ISMRM & ISMRT
ANNUAL MEETING & EXHIBITION
TORONTO
03-08 JUNE 2023

Berkeley
UNIVERSITY OF CALIFORNIA

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

Fundamentals of Deep Learning

Efrat Shimron

Deep learning is everywhere!

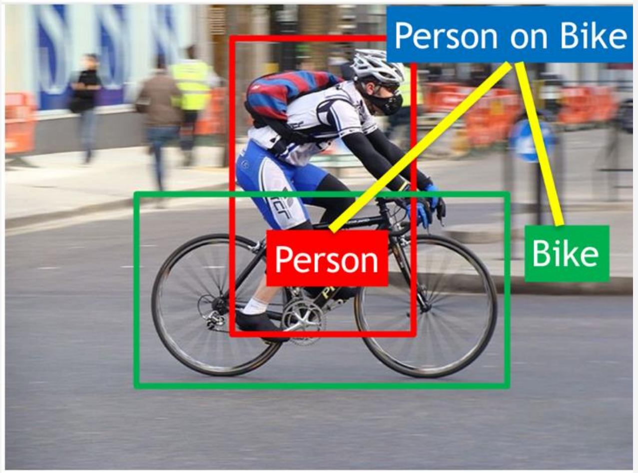
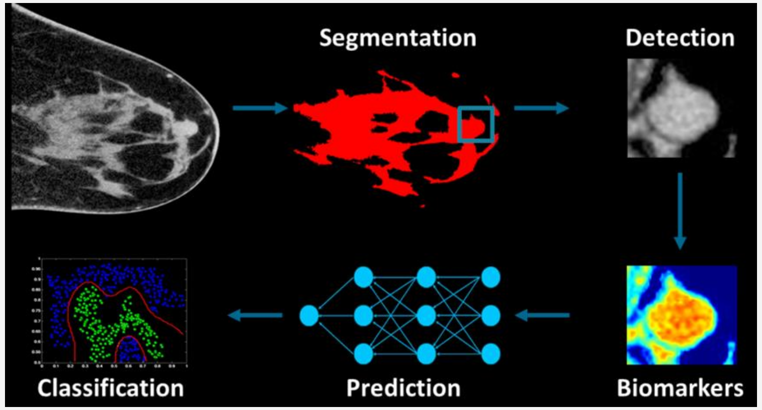
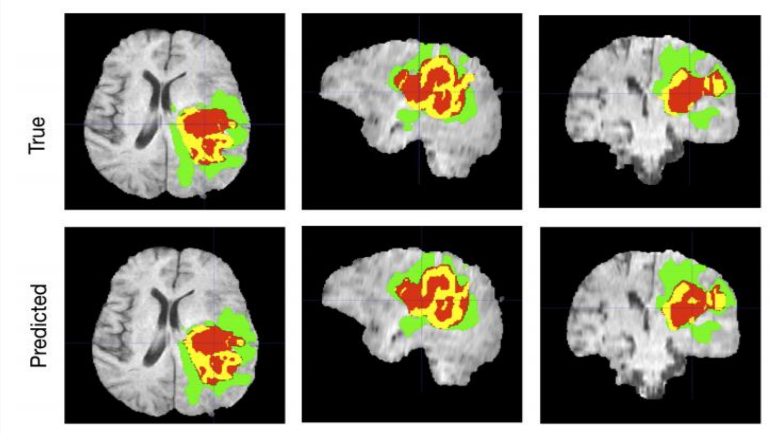


Image by Jin Clyde Monge

Mindy-support.com



http://axti.radboudimaging.nl/

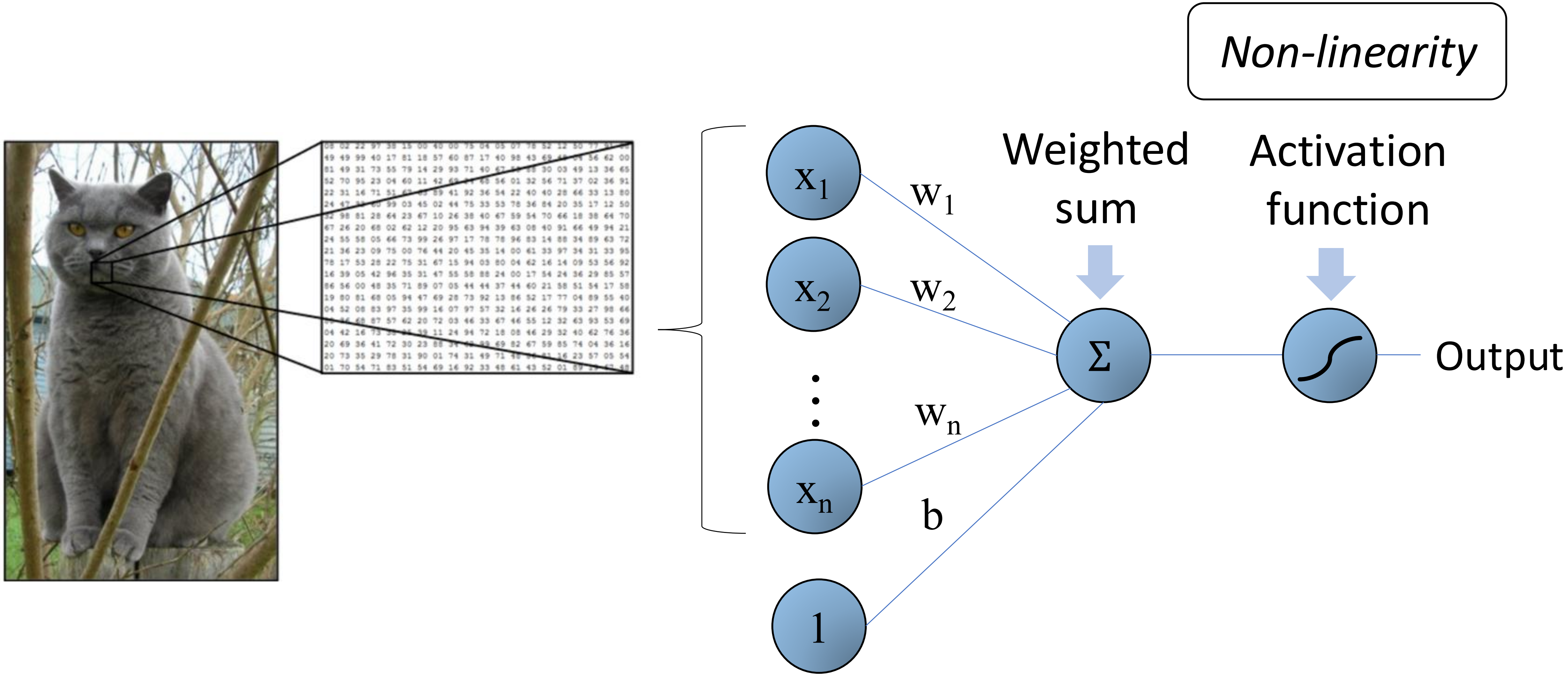


Nvidia.com

Outline

1. **Neural Networks: Basics**
2. Convolutional Neural Networks
3. Applications in MRI
4. Challenges and limitations

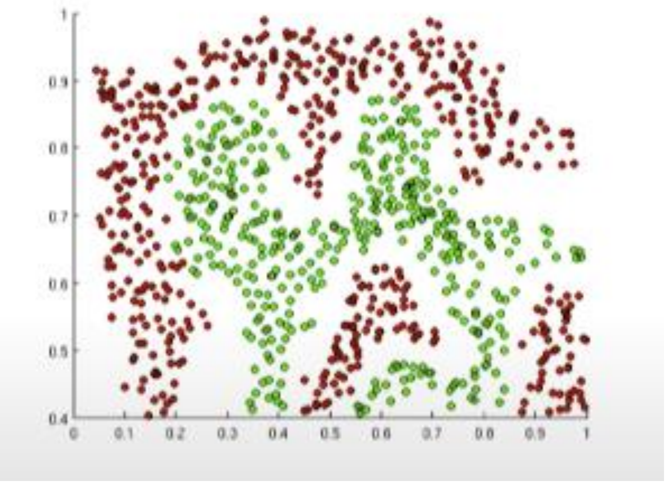
Artificial neuron



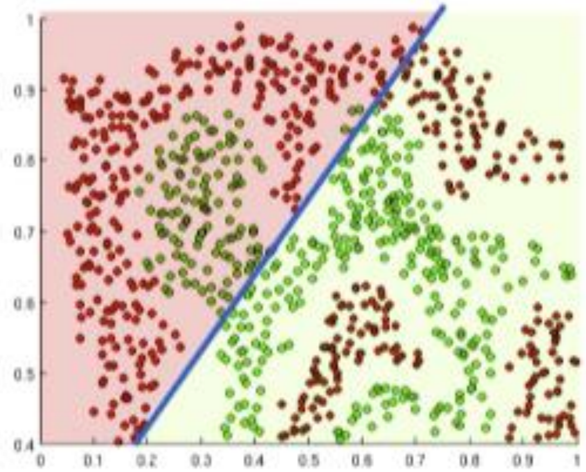
https://cs231n.github.io/classification/

Why non-linearity?

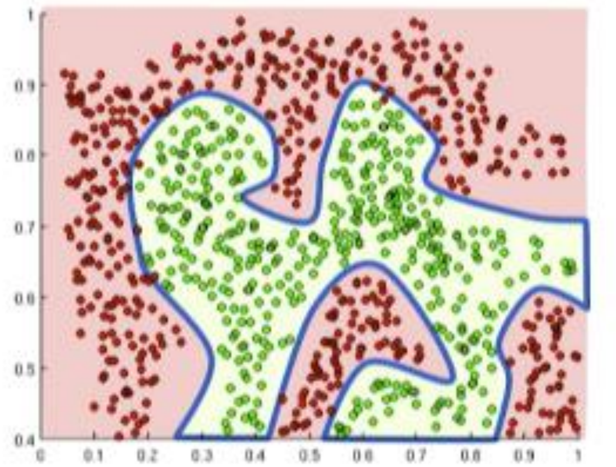
Can we separate the two groups?



