Machine learning: Neural networks

Bulat Ibragimov

bulat@di.ku.dk

Department of Computer Science
University of Copenhagen

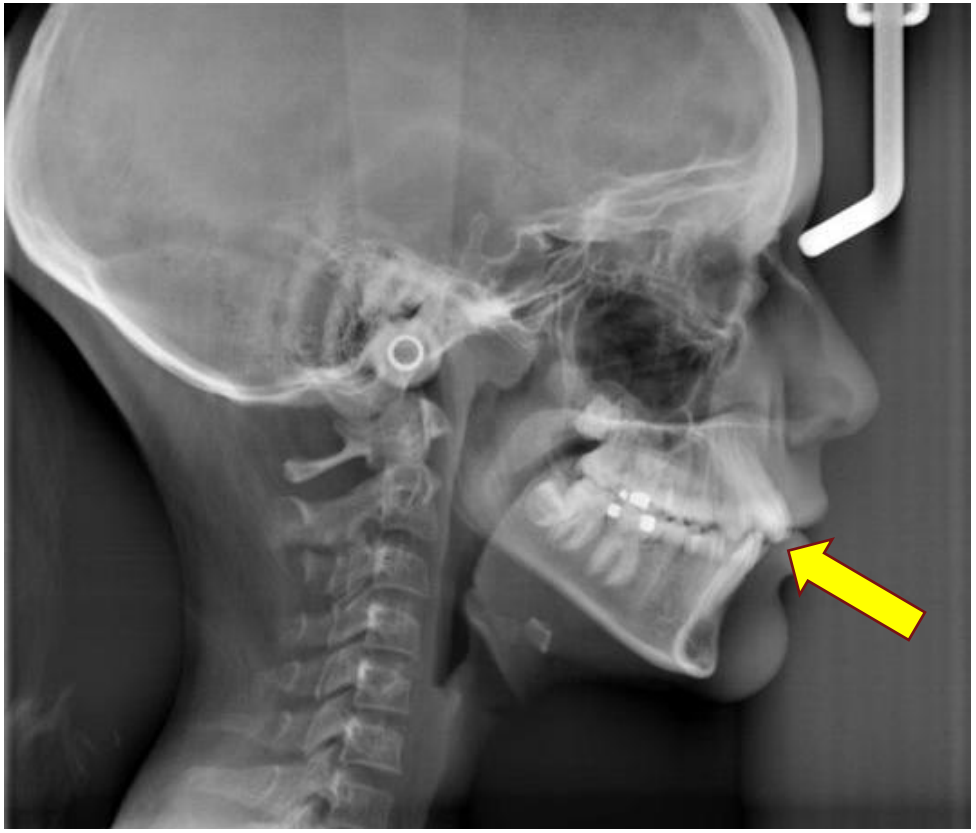
UNIVERSITY OF COPENHAGEN

Learning Objectives

- Convolution operation
- Backpropagation for convolution
- Convolutional neural network in keras
- Loss functions useful in medical imaging

Machine learning in medical image analysis

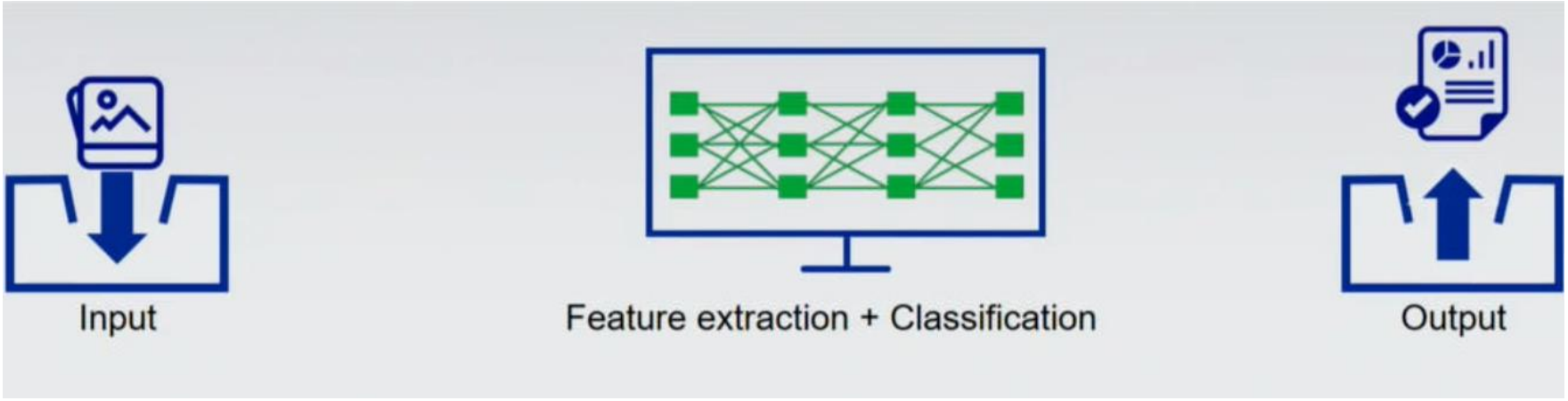
Which hand-crafted features will turn to be useful for a selected classifier?



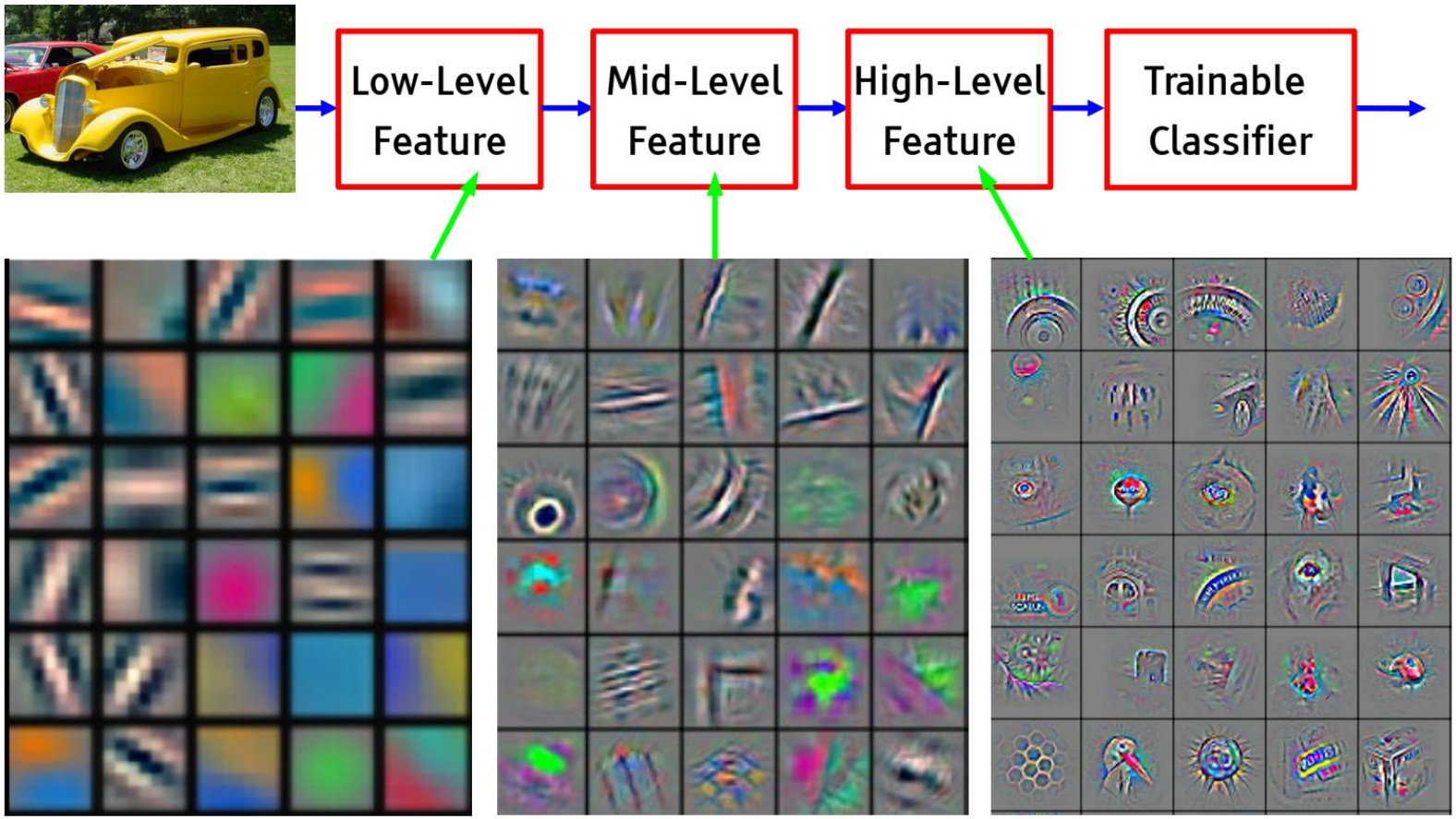
Deep learning in image analysis

Complex features are often needed

Deep learning



Deep learning in image analysis

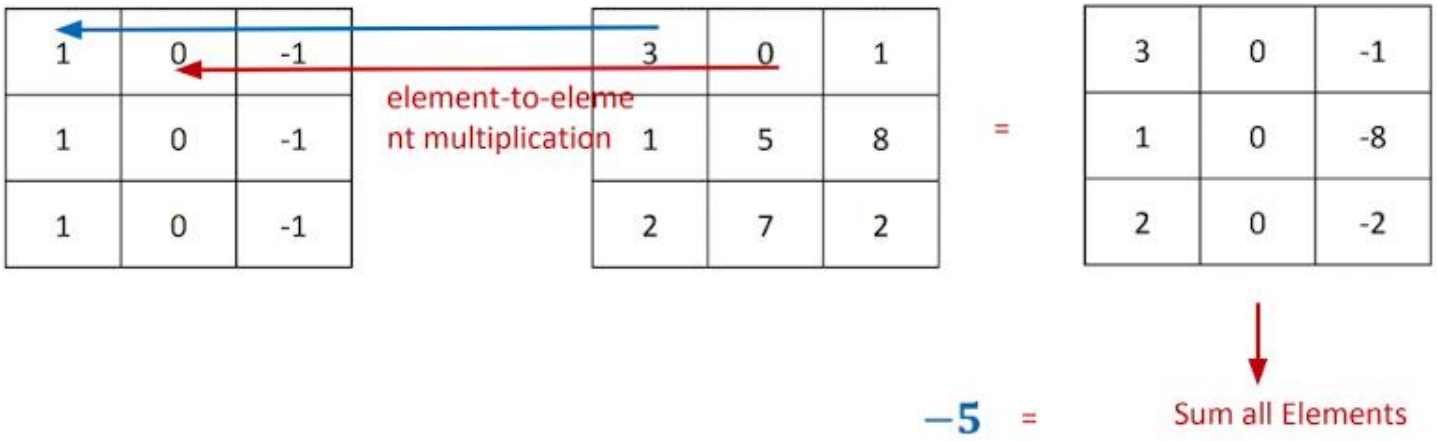


Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

LeCunn lecture https://drive.google.com/file/d/0BxKBnD5y2M8NVHRiVXBnOVpiYUk/view
Deep learning in radiology: an overview of the concepts and a survey of the state of the art

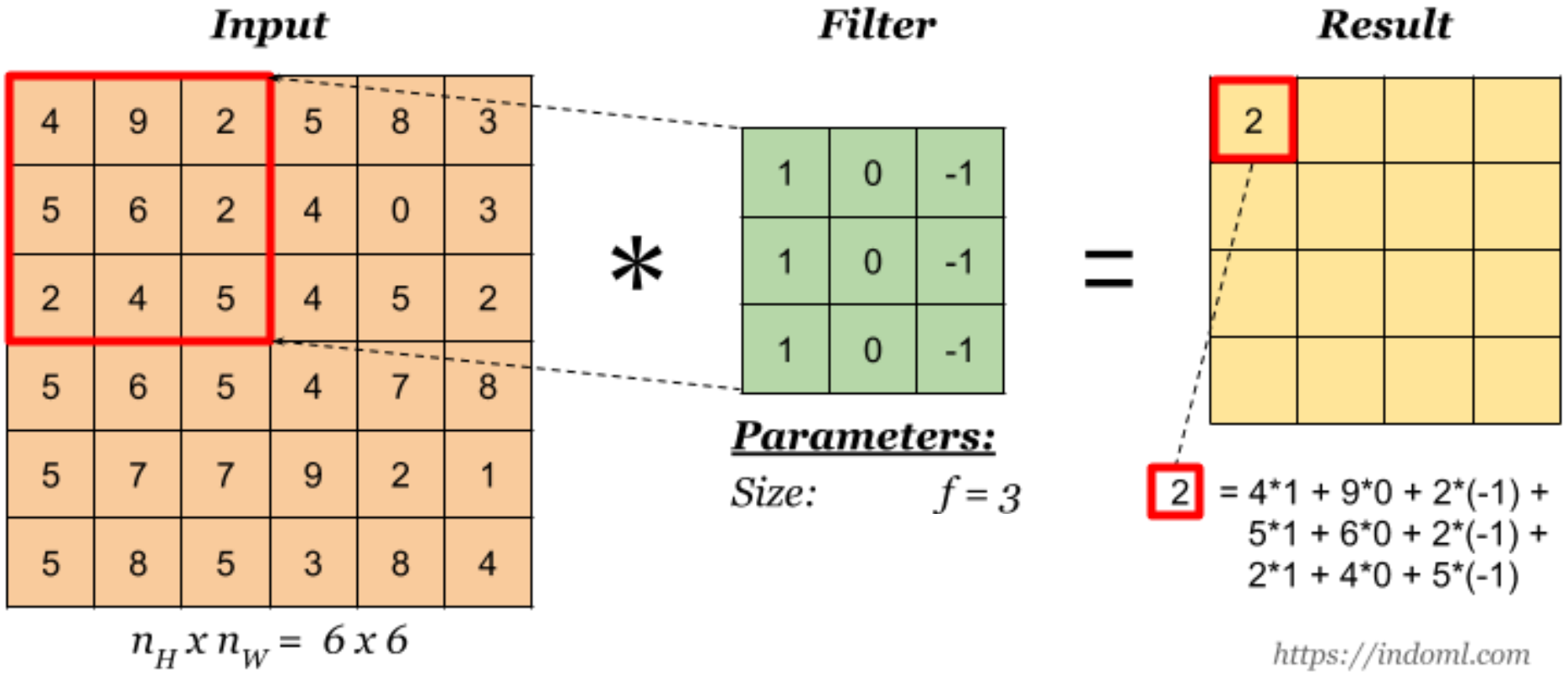
DL: convolution operation

Image **convolution** is an **element-wise multiplication** of two matrices **followed by a sum**



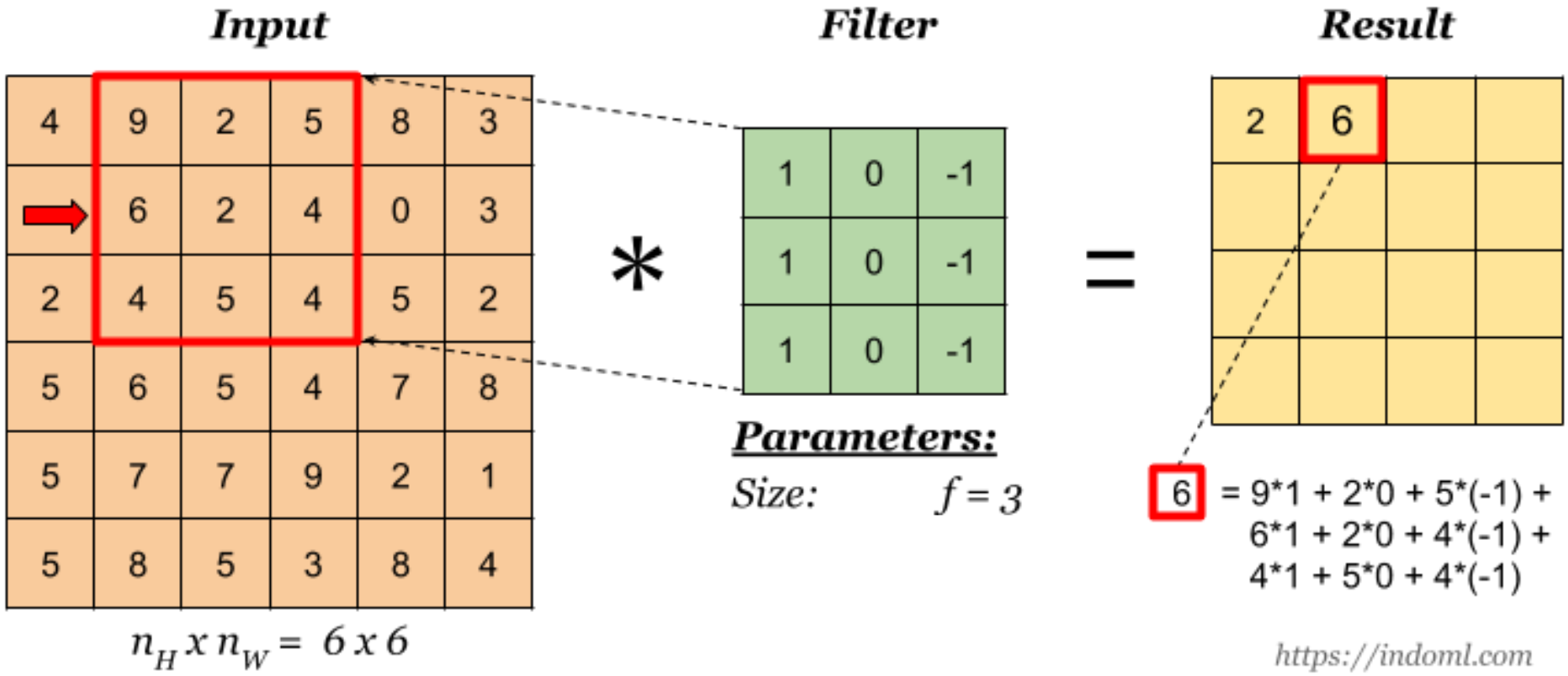
Brief CNNs: convolution

Elementwise multiplication of input signal and filter



Brief CNNs: convolution

Elementwise multiplication of input signal and filter



What if we do not want shrinking of the input size?

DL: convolution operation

What do these operations do?

Identity

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Sharpen

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

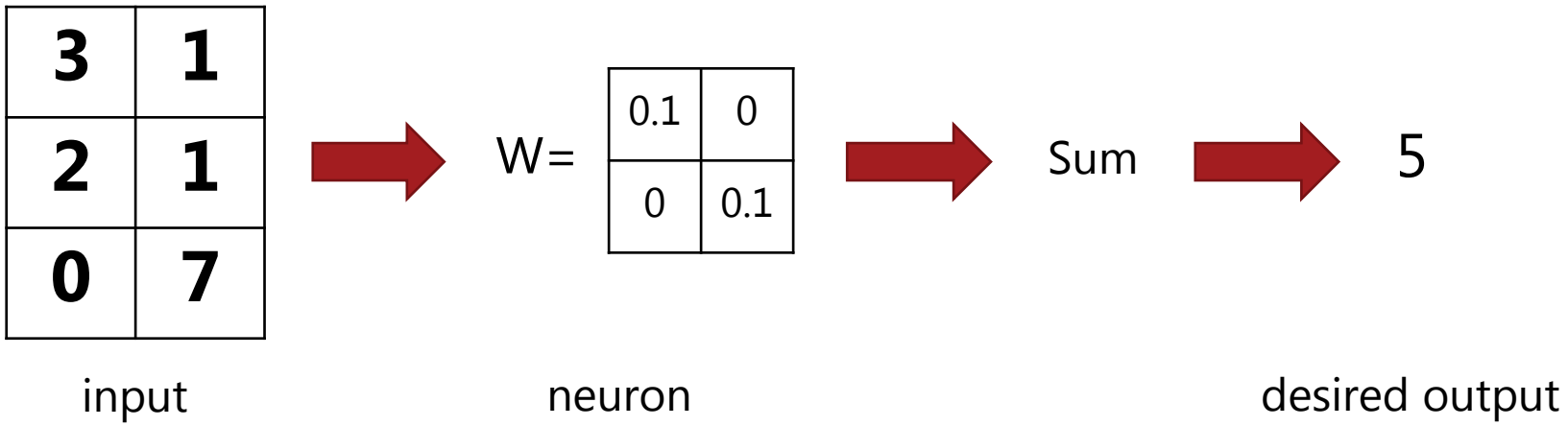
Blur

$$\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Convolution operation: backpropagation

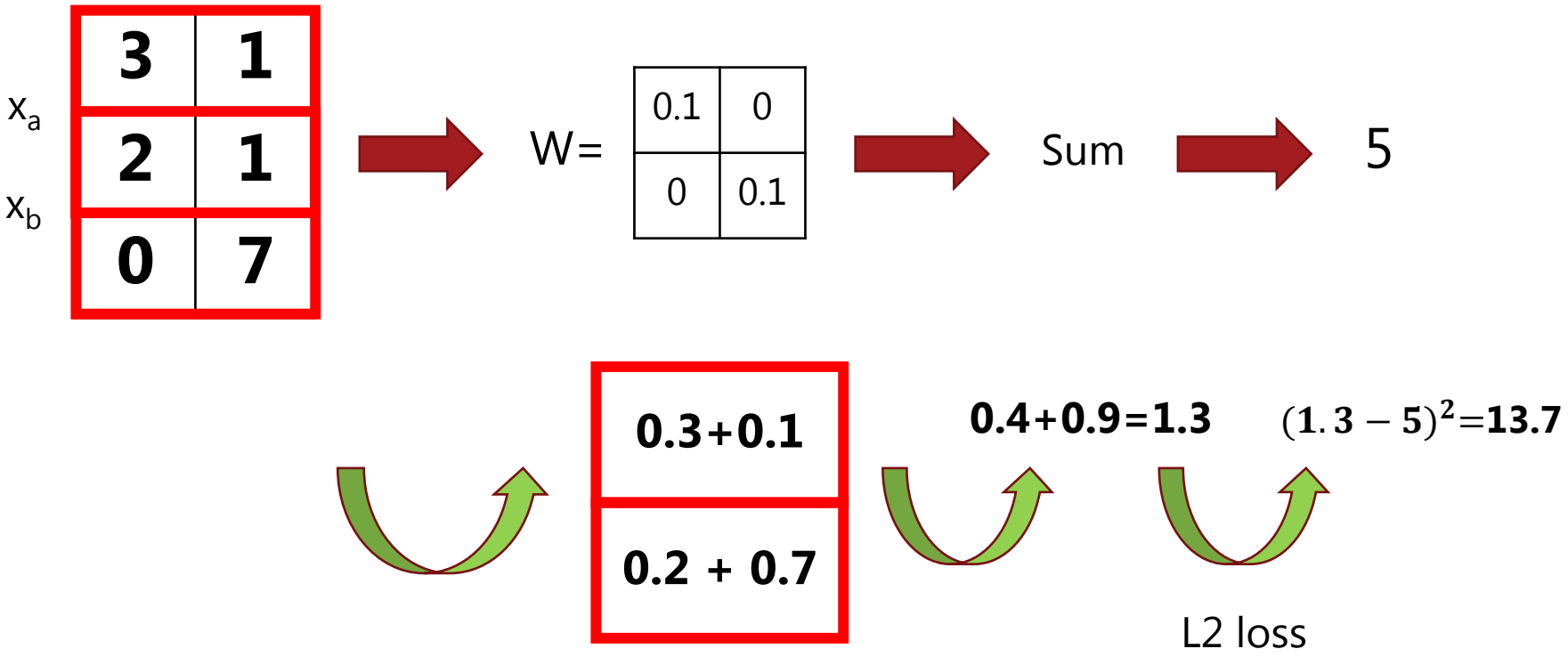
Network design:

