Clinical Applications of AI in Radiation Oncology and Physics

Nick Chng, PhD
Medical Physics

Disclosures

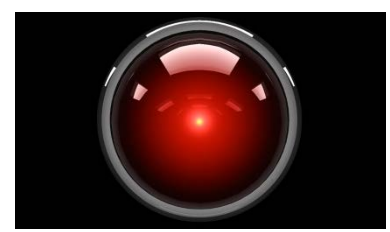
- None, except caveats about quality:
  - I am an enthusiastic amateur, not an AI researcher
  - I cannot compete with YouTube, and that’s a good thing!

What do we mean by AI?

- Artificial Intelligence (AI) vs Machine Learning
  – Lines are blurry, terms often used interchangeably
  – AI can be "general" or "narrow"

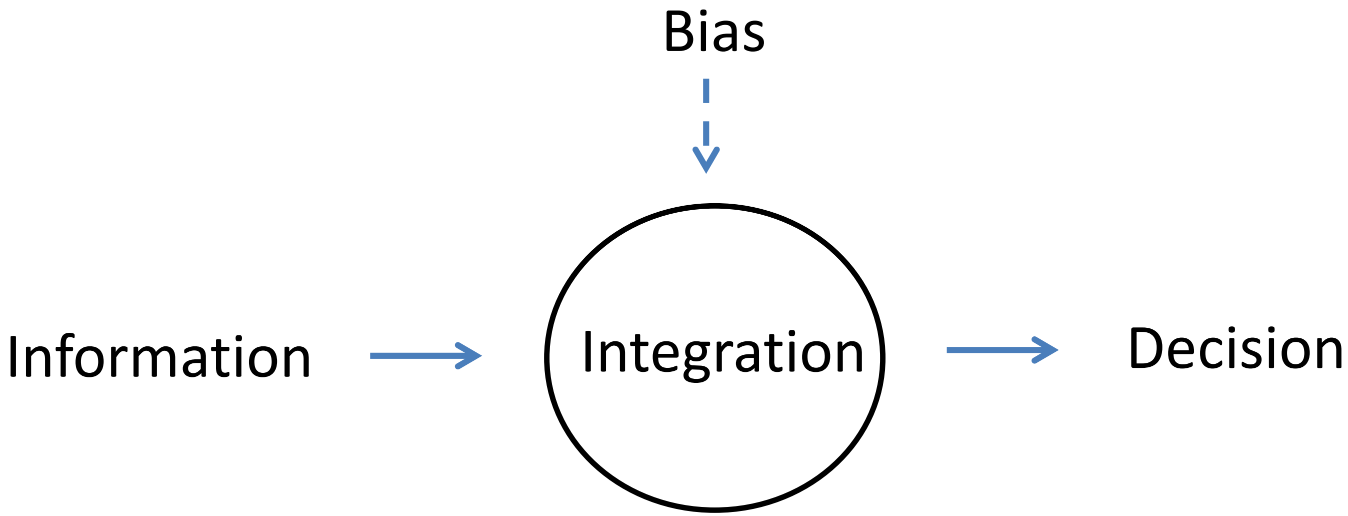
AlphaGo

MOBILEYE



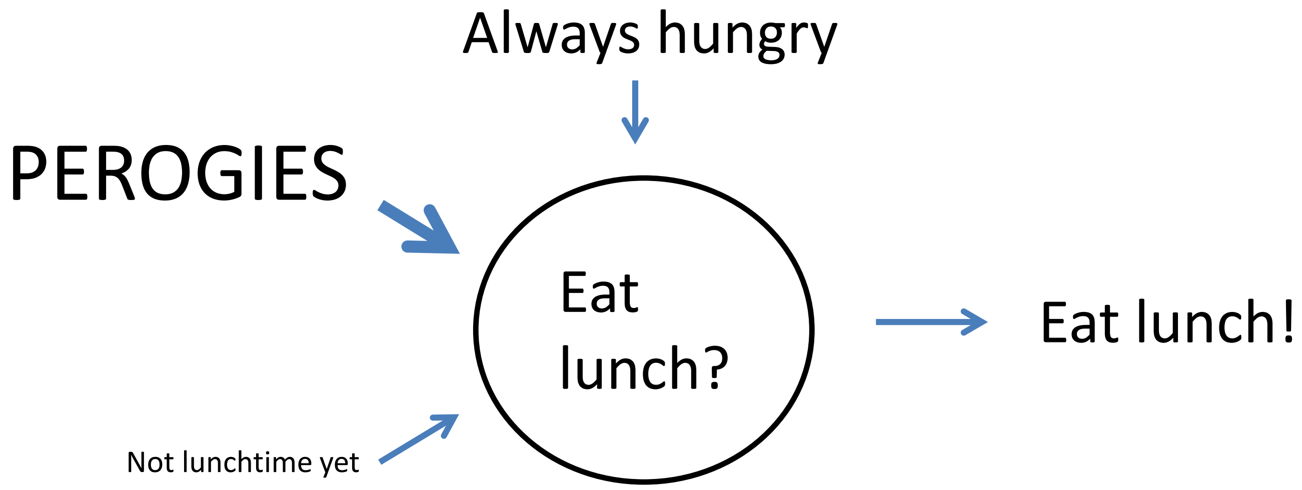
The Neuron

- An artificial **neuron** is an abstraction of a **decision process**

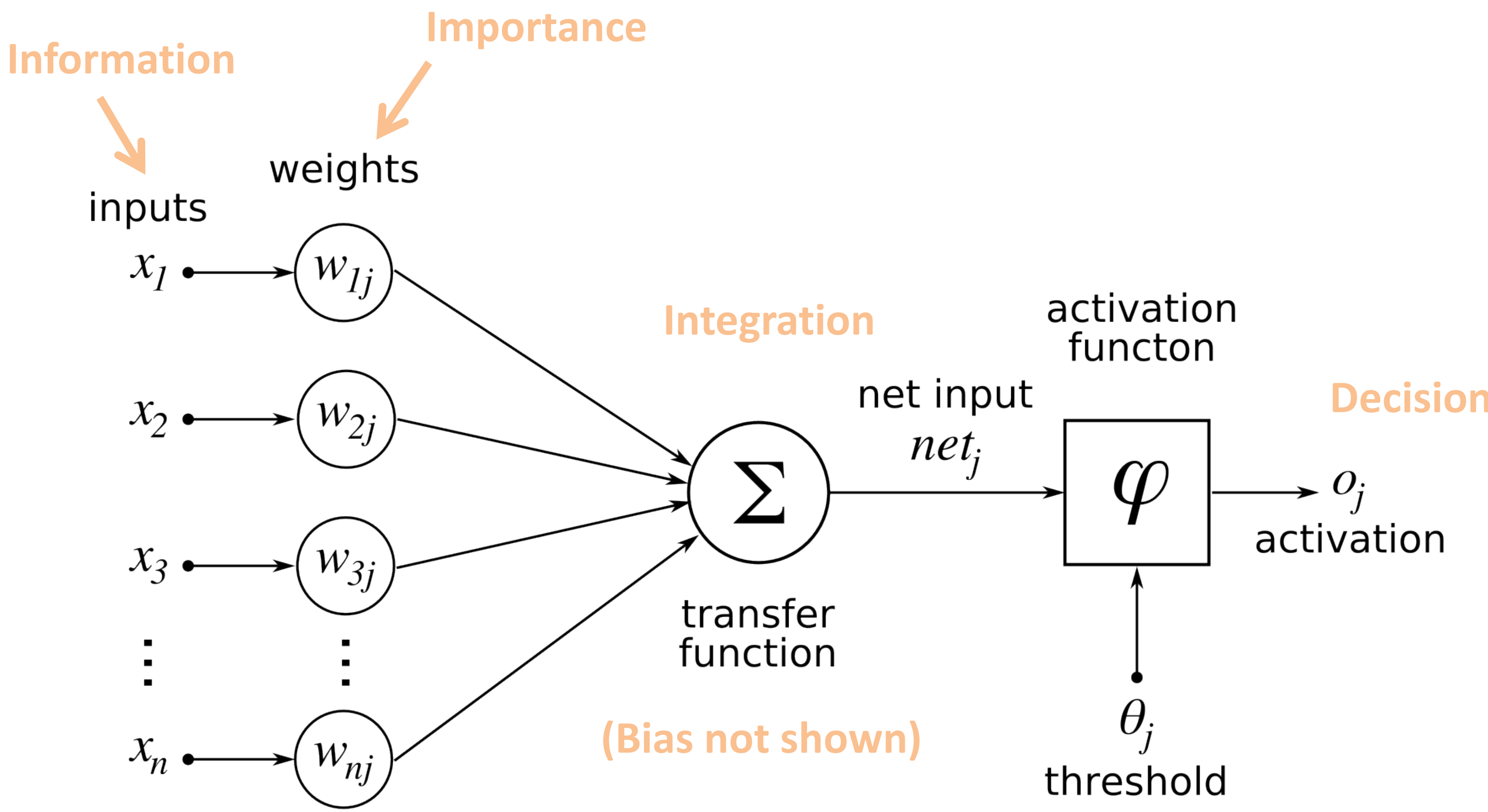


The Neuron

Not all information is of equal importance



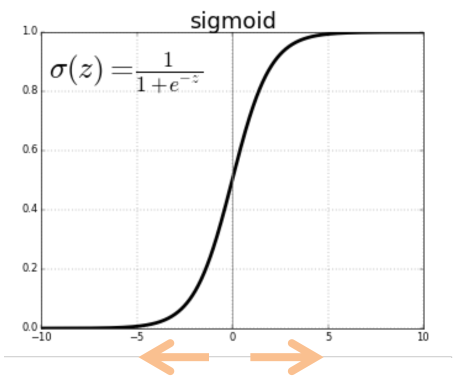
The Neuron

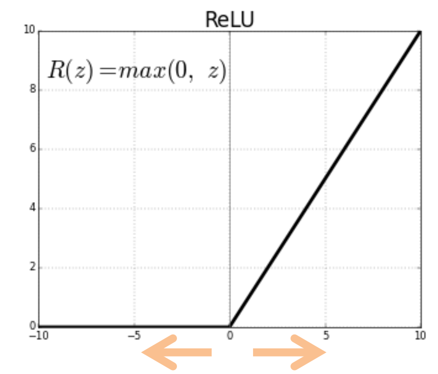


Information