Without non-linearity

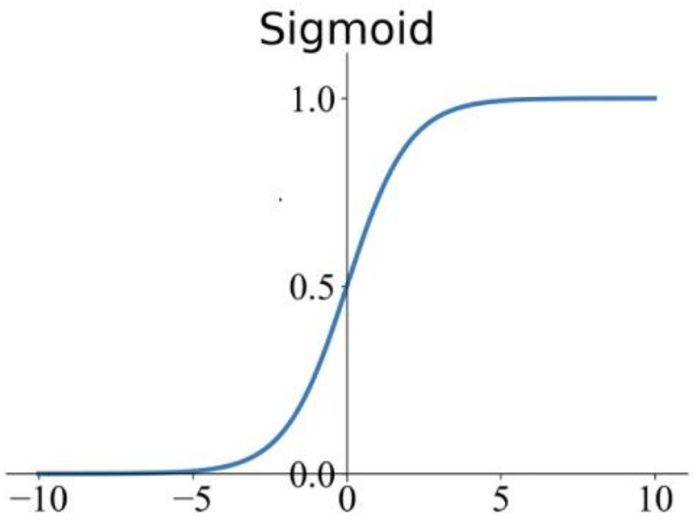


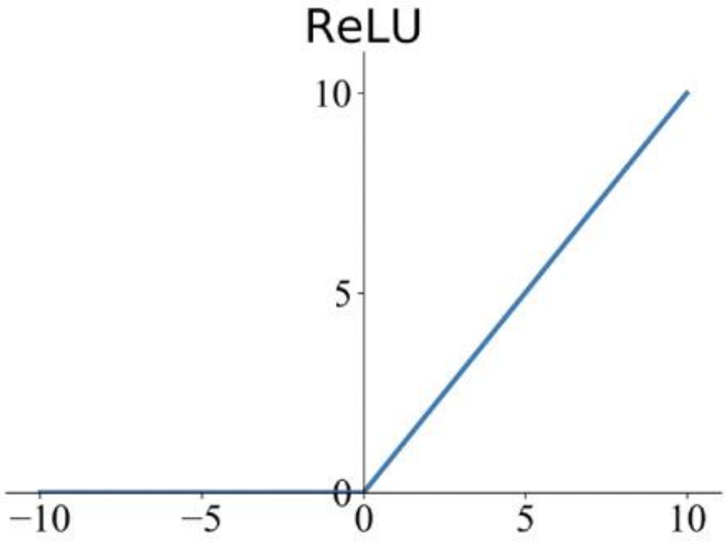
With non-linearity

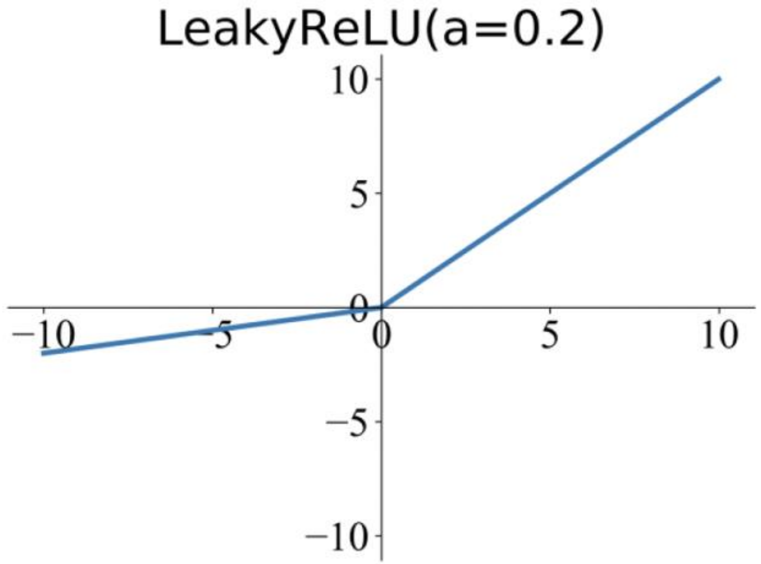


A. Amini, *Introduction to Deep Learning*, MIT course 6.S191 (2022)

Activation functions







$$f(z) = \frac{1}{1 + e^{-z}}$$

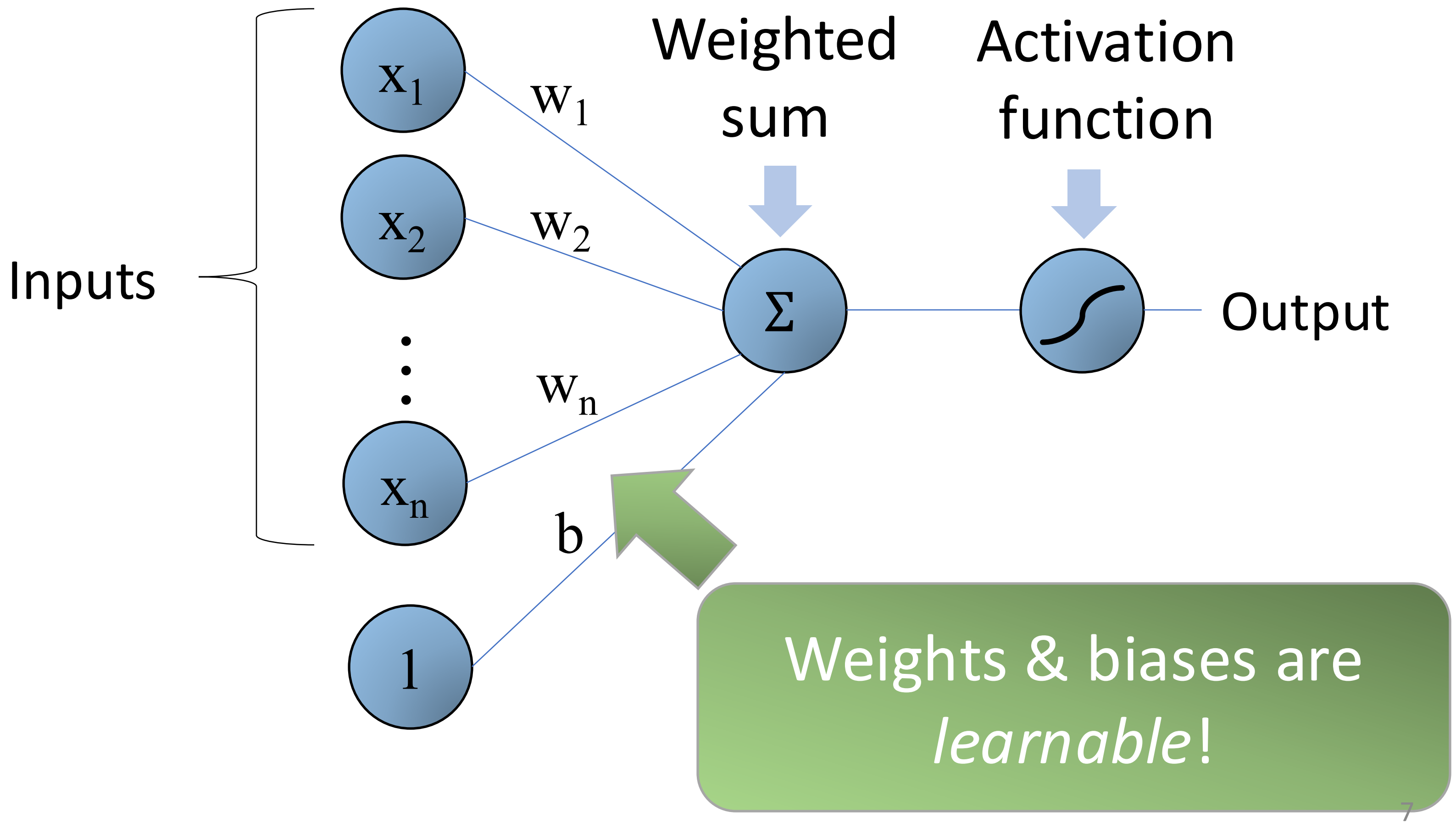
$$f(z) = \max(0, z)$$

$$f(z) = \begin{cases} z & z > 0 \\ az & z \leq 0 \end{cases}$$

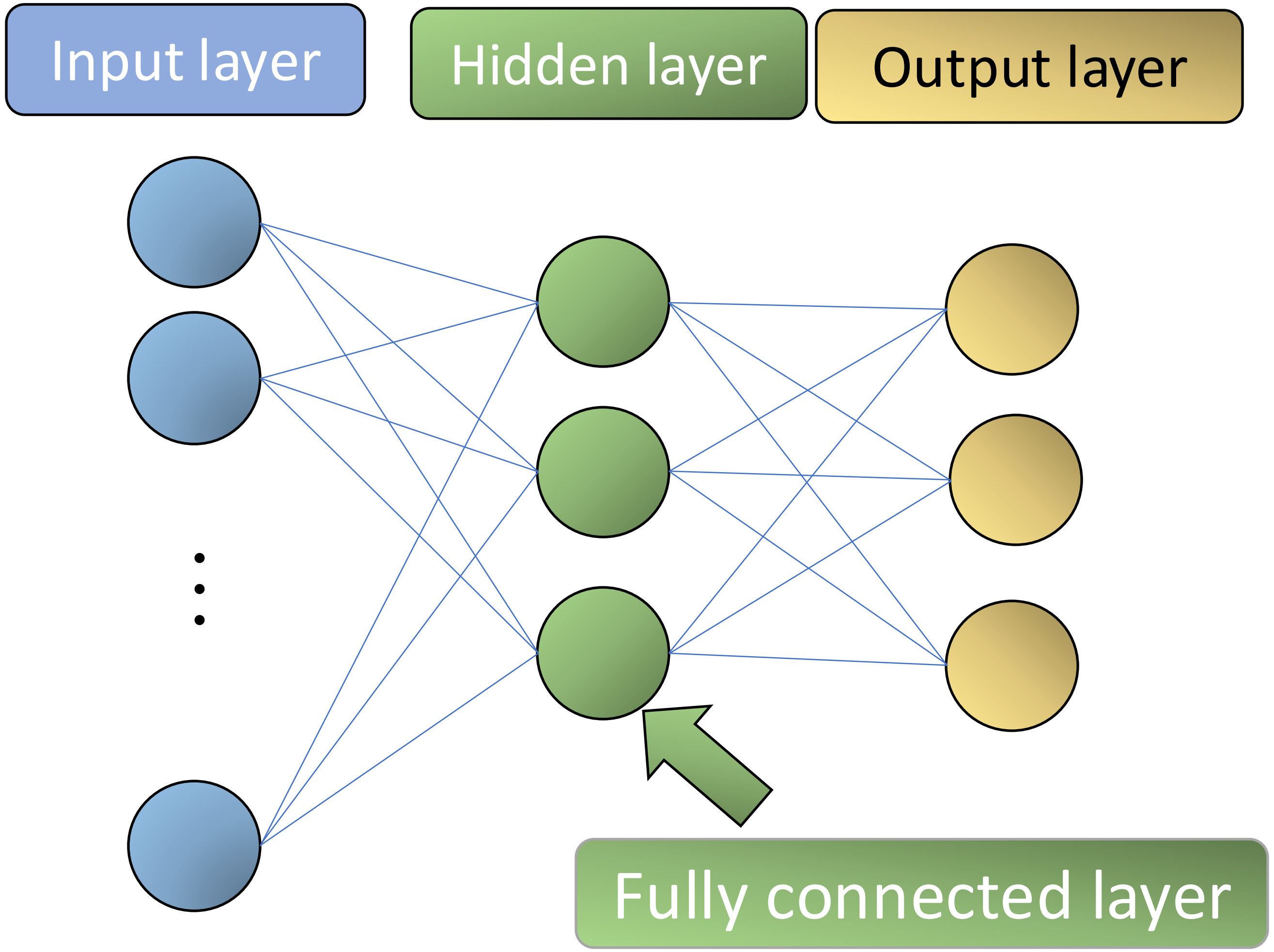
Figures from DOI:10.1103/PhysRevE.100.033308