- Input x is 3x2 array
- Neuron is 2x2 convolution with no bias
- Output y is 1D value
- Activation function σ is identity function



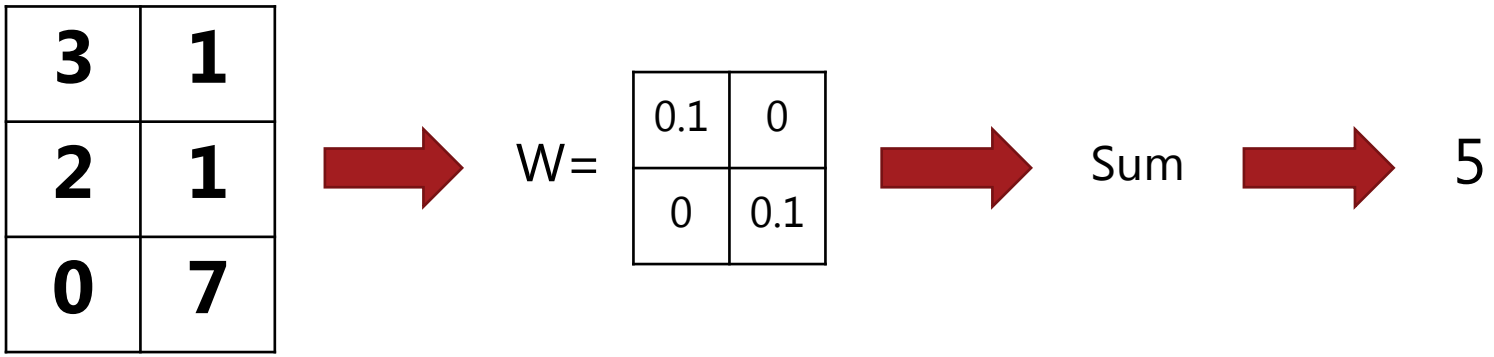
Convolution operation: backpropagation

Backward pass:



Convolution operation: backpropagation

Backward pass:



$$\frac{\partial Loss}{\partial g} = 2\sum(W(x_a) + W(x_b) - y_i)$$

$$\frac{\partial Loss}{\partial w_{0,0}} = 2x_{0,0}\sum(W(x_a) + W(x_b) - y_i)$$

$$+ 2x_{1,0}\sum(W(x_a) + W(x_b) - y_i)$$

13.7

Convolution operation: backpropagation

Backward pass:

$x_{0,0}=$ $x_{1,0}=$

3	**1**
2	**1**
0	**7**

$W(x_a) =$ **0.3+0.1**

$W(x_b) =$ **0.2 + 0.7**

$$\frac{\partial Loss}{\partial w_{0,0}} = 2x_{0,0}\sum(W(x_a)+W(x_b)-y_i) + 2x_{1,0}\sum(W(x_a)+W(x_b)-y_i) =$$

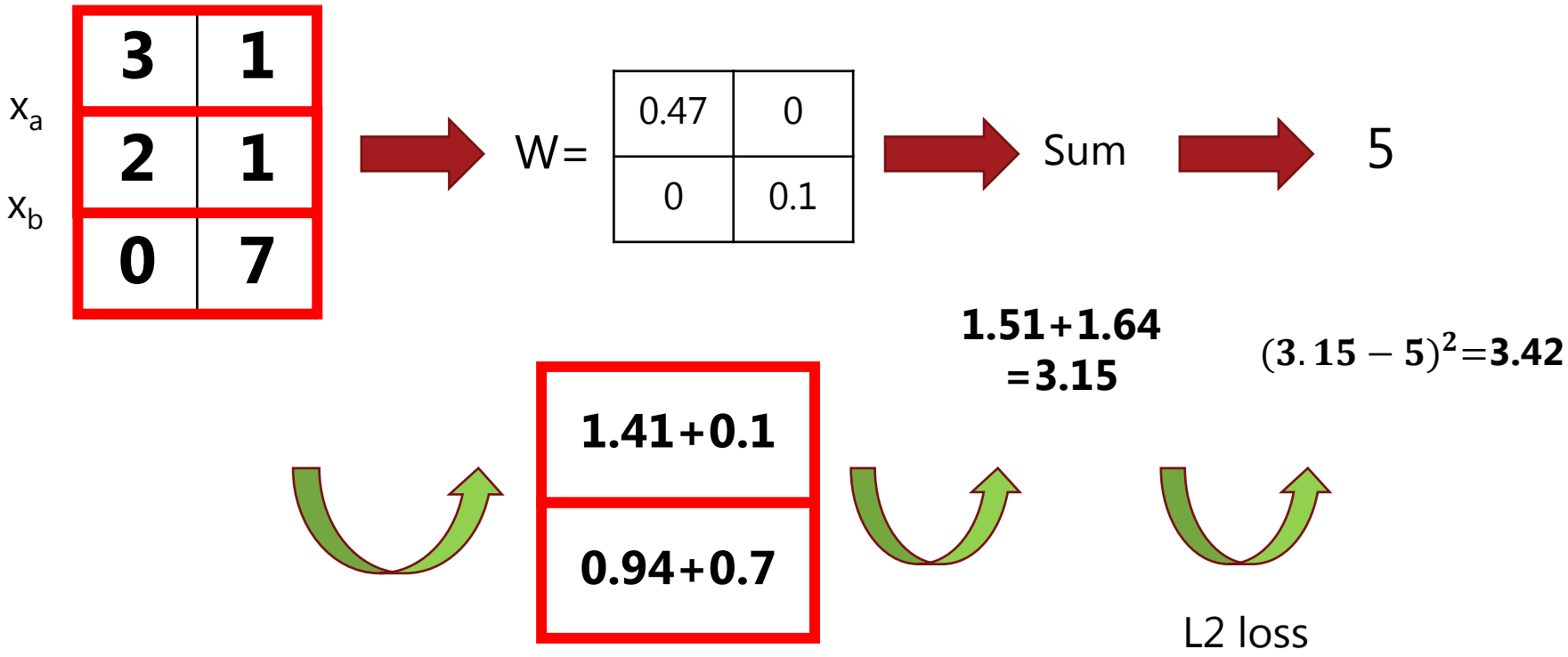
$$w_{0,0} \leftarrow w_{0,0} - \gamma\frac{\partial Loss}{\partial W} = 0.1 - 0.01(-37) = 0.47$$

W=

0.1	0
0	0.1

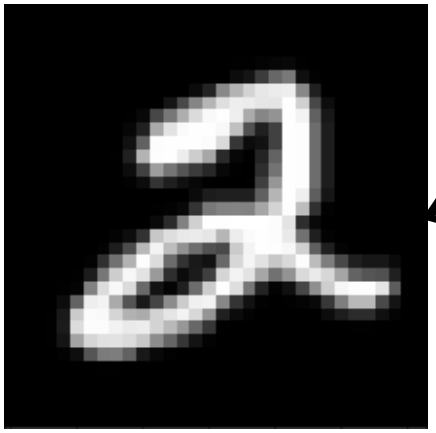
Convolution operation: backpropagation

Forward pass:



Convolution

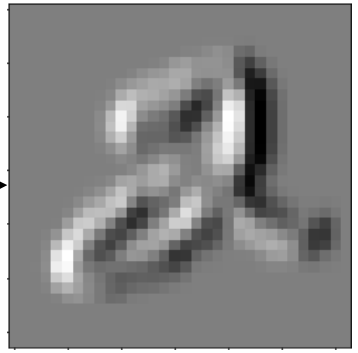
This is an example of convolution layer with filter size 3x3 and number of features 3



Input

-1	0	1
-1	0	1
-1	0	1

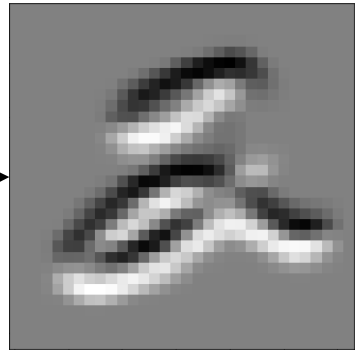
Filter1



Output1

1	1	1
0	0	0
-1	-1	-1

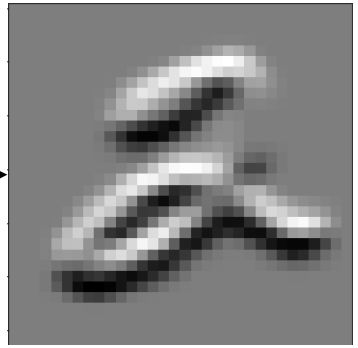
Filter2



Output2

-1	-1	-1
0	0	0
1	1	1

Filter3



Output3

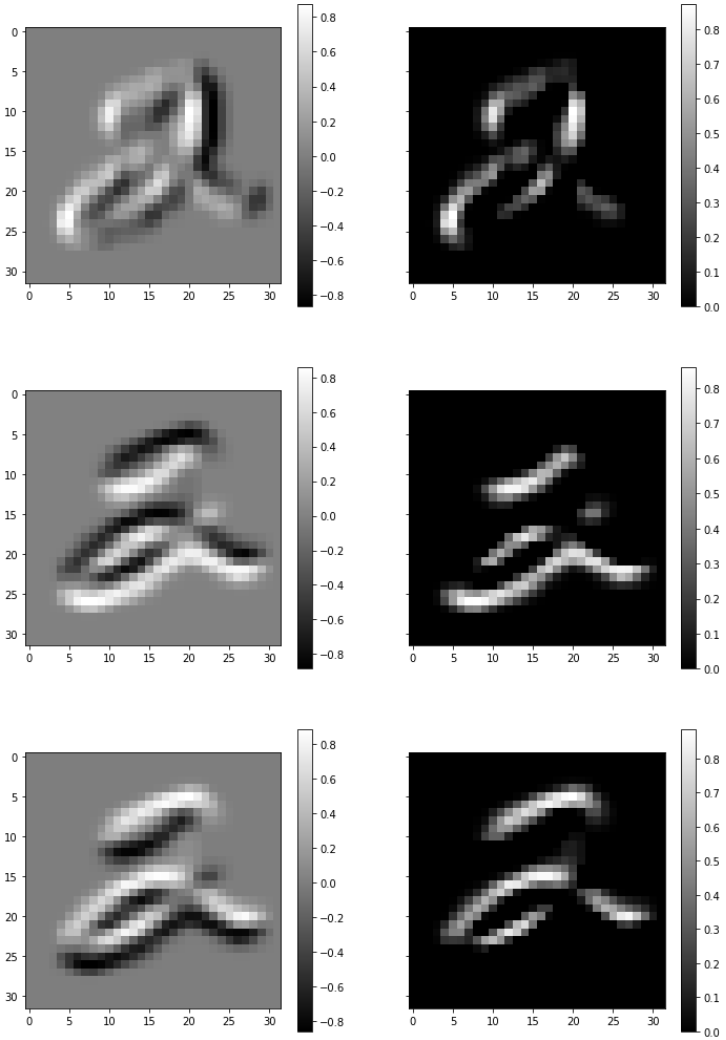
Convolution: activation function

N raw convolution layers can be transformed into one convolution layer, so there will be no depth in the network*

Activation function addresses this issue:

- Tanh(x)
- Sigmoid(x)
- ReLU(x)

$$ReLU(x) = max(0, x)$$



convolution Convolution + ReLU

*https://medium.com/the-theory-of-everything/understanding-activation-functions-in-neural-networks-9491262884e0

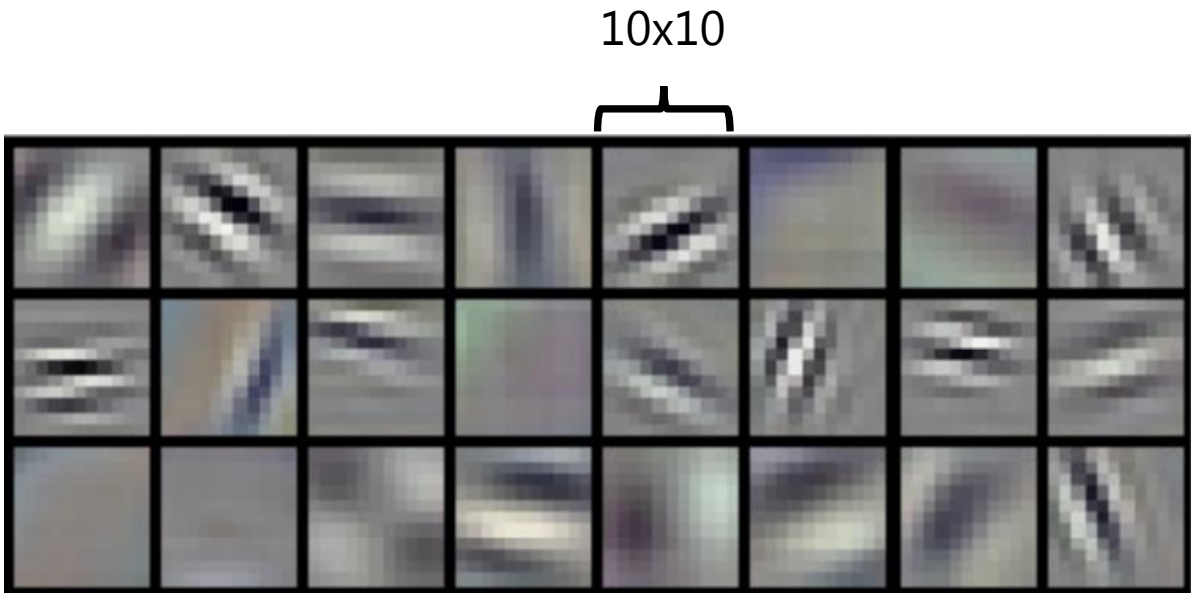
Convolution in keras

```
conv = Conv2D(24, 10, activation = 'relu', padding = 'same')(prev_Layer)
```

24 – number of features

10 – size of each feature

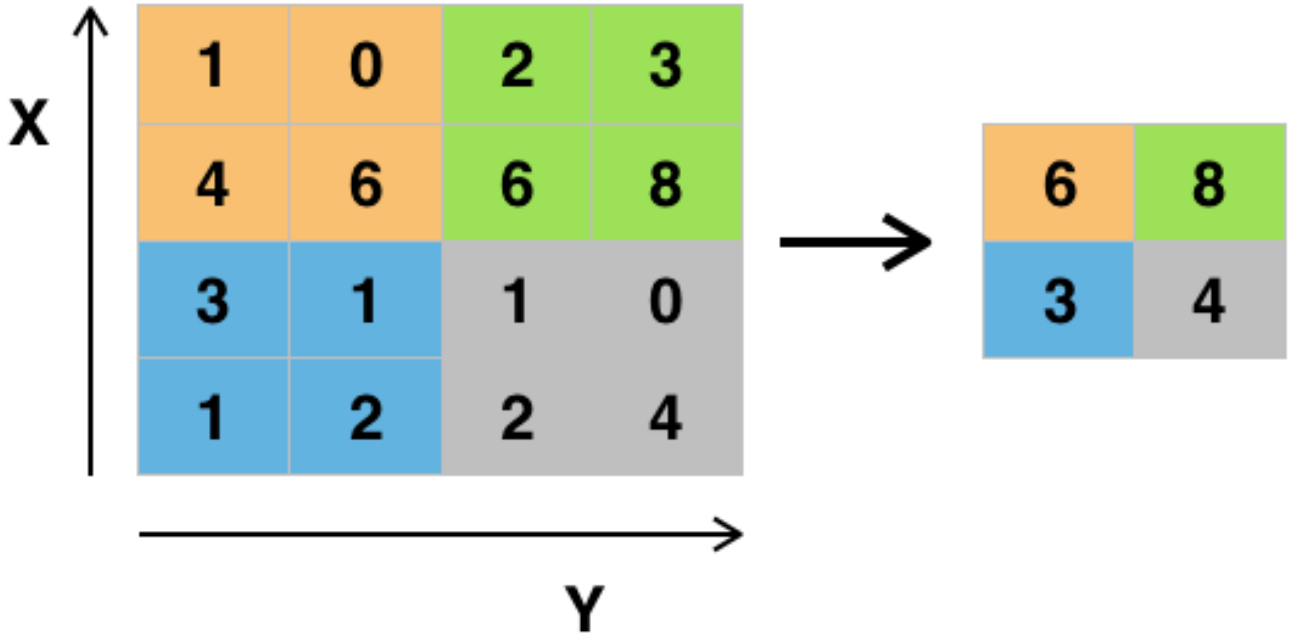
Relu - nonlinearity



Pooling

Pooling layers:

- Reduce dimensionality of the input
- Introduce some translation invariance

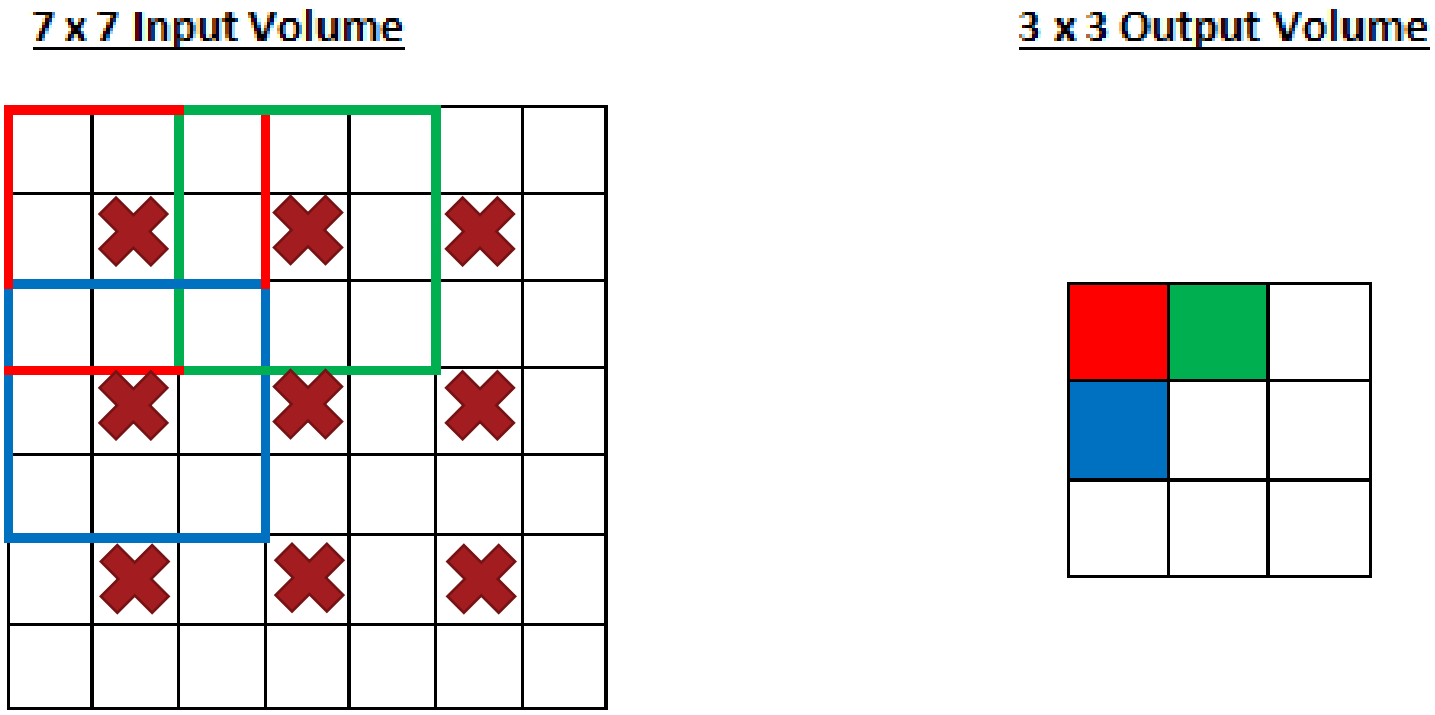


2x2 maxpooling

Convolution with stride

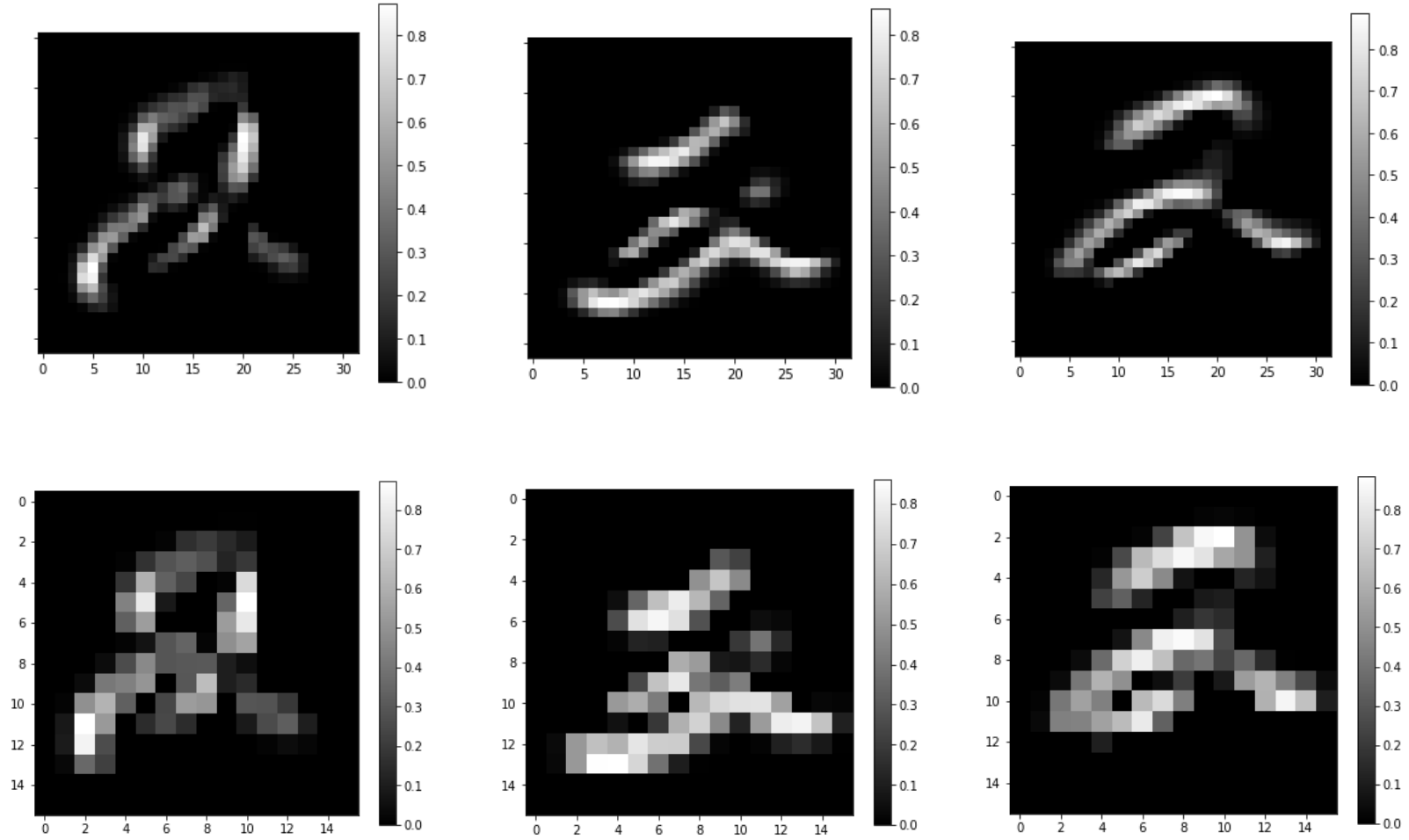
Convolution with stride >1:

- Reduce dimensionality of the input
- Introduce some translation invariance



Pooling in keras

pool = MaxPooling2D(pool_size=(2, 2))(prev_Layer)



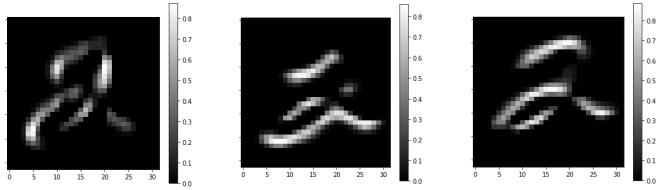
Stacking layers in keras

Let's say we have one testing input image of size 32x32:

input shape = (None, 32, 32, 1)

conv1 = Conv2D(x, 3, activation = 'relu', stride = 1, padding = 'same')(input)

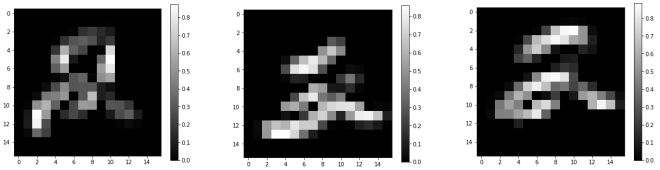
output shape = (None, 32, 32, x)



input shape = (None, 32, 32, x)

pool1 = MaxPooling2D(pool_size=(2, 2))(conv1)

input shape = (None, 16, 16, x)



input shape = (None, 16, 16, x)

conv2 = Conv2D(y, 3, activation = 'relu', stride = 1, padding = 'same')(pool1)

input shape = (None, 16, 16, y)

How do conv2 filters look like?

