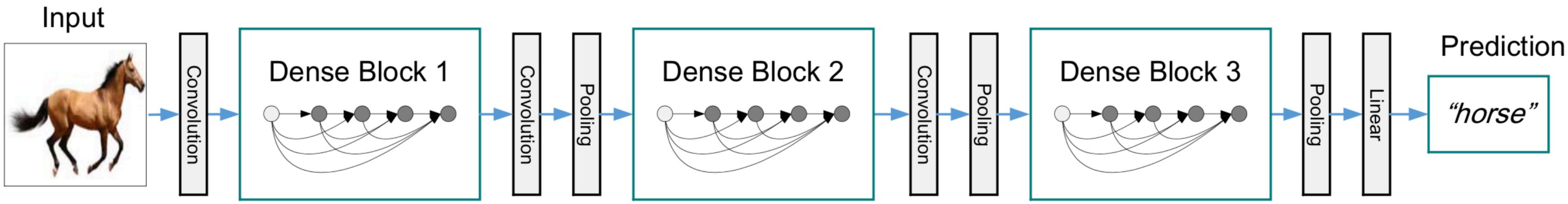


DenseNet



Huang et al., Densely Connected Convolutional Networks, CVPR, 2017

DenseNet



Huang et al., Densely Connected Convolutional Networks, CVPR, 2017

Challenges in medical image classification

- few training data
- no RGB images
- small lesions
- big images
- interpretability

Source: The Radiology Assistant : Bi-RADS for Mammography and Ultrasound 2013

Interpretability of predictions

?

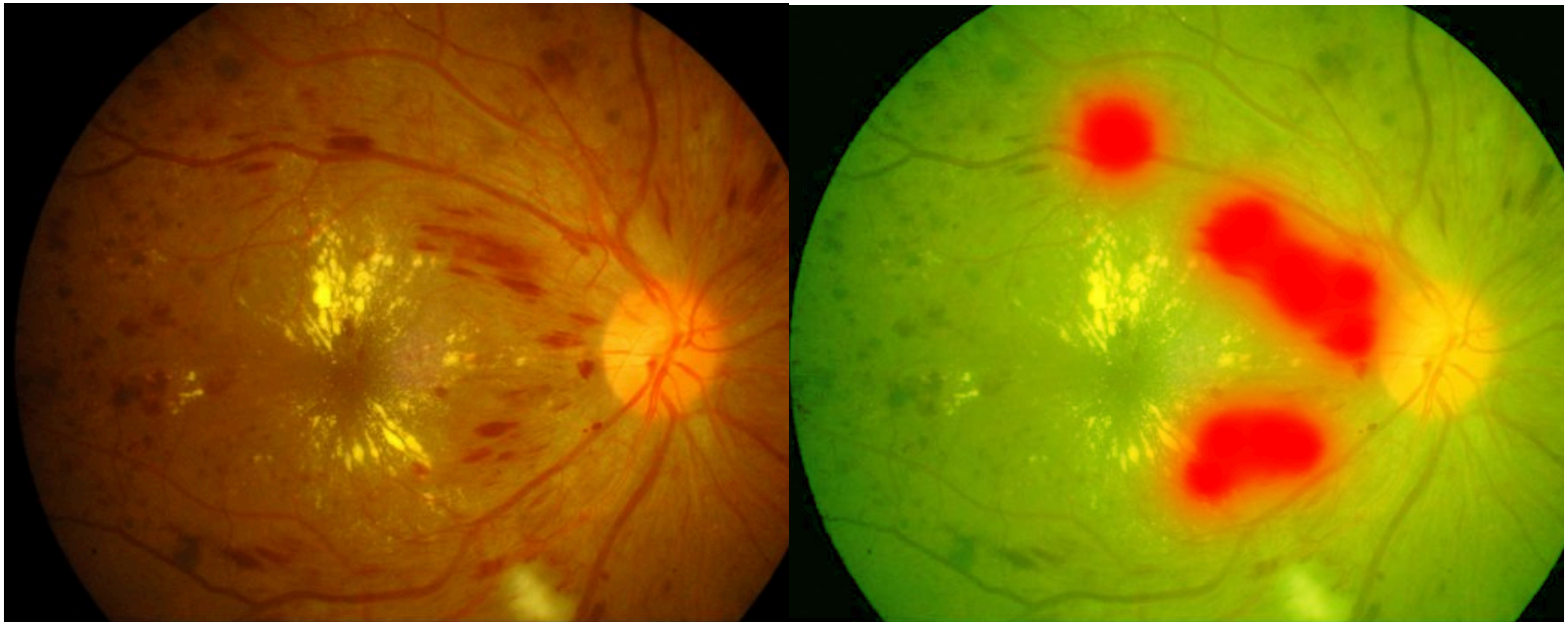
0.98 **p(diabetic)**

0.02 p(normal)

A deep neural network is often considered as a “black box”.