Artificial neuron

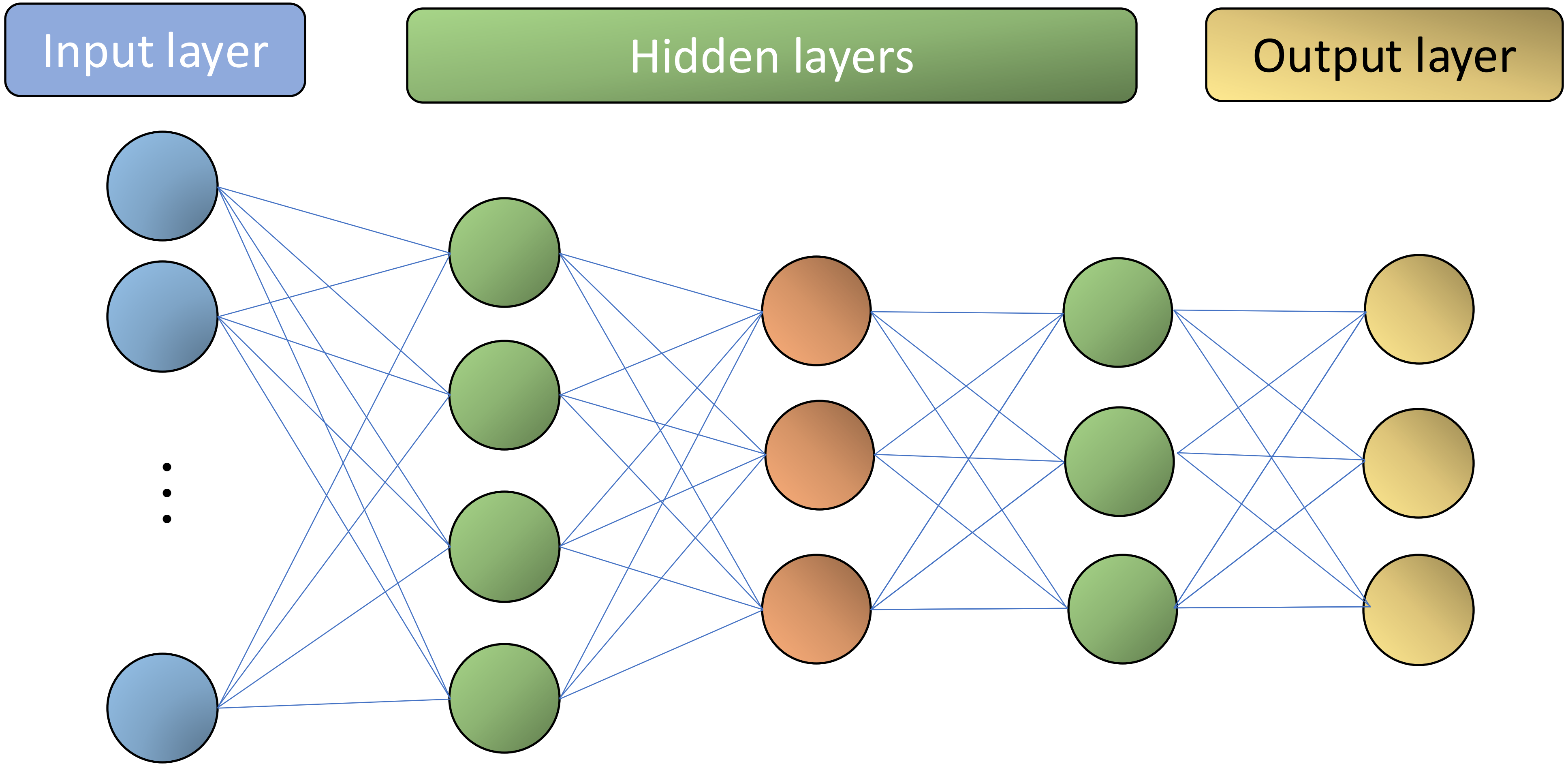
$$Y = f\left(\sum(input * weight) + bias\right)$$



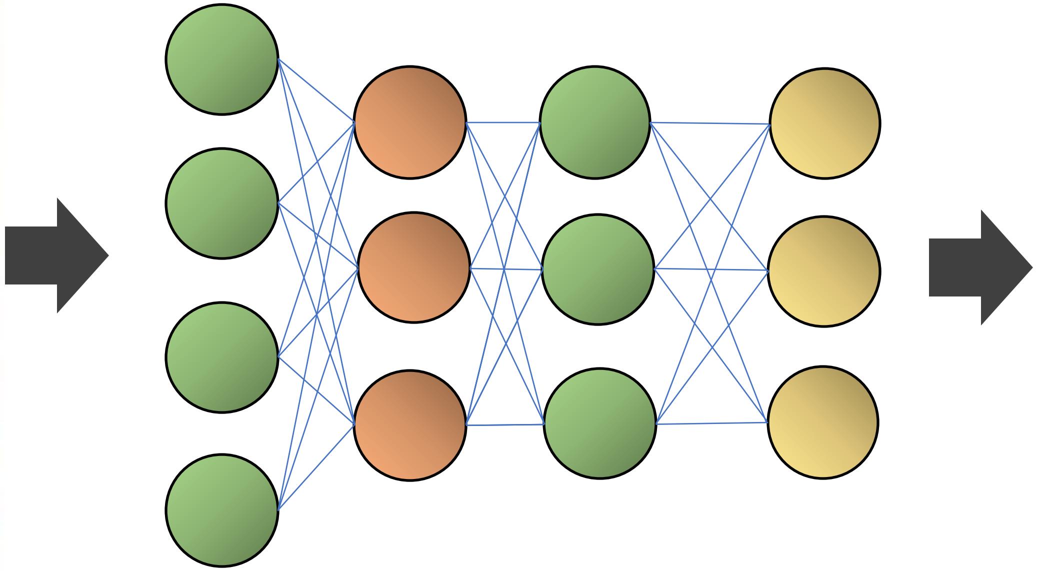
Simple neural networks



Simple neural networks



Setting up the problem



Bat
Plane
Boat
Car
Plane
...

https://www.cs.toronto.edu/~kriz/cifar.html

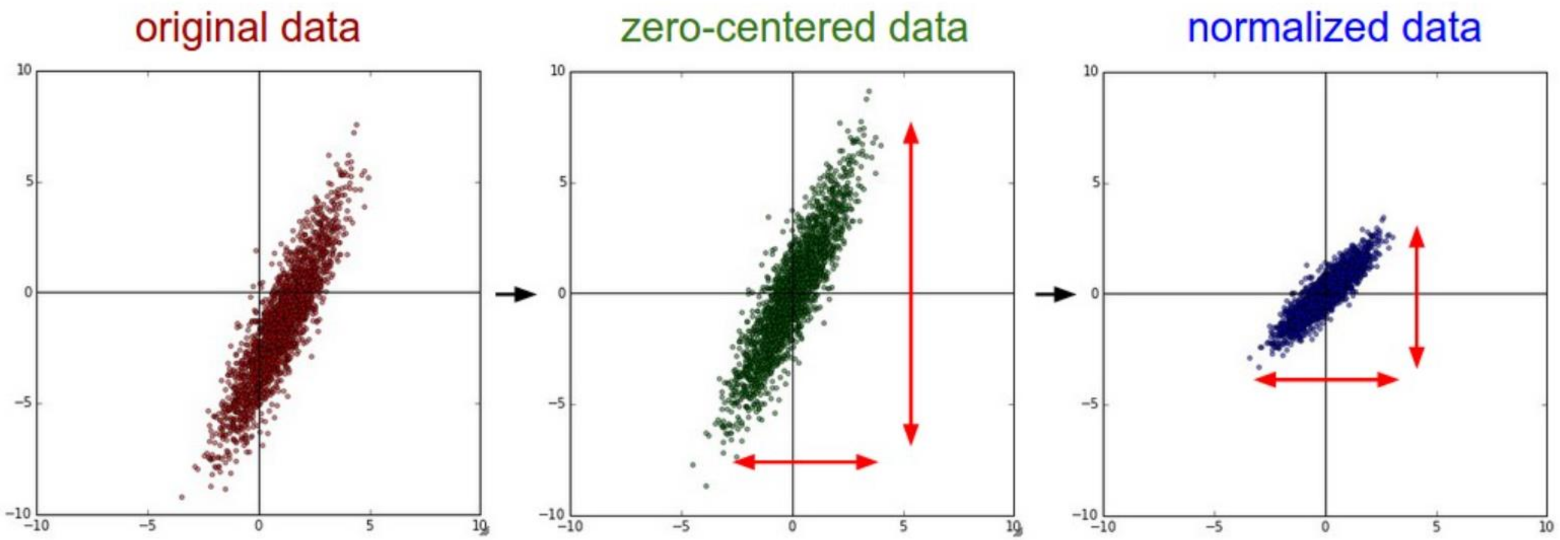
Setting up the problem

Database of input-output pairs

- ImageNet, CIFAR10, …

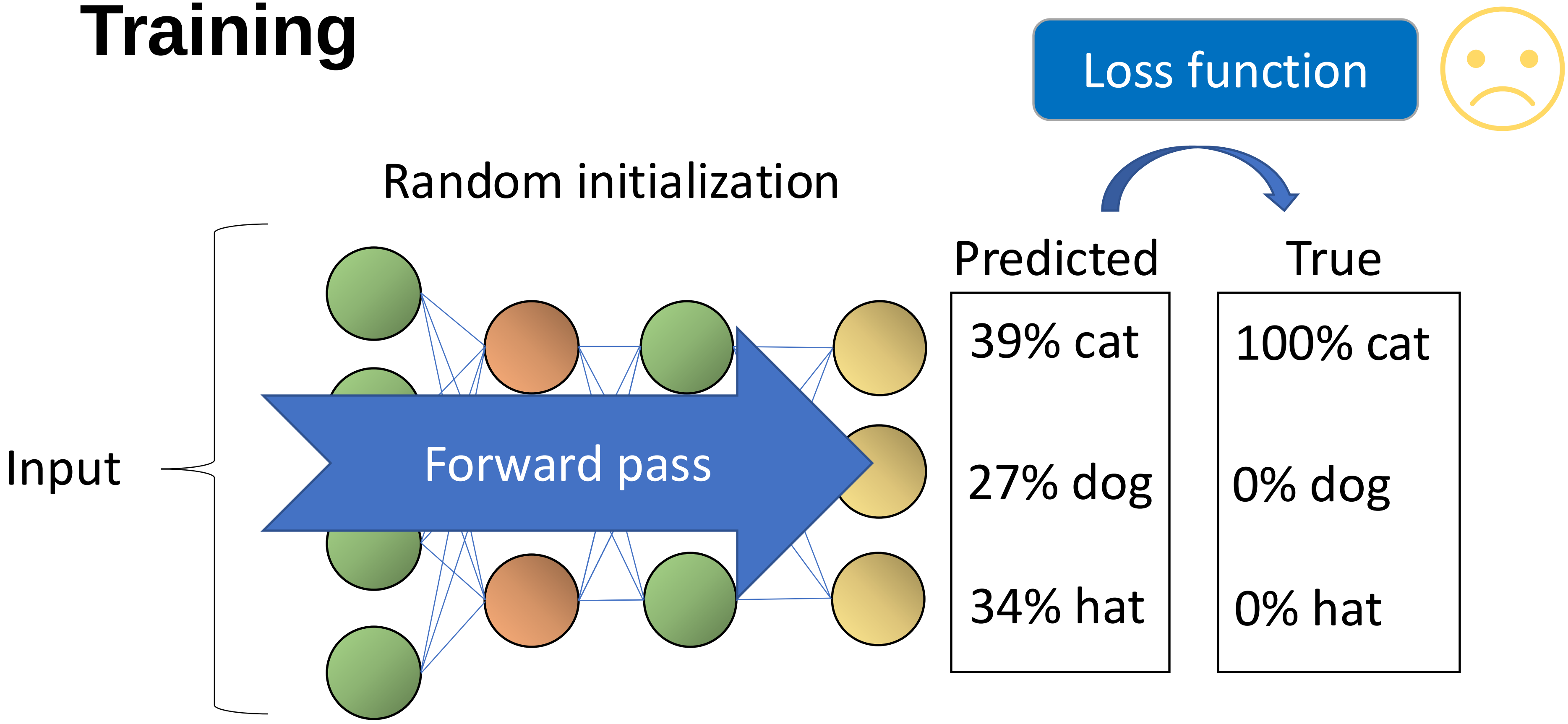
Initial processing steps

- subtract mean, normalize, …



https://cs231n.github.io/neural-networks-2/

Training


Loss function
Random initialization
Input
Forward pass
Predicted
39% cat
27% dog
34% hat
True
100% cat
0% dog
0% hat