Stacking layers in keras

- Each filter in conv2 works ALL x maps of pool1

Filter in conv2 of size (3x3x3)

Filter[:, :, 0] =

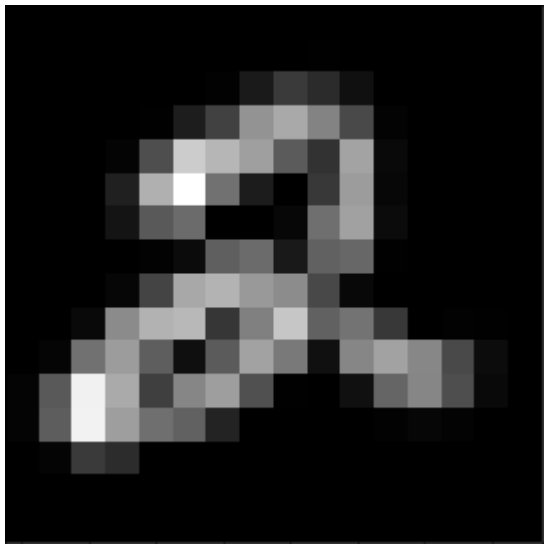
0	0	0
0	1	0
0	0	0

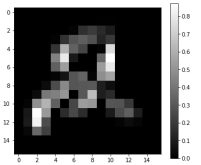
Filter[:, :, 1] =

0	0	1
0	0	0
0	0	0

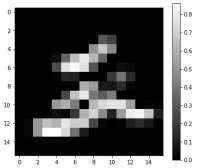
Filter[:, :, 2] =

0	0	0
0	0	0
0	0	1

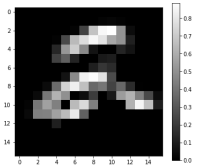




-1	0	1
-1	0	1
-1	0	1



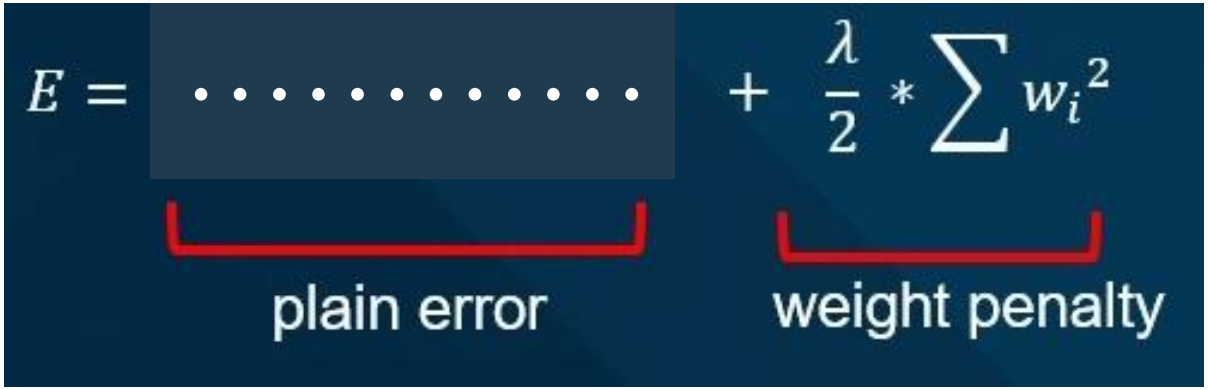
-1	-1	-1
0	0	0
1	1	1



1	1	1
0	0	0
-1	-1	-1

L2 regularization

- Regularization makes your network more robust



Weight penalty

$0.5\lambda\ (3\ *(-1)2\ +\ 3\ *\ (1)2)\ =3\lambda$

-1	0	1
-1	0	1
-1	0	1

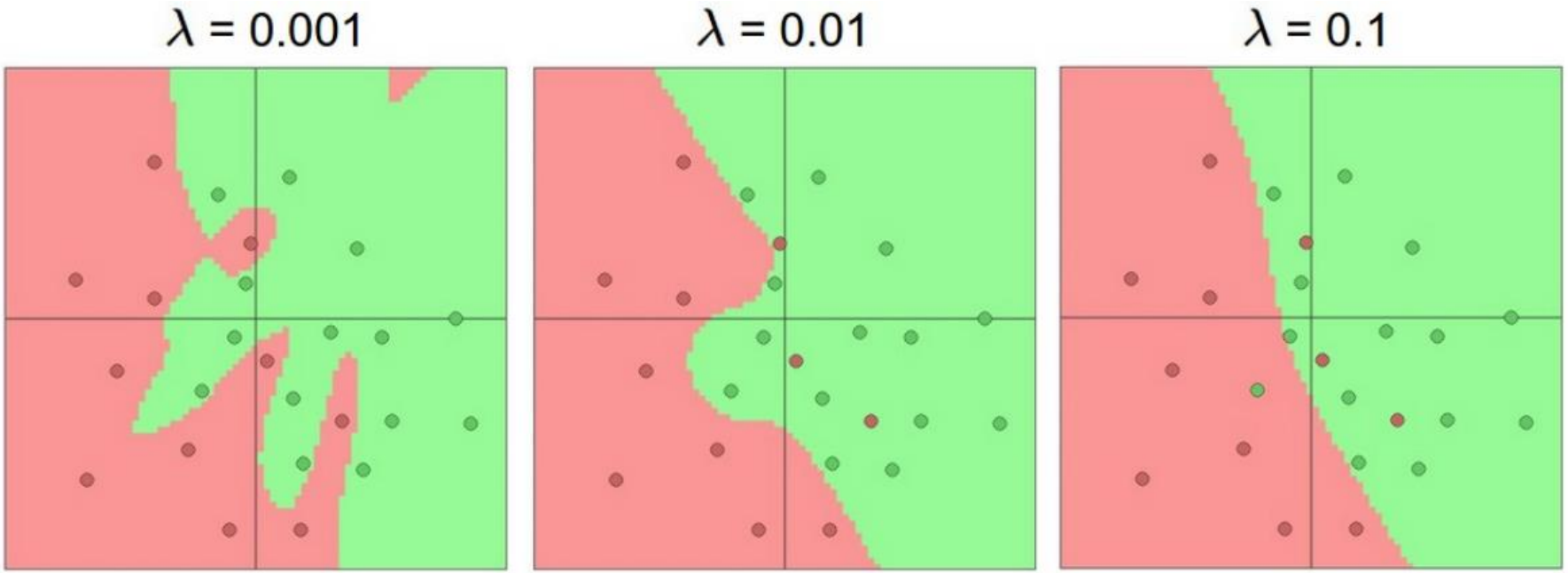
Weight penalty

$0.5\lambda\ (6\ *(-0.5)2\ +\ 3\ *\ (1)2)\ =2.25\lambda$

-0.5	-0.5	1
-0.5	-0.5	1
-0.5	-0.5	1

L2 regularization

- Regularization makes your network more robust



http://cs231n.github.io/neural-networks-1/

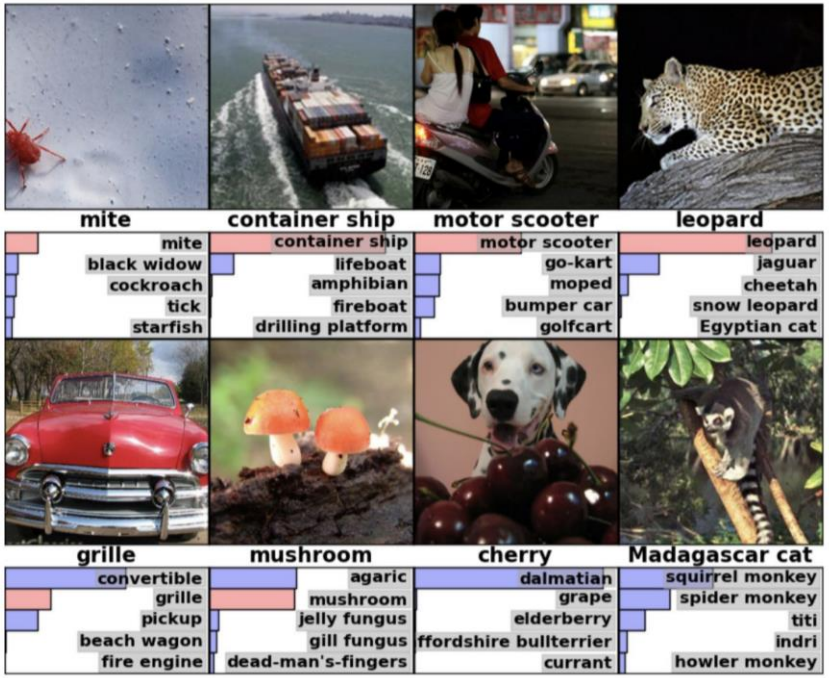
Deep learning in image analysis

Deep learning started its expansion from visual object classification

This work can be extrapolated to classification of medical images

IMAGENET

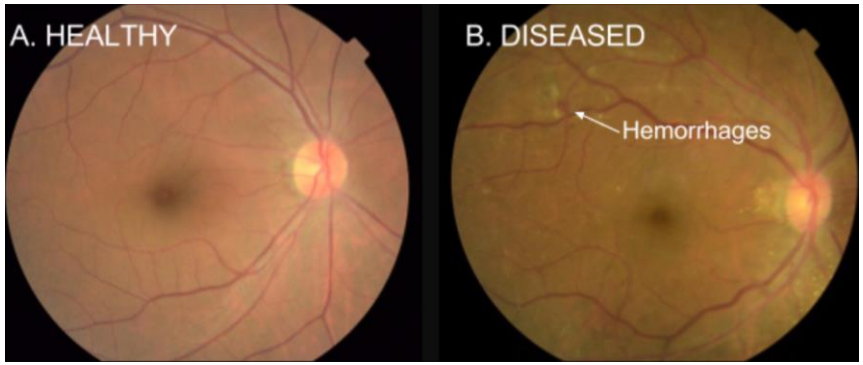
- 1,000 object classes (categories).
- Images:
  - 1.2 M train
  - 100k test.

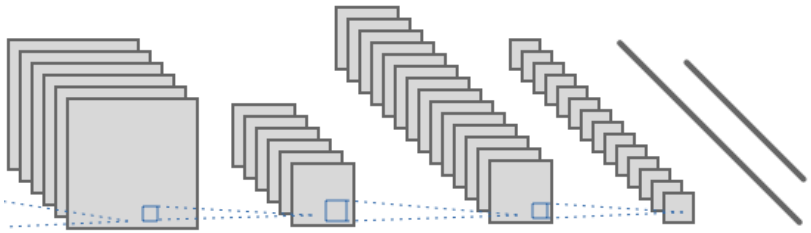


www.image-net.org/

Deep learning in medical imaging: image → classes

Classification into healthy and pathological cases





healthy

diseased

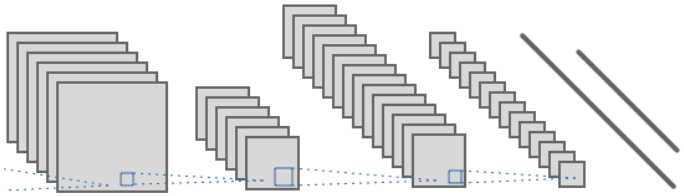
CNN for classification

*Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

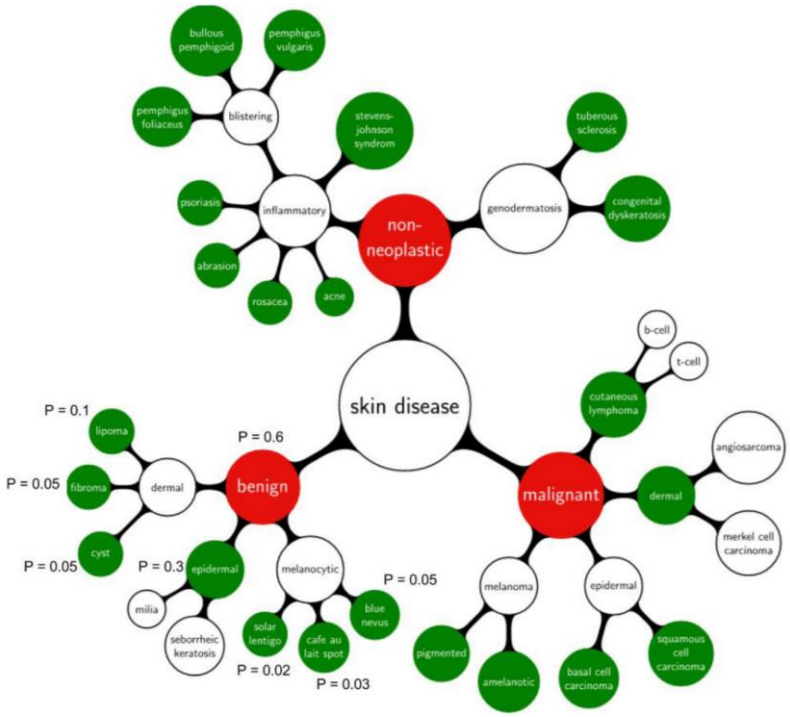
Deep learning in medical imaging: image → classes

Classification labels may form complex taxonomy:

- Aggregation of rare outputs
- Class imbalance problem



CNN for classification



*Dermatologist-level classification of skin cancer with deep neural networks

Deep learning in medical imaging: image → classes

Preservation of the input image size is not needed for classification task

Deep Visualization Toolbox

yosinski.com/deepvis

#deepvis


Jason Yosinski


Jeff Clune


Anh Nguyen


Thomas Fuchs


Hod Lipson

Cornell University

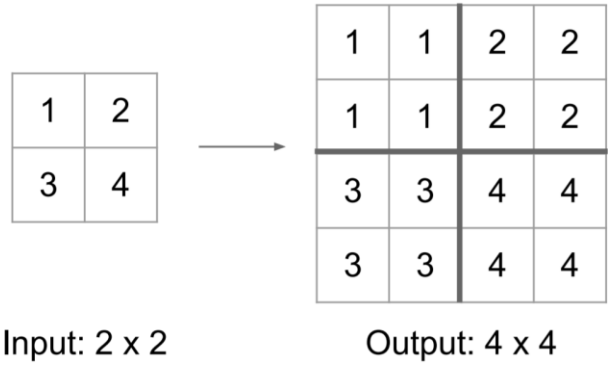
University of Wyoming

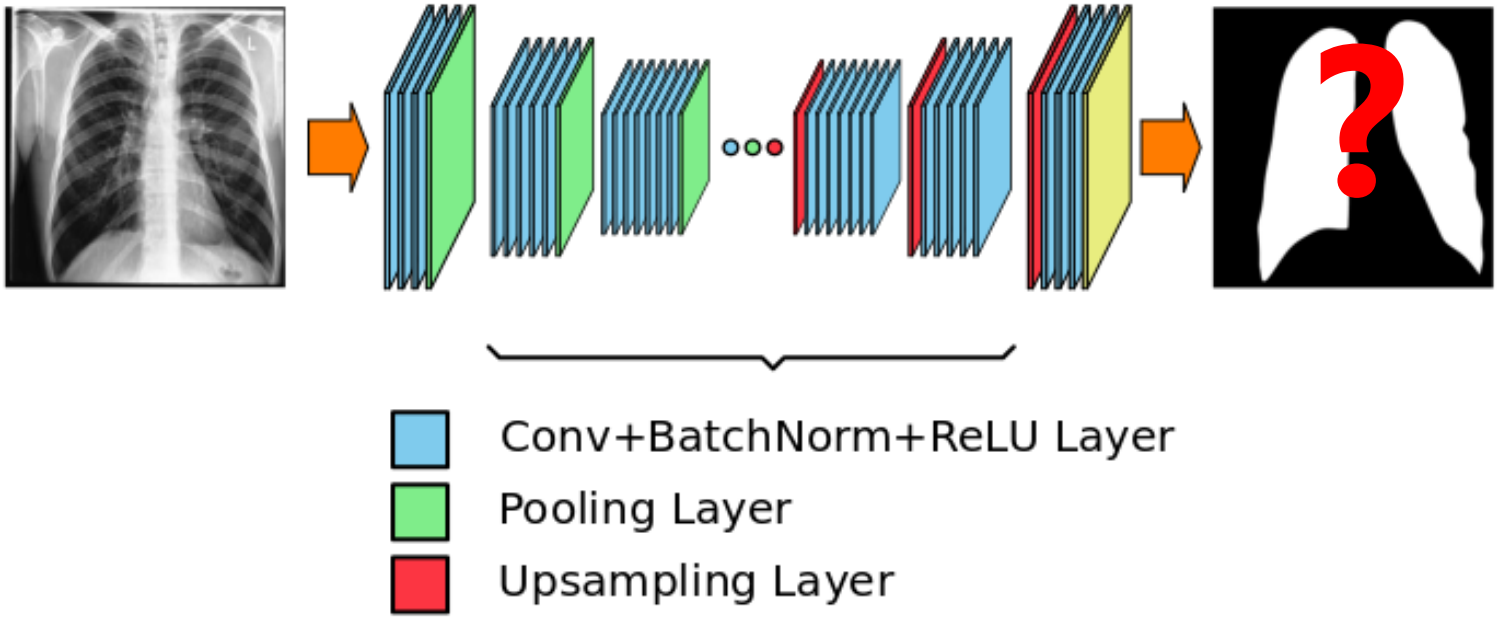
NASA Jet Propulsion Laboratory
California Institute of Technology

http://yosinski.com/

Deep learning in medical imaging: image → image

Upsampling layer

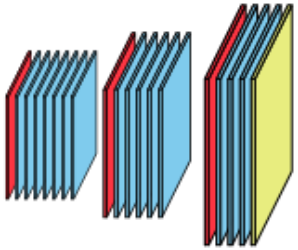


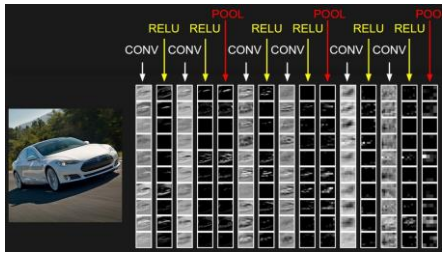


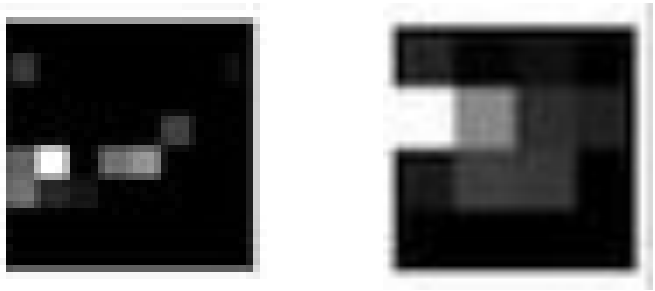
Deep learning in medical imaging: image → image

What happens with fine image details after upsampling?

Can we expect that network layers after upsampling will recover fine object border?







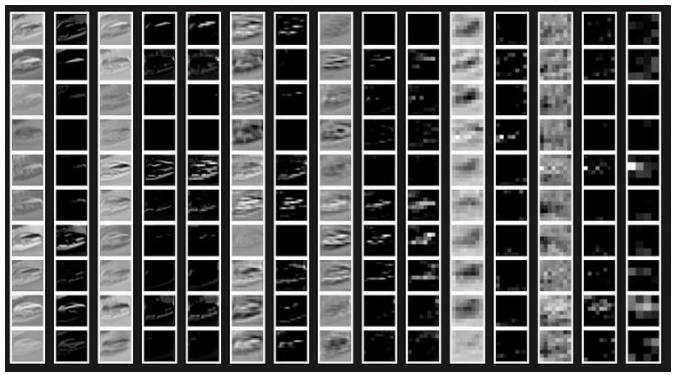
Can we recover side window borders from feature maps that look like this?

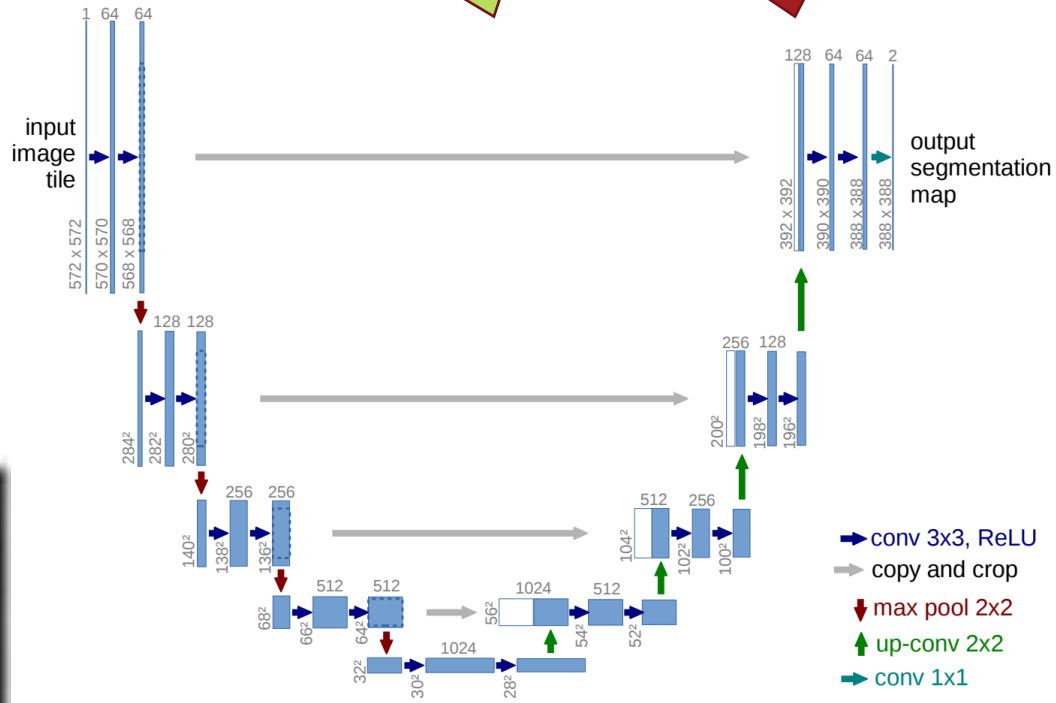
http://cs231n.stanford.edu/

Deep learning in medical imaging: image → image

UNet architecture for segmentation:

- Skip connections pass fine details from early layers to late layers





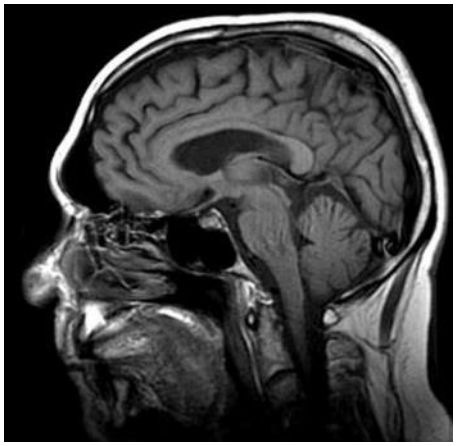
Can we recover side window borders from feature maps that look like this?