Importance
weights
inputs
x_1
x_2
x_3
x_n
w_{1j}
w_{2j}
w_{3j}
w_{nj}
Integration
Σ
transfer function
(Bias not shown)
net input
net_j
activation functon
φ
θ_j
threshold
Decision
o_j
activation

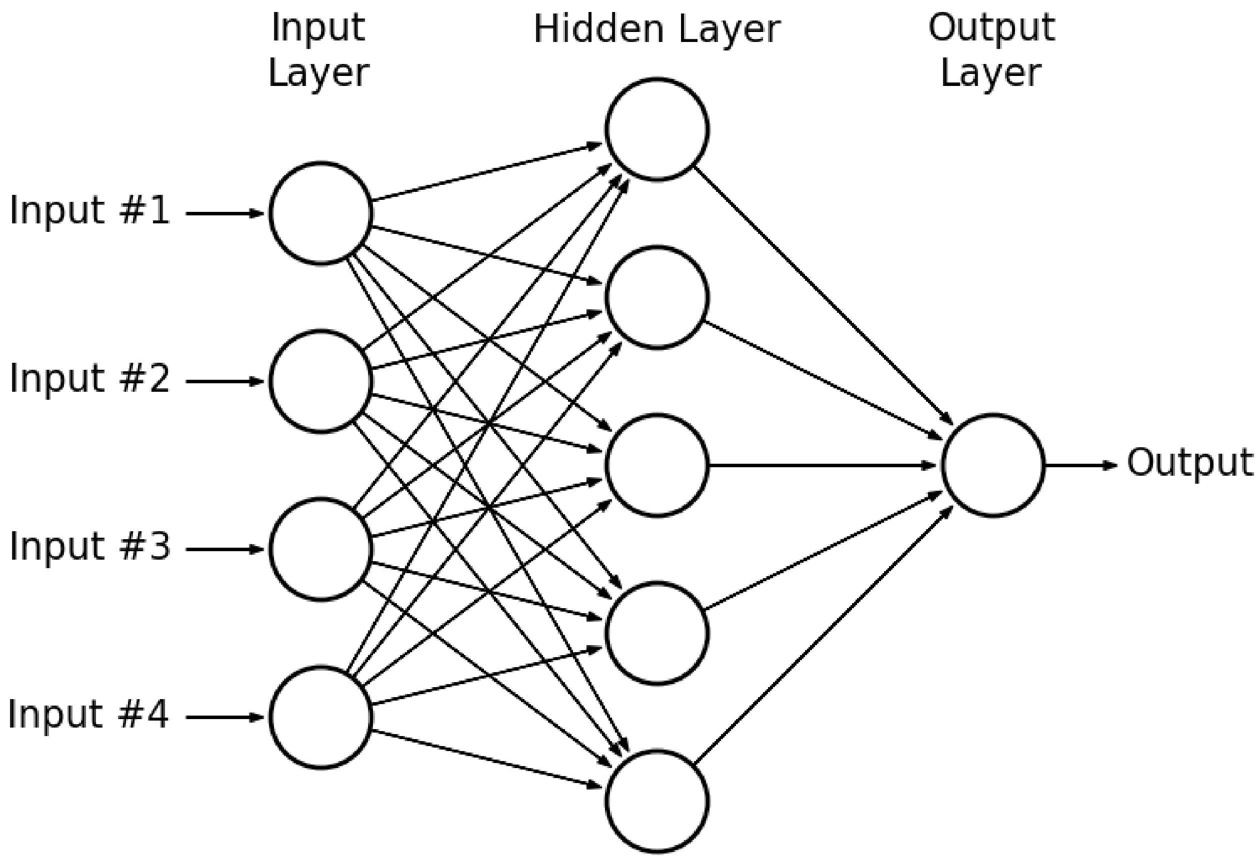
The Neuron

- Different activation models are used
  - **Sigmoid** vs **tanh** vs **ReLU**
  - **Bias** alters the magnitude of input necessary to activate



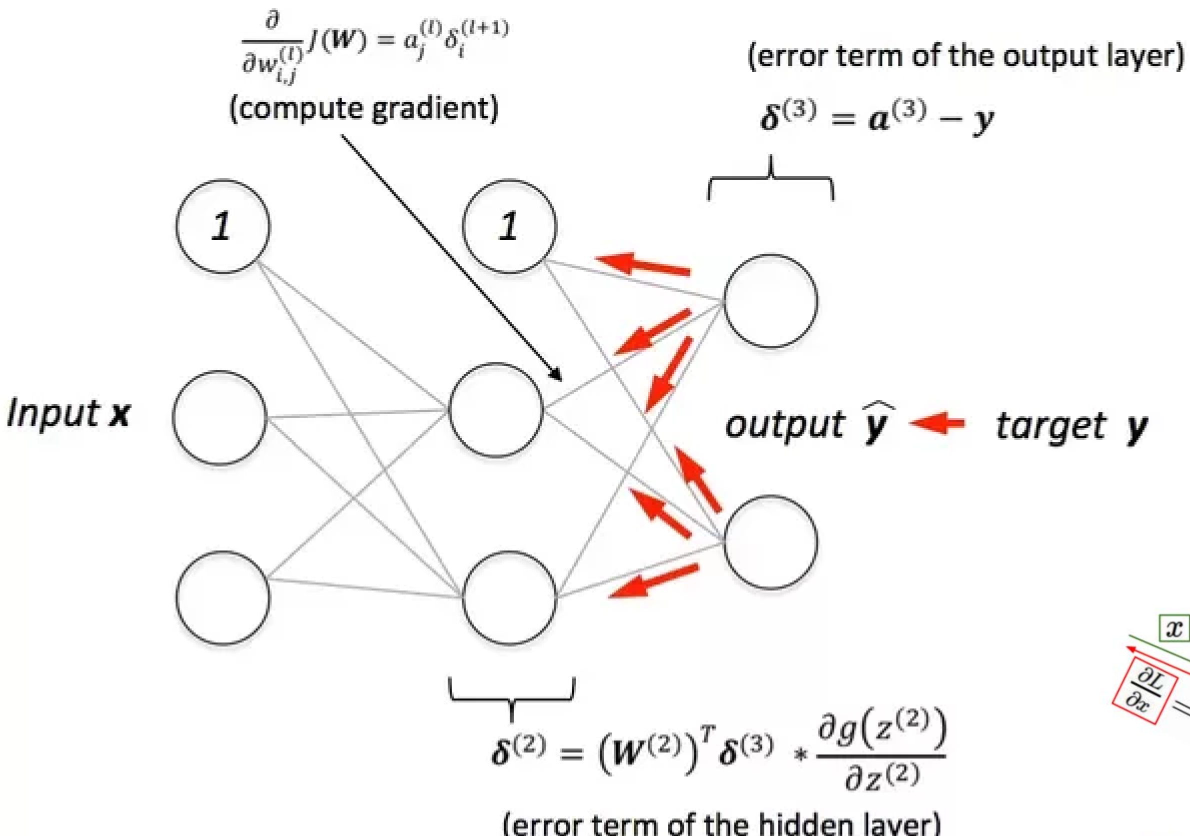


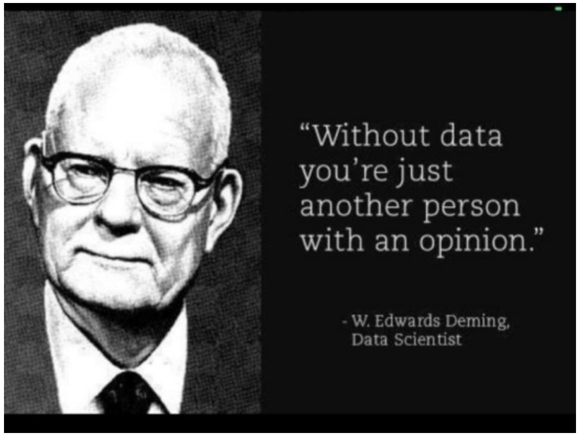
The Neural Network

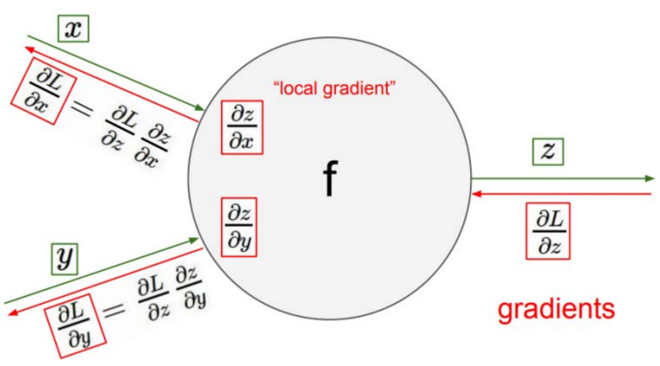


"Fully connected layers"

Training by Backpropagation







The Neural Network

- The number of hidden layers is a design decision
- **Deep networks** are simply networks with a lot of hidden layers
- **Any function** can be approximated by a single hidden layer of sufficient breadth
  - No guarantee it can be determined!