Training: Loss functions

Cross-entropy loss - useful for classification

$$\mathcal{L}_{CE} = -\sum_{i=1}^{n} t_i \log(p_i)$$

i=1,…,n class

t_i true probability of the label

p_i predicted probability

Mean Squared Error (MSE) loss - useful for regression

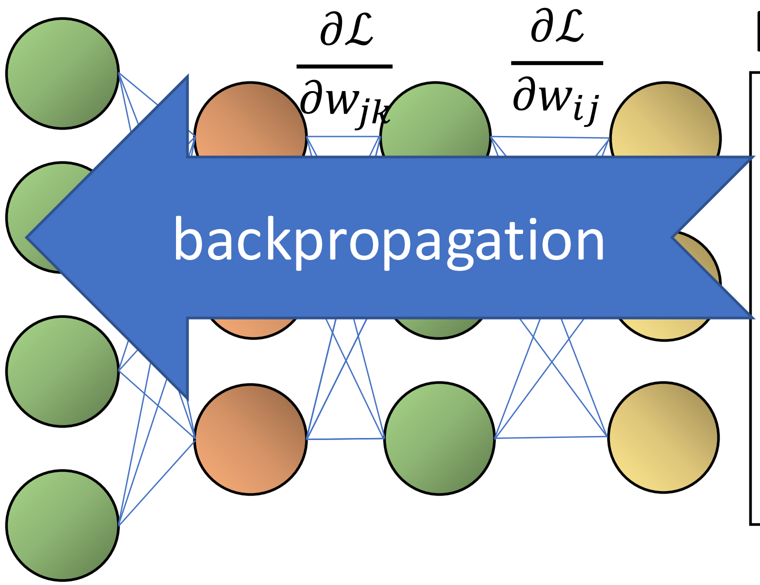
$$\mathcal{L}_{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2$$

There are many other loss functions!
MAE, VGG loss, …

Training

$$w \leftarrow w - \eta \frac{\partial \mathcal{L}}{\partial w}$$

Loss ($\mathcal{L}$)



Predicted	True
39% cat	100% cat
27% dog	0% dog
34% hat	0% hat

Backpropagation:

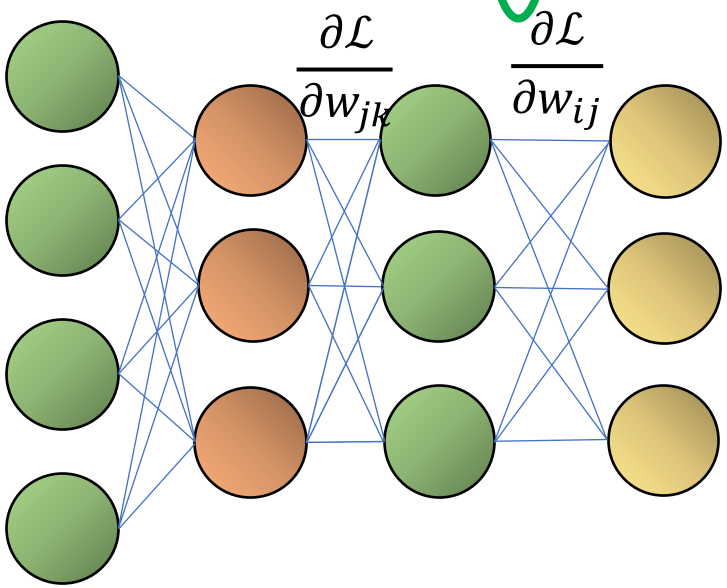
Computes the gradient of the loss w.r.t. each weight

Updates the weights to minimize the loss

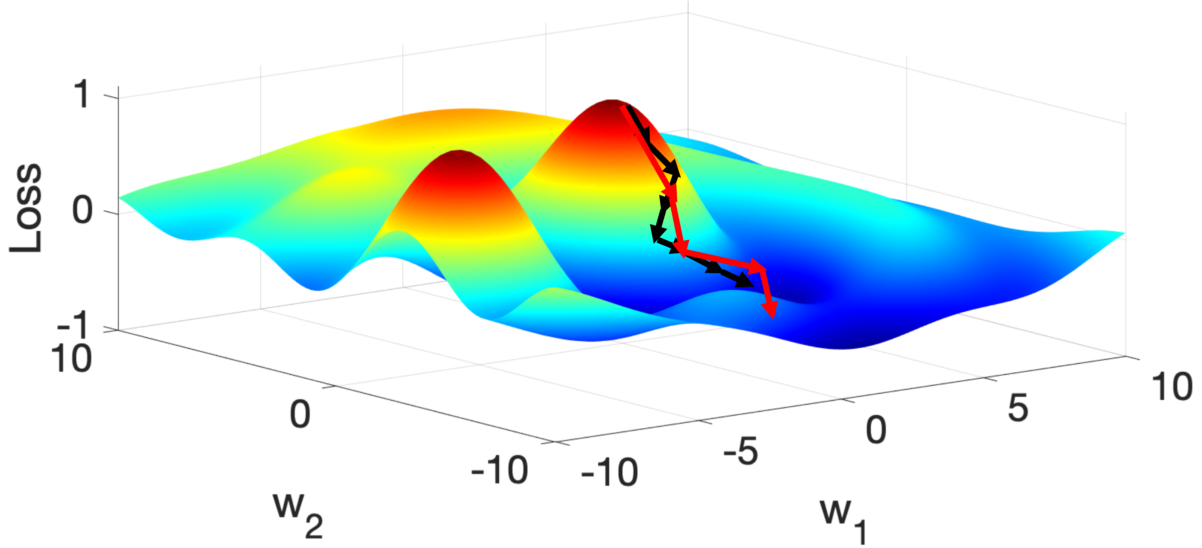
Training

Learning rate

$$\eta \frac{\partial \mathcal{L}}{\partial w}$$



$\mathcal{L}(w_1, w_2)$



Optimization algorithms:
Gradient Descent, Adam, …

TensorFlow PyTorch

Training

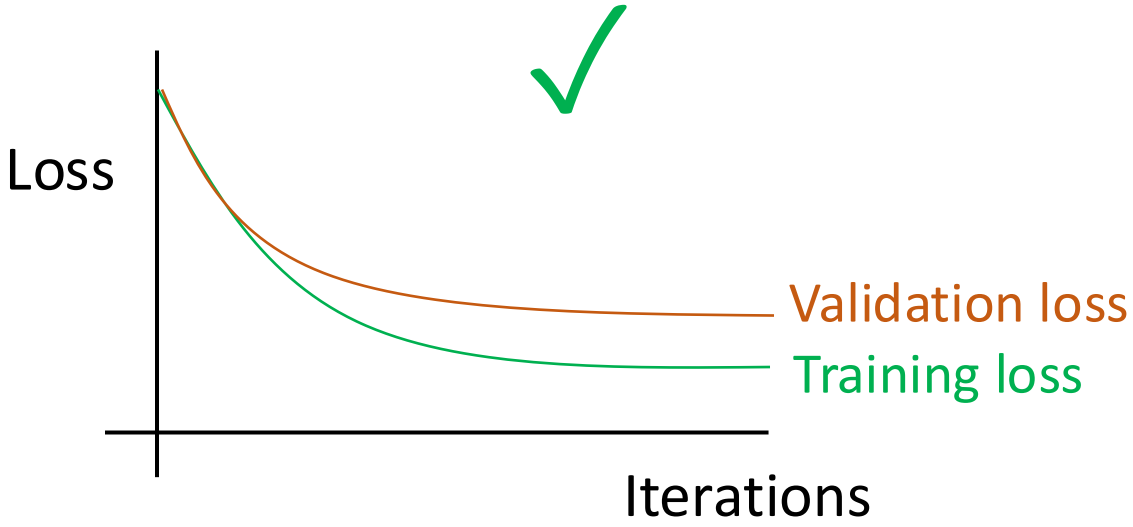
Split data: train / validation / test

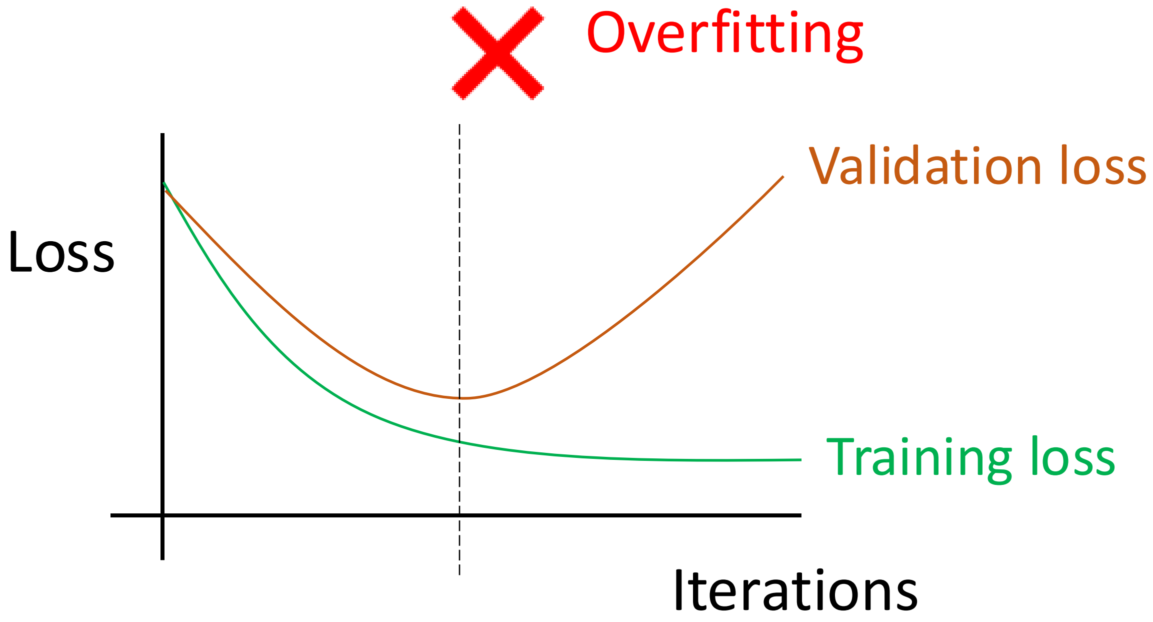
For example: 70% / 15% / 15%

What's the difference between validation and test?

- Validation data –
  - Used *during training* for model evaluation
  - Enables hyperparameter & architecture optimization
- Test data –
  - Should not be accessed during training!