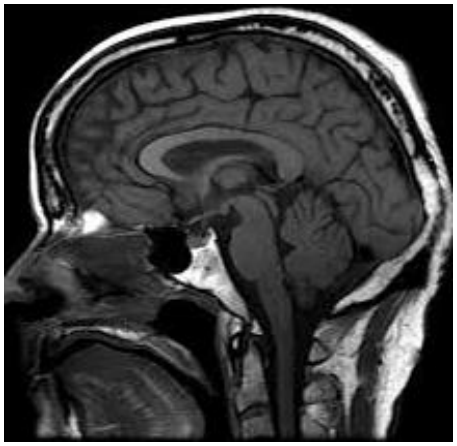
Medical image analysis vs Computer vision

Medical images are better standardized:

- Similar imaging protocols (resolution, slice thickness, etc.)
- Standardized patient positioning

Natural images of the same object look very different



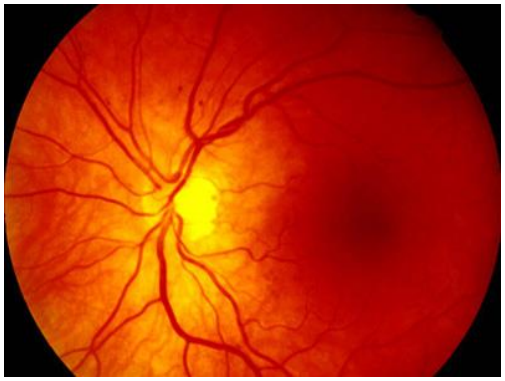


Brain MRIs are way more similar

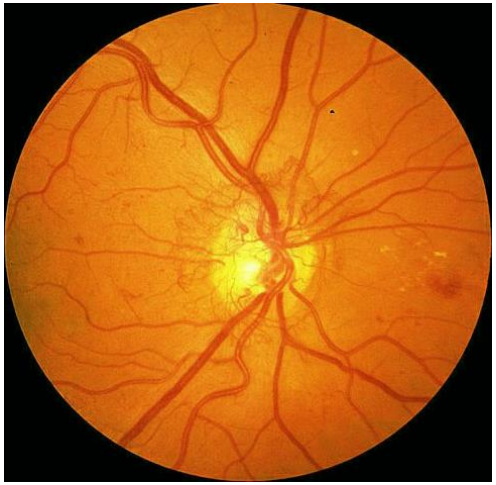
Medical image analysis vs Computer vision

The difference between natural images of different types is also better pronounced than for medical images

Cats are very different from planes



Healthy retina

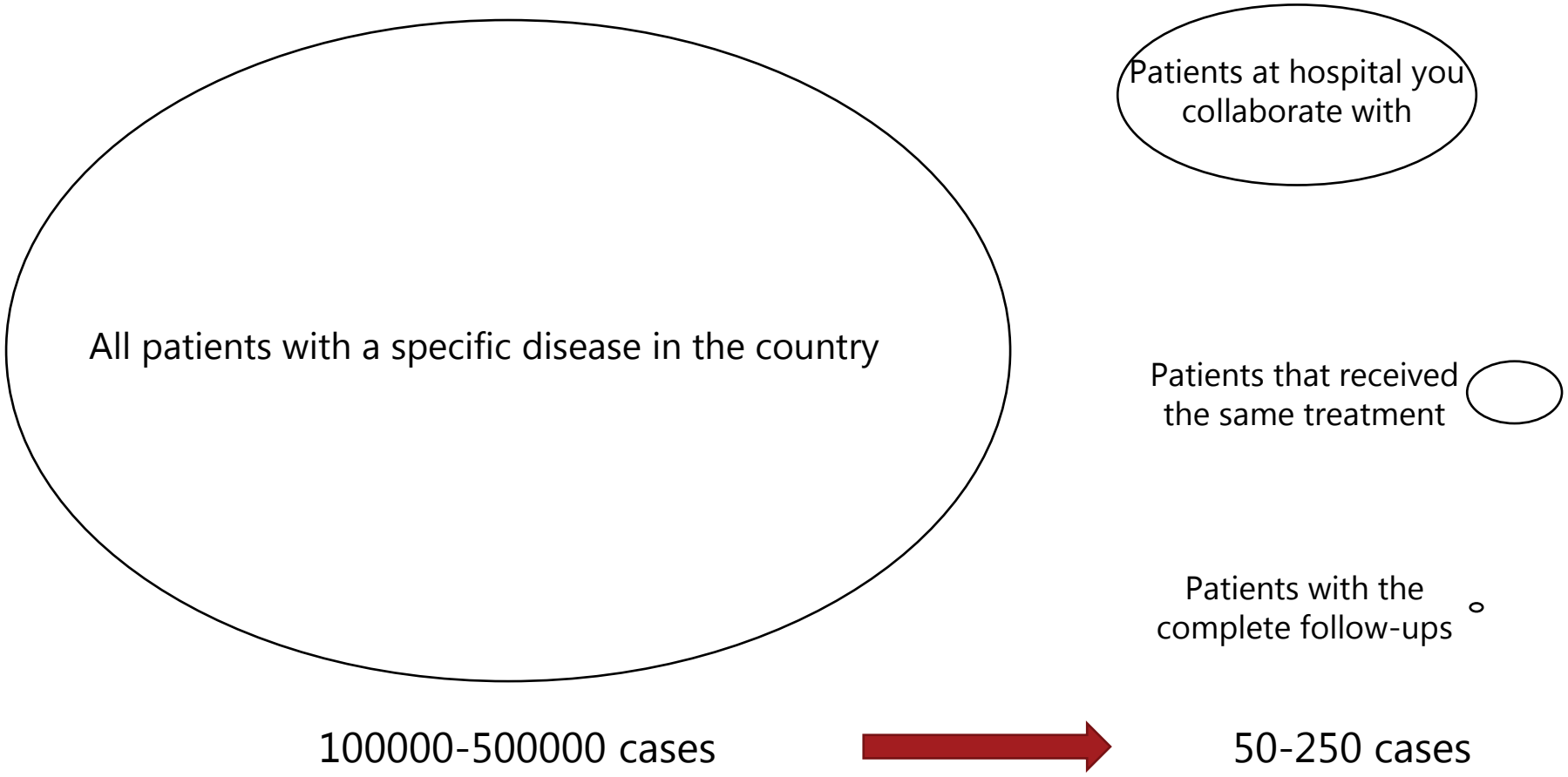


Diabetic retina

Medical image analysis vs Computer vision

The number of available medical images is by several orders of magnitude lower than of natural images:

- Google ‘images of cats’ and get 25000 images in the first link



Training-validation-testing

Training images – training features of the model

Validation images - training hyperparameters of the model

Testing images – used at the very end, when the model is locked and will not be anymore edited

Training images

Validation images

Testing images

Loss: Binary cross-entropy

```
model.compile(optimizer = Adam(lr = 1e-4), loss = 'binary_crossentropy',
metrics = ['accuracy'])
```

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i))$$

True label	Prediction	Loss
1	0.3	1.2

-(1 * ln(0.3)+(1-1)*ln(1-0.3)) =
-ln(0.3) = 1.204

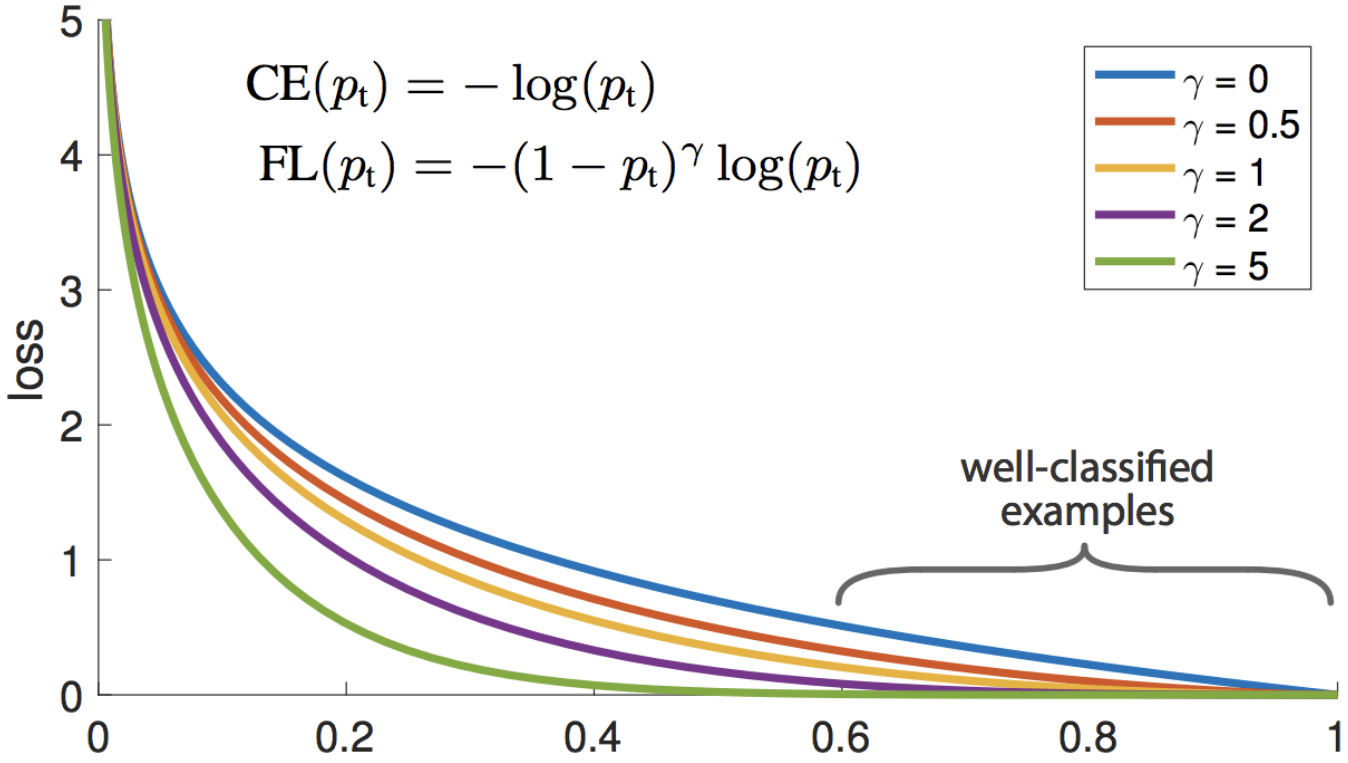
True label	Prediction	Loss
1	0.9	0.1

-(1 * ln(0.9)+(1-1)*ln(1-0.1)) =
=-ln(0.9) = 0.105

Focal loss

Class imbalance (large size of the background) is a problem for binary cross-entropy

Focal loss focuses on difficult-to-classify cases



https://arxiv.org/pdf/1708.02002.pdf

Dice loss

We can use Dice coefficient to compute loss

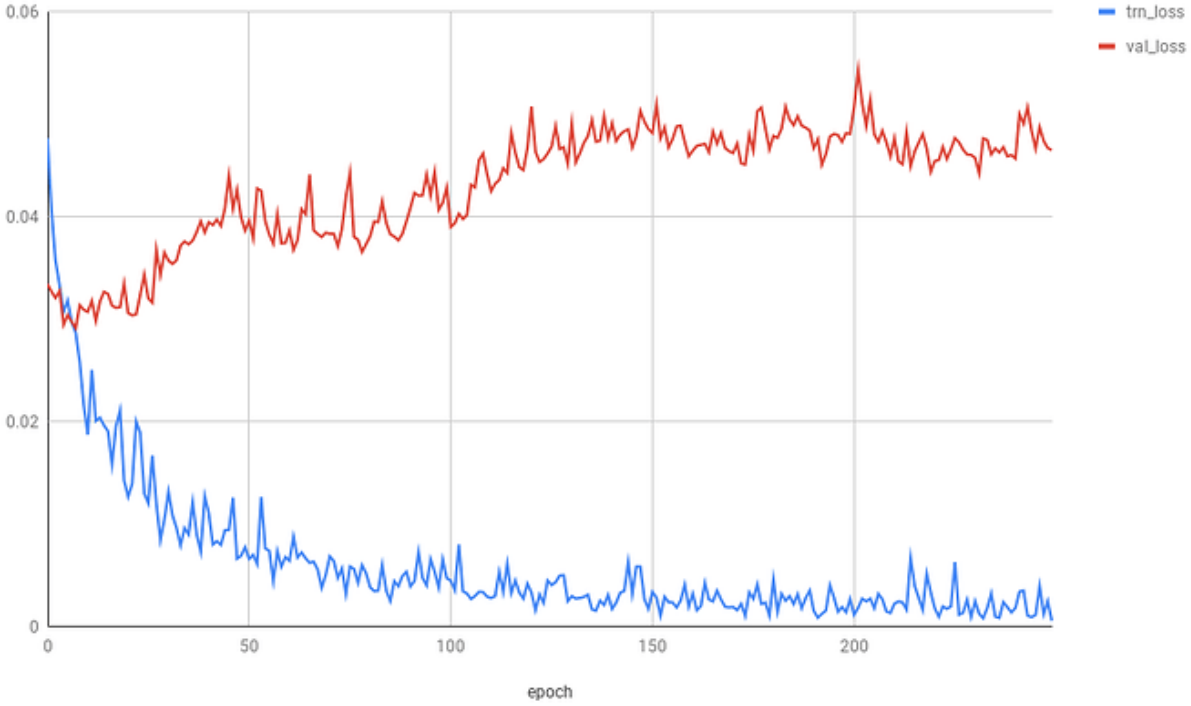
No problem with background size

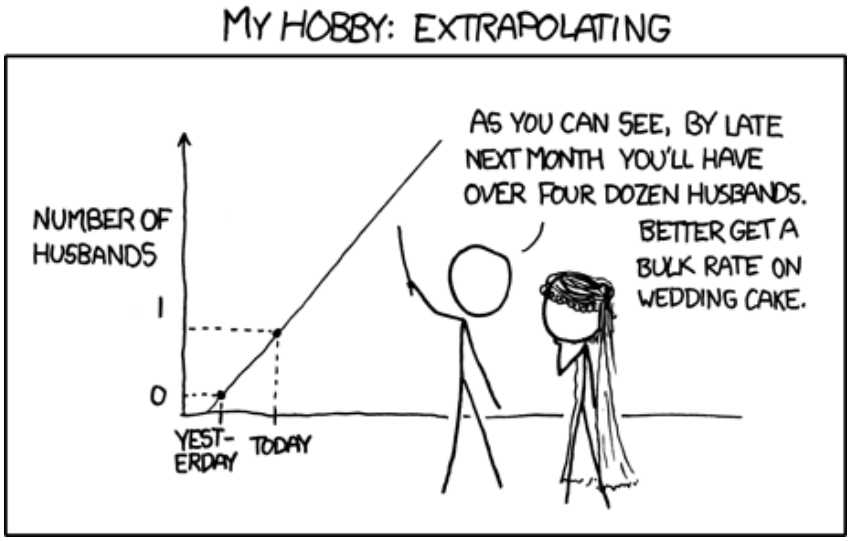
```
def __dice_coefficient(y_true, y_pred, smooth=1.):
    y_true_f = keras.flatten(y_true)
    y_pred_f = keras.flatten(y_pred)
    intersection = keras.sum(y_true_f * y_pred_f)
    return (2. * intersection + smooth) / (keras.sum(y_true_f) + keras.sum(y_pred_f) + smooth)

def __dice_coefficient_loss(y_true, y_pred):
    return 1 - NN_Architectures.__dice_coefficient(y_true, y_pred)
```

Overfitting

Ways to reduce overfitting:

- Reducing the number of features in convolution layers
- Regularization
  - Dropout
  - L2 regularization
- Data augmentation





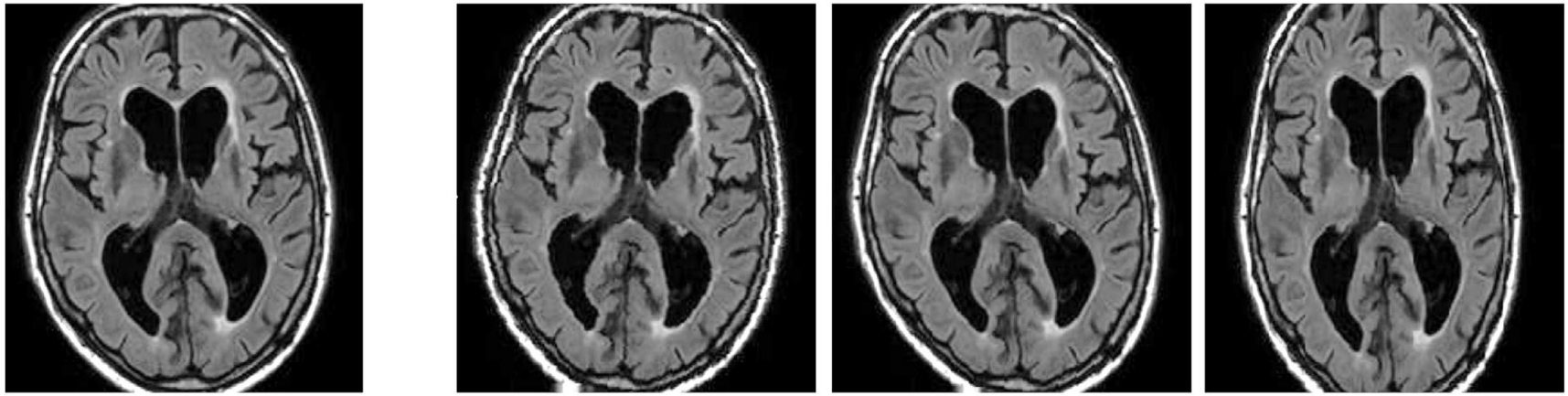
Questions?

Medical image augmentation

Image augmentation is useful:

- To enrich training database
- To address the class imbalance problem

Affine transformations (anatomical restrictions may apply)



original slice | rotation | shear mapping | scaling

*Fully Convolutional Network Ensembles for White Matter Hyperintensities Segmentation in MR Images

Medical image augmentation

Intensity transformations:

- Noise
- Blur
- Contrast
- etc

Original image

Random brightness

Random gamma value

Random Gaussian noise

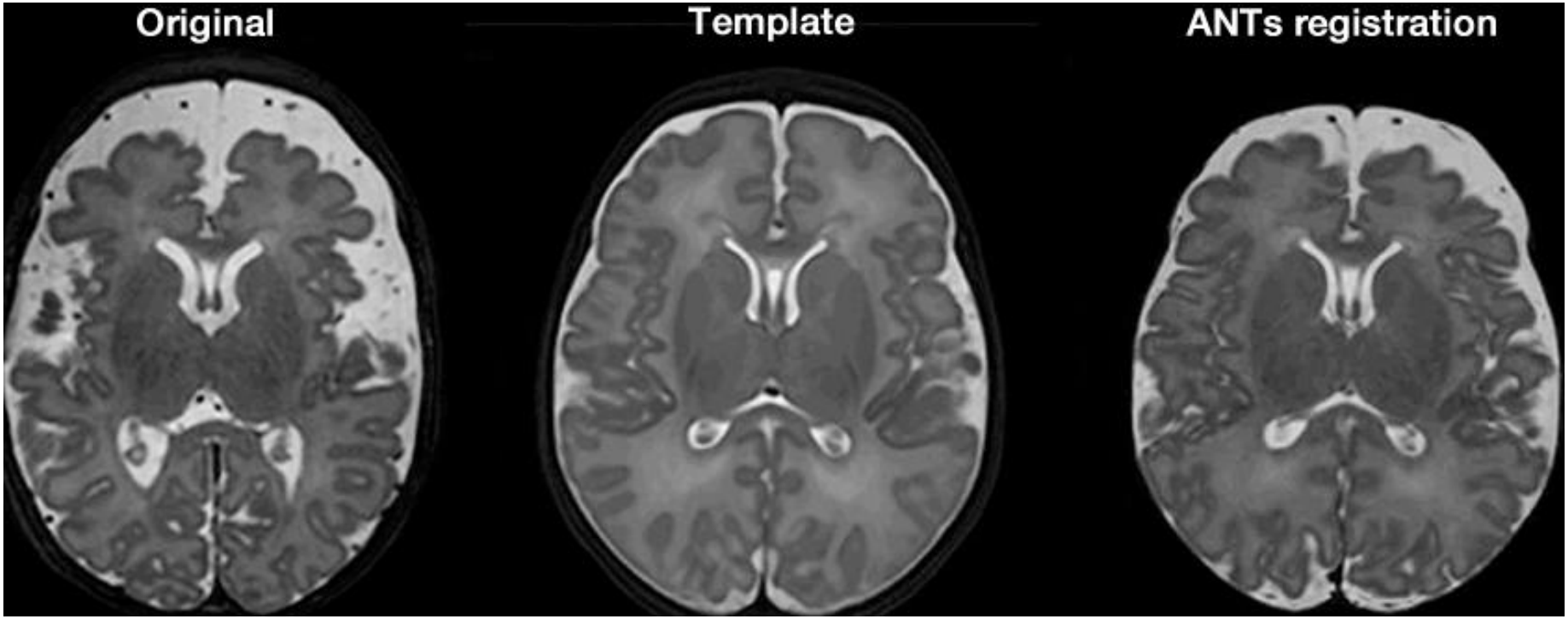
Random Gaussian blur

*Data Augmentation for Chest Pathologies Classification

Medical image augmentation

Registration-based augmentation:

- Very useful for medical in contrast to natural images



Medical image augmentation

Style transform simultaneously preserves both:

- Content of the target image (neuron activations – critical-to-network information that leads to content image classification)
- Style from the style image (pairwise activation covariances – the neurons that get activated together, i.e. consistent appearance patterns)

https://deepdreamgenerator.com/

Medical image augmentation

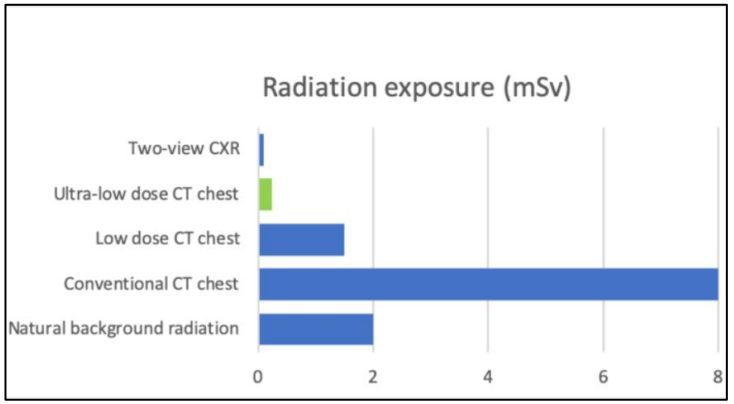
Style transform for medical image augmentation

What if content image is for a healthy case,
while the style image is for a pathological case?