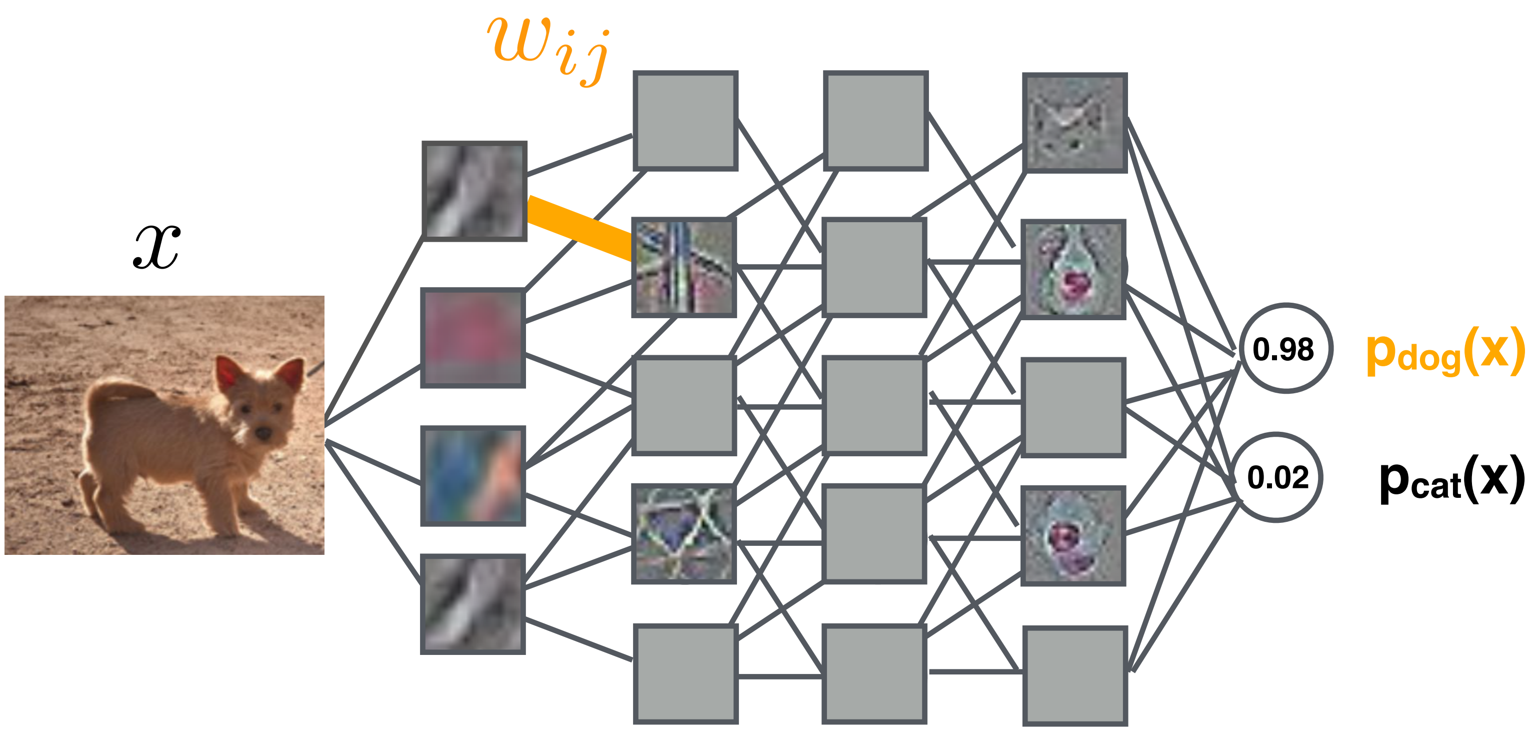
Interpretability of predictions

"What parts of the input image affect the decision?"