Training

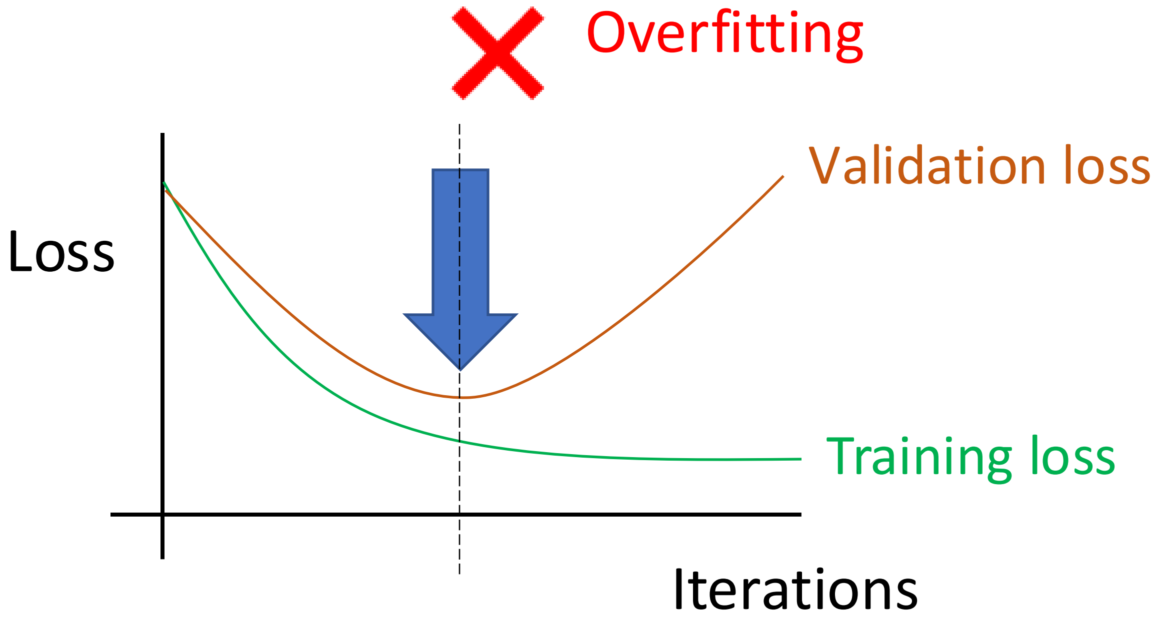

Loss
Validation loss
Training loss
Iterations


Overfitting
Loss
Validation loss
Training loss
Iterations

Training

How to avoid overfitting:

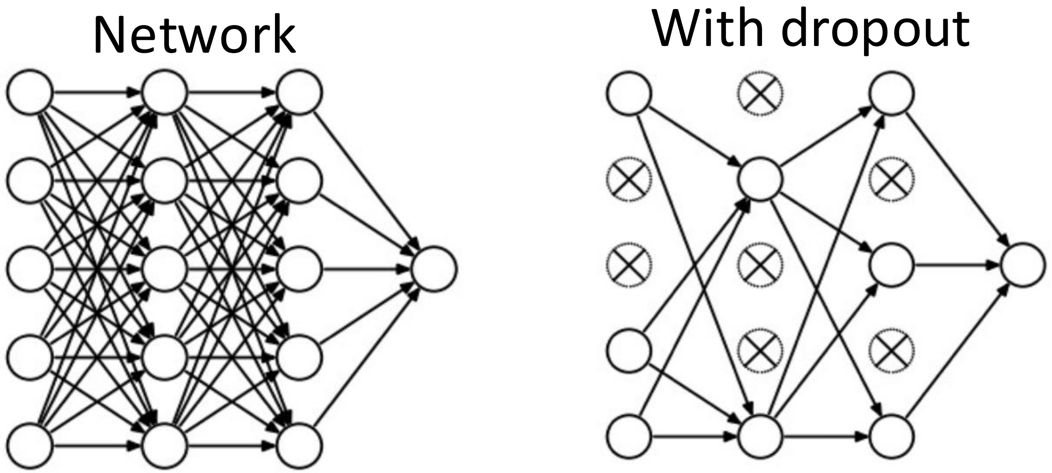
- Early stopping
- Data augmentation – rotations, scaling etc.



Training

How to avoid overfitting:

- Early stopping
- Data augmentation – rotations, scaling etc.
- Dropout
    - Some layers are randomly dropped



Srivastava, et al. (2014). *The journal of machine learning research*.

Training

How to avoid overfitting:

- Early stopping
- Data augmentation – rotations, scaling etc.
- Dropout
- Regularization

$$\mathcal{L} = Error(y, \tilde{y}) + \mathcal{R}(w)$$

- L1 regularization

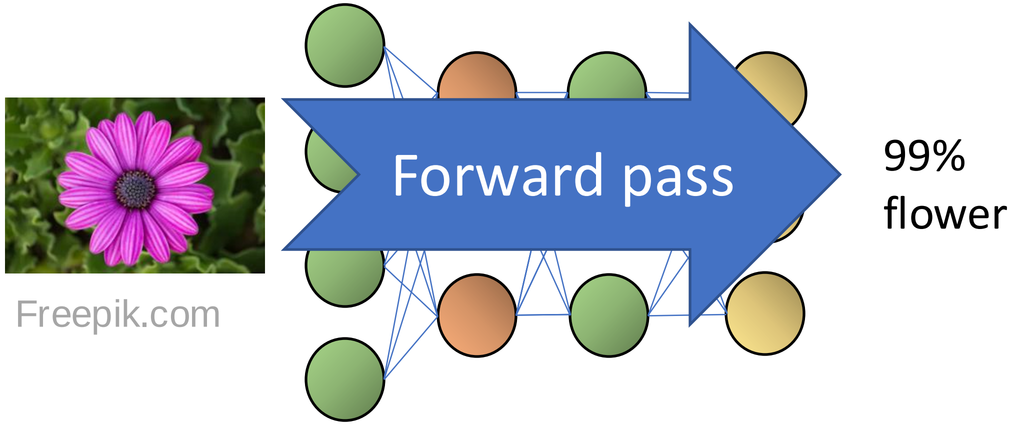
$$\mathcal{L} = Error(y, \tilde{y}) + \lambda \sum_i |w_i|$$

- L2 regularization

$$\mathcal{L} = Error(y, \tilde{y}) + \lambda \sum_i w^2$$

Test

- Test data – new input
- No dropout



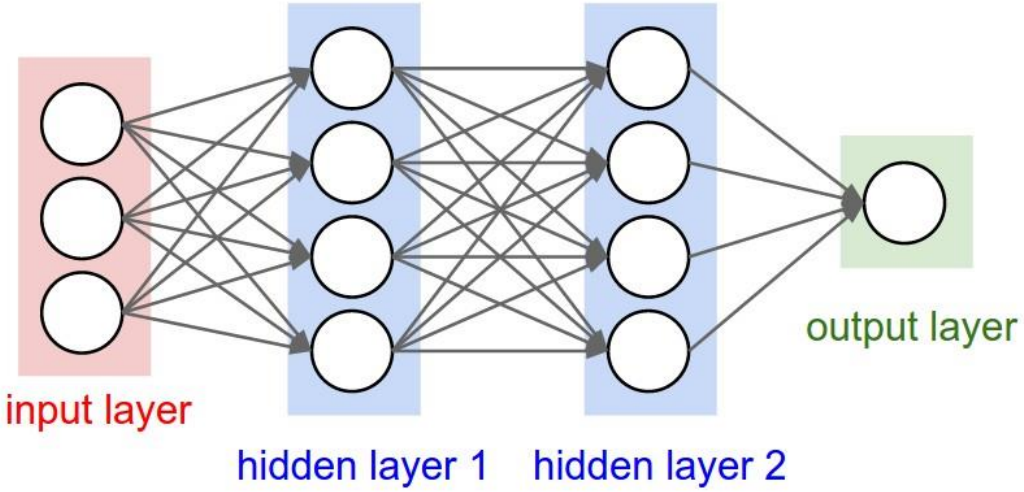
Outline

1. Neural Networks: Basics
2. **Convolutional Neural Networks**
3. Applications in MRI
4. Challenges and limitations

Why CNNs?

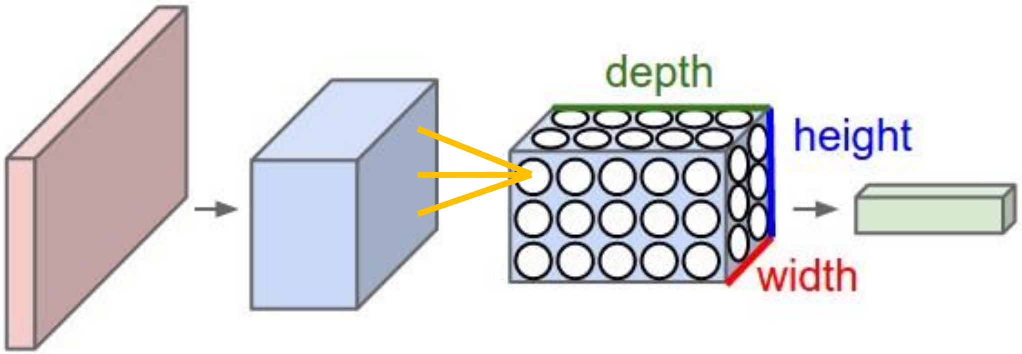
- Fully connected neural networks don’t scale well to large images
  - Typical image: 300x200 pixels x 3 channels (RGB)
  - Full connectivity – computationally demanding, wasteful
- CNNs are more suited to images