*Data augmentation for improving deep learning in image classification problem

Image enhancement: dose/quality problem

- The dose of CT imaging has dramatically reduced during the last 10 years
- The number of CTs per capita increased 3 times in the last 10 years



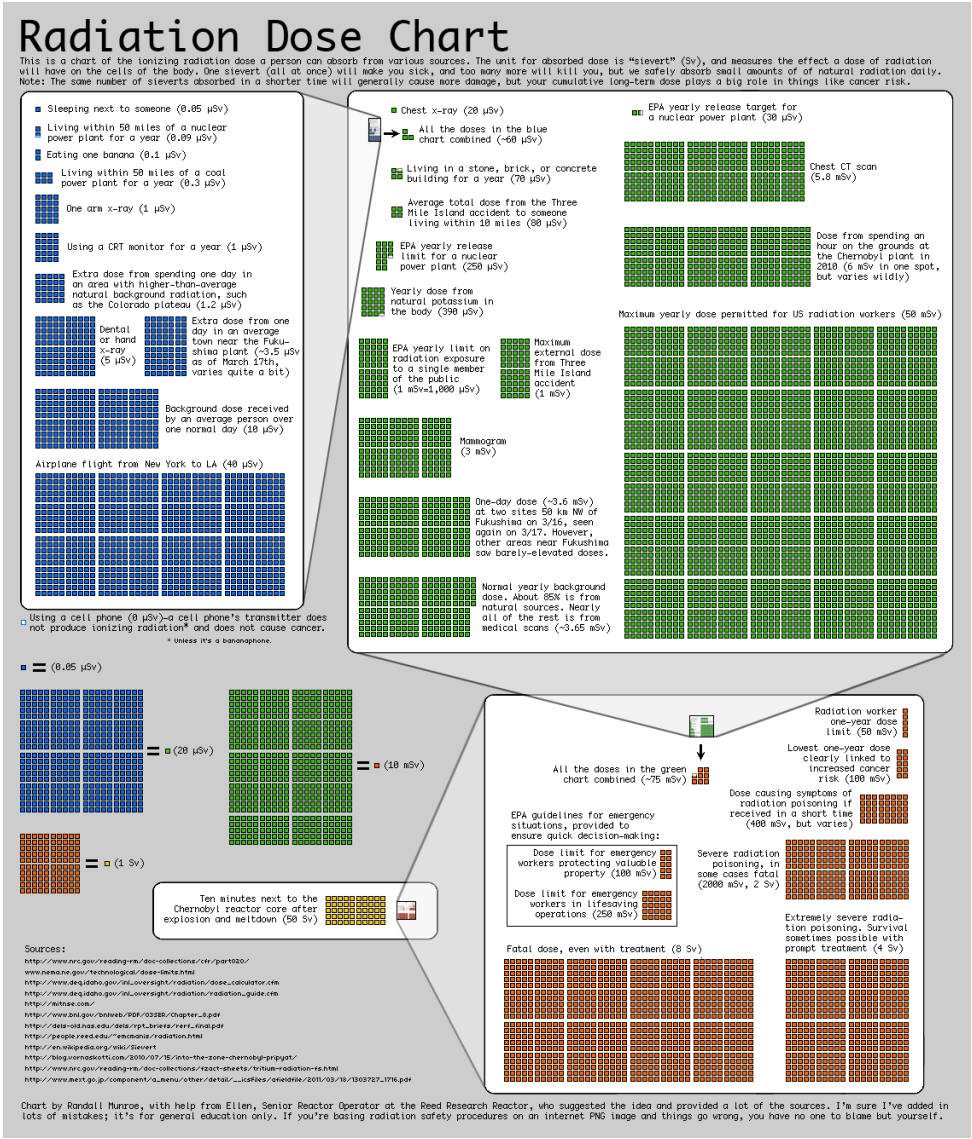
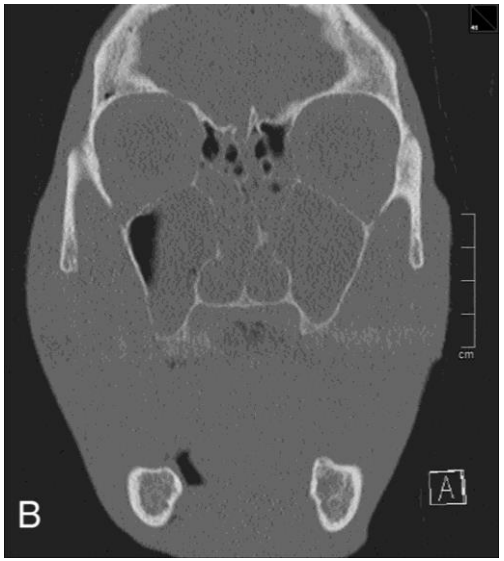
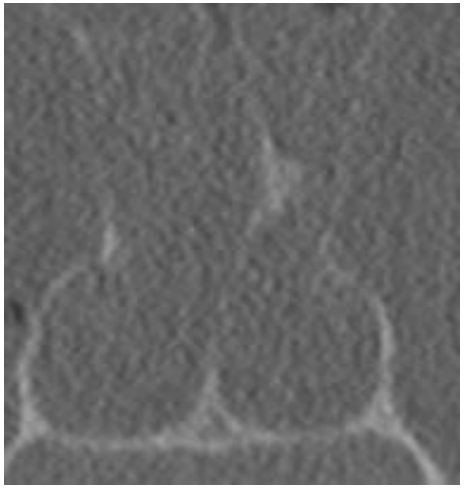


Image enhancement: dose/quality problem



Dose reduction results in reduction of the image quality



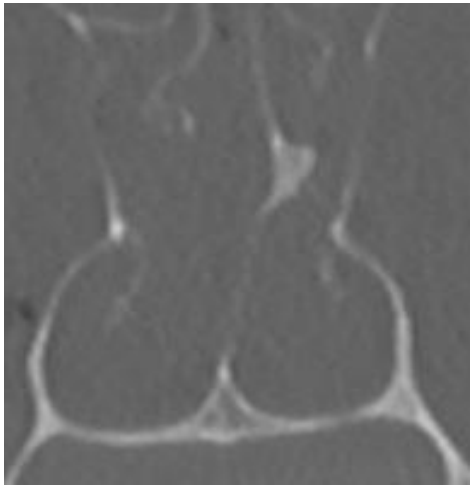
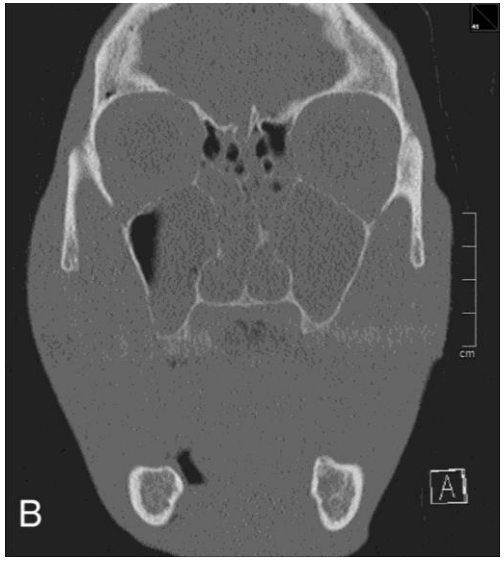


Image enhancement: dose/quality problem

- Input: low-resolution images
- Output: corresponding high-resolution images

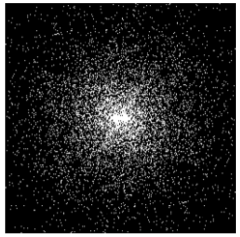
- The input can be generated from high-resolution images by image blurring and adding noise



Encoder-decoder NN

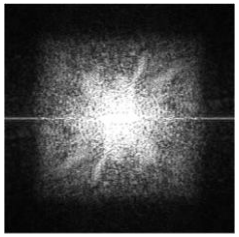
Image enhancement

We can use raw data instead of reconstructed medical images



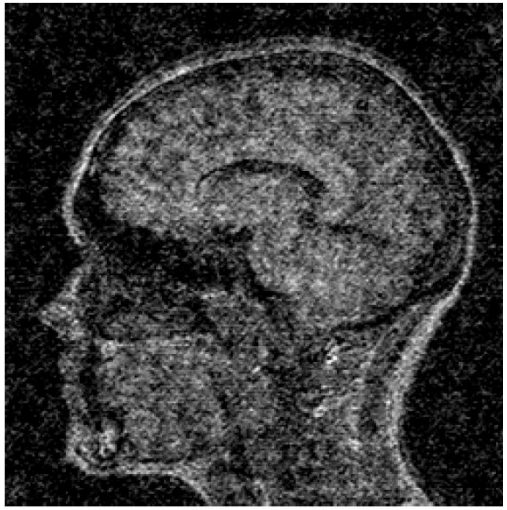
Undersampled spectrum

Encoder-decoder NN

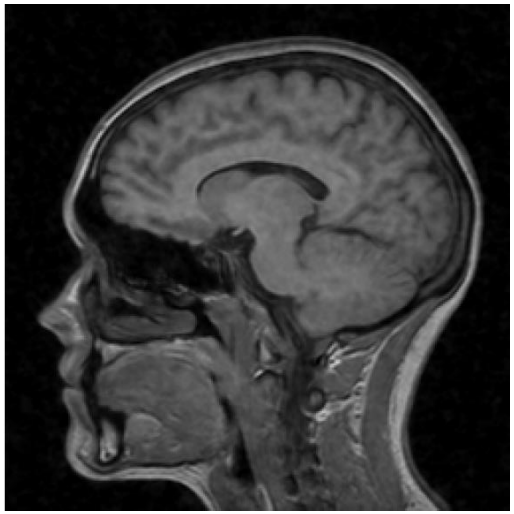


High-resolution spectrum

MRI from undersampled raw data



Reconstructed MRI



Ground truth

*Noise2Noise: Learning Image Restoration without Clean Data

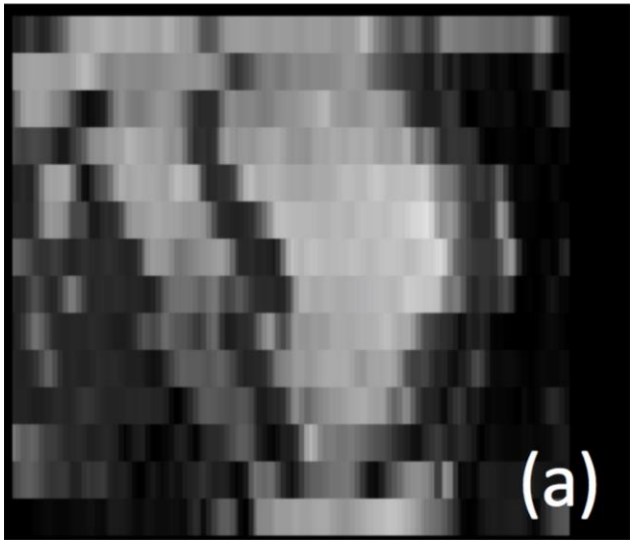
Image enhancement: improved resolution

3D imaging takes time:

- Patients move slightly during acquisition
- Movement of some organs cannot be controlled, e.g. heart
- More motion –more blur

How to image moving organs?

- Reduce overall image quality
- Sacrifice one image dimension

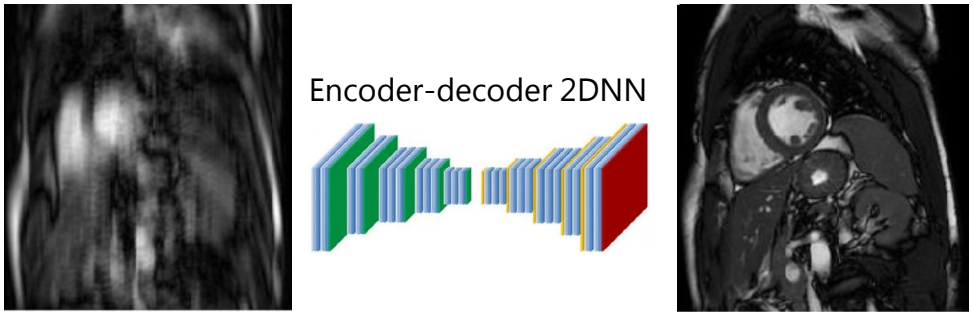


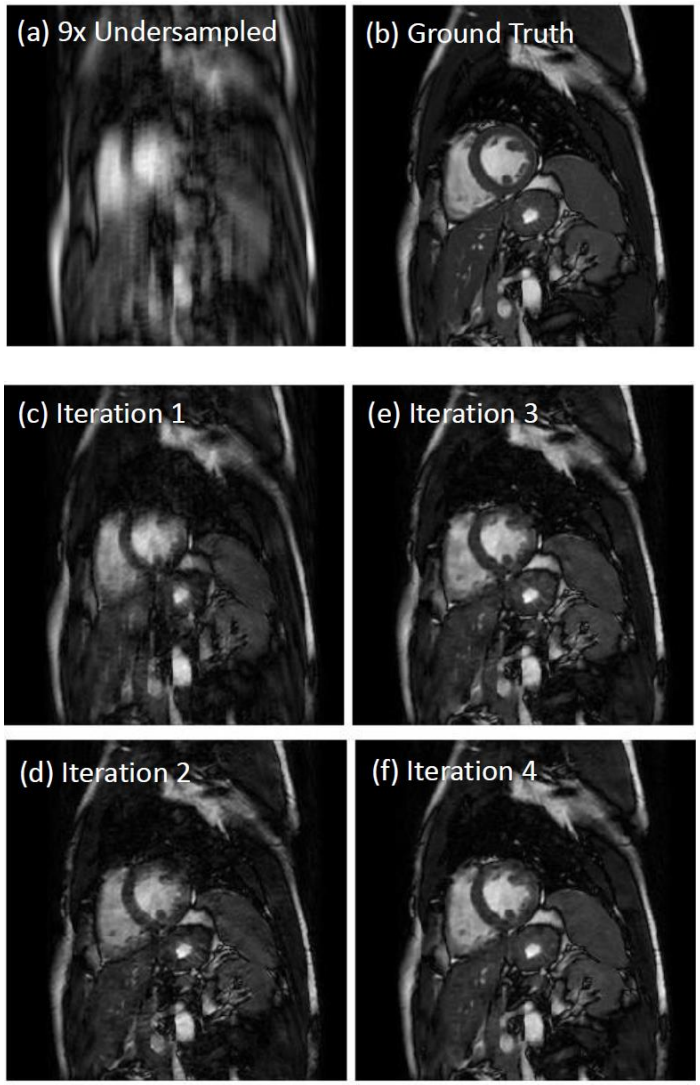
One option – sacrifice one dimension

Image enhancement : improved resolution

One solution:

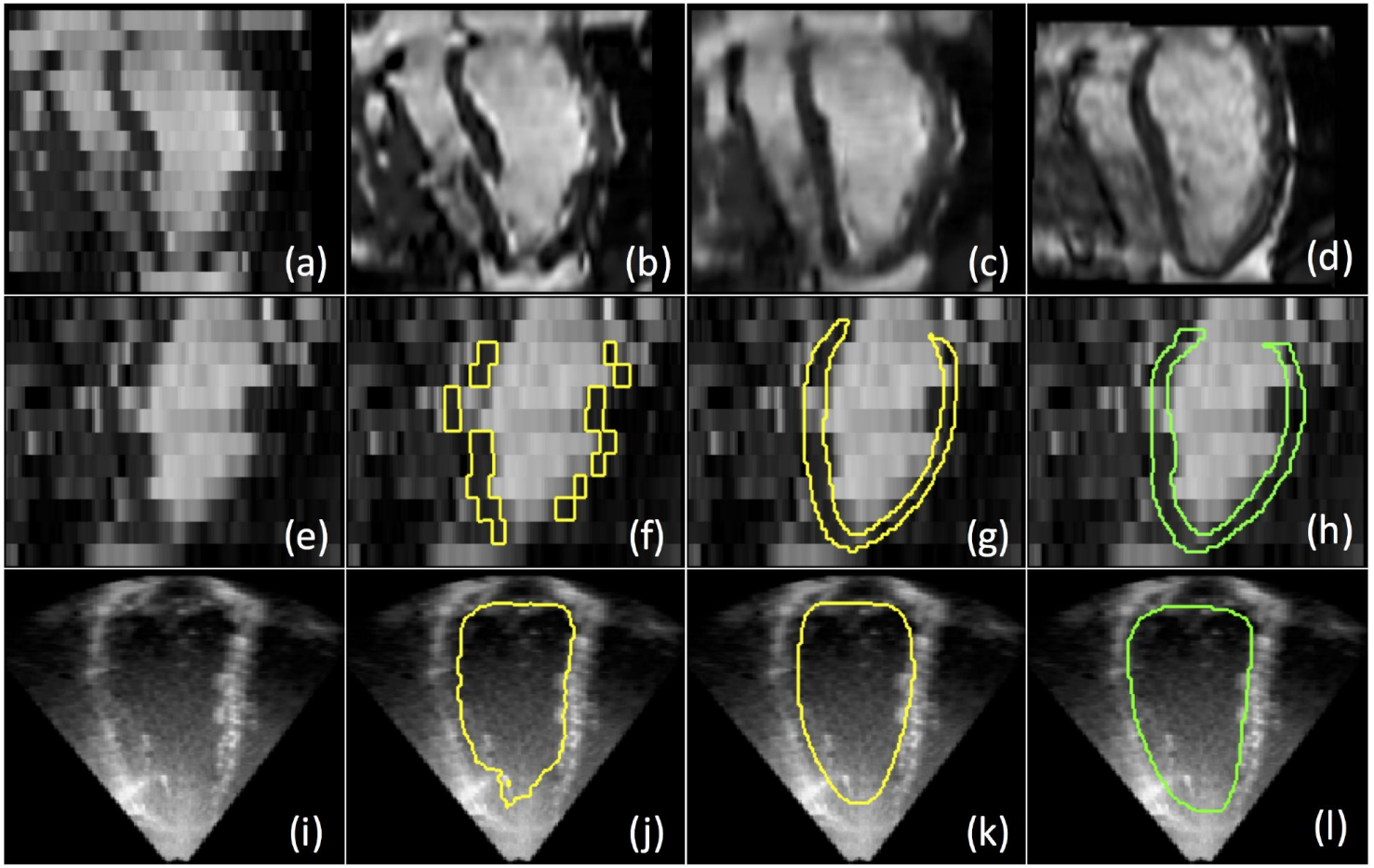
- Acquire axially(coronally/sagittally)-reconstructed images or high-resolution images
- Artificially vandalize them
  - Undersample spectrum
  - Reduce resolution in one dimension
- Train the network to recover the original images





*Convolutional Recurrent Neural Networks for Dynamic MR Image Reconstruction

Image enhancement: improved resolution



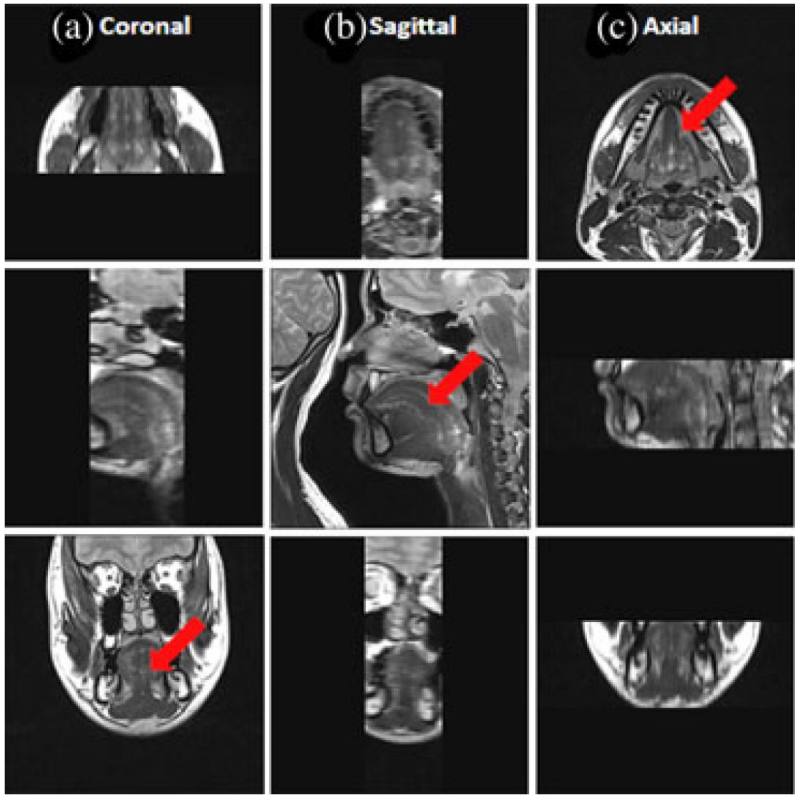
Low-resolution CNN version 1 CNN version 2 Original

* Anatomically Constrained Neural Networks (ACNN): Application to Cardiac Image Enhancement and Segmentation

Image enhancement: improved resolution

Tongue image analysis:

- Training: we have 3 orthogonal images that, however, do not cover the whole 3D volume



