Medical Image Analysis

C4M

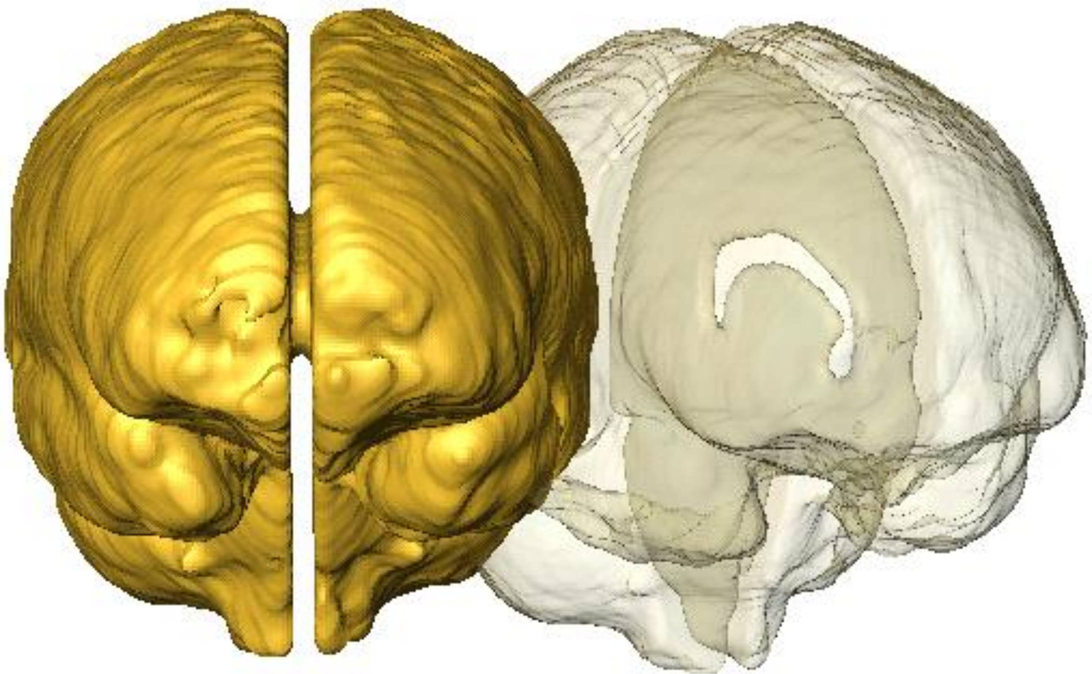
Chris McIntosh

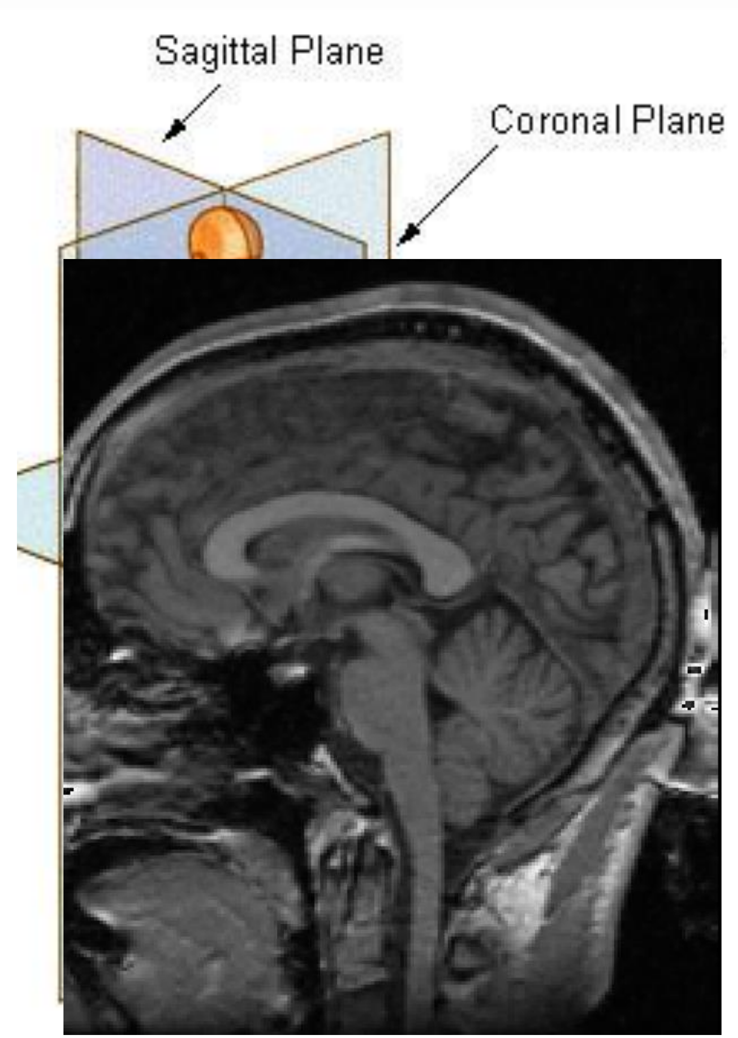
Toronto General
Toronto Western
Princess Margaret
Toronto Rehab
COURAGE LIVES HERE