Regular network



- Full connectivity

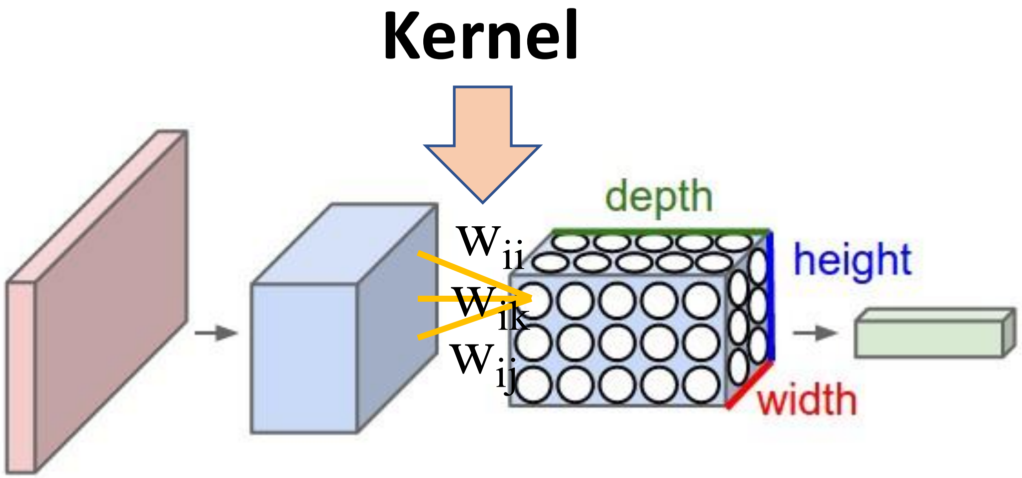
CNN



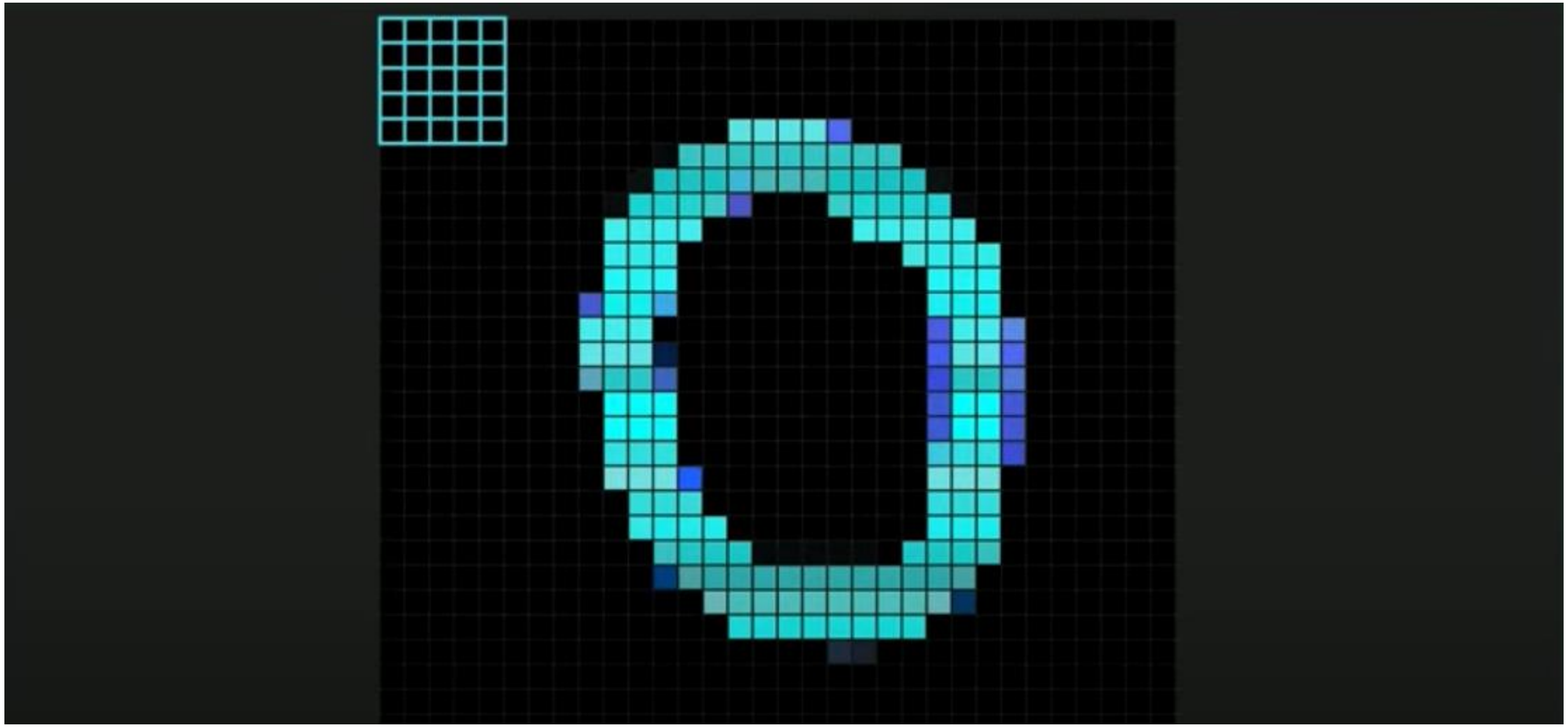
- Connections to only *a few* neurons
- Layers have depths (3D structure)
- Computationally efficient
- Suitable to high-dimensional data

https://cs231n.github.io/convolutional-networks/

- **Conv layer**
  - Computes a sliding dot-product of the input and the kernel (weights matrix)
  - Produces a feature map



- **Conv layer**

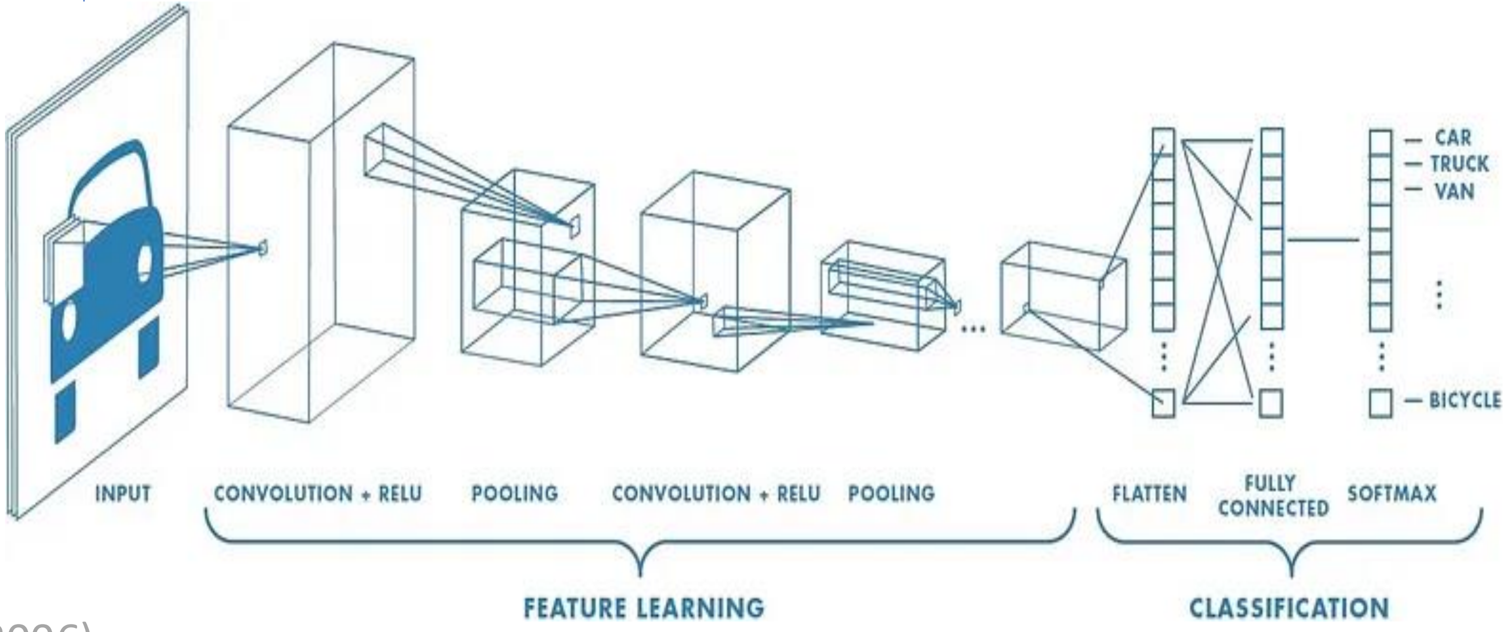


Futurology https://www.youtube.com/watch?v=pj9-rr1wDhM

Convolutional Neural Networks (CNNs)

- **Conv layer**
- **ReLU layer**
  - Element-wise activation function
- **Pooling layer**
  - Down-sampling operation
  - Example: max pooling
  - Benefit: translation invariance
- **Fully Connected layer**

Repeated several times



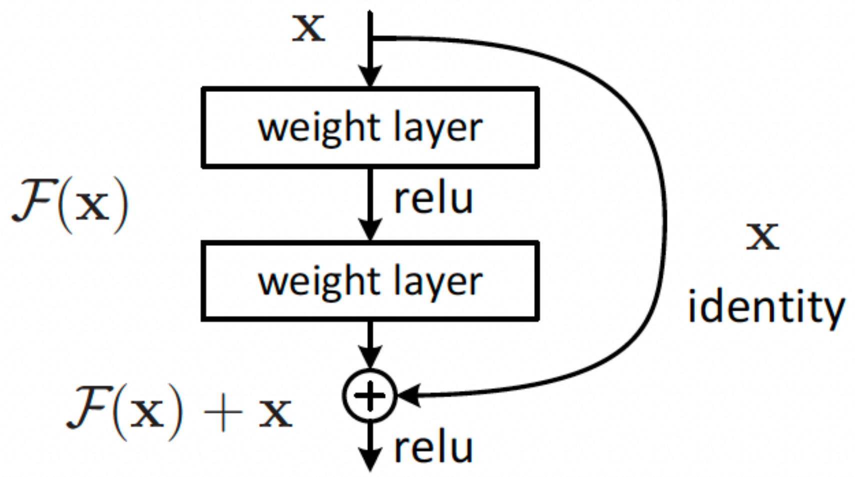
S. Saha *towardsdatascience.com*

Goodfellow et al., *Deep Learning*, MIT Press (2006)

Basic architectures for medical imaging

ResNet

- Very deep NNs are hard to train due to vanishing gradients
- Introduced
  - Residual blocks
  - Skips connections
- Enables training very deep networks → excellent performance (SoTA in 2015)

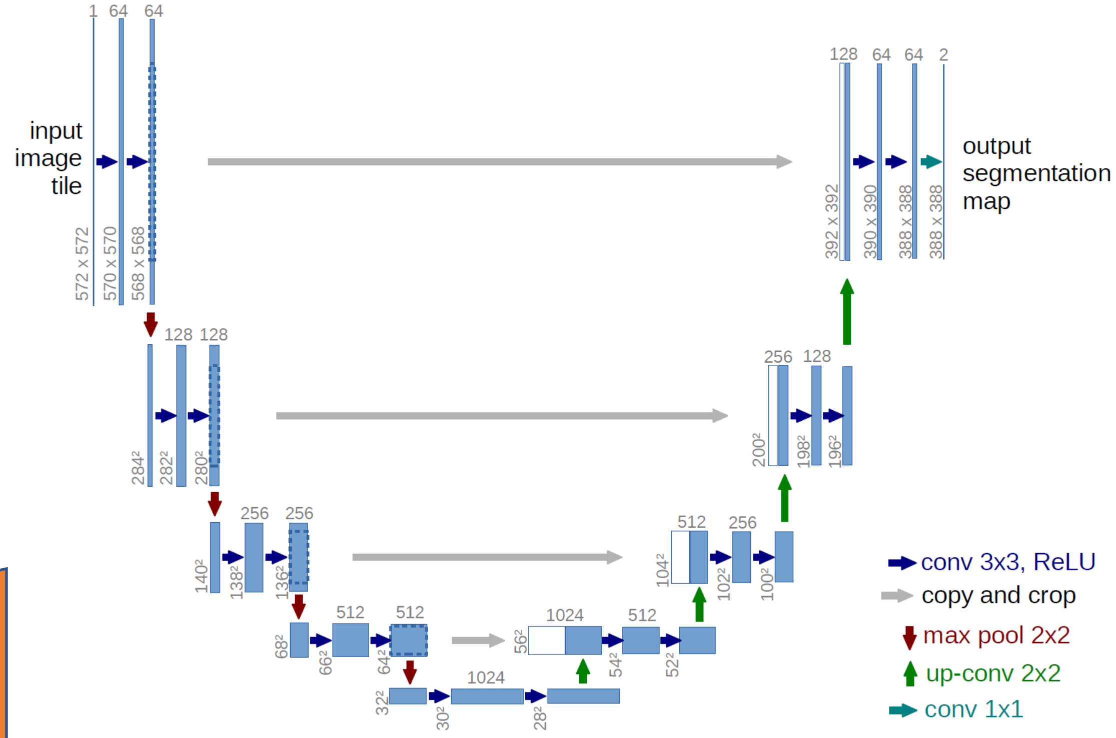


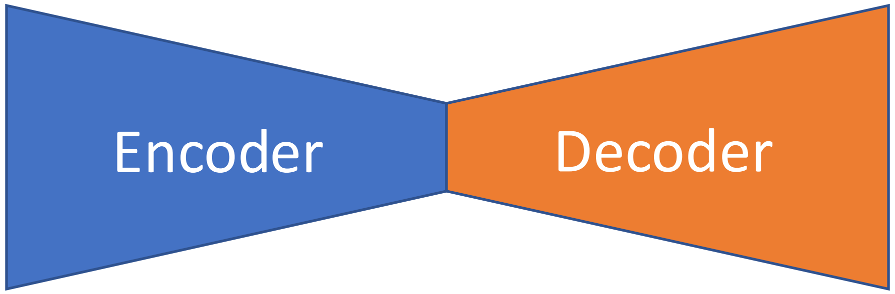
He, Kaiming, et al. "Deep residual learning for image recognition." *CVPR* 2016

Basic architectures for medical imaging

U-NET

- CNN, encoder-decoder
- Down-sampling & up-sampling
- Skip connections
- Developed for segmentation
- Highly popular





Ronneberger, et al. "U-net: Convolutional networks for biomedical image segmentation." *MICCAI 2015*

There are many other architectures!