Convolutional Neural Networks (CNNs)

- Computer vision includes many tasks relevant to diagnosis and treatment
  - Object classification
  - Image processing
  - Segmentation
- Convolution of the input images with various filters produces feature maps
- The attached neural network learns which combination of features are relevant

Convolutional Neural Networks (CNNs)

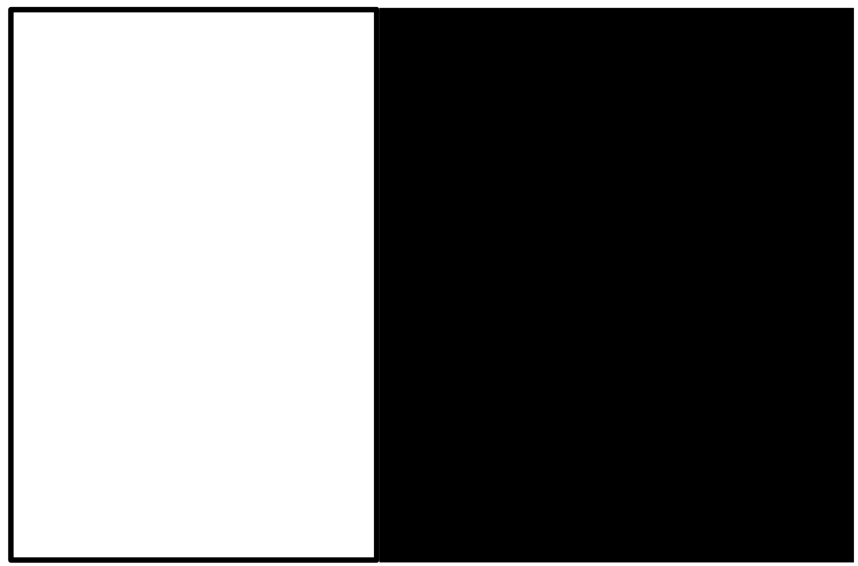
- CNNs are popular because:
  - Convolution is embarrassingly parallel
  - Synergy with computer graphics hardware
  - Infrastructure exists to collect huge datasets (phones, social media, IoT)
  - Training can be crowd-sourced

How CNNs work in 5 slides…

Edge detection

-1	0	1
-2	0	2
-2	0	1

Sobel Filter



The convolution filter

For example - Edge detection

-1	0	1
-2	0	2
-2	0	1

Sobel Filter

0	0	0	1	1	1	1
0	0	0	1	1	1	1
0	0	0	1	1	1	1
0	0	0	1	1	1	1
0	0	0	1	1	1	1
0	0	0	1	1	1	1

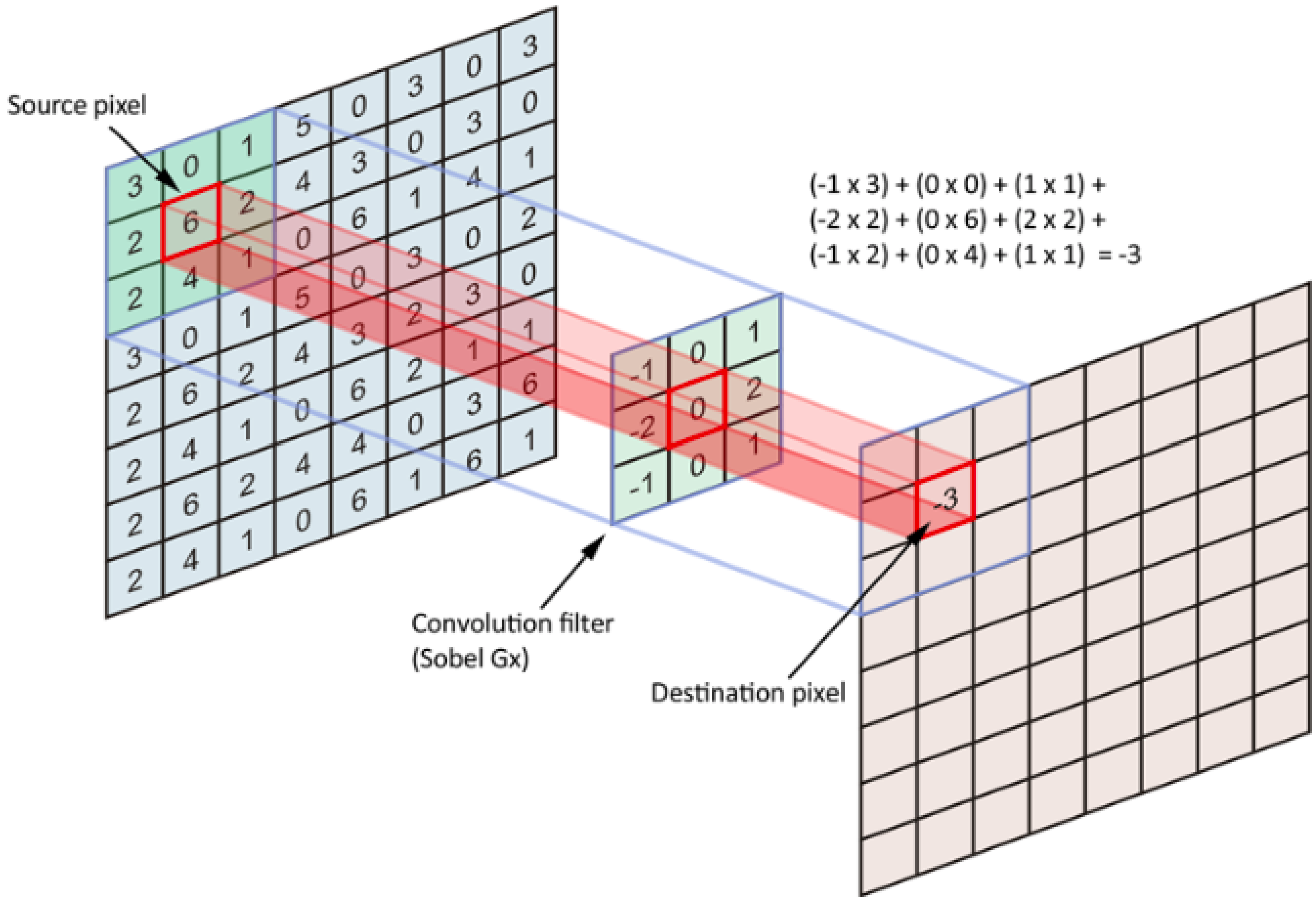
Convolutional Neural Networks

-1	0	1
-2	0	2
-2	0	1

Sobel Filter

0	0	0	1	1	1	1
0	0	0	1	1	1	1
0	0	0	1	1	1	1
0	0	0	1	1	1	1
0	0	0	1	1	1	1
0	0	0	1	1	1	1

Result of filter is the sum of the corresponding elements of the filter and image


Source pixel
(-1 x 3) + (0 x 0) + (1 x 1) +
(-2 x 2) + (0 x 6) + (2 x 2) +
(-1 x 2) + (0 x 4) + (1 x 1) = -3
Convolution filter
(Sobel Gx)
Destination pixel

Convolutional Neural Networks

-1	0	1
-2	0	2
-2	0	1

*

0	0	1
0	0	1
0	0	1

=

0	0	1
0	0	2
0	0	1

→

4

Output

0	0	0	1	1	1	1
0	0	0	1	1	1	1
0	0	0	1	1	1	1
0	0	0	1	1	1	1
0	0	0	1	1	1	1
0	0	0	1	1	1	1

Input Image

→

0	0	0	0	0	0	0
0	0	4	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0

Output Image

Convolutional Neural Networks

Convolution is the result of the application of the filter across the image

0	0	0	1	1	1	1
0	0	0	1	1	1	1
0	0	0	1	1	1	1
0	0	0	1	1	1	1
0	0	0	1	1	1	1
0	0	0	1	1	1	1