UNIVERSITY OF TORONTO

Medical Imaging

- Enable clinicians to examine anatomy in-vivo (without extracting it)
- Some major areas for computer science
  - Data visualization
  - Image Analysis
  - Computer assisted diagnosis (CAD)
  - Disease understanding





chris.mcintosh@rmp.uhn.on.ca

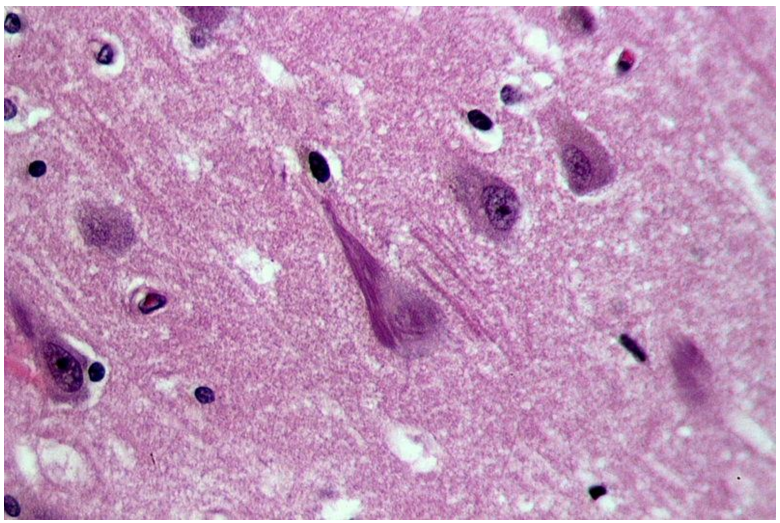
Acquisition

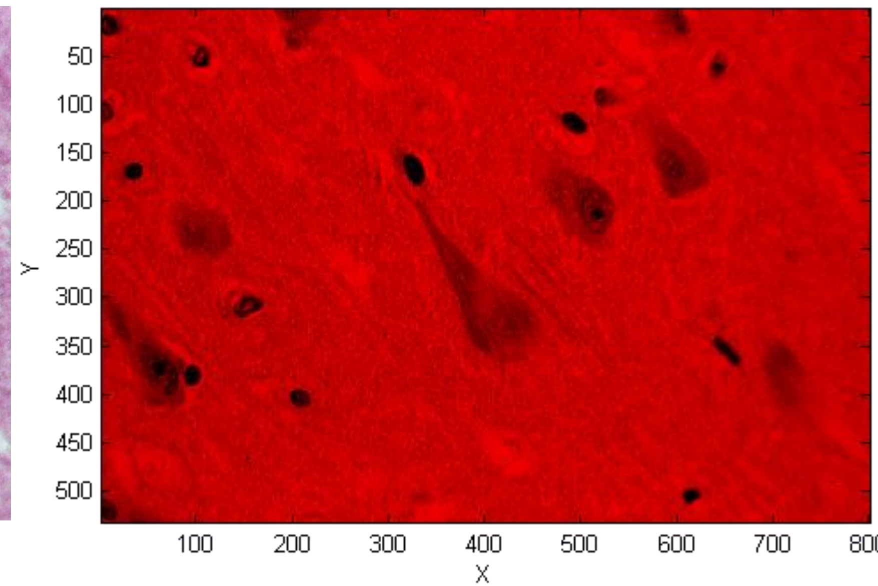
- Many different types of signals
  - Colour (dermatology, pathology)
  - Radiodensity (structural, geometrical)
  - Radioactive isotope uptake over time (functional)
  - Water and fat (soft-tissue structural)
  - Water diffusion over time (soft-tissue functional)
  - High frequency sound wave refraction (Ultrasound)