- AlexNet (2012)
- Recurrent Neural Networks (RNNs)
- Long Short Term Memory (LSTM)
- VGG
- AutoEncoders
- Generative Adversarial Networks (GANs)
- Diffusion models
- Vision Transformers

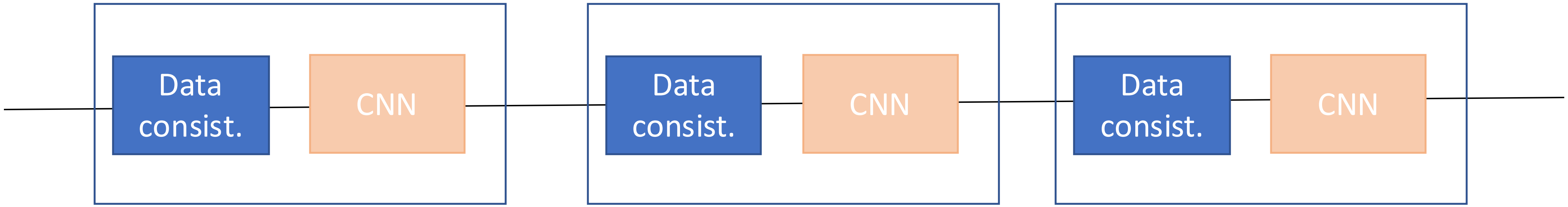
Lundervold, et al." *Zeitschrift für Medizinische Physik* (2019); Alzubaidi et al. *Journal of big Data* (2021)

Outline

1. Neural Networks: Basics
2. Convolutional Neural Networks
3. **Applications in MRI**
4. Challenges and limitations

Image reconstruction

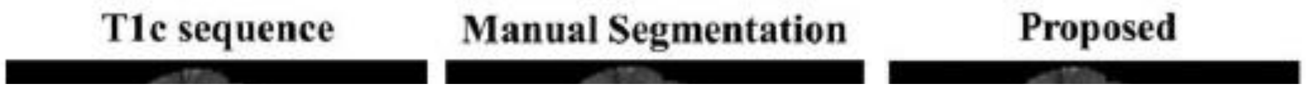
- Aim: scan acceleration by reconstruction from sub-sampled k-space data
- Early DL approaches: **data-driven,** image-to-image
- Current focus: **physics-guided/model-based iterative methods**
  - Unrolled network:

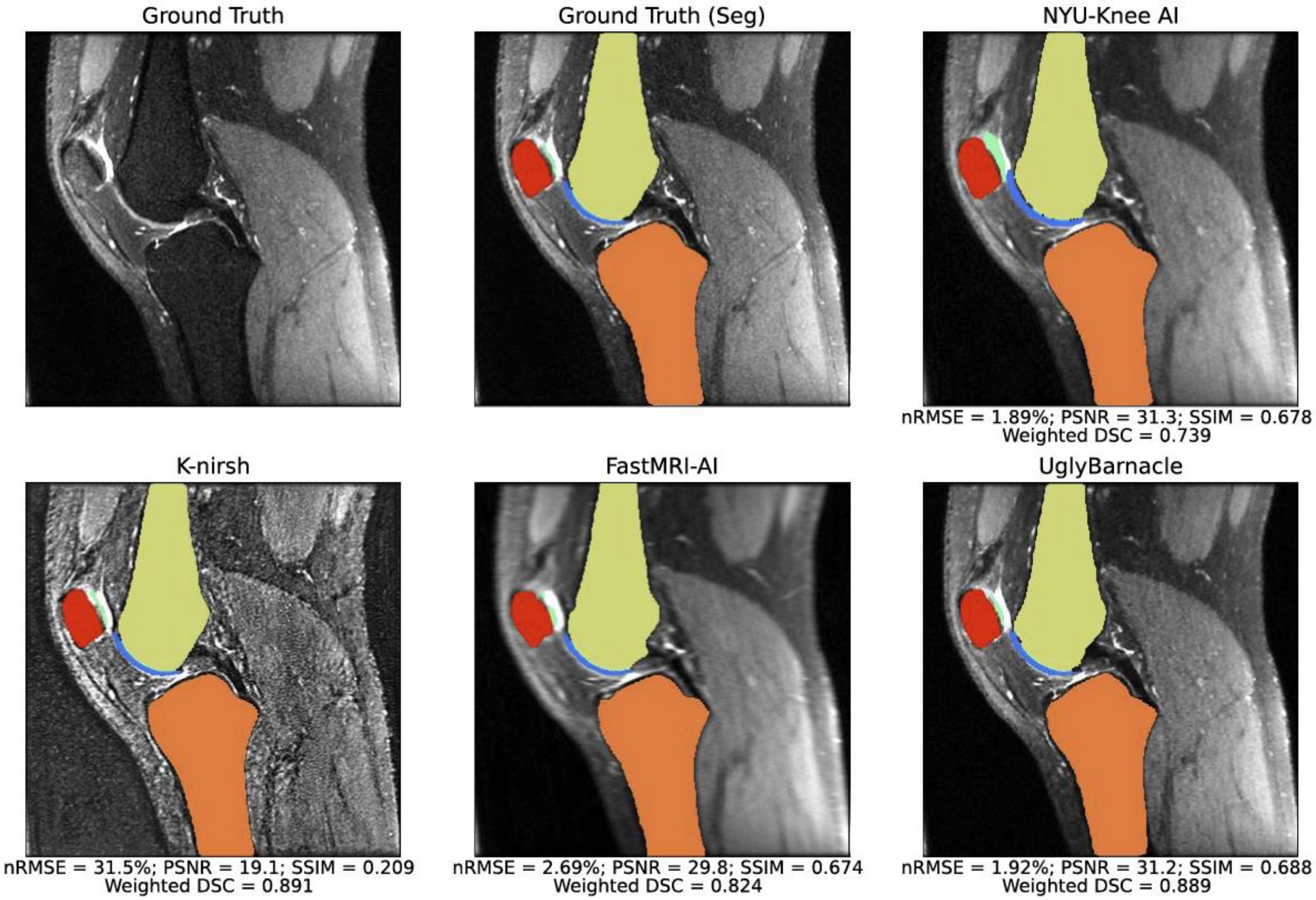


Yang et al., NIPS 2016; Wang et al., ISBI 2016; Schlemper et al., IEEE TMI, 2016; Zhu et al., Nature 2017; Hammernik et al., MRM 2018; Aggarwal et al. IEEE TMI 2018; Jalal et al., NeurIPS 2021

Segmentation

Brain tumor segmentation using CNNs [1]



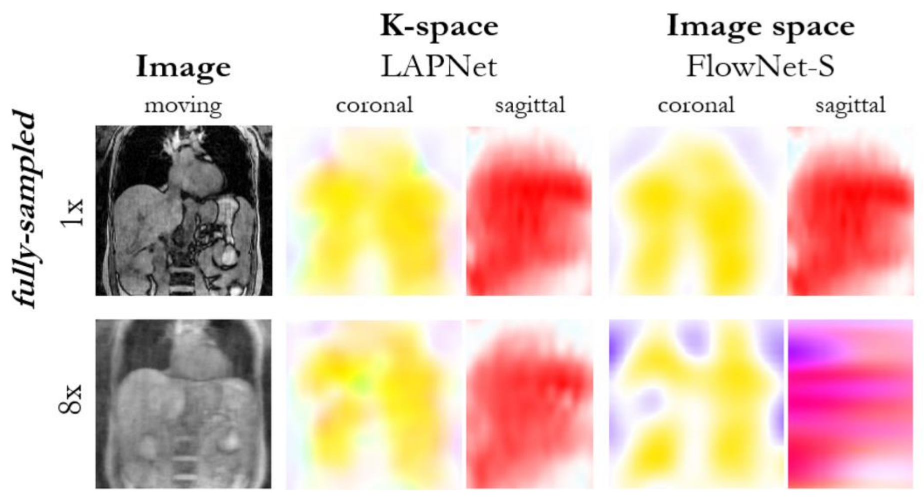


K2S Challenge: from 8x under-sampled k-space to segmentation[2]

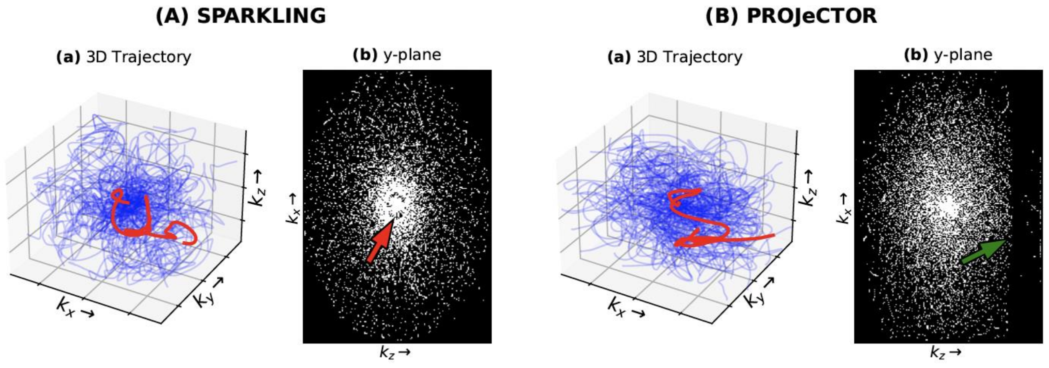
[1] Pereira, *IEEE TMI* (2016); [2] Tolpadi et al., Bioengineering (2023)

Other applications

- Image registration
- Motion correction
- Automated pulse sequence design
- Protocol optimization
- Contrast synthesis
- Quantitative MRI
- Classification