Input Image

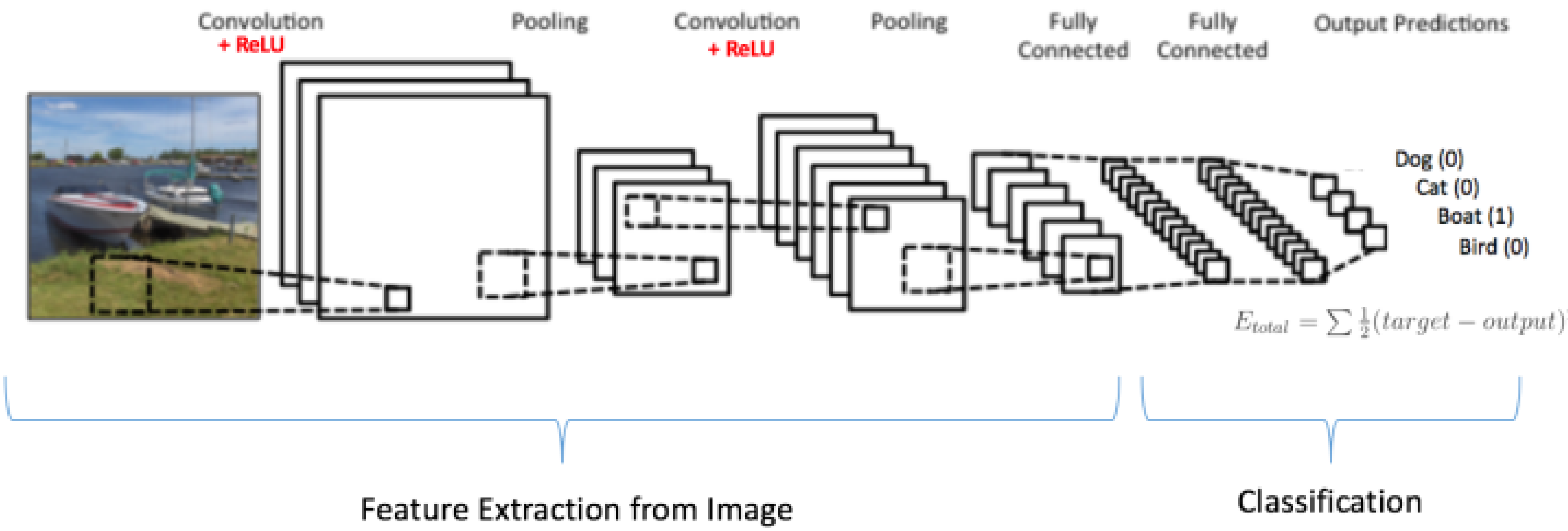
0	0	4	4	0	0	0
0	0	4	4	0	0	0
0	0	4	4	0	0	0
0	0	4	4	0	0	0
0	0	4	4	0	0	0
0	0	4	4	0	0	0

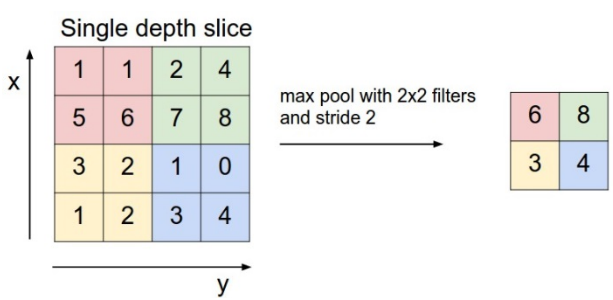
Output Image (after convolution)

Convolutional Neural Networks

Edge detection

CNNs: Classification





"The *pooling* operation used in convolutional neural networks is a big mistake and the *fact* that it works *so well is a disaster*."

George Hinton

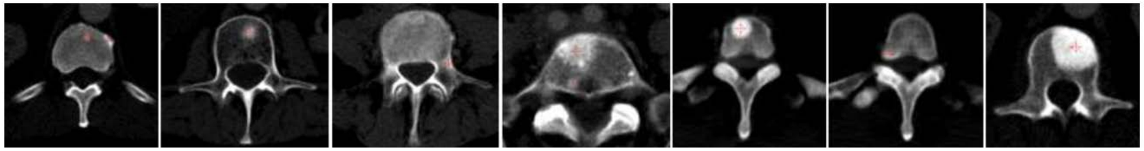


Figure 1: Examples of sclerotic metastases as detected by the CADe candidate generation step (red mark).

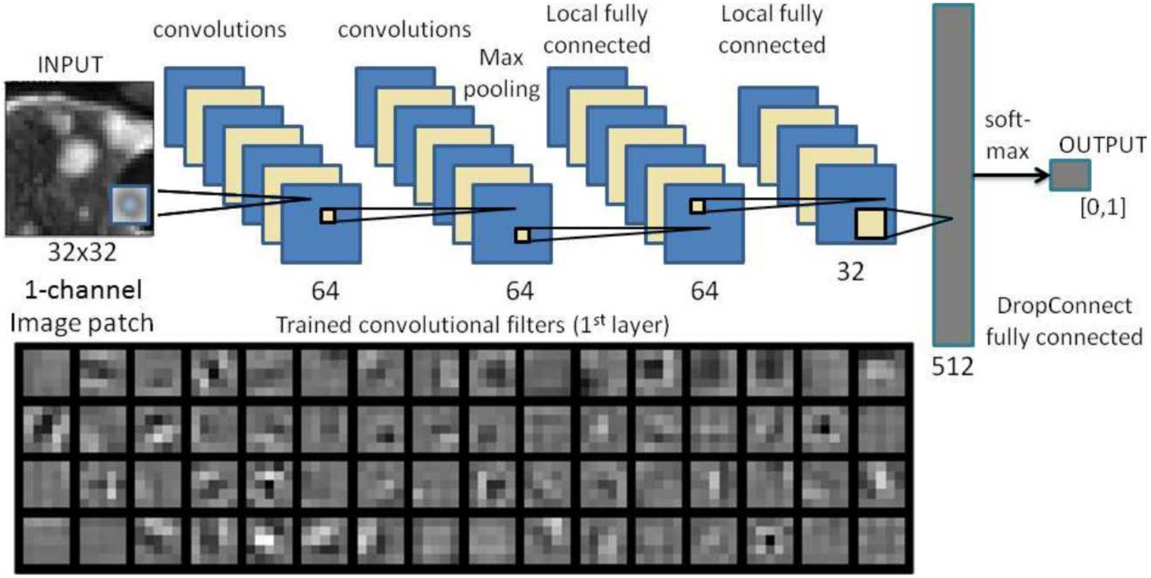
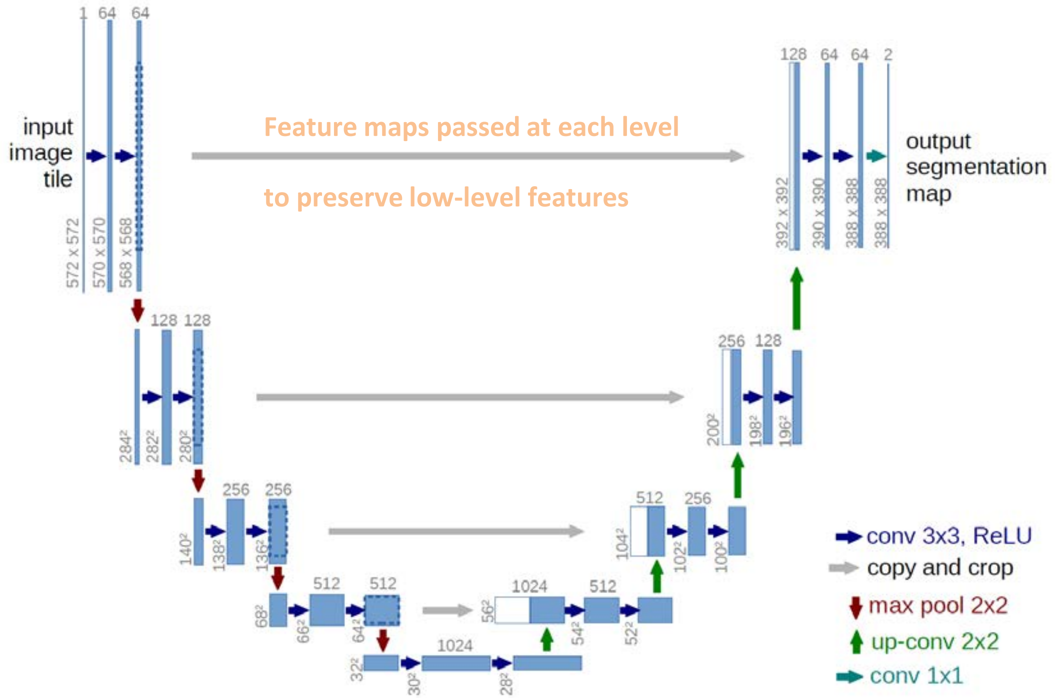


Figure 3: The proposed convolution neural network consists of two convolutional layers, max-pooling layers, locally fully-connected layers, a DropConnect layer, and a final 2-way softmax layer for classification. The number of filters, connections for each layer, and the first layer of learned convolutional kernels are shown.