Gulshan et al., Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs, 2016

Recap: Training via Backpropagation

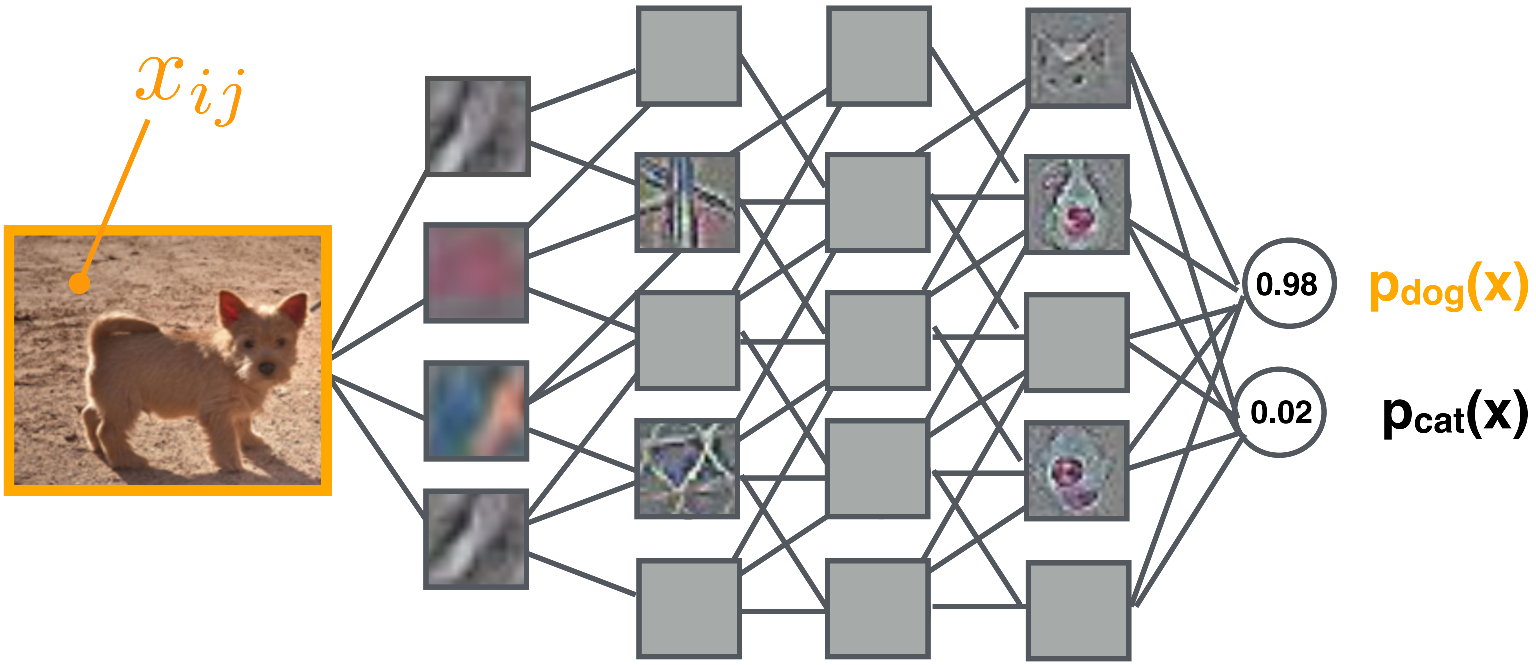


$$c = -log(p_{dog}(x)) \longrightarrow \frac{dc}{dw_{ij}}$$

Slides by courtesy of Paul Jäger

Saliency maps

"What parts of the input image affect the decision?"