Kustner et al., IEEE TMI 2021



Chaithya & Ciuciu, Bioengineering, 2023

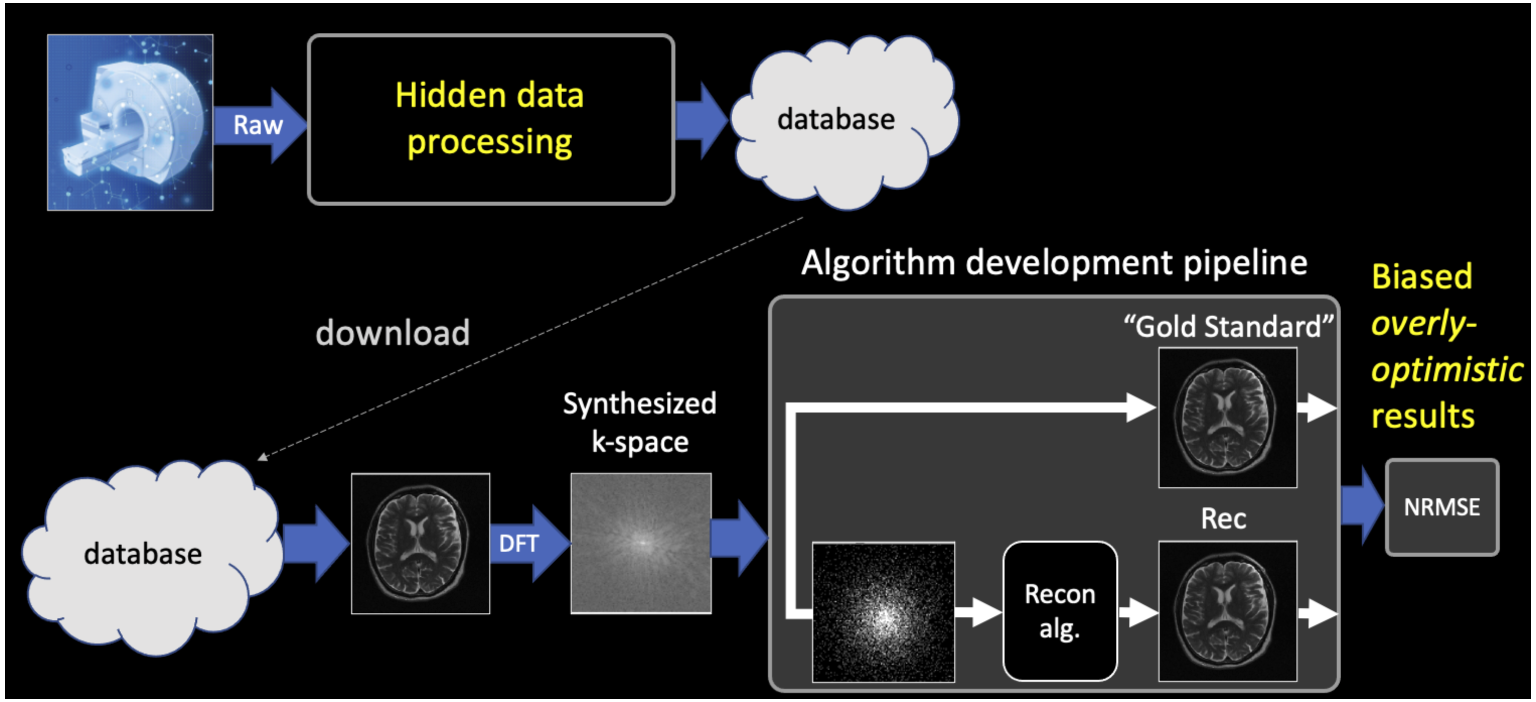
Reviews: Lundervold, et al. *Zeitschrift für Medizinische Physik* (2019); Mazurowski et al., *JMRI* (2019); Alzubaidi et al. *Journal Bf big Data* (2021); Hammernik et al., *Sig Proc. Mag.* (2023) Spieker et al. *arXiv* (2023)

Outline

1. Neural Networks: Basics
2. Convolutional Neural Networks
3. Applications in MRI
4. **Challenges and limitations**

Limited data availability

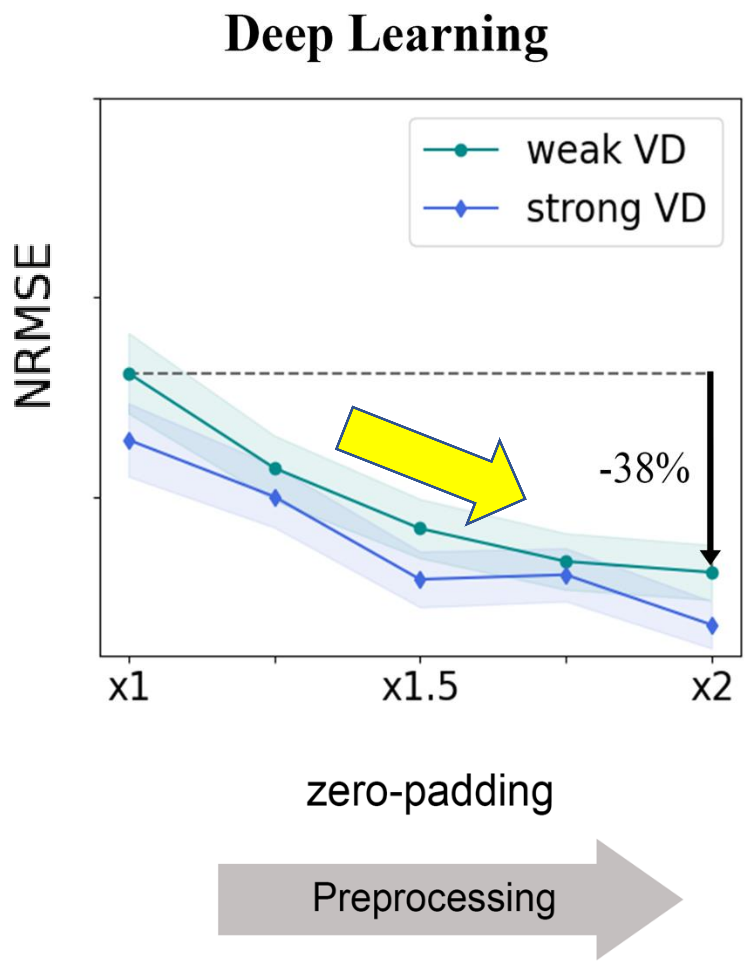
- DL is data-hungry, but training data are scarce
- Common workaround: ”off-label” data use
  - Biased, overly optimistic results



Shimron et al., *PNAS* "Data Crimes" 2022

Limited data availability

- DL is data-hungry, but training data are scarce
- Common workaround: ”off-label” data use
    - Biased, overly optimistic results
    - Algorithmic failure for real-world data



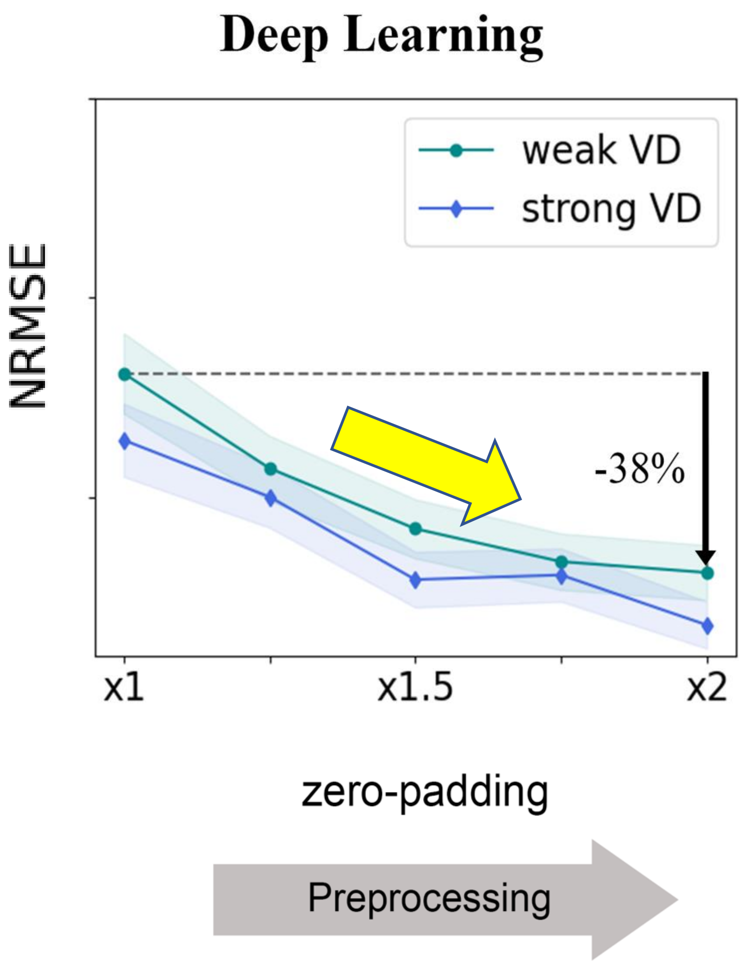
Shimron et al., *PNAS* “Data Crimes” 2022

Limited data availability

- DL is data-hungry, but training data are scarce

Practical Solutions

- Augmentation
- Transfer learning
- Pre-training on other data & fine-tuning
- Training on synthesized/simulated data
- If using processed data: *report* the preprocessing

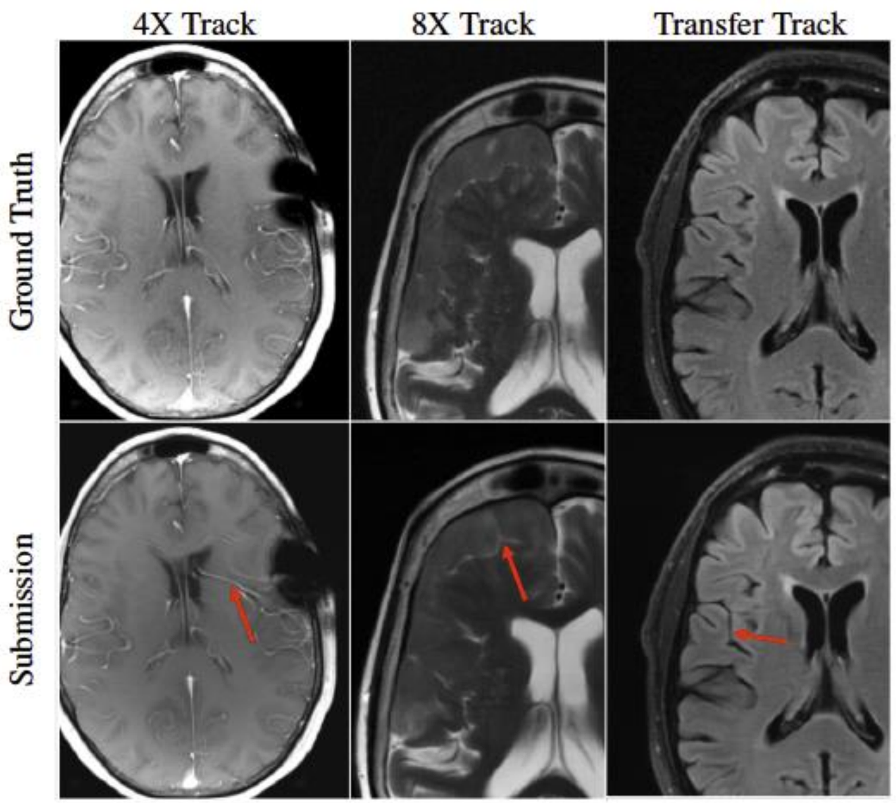


Shimron et al., *PNAS* "Data Crimes" 2022

Hallucinations

- DL can produce *hallucinated structures which look realistic* – hard to detect!
- Open problem
- Uncertainty estimation may help

2nd FastMRI challenge



Muckley et al., arXiv, 2020

Summary

Deep learning is very powerful

Many opportunities for novelty and breakthroughs

However, problems must be addressed!

My Lab has Open positions!

Technion

Team

Join us

40

Our recent publications

Review paper

Deep learning for accelerated and robust MRI reconstruction

Reinhard Heckel, Mathews Jacob, Akshay Chaudhari, Or Perlman, Efrat Shimron, (MAGMA 2024)

K-band: Self-supervised MRI Reconstruction via Stochastic Gradient Descent over K-space Subsets

Frederic Wang, Han Qi, Alfredo De Goyeneche, Michael Lustig, Efrat Shimron (arXiv 2023)

Implicit data crimes: Machine learning bias arising from misuse of public data

Efrat Shimron, Jonathan Tamir, Ke Wang, Michael Lustig