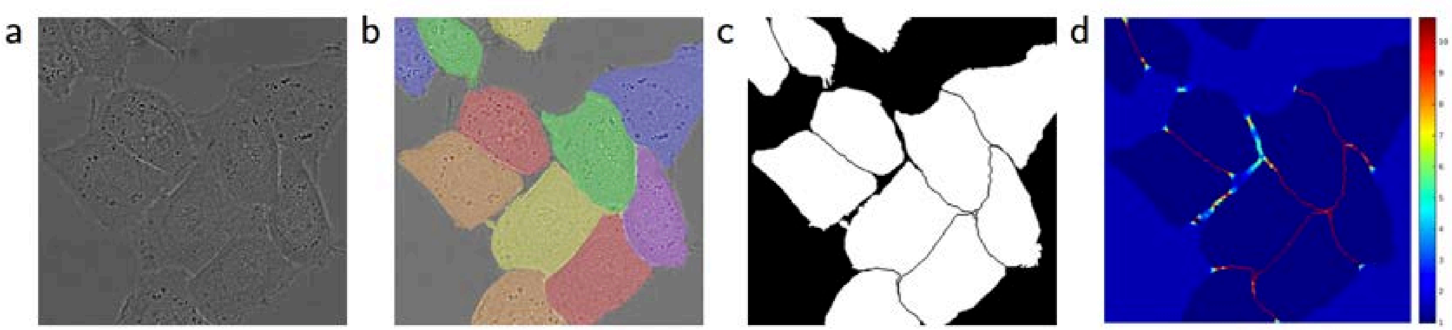
reject difficult false positives while preserving high sensitivities. We validate the approach on CT images of 59 patients (49 with sclerotic metastases and 10 normal controls). The proposed method reduces the number of FP/vol. from 4 to 1.2, 7 to 3, and 12 to 9.5 when comparing a sensitivity rates of 60%, 70%, and 80% respectively in testing. The Area-Under-the-Curve (AUC) is 0.834.

Detection of Sclerotic Spine Metastases via Random Aggregation of Deep Convolutional Neural Network Classifications (Roth et al. 2014)

CNN: U-nets for segmentation



Input | Ground truth masks | Segmentation map | Pixel-wise loss weights



"U-Net: Convolutional Networks for Biomedical Image Segmentation" Ronneberger et al. (Winner of the Cell Tracking Challenge at ISBI 2015)

CNN: U-nets for dose prediction

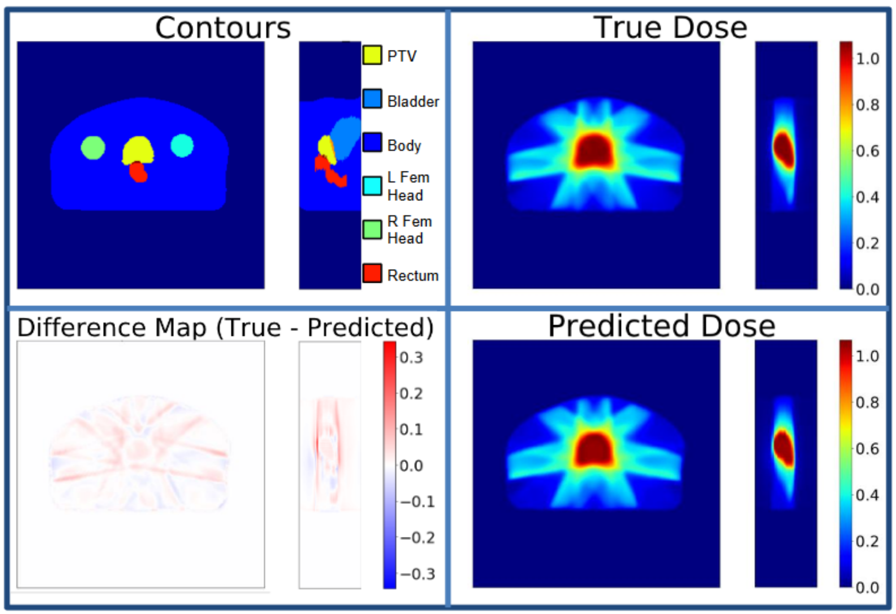


Figure 7: Contours of the planning target volume (PTV) and organs at risk (OAR), true dose wash, predicted dose wash, and difference map of an example patient.

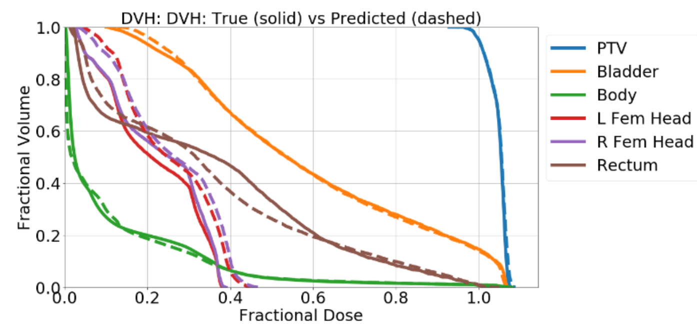


Figure 8: Example of typical dose volume histogram (DVH) comparing true dose and predicted dose for one patient.

As a typical prediction example from the U-net model, Figure 7 shows the input contours, true and predicted dose washes, and a difference map of the two doses for one patient. On average, the dose difference inside the body was less than 1% of the prescription dose, shown in Table 1. Figure 8 shows the DVH of one of the example test patients. Visually on the DVH, one can see that the U-net tends to predict a similar PTV dose coverage with minimal errors in the dose prediction to the OARs.

Dose Prediction with U-net: A Feasibility Study for Predicting
Dose Distributions from Contours using Deep Learning on Prostate IMRT Patients (Nguyen et al. 2017)

NPC target segmentation

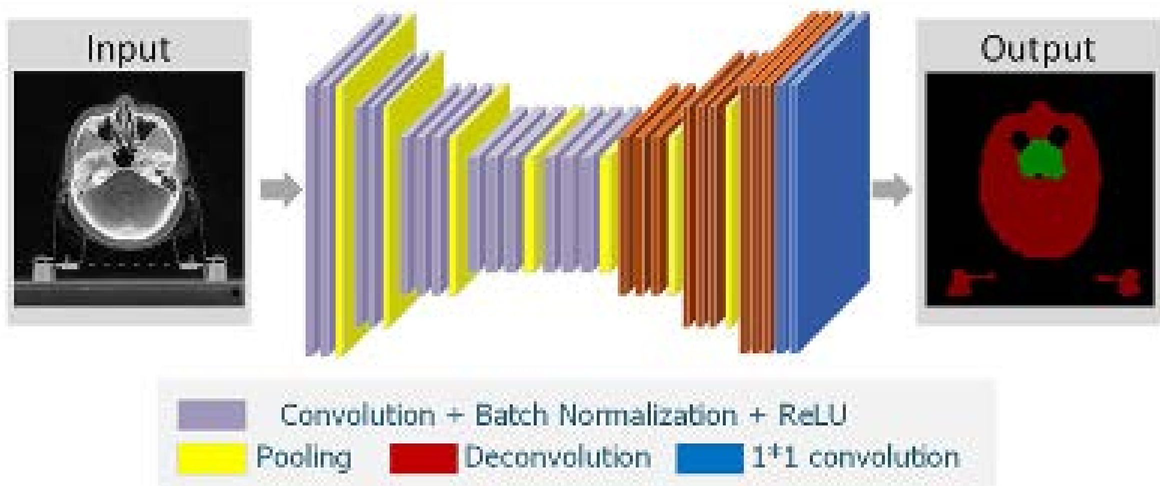


FIGURE 1 | Overall framework of the proposed algorithm.

TABLE 1 | Dice similarity coefficient (DSC) and Hausdorff distance for nasopharynx gross tumor volume (GTVnx), metastatic lymph node gross tumor volume (GTVnd), and clinical target volume (CTV).

	DSC (%)			Hausdorff distance (mm)		
Region of interest	**CTV**	**GTVnx**	**GTVnd**	**CTV**	**GTVnx**	**GTVnd**
Deep deconvolutional neural network	82.6	80.9	62.3	6.9	5.1	25.8
VGG-16	73.7	72.3	33.7	11.1	7.7	51.5

Deep Deconvolution Neural Network for Target Segmentatino of Nasopharyngeal Cancer in Planning Computed Tomographi Images. Men et al. 2017