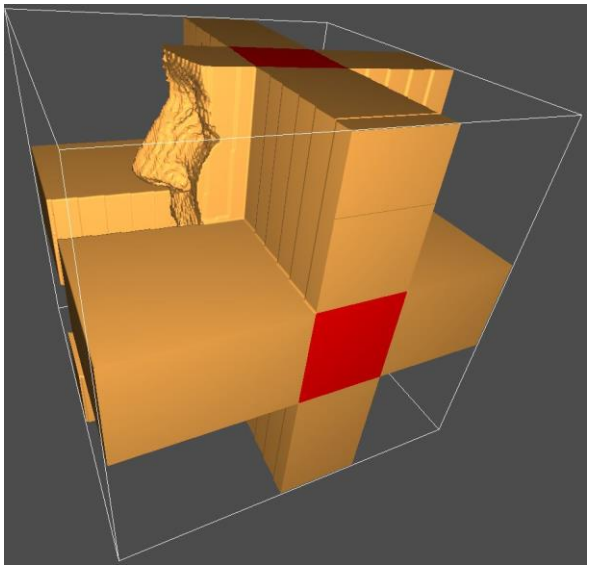
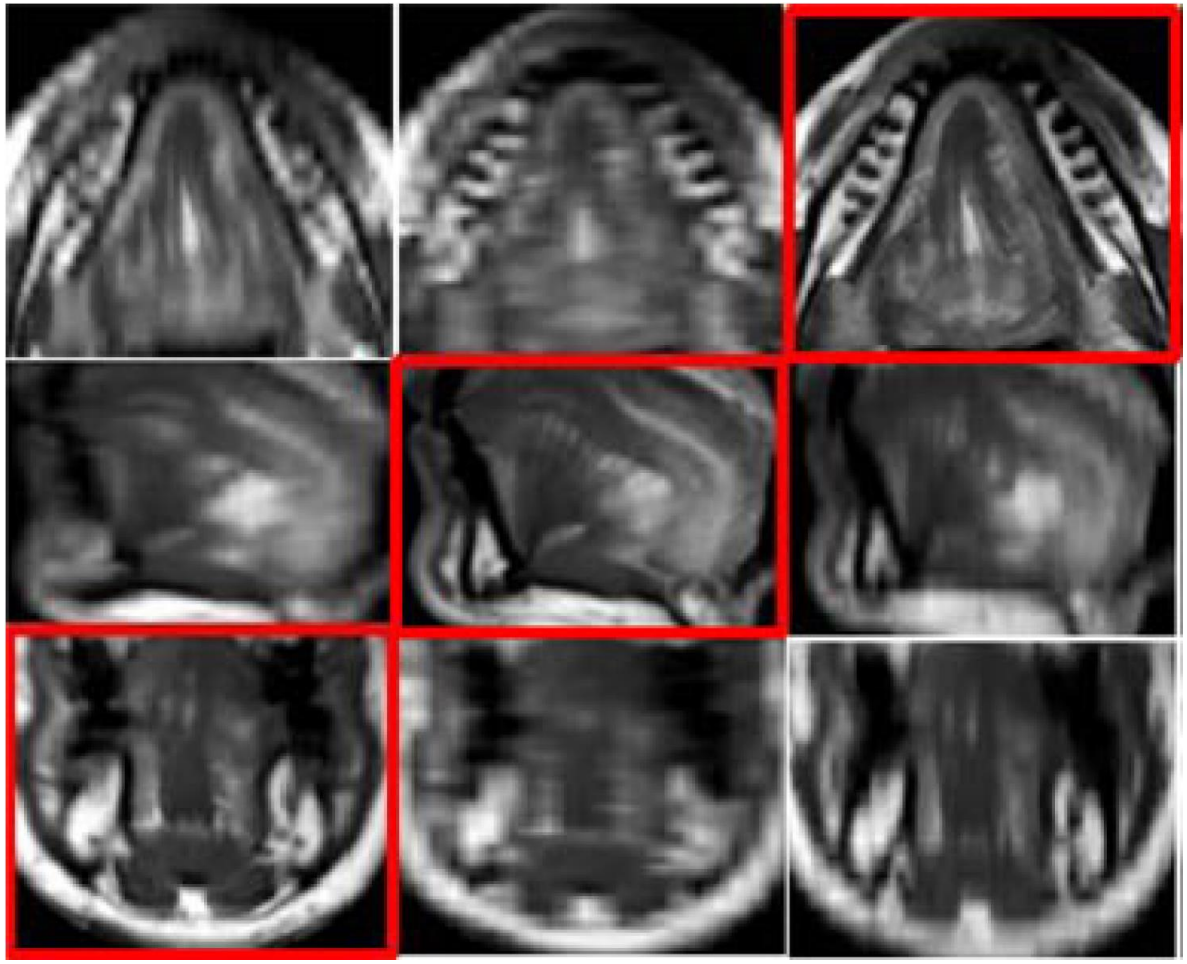


Image enhancement: improved resolution

Tongue image analysis:

- Use image registration to align images
- Merge aligned images into one
- Train CNN to recover the merged image from individual orthogonal image



* Reconstruction of high-resolution tongue volumes from MRI

Image enhancement: artifacts removal

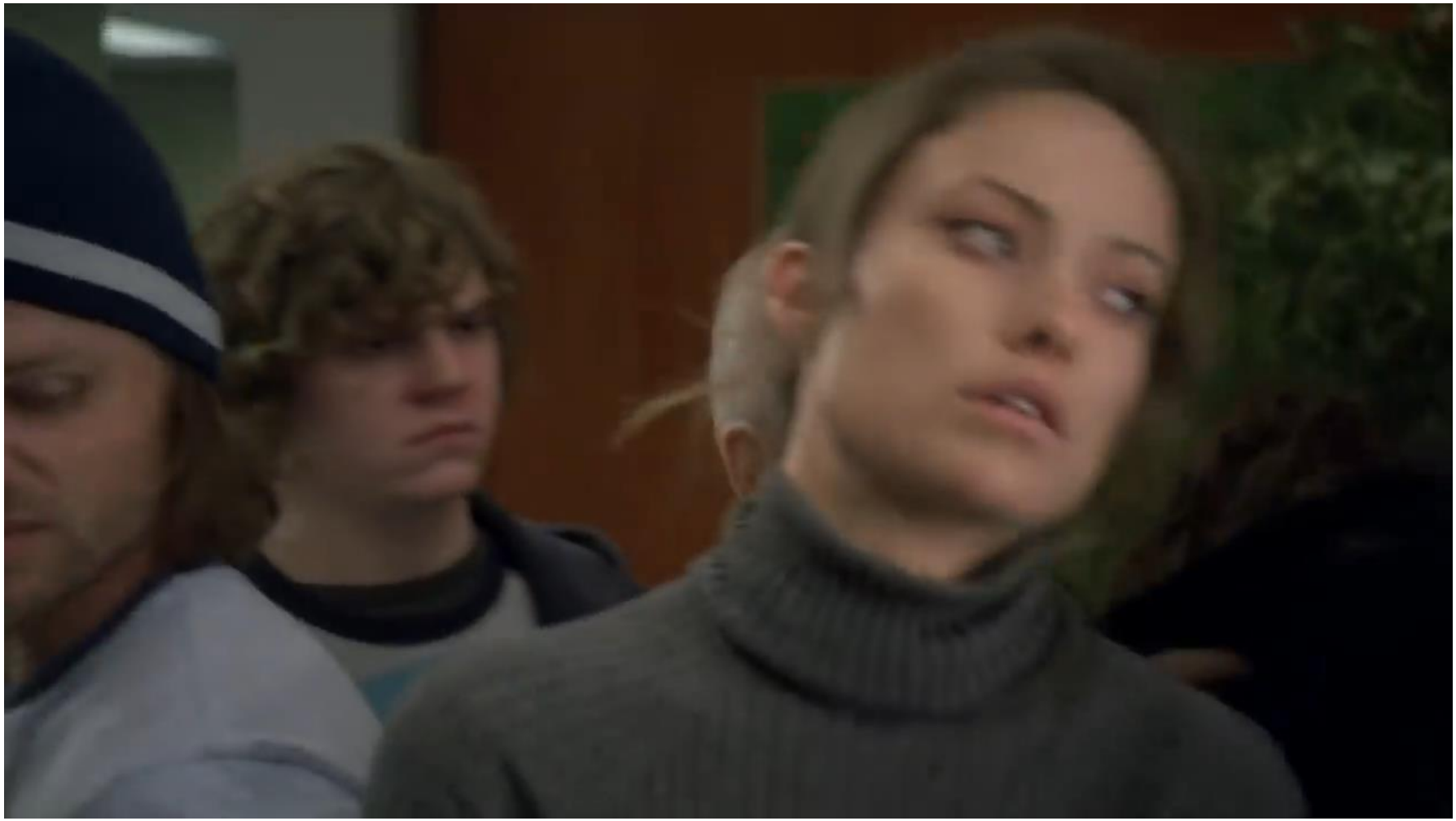
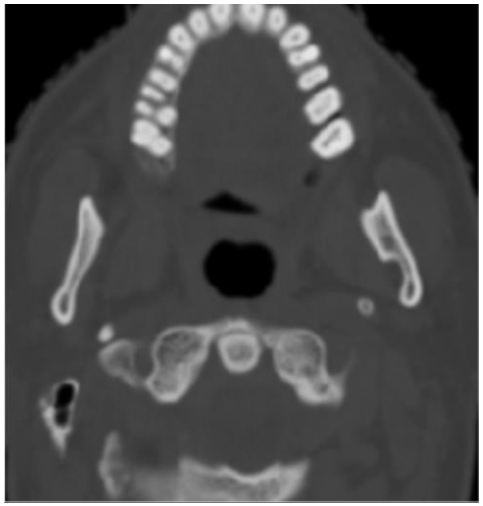
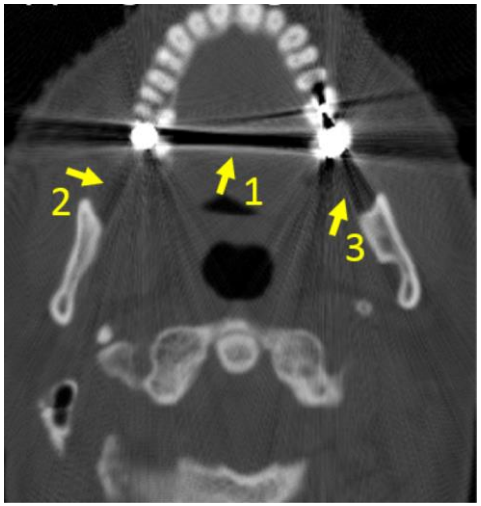


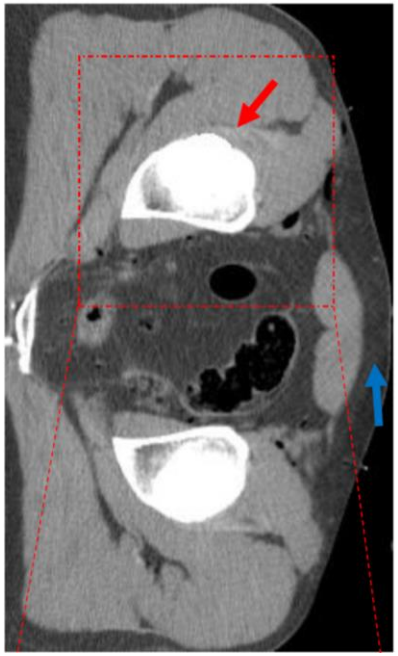
Image enhancement: artifacts removal

Image metal artifacts:

- Dental metal crowns
- Femoral head implants
- Pedicle screws
- etc.







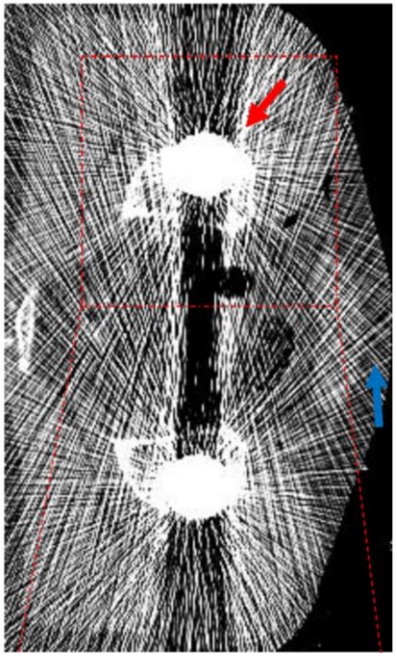
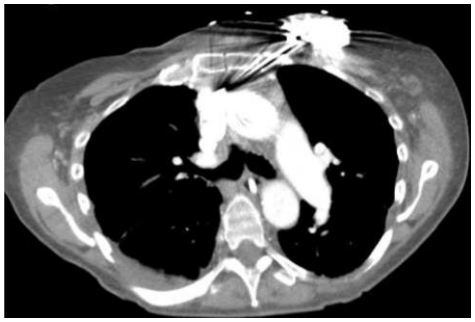


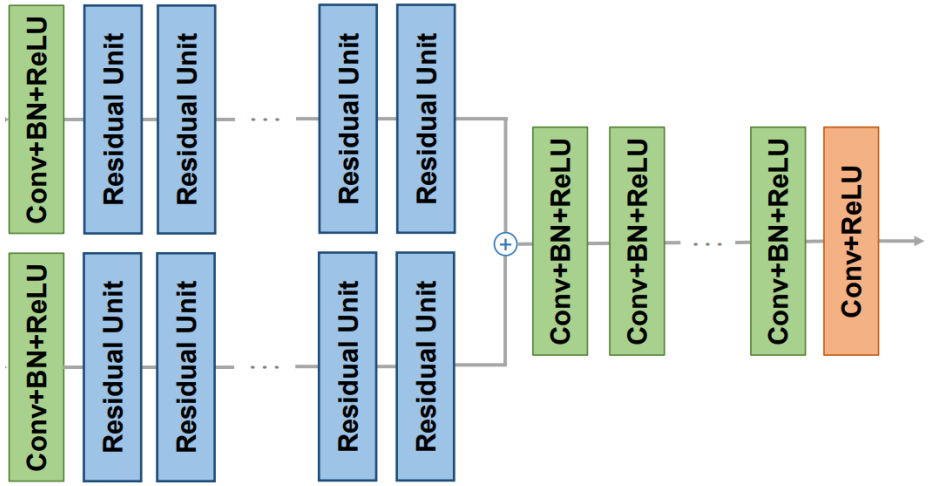
Image enhancement: artifacts removal

We can use only CT, only sinogram or both sinogram and CT image with artifacts



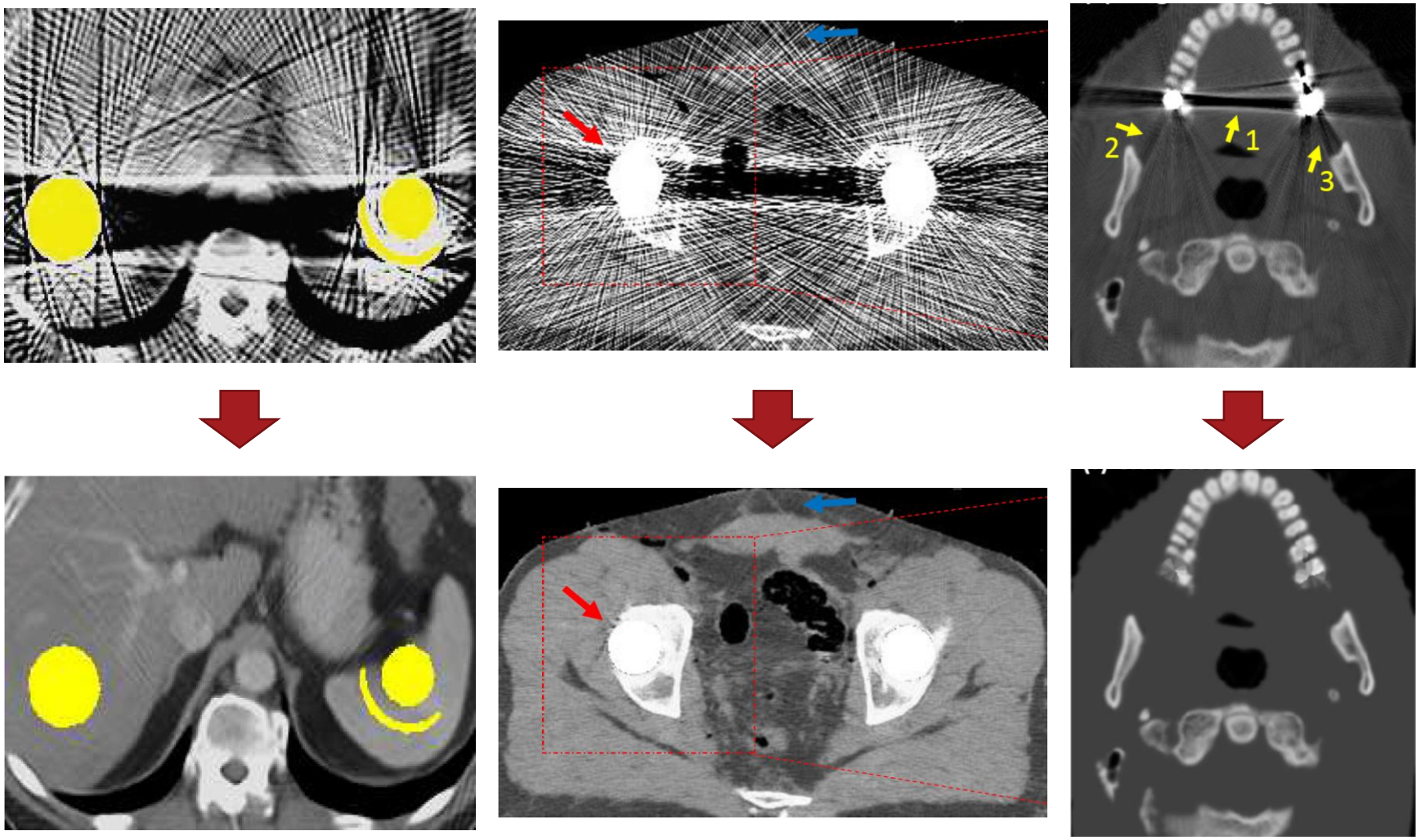
Original Image

Sinogram



CT image without artifacts

Image enhancement: artifacts removal



*DuDoNet: Dual Domain Network for CT Metal Artifact Reduction
Deep Neural Network for CT Metal Artifact Reduction with a Perceptual Loss Function
Convolutional Neural Network based Metal Artifact Reduction in X-ray Computed Tomography

Image synthesis

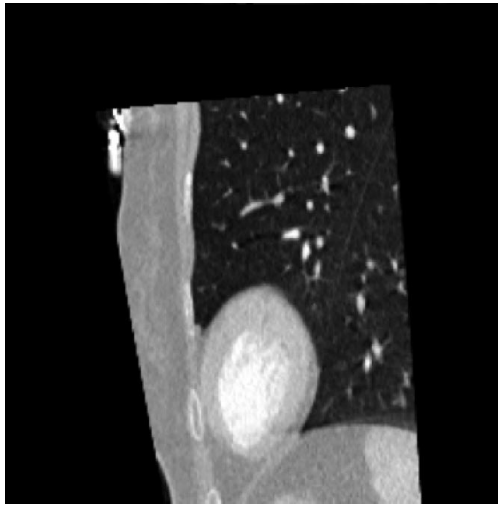
Artificially generate one image modality from another on for the same patient

How can this be useful?

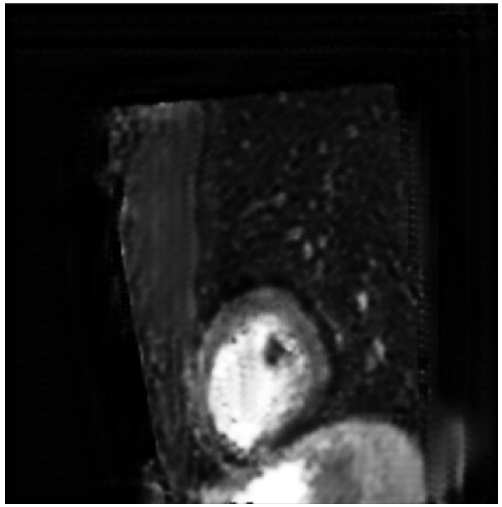
Image synthesis

Example:

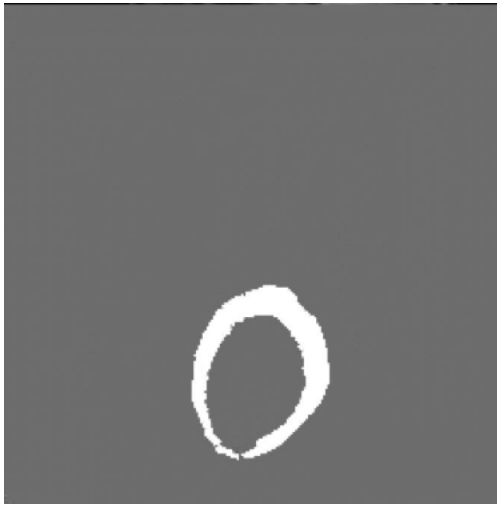
- Let's say we have an algorithm for segmentation /classification /detection that requires both CT and MR
- A new patient has only MR
- Using image synthesis we can generate artificial CT



real CT



synthetic MR



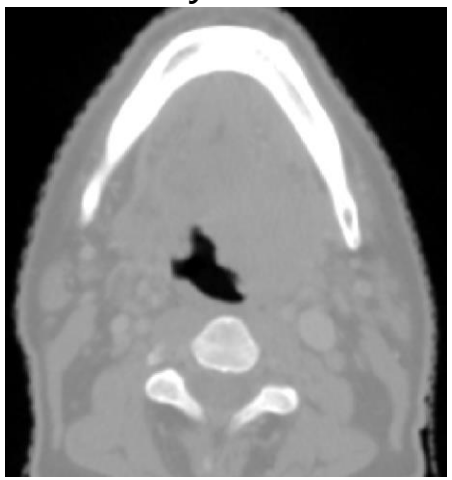
ventricle mask

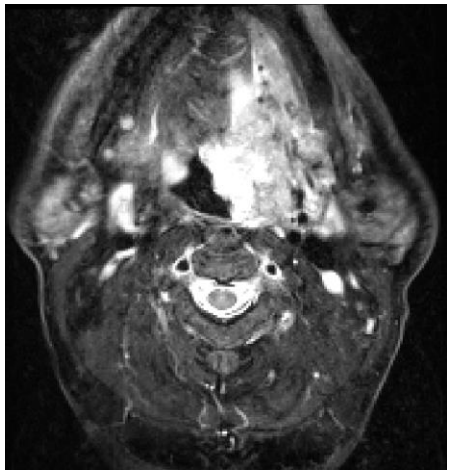
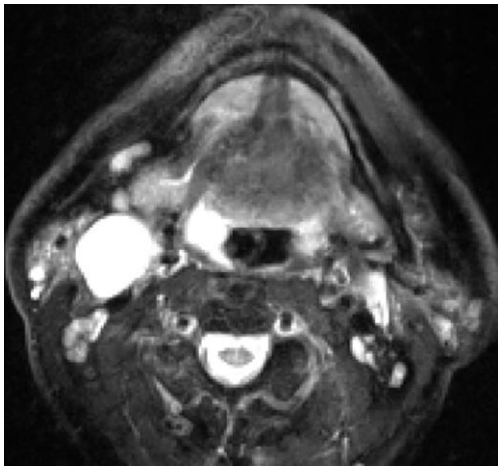
Adversarial Image Synthesis for Unpaired Multi-Modal Cardiac Data

Segmentation through synthesis

Example:

- Most of head and neck tumors are very poorly visible on CT images but very visible on MR T2 images

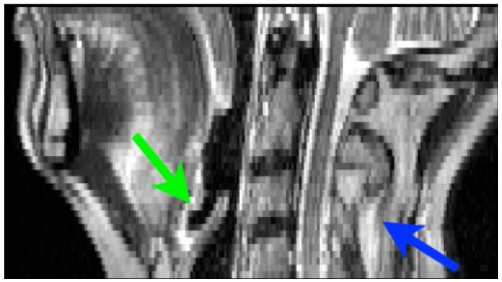


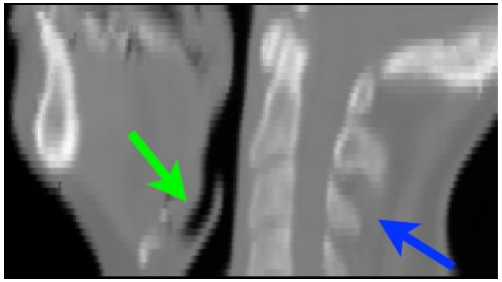
Segmentation through synthesis

- Register MRs to CT images for healthy subjects
- Train network to synthesize MR images from healthy CT
- Apply network to synthesize MR from CT with tumor. This MR will have no tumor because network does not know the tumor appearance and tumor is not very visible on CT
- The difference between original and synthetic MR will be the tumor

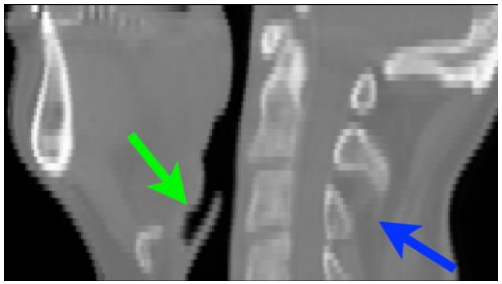
Head-and-neck CT synthesis from MR images