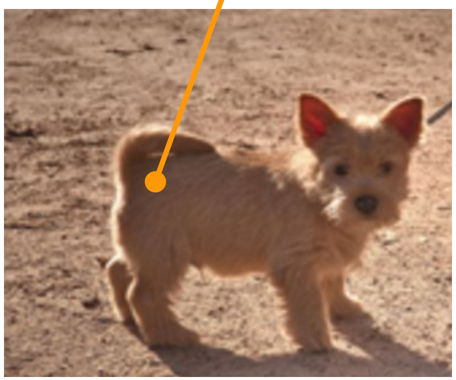


"backprop into image": $\dfrac{dp_{dog}(x)}{dx_{ij}}$

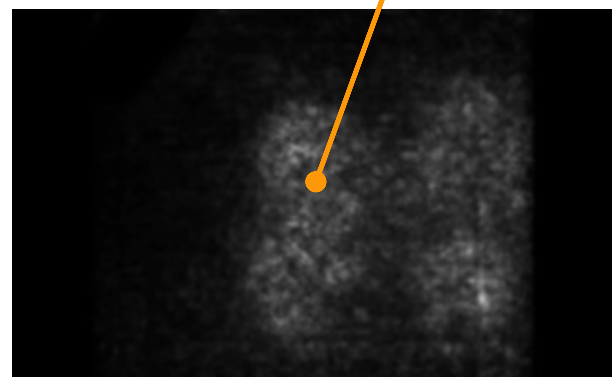
Slides by courtesy of Paul Jäger

Saliency maps

$$x_{ij}$$

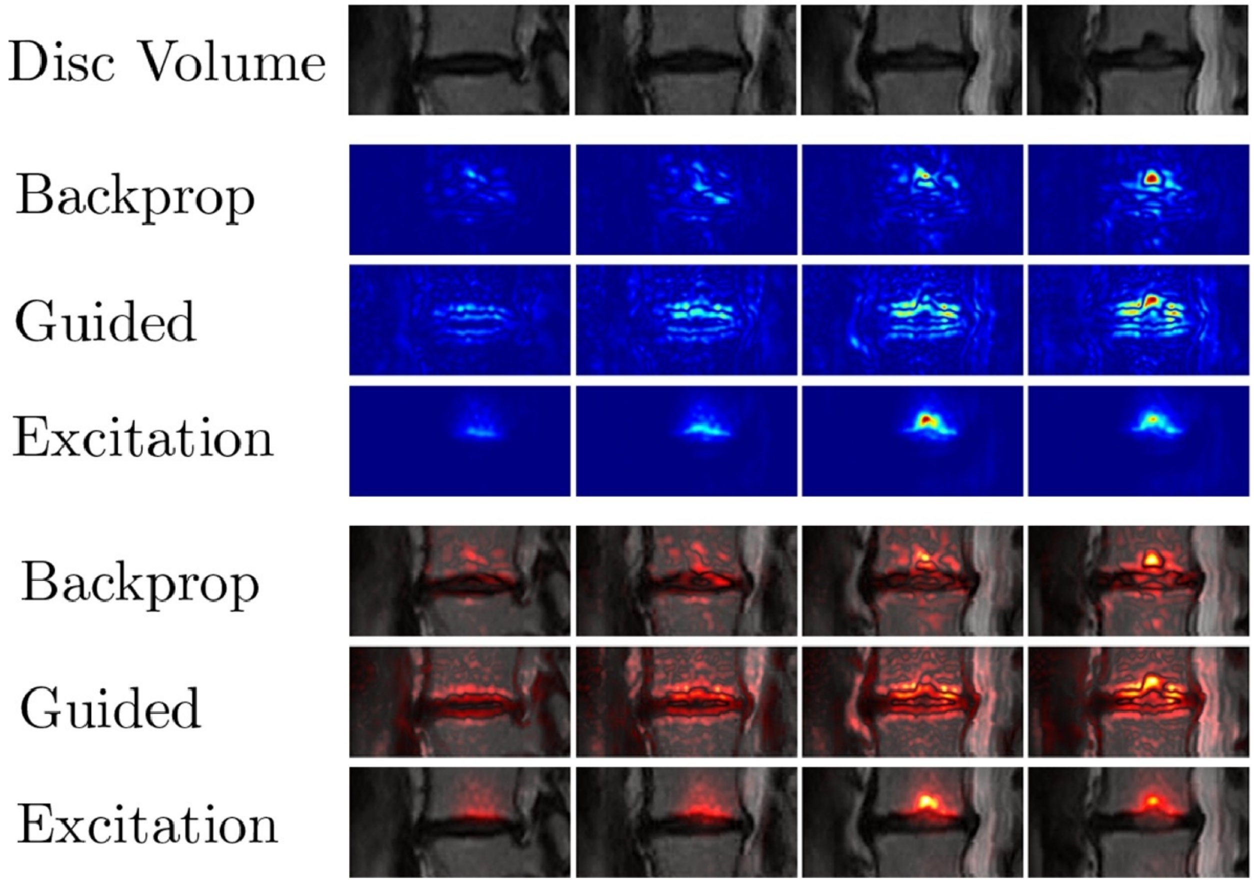


$$\frac{dp_{dog}(x)}{dx_{ij}}$$



Slides by courtesy of Paul Jäger

Interpretability of predictions



Jamaludin et al., SpineNet: Automated classification and evidence visualization in spinal MRIs, Medical image analysis, 2017

Questions

Backup

Advanced: Saliency via Perturbation

"Interpretable Explanations of Black Boxes by Meaningful Perturbation"
Ruth et al., arXiv, 2018

Trick: Backprop into a mask **m** multiplied with the image to be the "minimal destroying region".

$$\frac{d[w * (x * m)]}{dm} = w * x$$

Saliency via Perturbation

network training:

$$w_{ij}^{'} = w_{ij} - \alpha \frac{dc_{dog}(p(x))}{dw_{ij}}$$

vs

image perturbation:

$$m_{ij}^{'} = m_{ij} - \alpha \frac{dc^{*}(p_{dog}(x), m)}{dm_{ij}}$$

object: find the smallest destroying region.