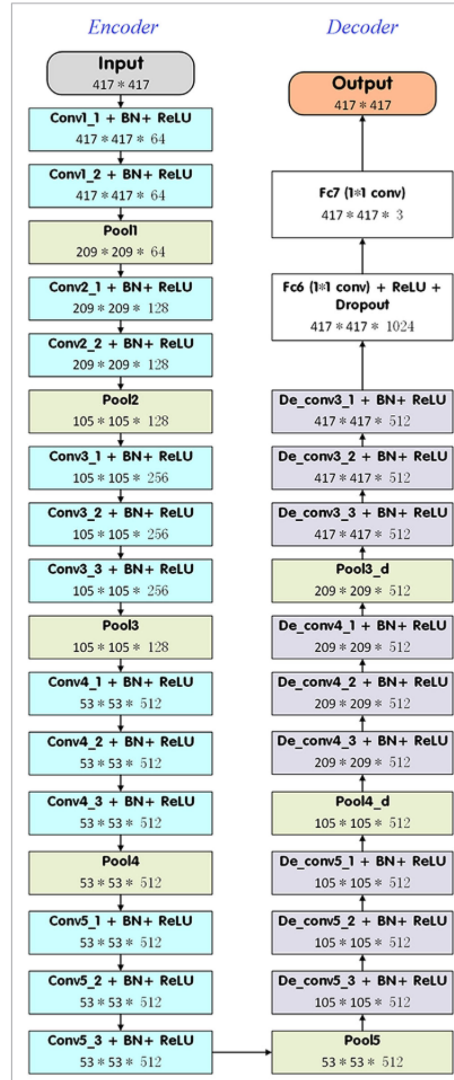
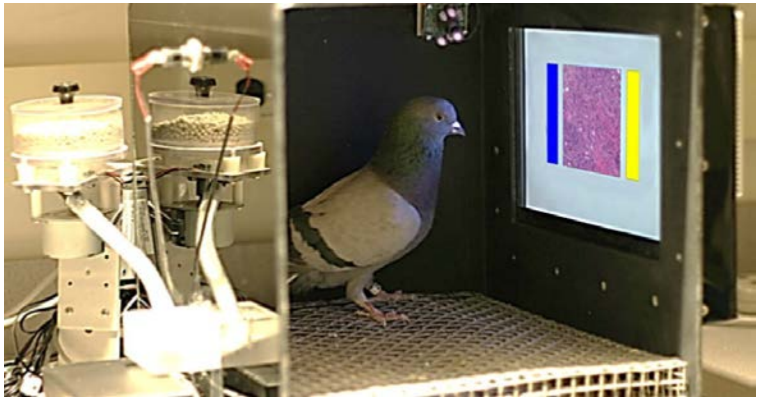


FIGURE 2 | The detailed architecture of deep deconvolutional neural network.

Implementations

- A number of frameworks are available
  - TensorFlow, Caffe, PyTorch, Keras
- Python based systems are easier to get started with
- Matlab has a Caffe interface
- Pigeons!



Pigeons (*Columba livia*) as Trainable Observers of Pathology and Radiology Breast Cancer Images. Levenson et al. 2015

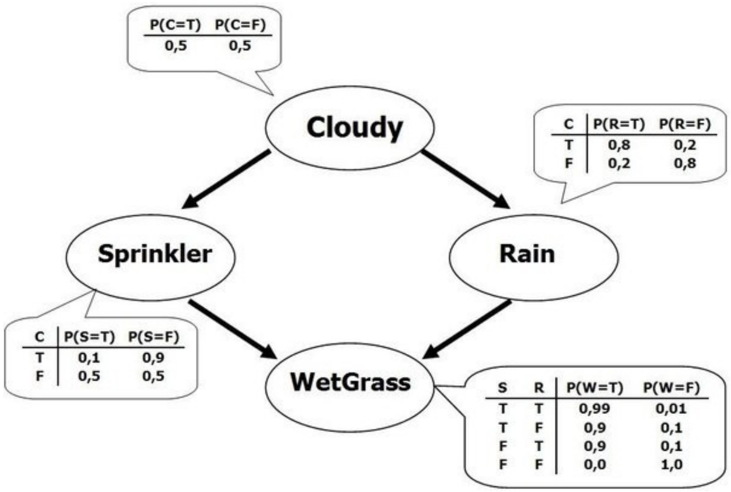
Bayesian Networks

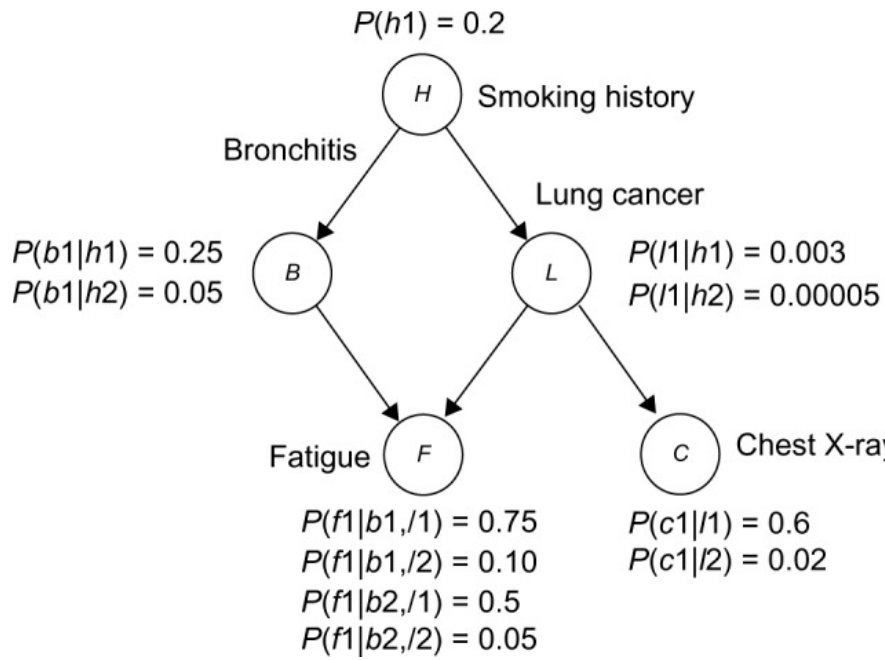
- Bayesian networks are graphs that encode the **probabilistic dependencies between states**
- They are not neural networks but can be used for inference
- Graph structure is critical and set out in advance

"Without an opinion, you're just another person with data"

-Something I believe the Rev. Bayes might have said

Bayesian Networks





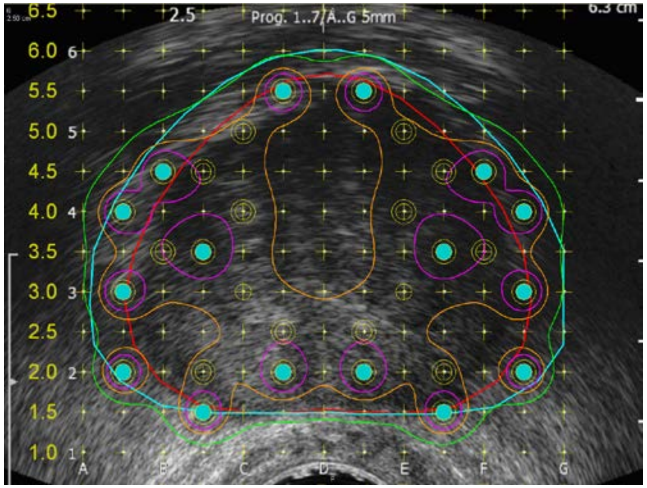
Bayes nets can be used to simplify causal systems when conditional independence exists (or can be reasonably approximated)

Bayesian Networks

- Bayesian networks can be trained to encode the probability of various system states
- The conditional probabilities of each node are updated based the frequency that that state exists in the training data
- Unlike neural networks, a priori beliefs can be encoded to make them more robust to error
- Used for post-cancer survivorship models, kidney transplant decision-making, and regional lymph node status

Bayes' Net for prostate brachytherapy planning

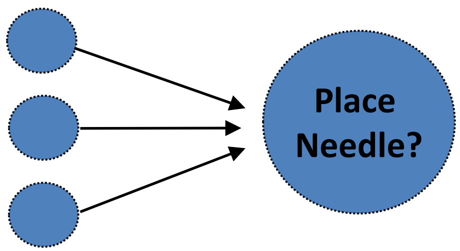
- Insertion of radioactive I-125 seeds into prostate via transperineal needles
- Planning challenge is to find the distribution of seeds that adequately treats the prostate with maximal simplicity
- Bayes net was trained on the distribution of 145 past patients, encoding the most common needle distributions.



The SOURCE Network

In SOURCE, the placement likelihood of each needle is represented by a node…

…informed by a set of ***contour attributes***… (evidence)



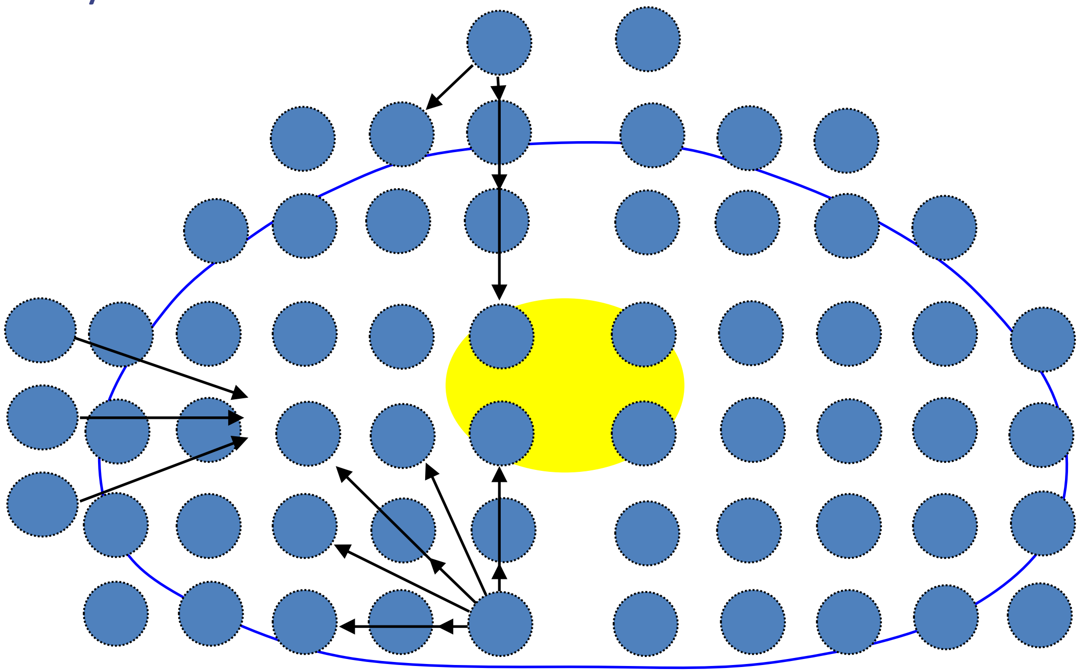
BC Cancer Agency
CARE + RESEARCH
An agency of the Provincial Health Services Authority