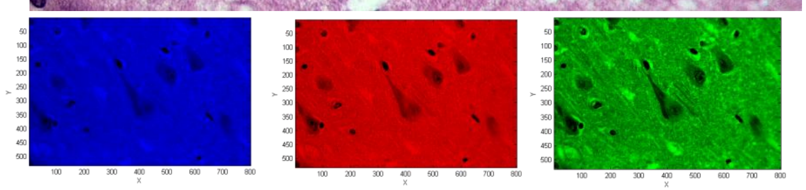
chris.mcintosh@rmp.uhn.on.ca

Colour Images

- Acquired by a camera, optionally with aid
  - Microscope, Dermatoscope, Endoscope, etc.
- Composed of 3 data channels yields an MxNx3 array
- Each discrete element is a pixel on the X, and Y axis







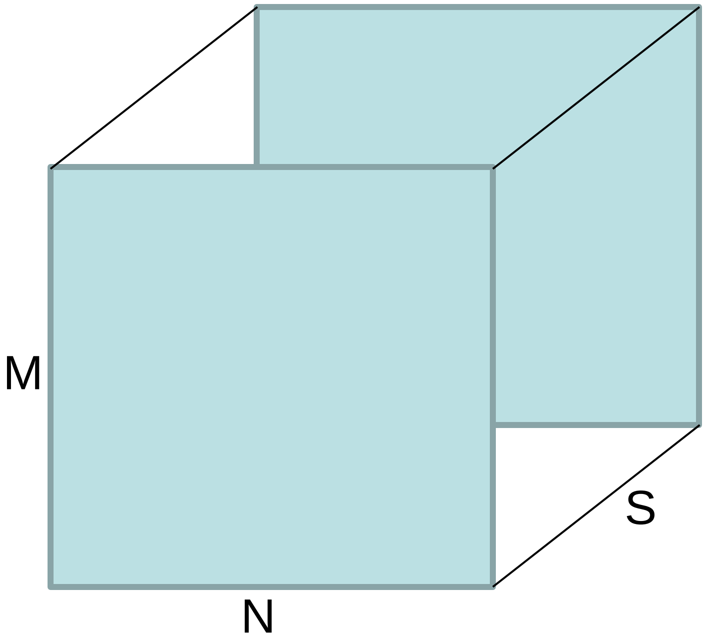
*Channels are colored for illustration only each channel is a scalar

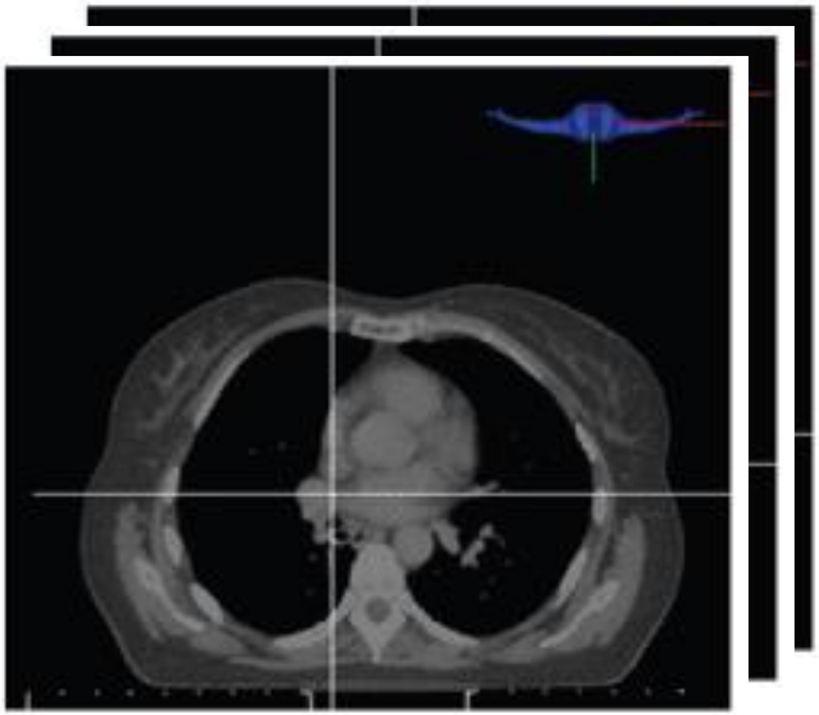
Original cell picture: Patho under CC BY-SA 3.0