Original MR



Original CT

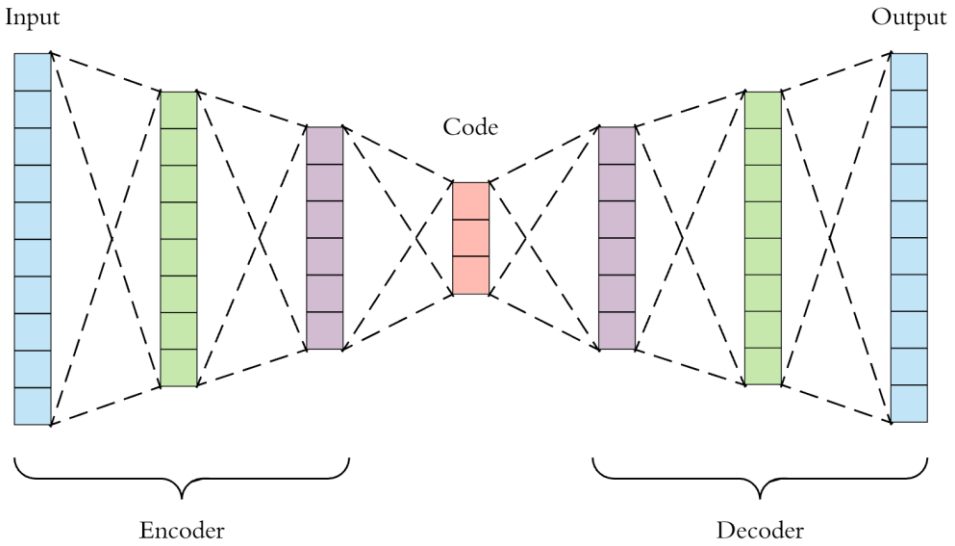


Synthetic CT

*Robust CT Synthesis for Radiotherapy Planning: Application to the Head & Neck Region

Segmentation through synthesis

Autoencoders:



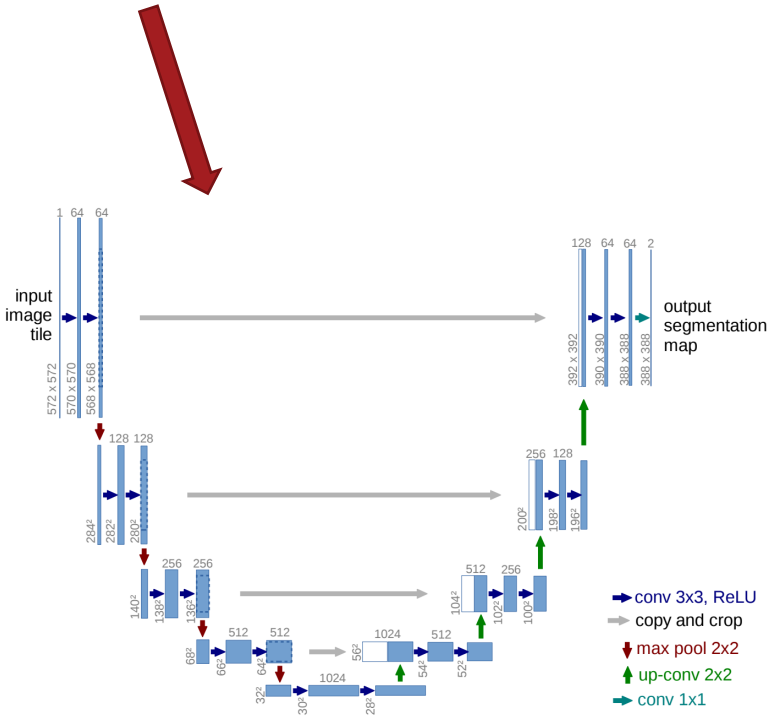
Autoencoders can be used for:

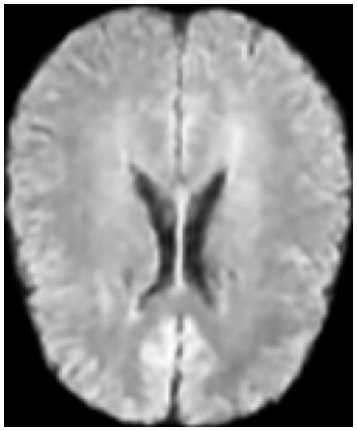
- Data compression
- Data denoising

Segmentation through synthesis

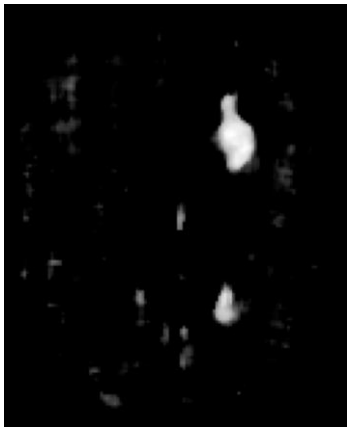
Let's train an autoencoder on healthy brain images

Pathological test image





Synthetic image

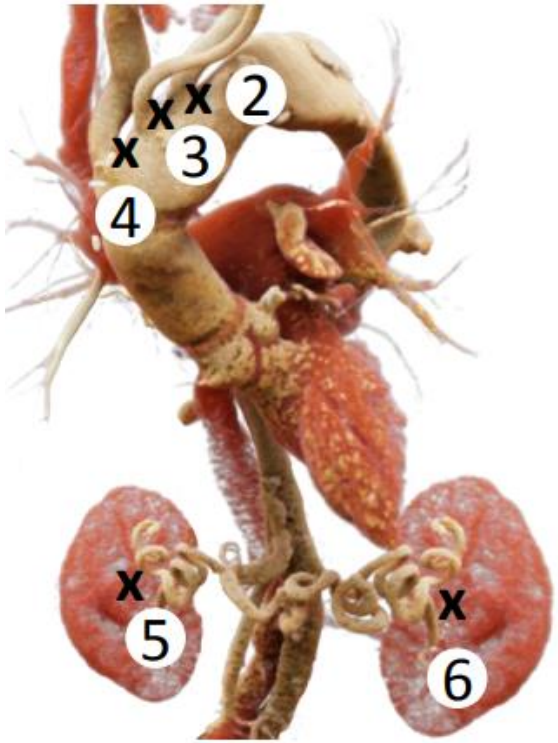


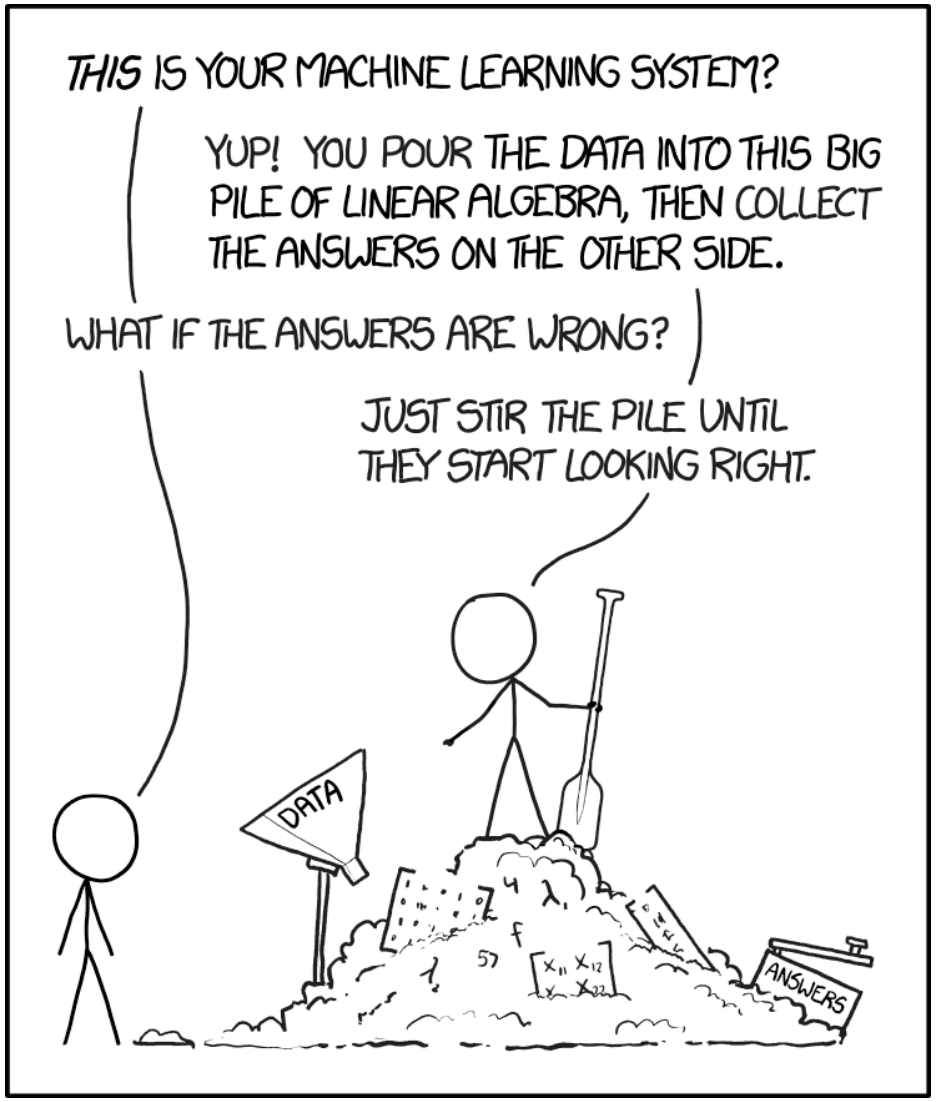
Difference between real and synthetic image

*Deep Autoencoding Models for Unsupervised Anomaly Segmentation in Brain MR Images

Landmarking

Can we effectively solve all segmentation/ classification /diagnosis problems in medical imaging following this schema?

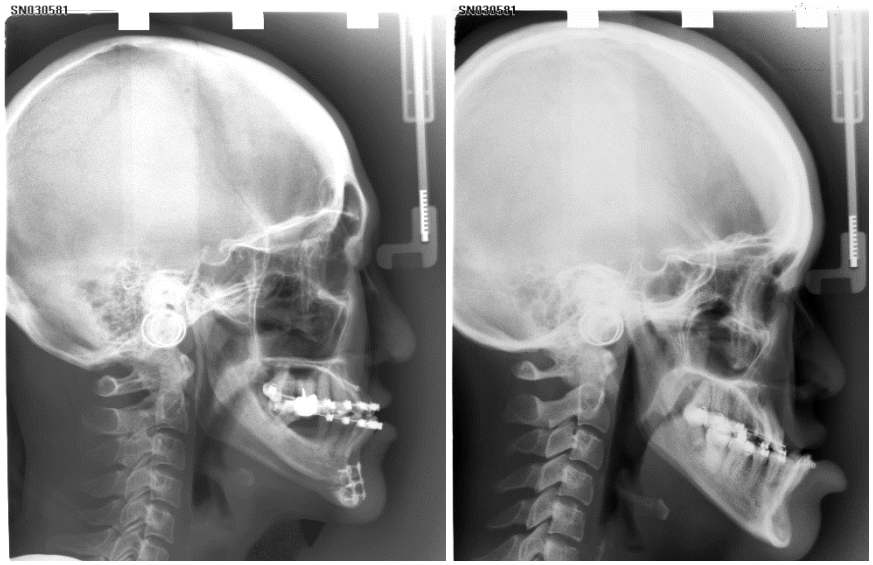




* from Siemens research 2018

Cephalometry

Can we effectively solve all segmentation/ classification /diagnosis problems in medical imaging following this schema?



100% of agreement

Can you guess what kind of measurements were used to compute the agreement?

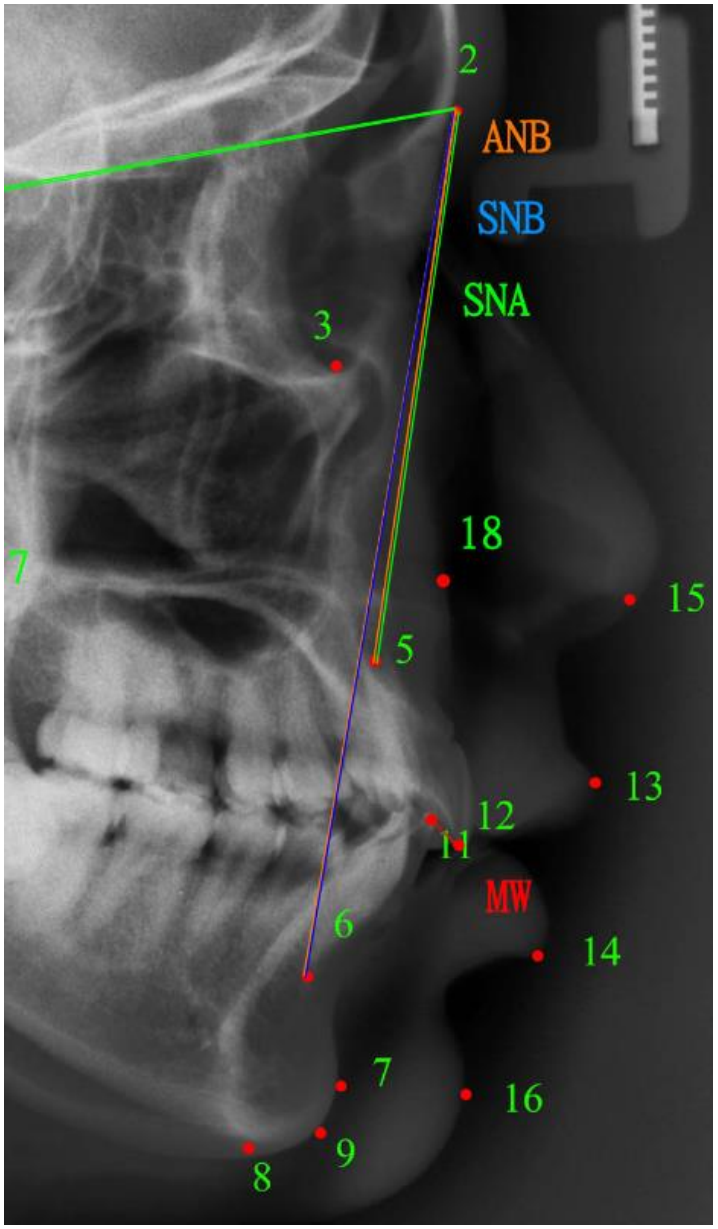
38% of agreement

Cephalometry

Cephalometry is based on computing angles and distances from head X-ray

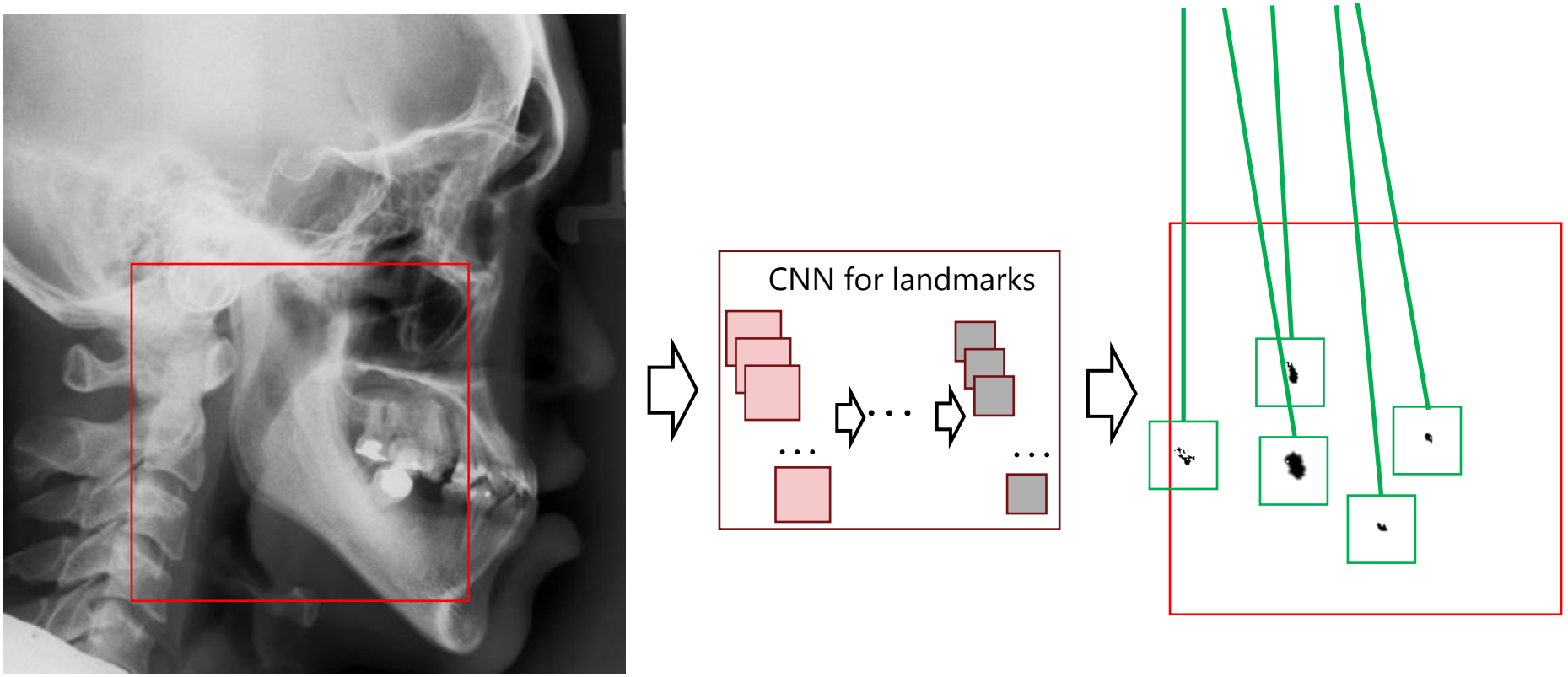
ANB angle:

- [3.7° – 5.2°] – normal range
- >5.2° - pathology type I
- < 3.7 ° - pathology type II



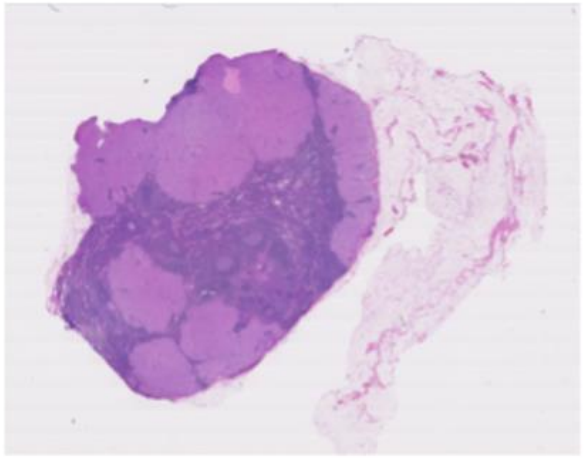
Cephalometry

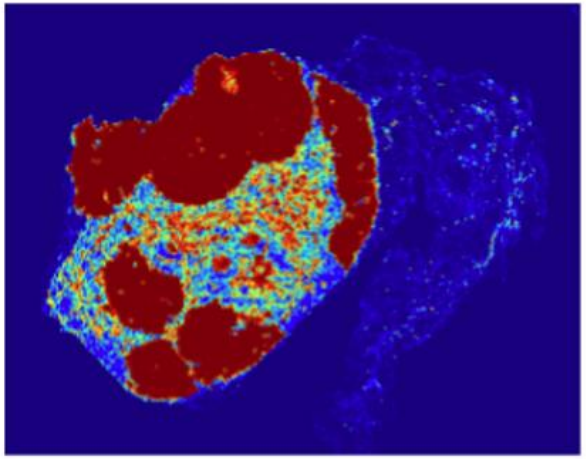
A good idea is to use CNN to detect landmarks and not to predict pathologies that are based on landmarks

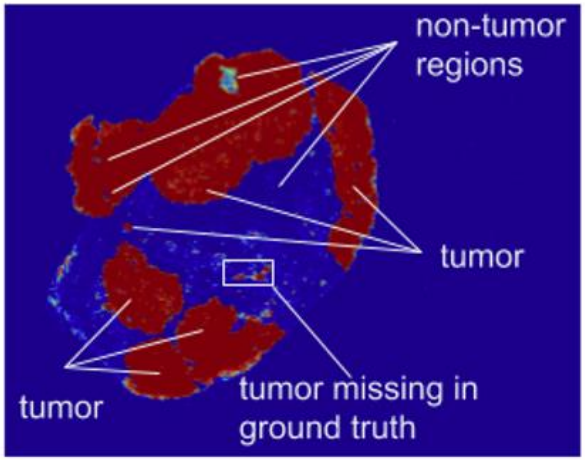


Disease diagnosis

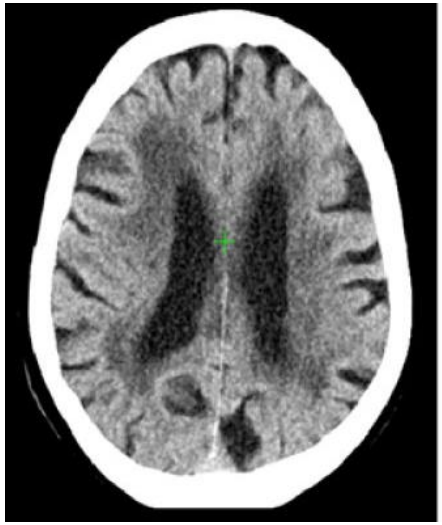
CNN for detection and segmentation pathologies

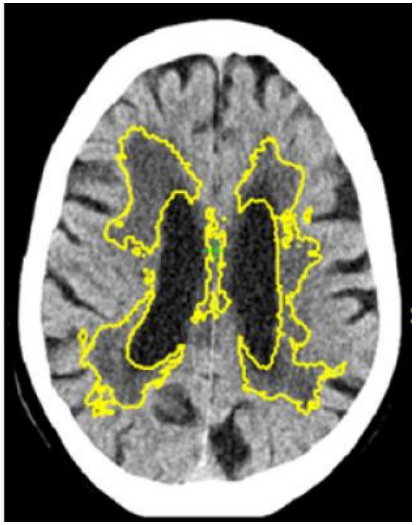






https://ai.googleblog.com/2017/03/assisting-pathologists-in-detecting.html

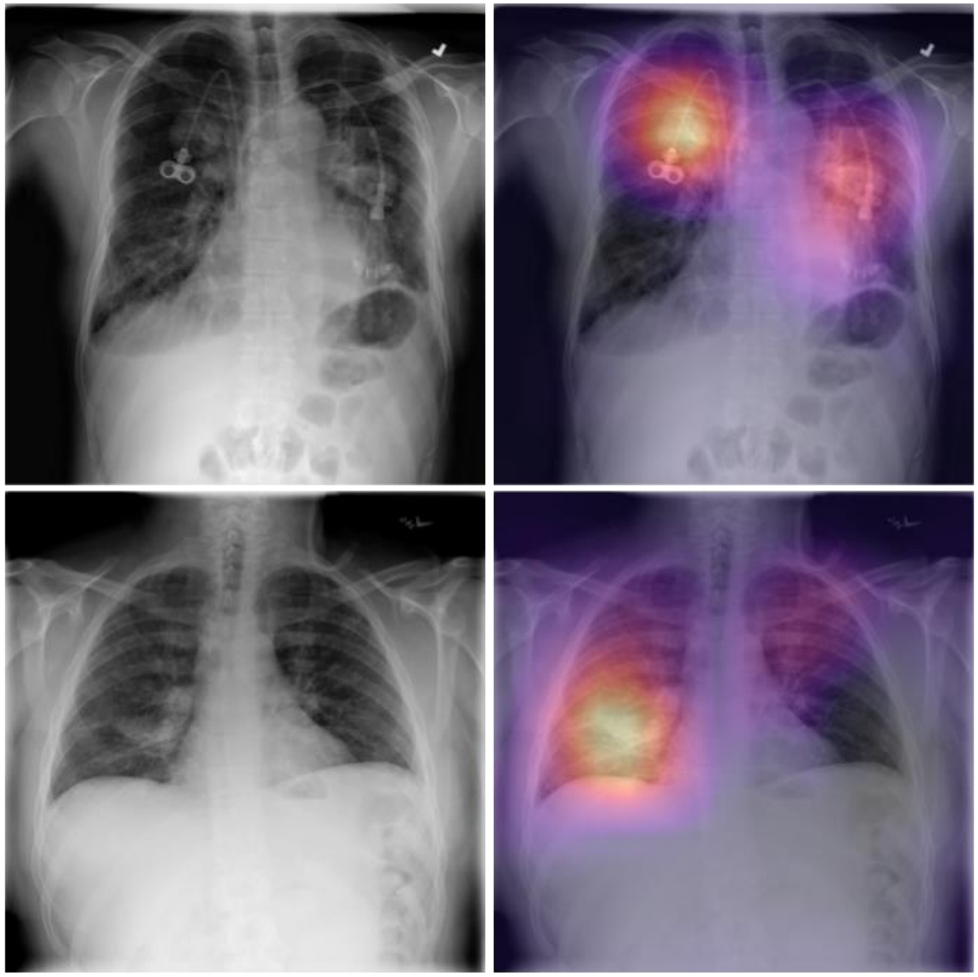




https://www.imperial.ac.uk/news/186108/artificial-intelligence-improves-stroke-dementia-diagnosis/