* For simplicity, not all dependencies may be illustrated

Bayesian Network in SOURCE

In SOURCE, the placement likelihood of each needle is represented by a node…

…with a likelihood characterized by a set of ***contour attributes…***



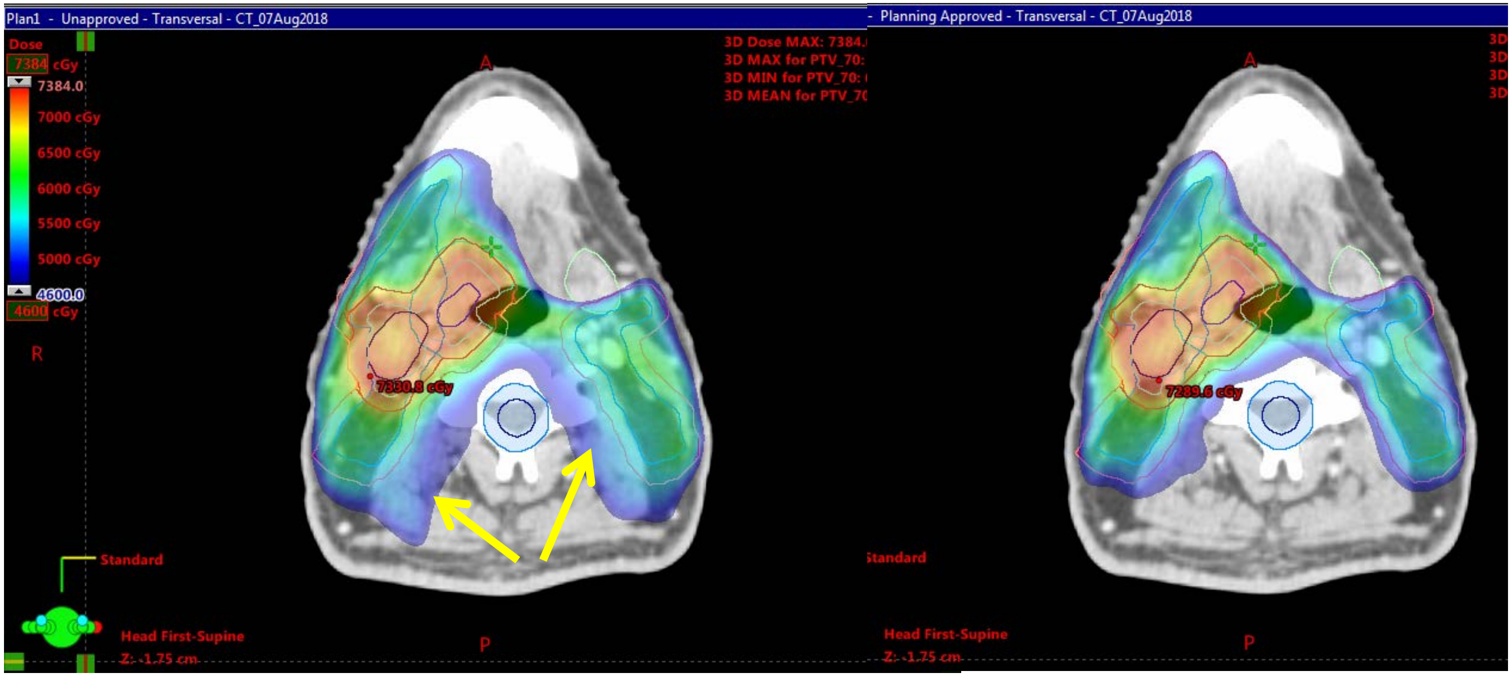
… and the existence of other needles in the proposed plan (at any iteration)

BC Cancer Agency
CARE + RESEARCH
An agency of the Provincial Health Services Authority

* For simplicity, not all dependencies may be illustrated

Varian RapidPlan™

- First clinical implementation of “Knowledge based radiotherapy planning”
- Predicts achievable dose parameters based on features and trade-offs in past cases
- Not really “AI” – does not learn best representation



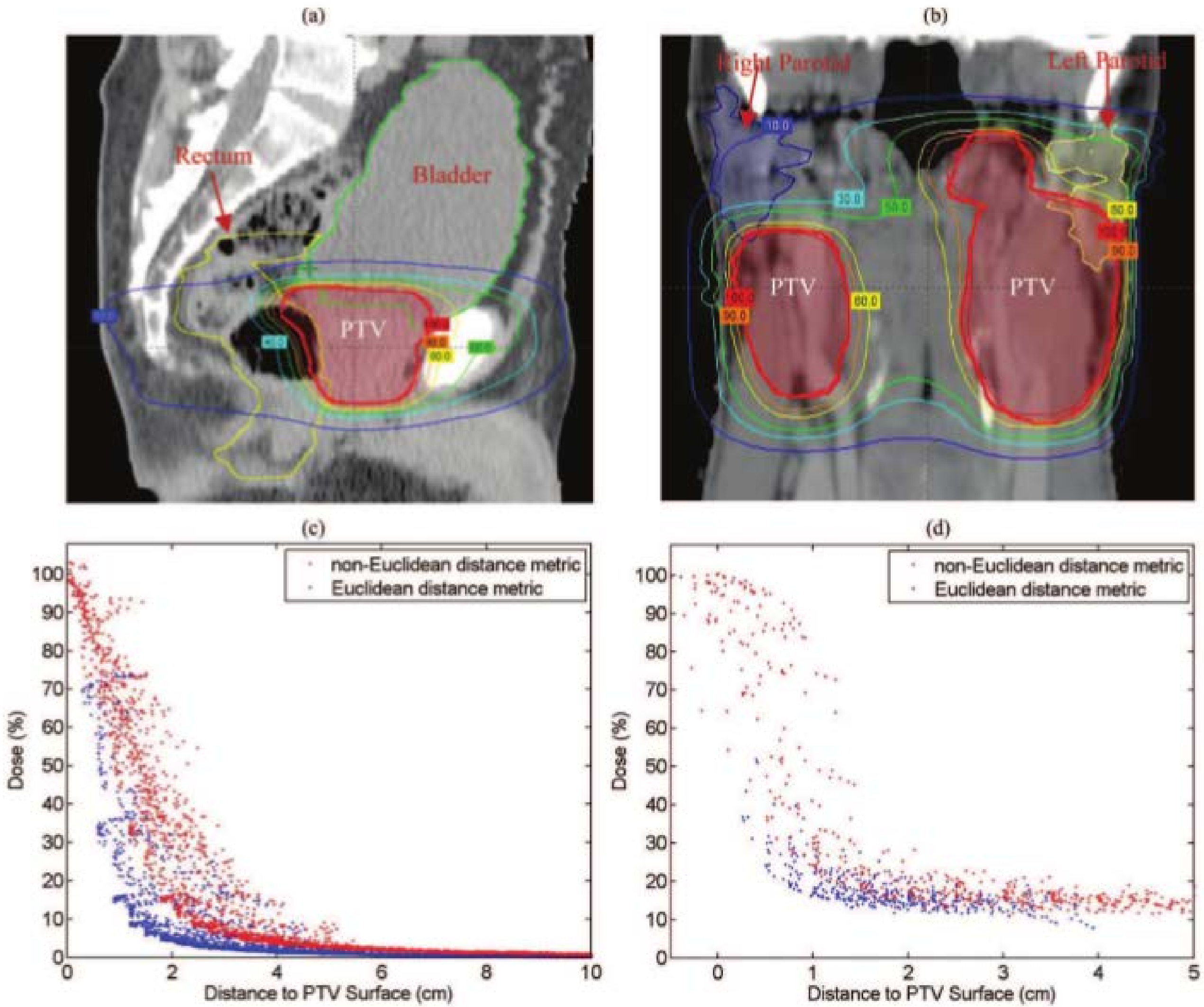
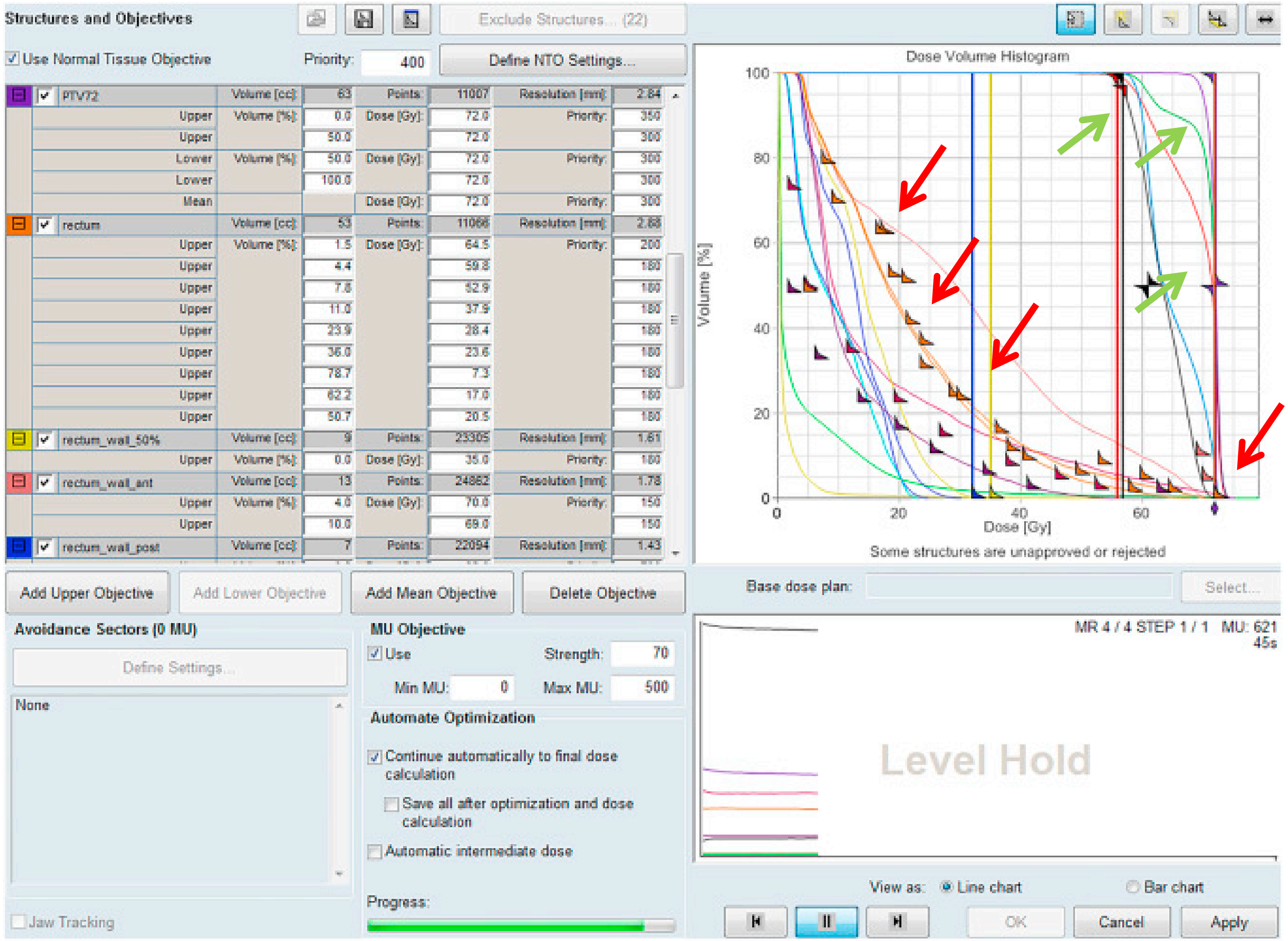