$$c^{*} = \lambda_1 \|m\| + p_{dog}(\phi(x; m)) + \lambda_2 TV(m)$$

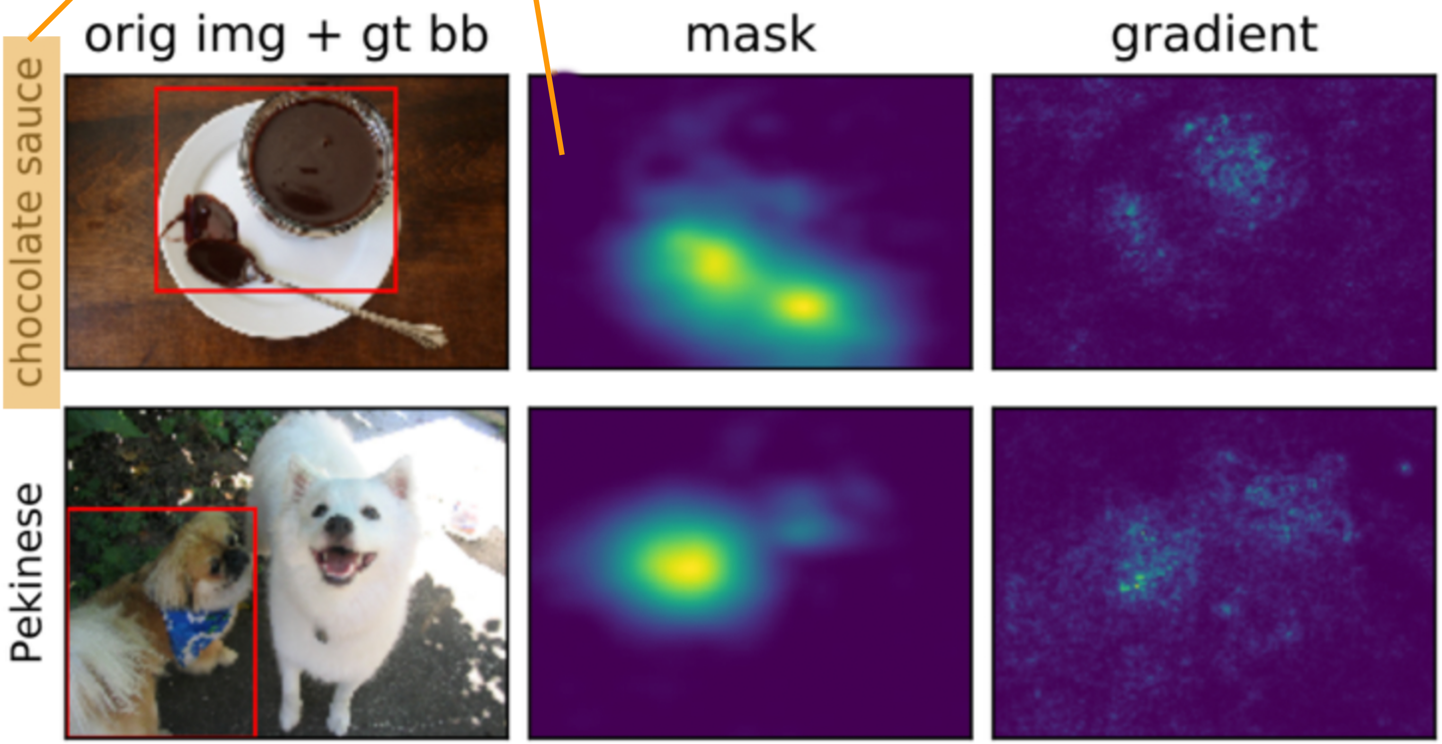
Saliency via Perturbation

Avoid high frequency artefacts by enforcing a smooth structure:

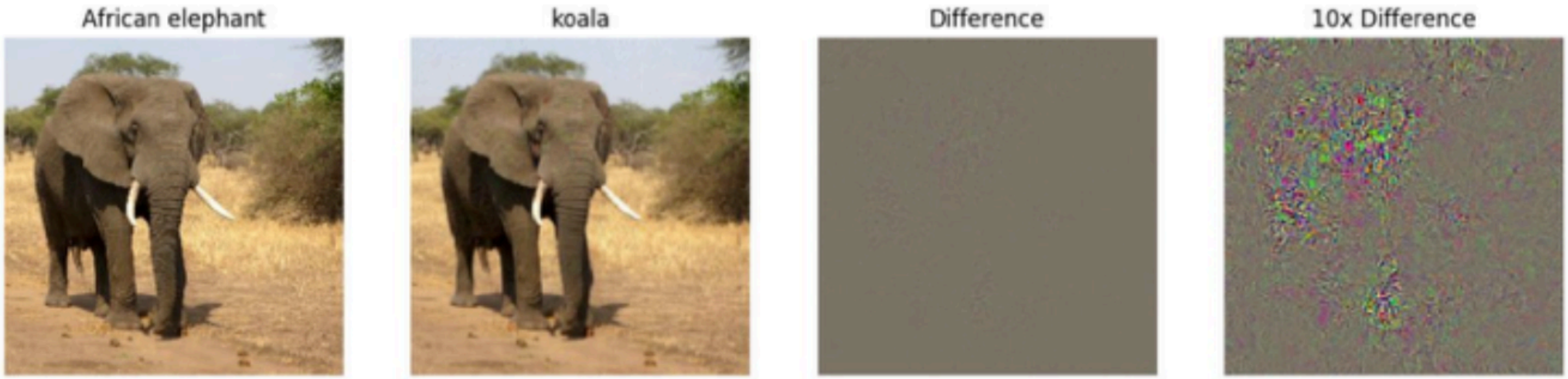
$$c^* = \lambda_1 \|1 - m\| + p_{dog}(\phi(x;m)) + \boxed{\lambda_2 TV(m)}$$

Saliency via Perturbation

result: ability to verify the underlying functionality



Why can CNNs be fooled so easily?



(source: Fei-Fei Li & Justin Johnson & Serena Young, cs231n 2017, Lecture 12)

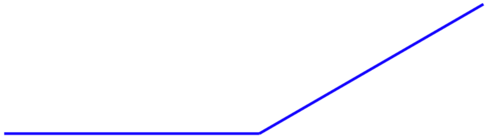
- Trake wrong class probability as cost function
- Backprop into image -> Gradients for optimal “fooling”
- Optimization on image pixels

Why can CNNs be fooled so easily?

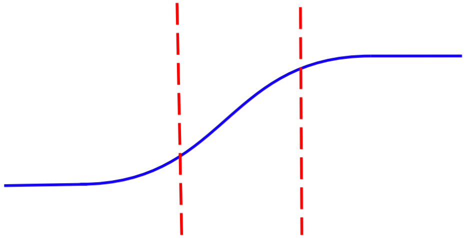
“Primary cause of NN’s vulnerability
to adversarial perturbations is their [piecewise]
linear nature”

(Explaining and Harnessing Adversarial Examples, Goodfellow et al., 2015)

ReLU



Sigmoid



(source: Ian Goodfellow, cs231n 2017, Lecture 16)