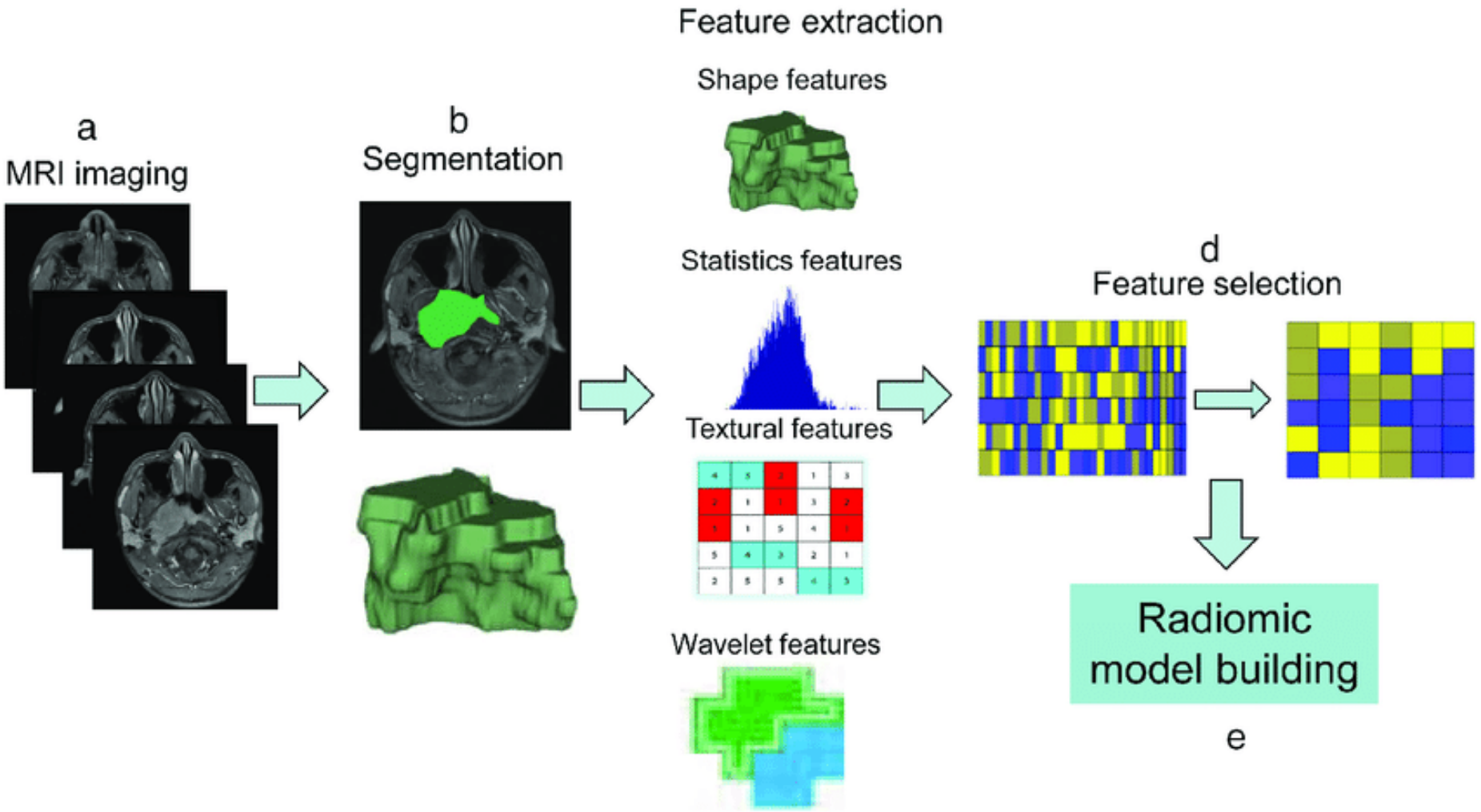
Disease diagnosis

- Diagnosis with explicit segmentation
- CNNs do not know where to look, but know which pathology should be present
- Heatmaps from CNNs tells us if CNNs look at meaningful spots

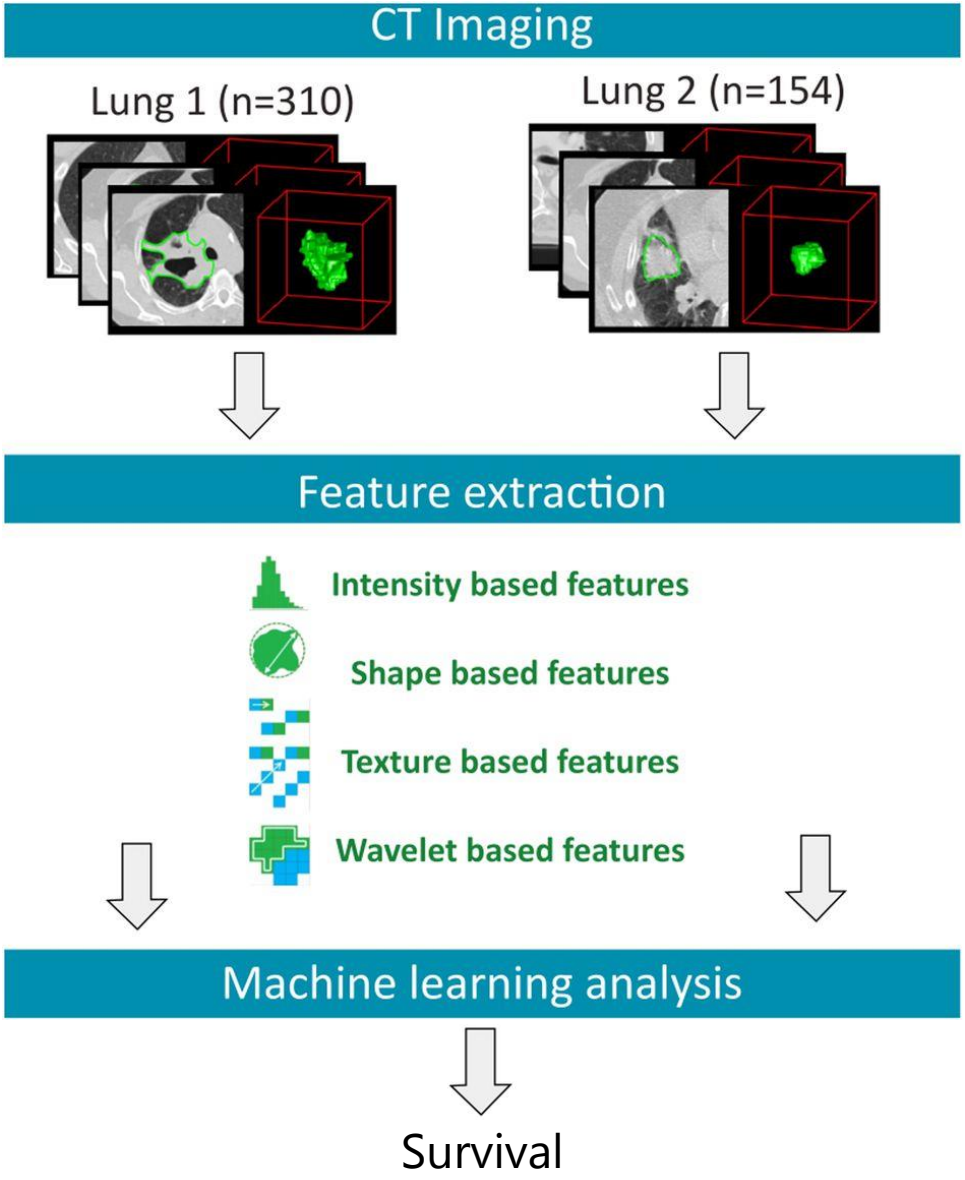


https://stanfordmlgroup.github.io/projects/chexnext/

Radiomics

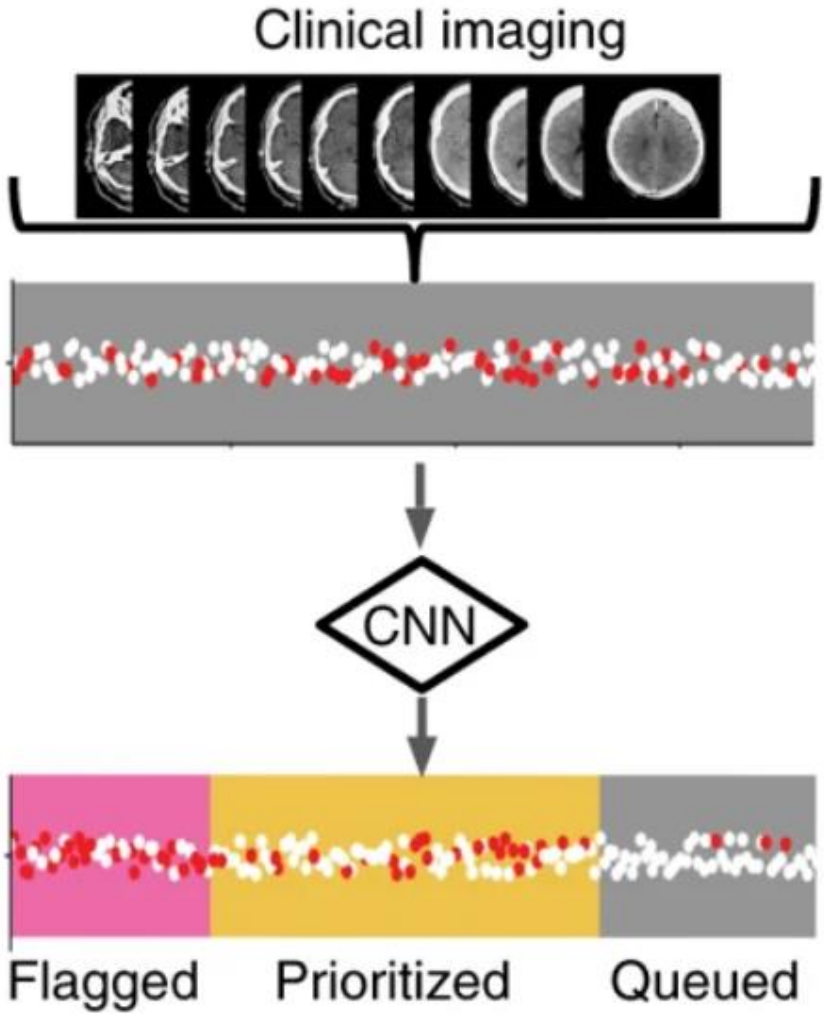

a
MRI imaging
b
Segmentation
Feature extraction
Shape features
Statistics features
Textural features
Wavelet features
d
Feature selection
Radiomic model building
e

Radiomics


CT Imaging
Lung 1 (n=310)
Lung 2 (n=154)
Feature extraction
Intensity based features
Shape based features
Texture based features
Wavelet based features
Machine learning analysis
Survival

Triaging

- Instead of automated diagnosis, machine learning can be used to sort patients
- Urgent patients will be shown first to physicians
- Safer, final decision is on human and not machine

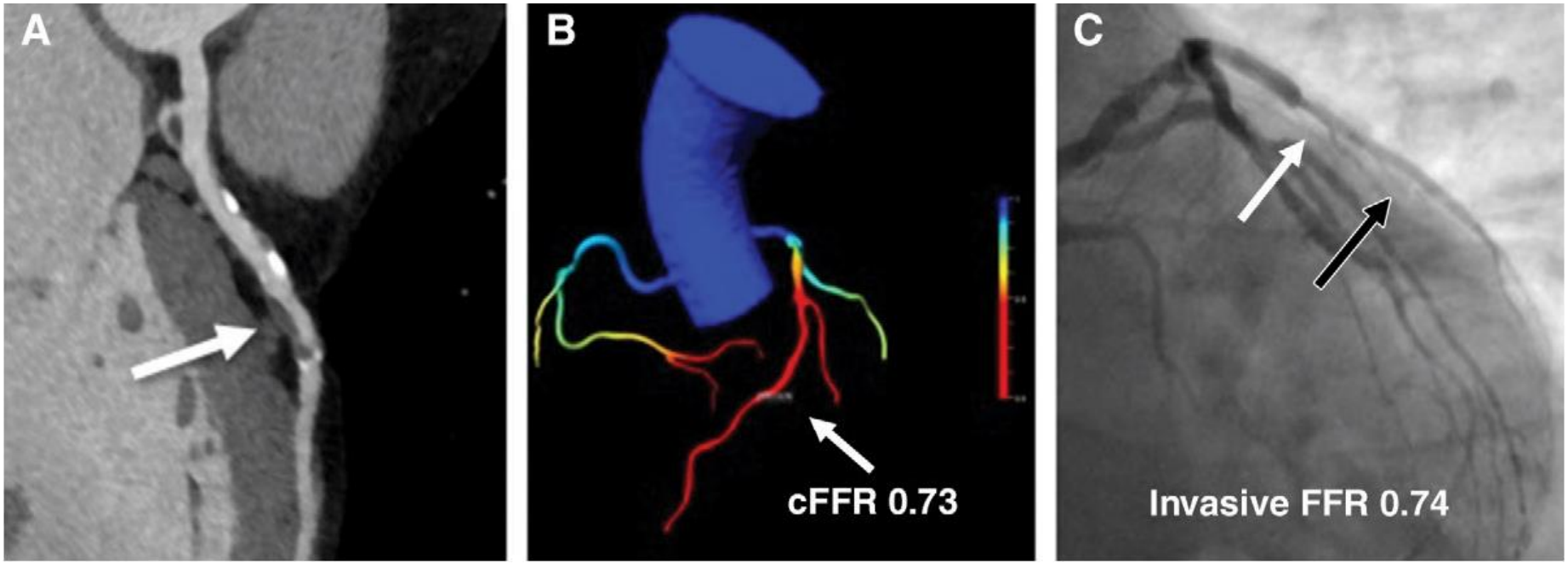


https://www.nature.com/articles/s41591-018-0147-y

Fractional Flow Reserve

- Use machine learning to segment heart vessels

- Estimate the blood flow through vessels using physical models of machine learning



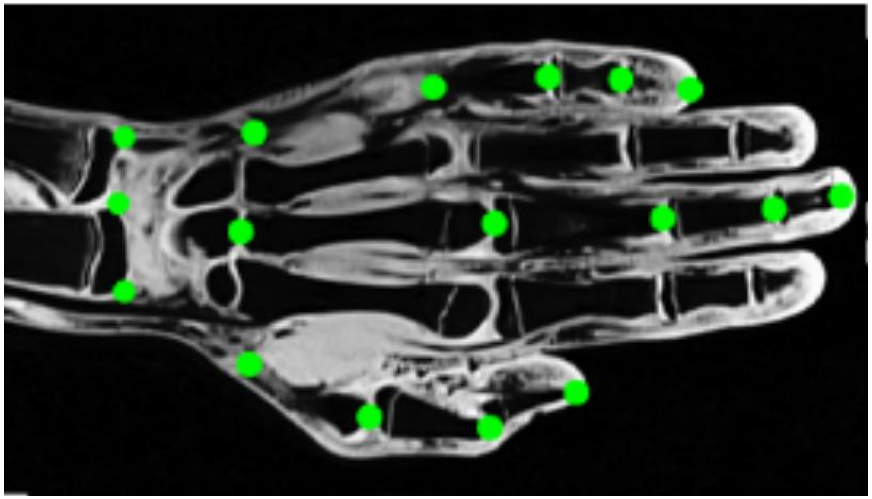
https://www.ncbi.nlm.nih.gov/pubmed/28926310

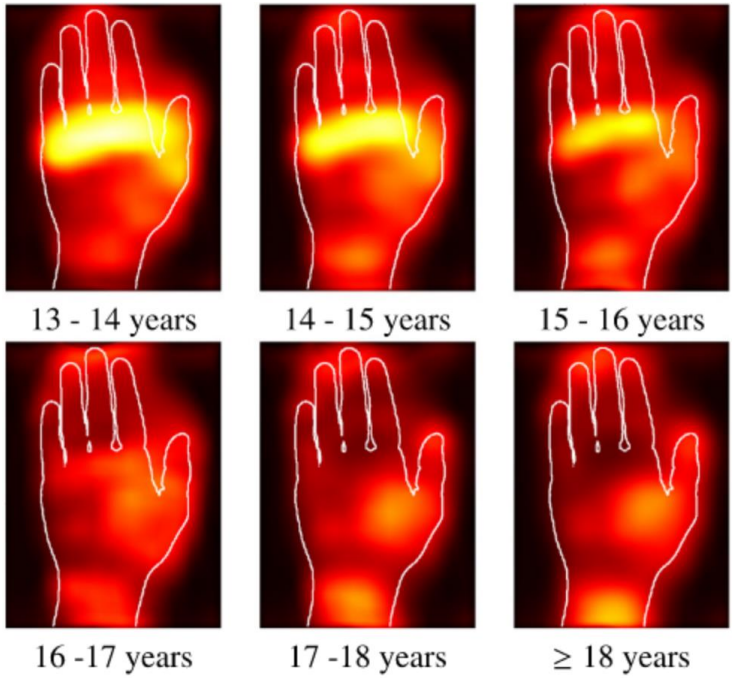
Age estimation

Problem:

- A person claims he is younger than 15 years?
- No reliable document is available

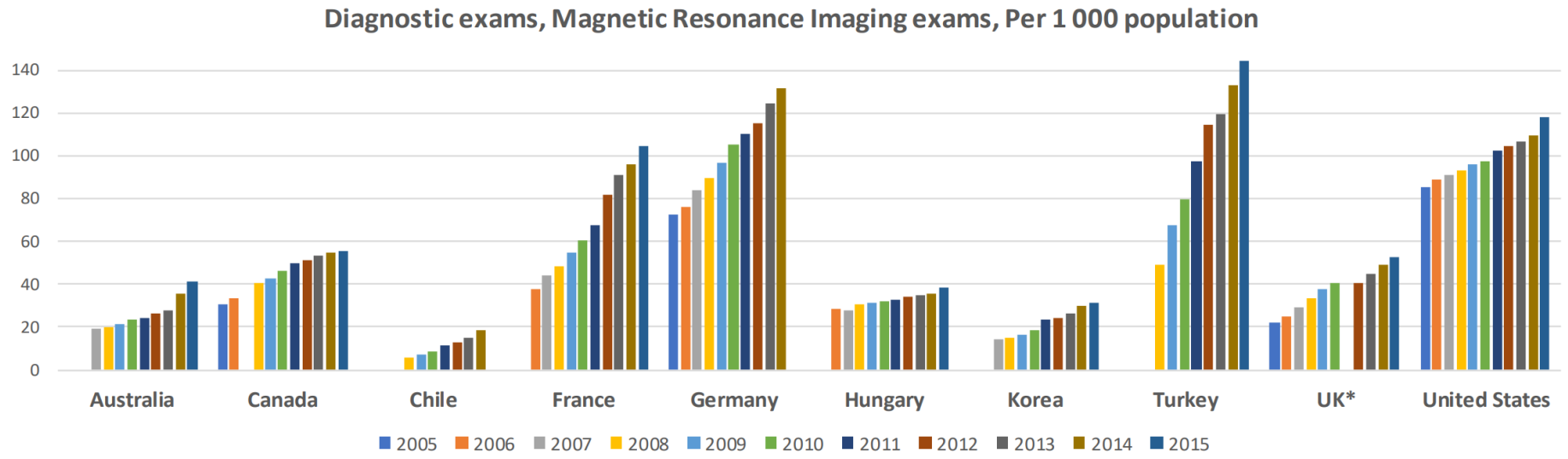
How to approximate the real age?

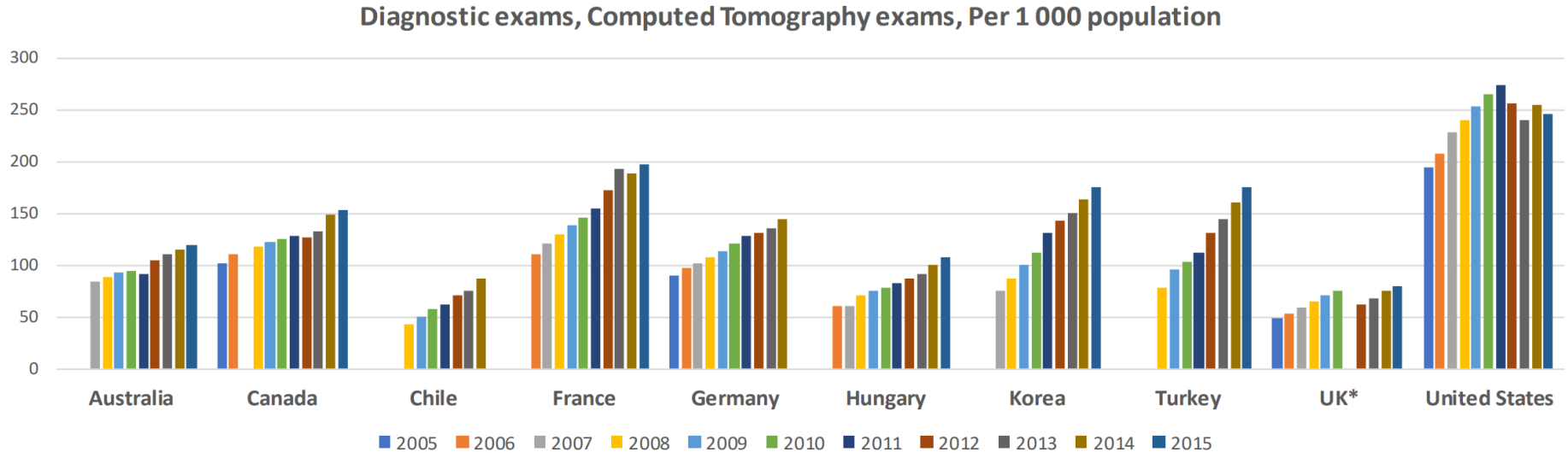




*Automated age estimation from MRI volumes of the hand, 2019

Potential for computer-aided medical image analyses





Questions?