chris.mcintosh@rmp.uhn.on.ca

Radiodensity

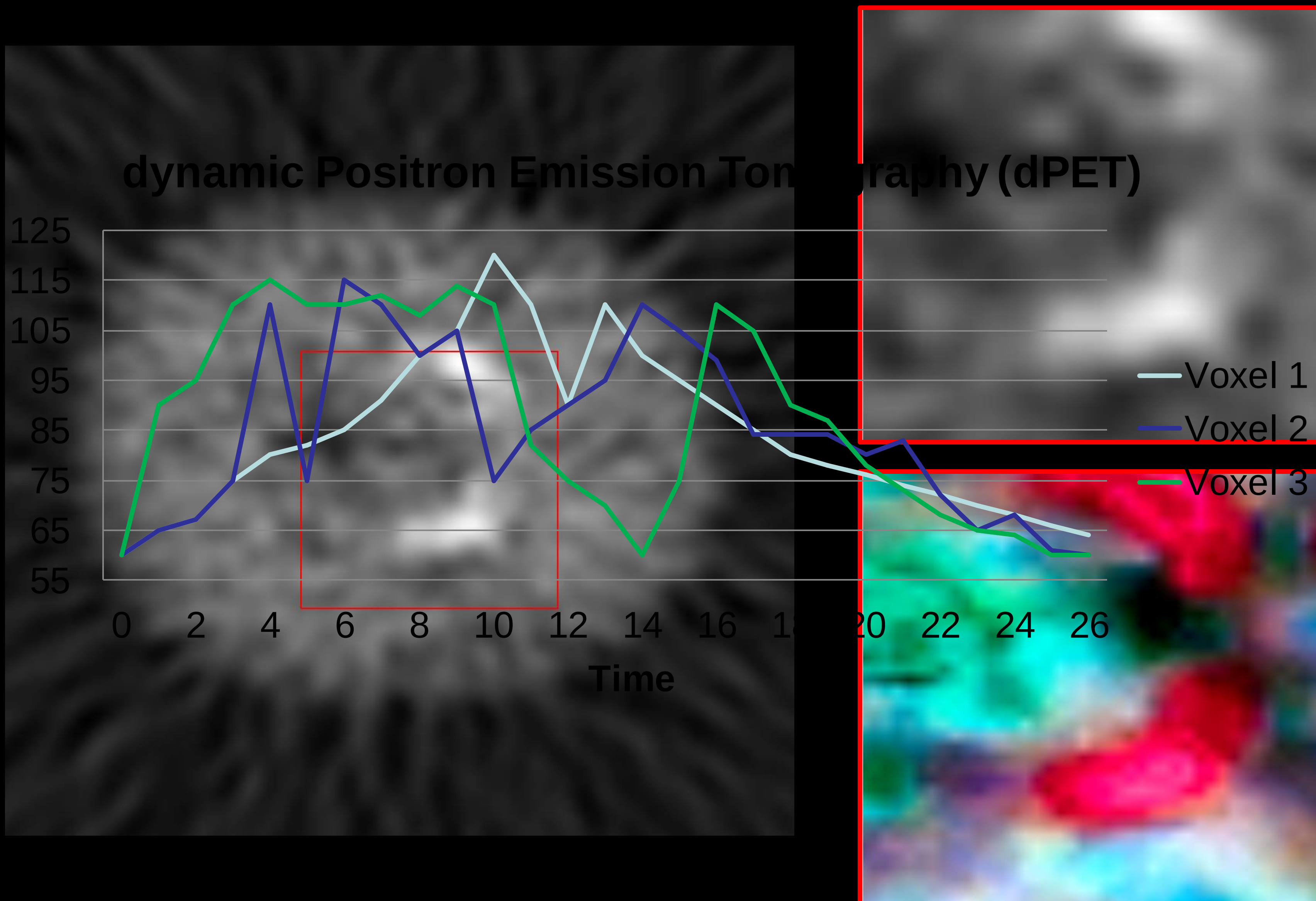
- CT Imaging
- MxNxS where S is the number of slices
- Each element is now a voxel in (X,Y,Z)