FIG. 1. (a) Sagittal CT image of a prostate plan showing the contours of PTV, bladder and rectum overlaid with isodose lines. (b) Coronal CT image of a HN plan showing the contours of PTV, left and right parotids overlaid with isodose lines. (c) and (d) Scatter plots of the correlation between dose and distance to PTV surface by the Euclidean distance metric and the non-Euclidean distance metric for the voxels inside (c) bladder in the prostate plan and (d) right parotid in the HN plan. Note the spread of dose-distance correlation is reduced by the non-Euclidean distance metric.

Quantitative analysis of the factors which affect the interpatient organ-at-risk dose sparing variation in IMRT plans. Yuan et al. 2012



Structures and Objectives
Exclude Structures... (22)
Use Normal Tissue Objective
Priority: 400
Define NTO Settings...
PTV72 Volume [cc]: 63 Points: 11007 Resolution [mm]: 2.84
Upper Volume [%]: 0.0 Dose [Gy]: 72.0 Priority: 350
Upper 50.0 72.0 300
Lower Volume [%]: 50.0 Dose [Gy]: 72.0 Priority: 300
Lower 100.0 72.0 300
Mean Dose [Gy]: 72.0 Priority: 300
rectum Volume [cc]: 53 Points: 11066 Resolution [mm]: 2.88
Upper Volume [%]: 1.5 Dose [Gy]: 64.5 Priority: 200
Upper 4.4 59.8 180
Upper 7.8 52.9 180
Upper 11.0 37.9 180
Upper 23.9 28.4 180
Upper 36.0 23.6 180
Upper 78.7 7.3 180
Upper 62.2 17.0 180
Upper 50.7 20.5 180
rectum_wall_50% Volume [cc]: 9 Points: 23305 Resolution [mm]: 1.61
Upper Volume [%]: 0.0 Dose [Gy]: 35.0 Priority: 180
rectum_wall_ant Volume [cc]: 13 Points: 24862 Resolution [mm]: 1.78
Upper Volume [%]: 4.0 Dose [Gy]: 70.0 Priority: 150
Upper 10.0 69.0 150
rectum_wall_post Volume [cc]: 7 Points: 22094 Resolution [mm]: 1.43
Dose Volume Histogram
Volume [%]
100
80
60
40
20
0
0
20
40
60
Dose [Gy]
Some structures are unapproved or rejected
Add Upper Objective
Add Lower Objective
Add Mean Objective
Delete Objective
Base dose plan:
Select...
Avoidance Sectors (0 MU)
Define Settings...
None
MU Objective
Use
Strength: 70
Min MU: 0
Max MU: 500
Automate Optimization
Continue automatically to final dose calculation
Save all after optimization and dose calculation
Automatic intermediate dose
Progress:
Jaw Tracking
MR 4 / 4 STEP 1 / 1 MU: 621
45s
Level Hold
View as: Line chart
Bar chart
OK
Cancel
Apply

Thanks for coming!

- Many elements of this talk were plundered from other, better tutorials on deep learning. It's hard to compete with YouTube these days!
- If you want to get started there are many excellent and interactive Python / PyTorch tutorials
- Happy to answer questions, and can provide references by email.

Logistic regression as a Neural Network

Logistic Regression

$z = b + a_1x_1 + a_2x_2 + a_3x_3$
$p = 1.0 / (1.0 + e^{-z})$

Ex:

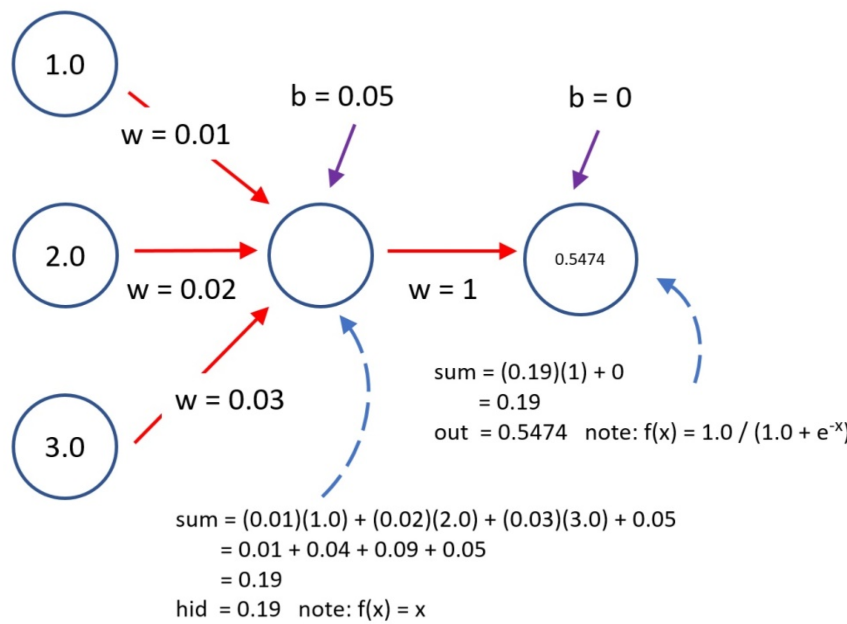
$w_1 = 1.0 \quad a_1 = 0.01$
$w_2 = 2.0 \quad a_2 = 0.02$
$w_3 = 3.0 \quad a_3 = 0.03$
$b = 0.05$

$$z = (0.05) + (0.01)(1.0) + (0.02)(2.0) + (0.03)(3.0) = 0.05 + 0.01 + 0.04 + 0.09 = 0.19$$

$$p = 1.0 / (1.0 + e^{-0.19}) = 0.5474 \text{ (predicted class = 1)}$$

Neural Network

single hidden layer, identity activation $f(x) = x$
single output node, logistic sigmoid activation $f(x) = 1 / (1 + e^{-x})$



https://jamesmccaffrey.wordpress.com/2017/07/01/a-neural-network-equivalent-to-logistic-regression/