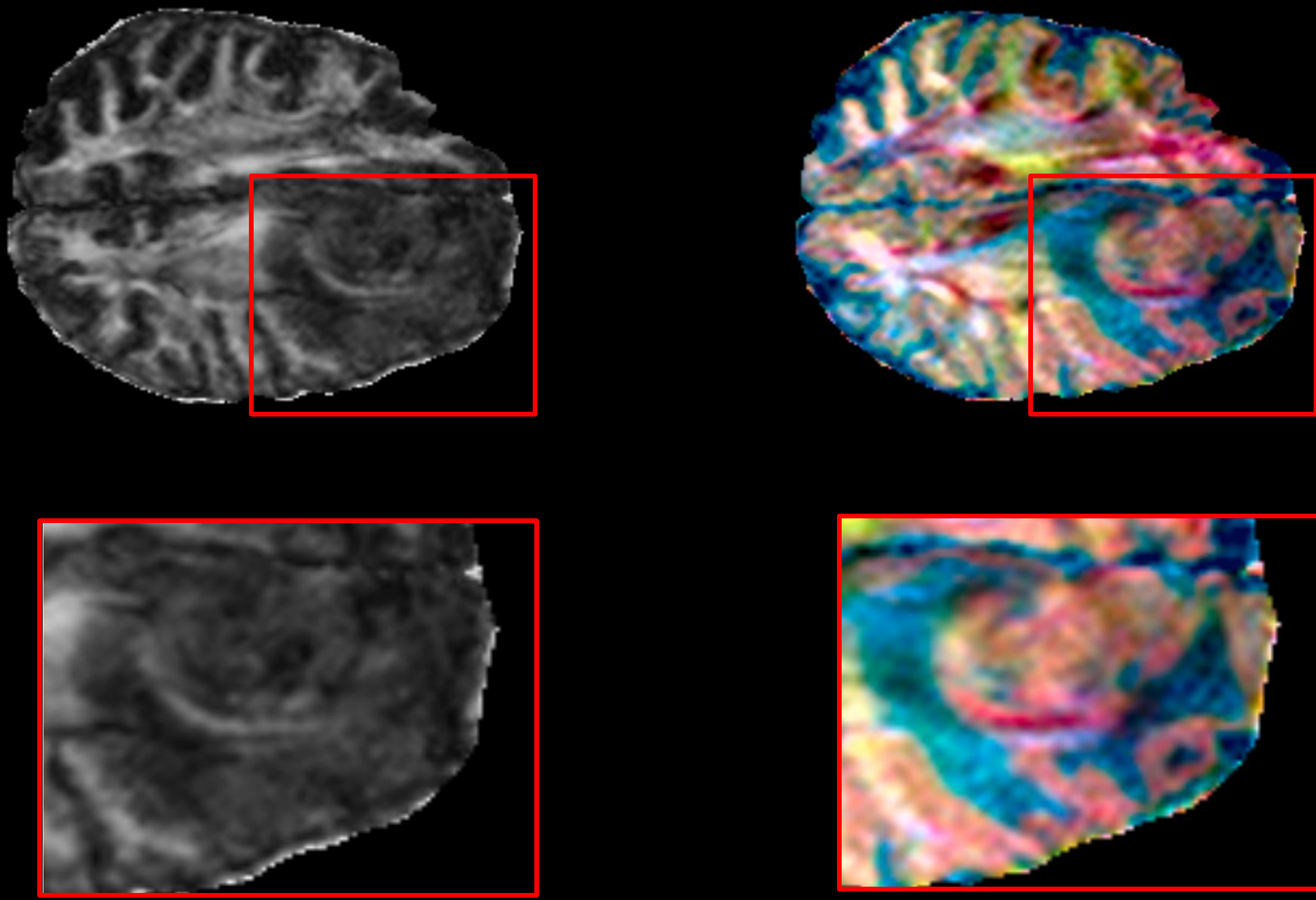
chris.mcintosh@rmp.uhn.on.ca

Positron Emission Tomography



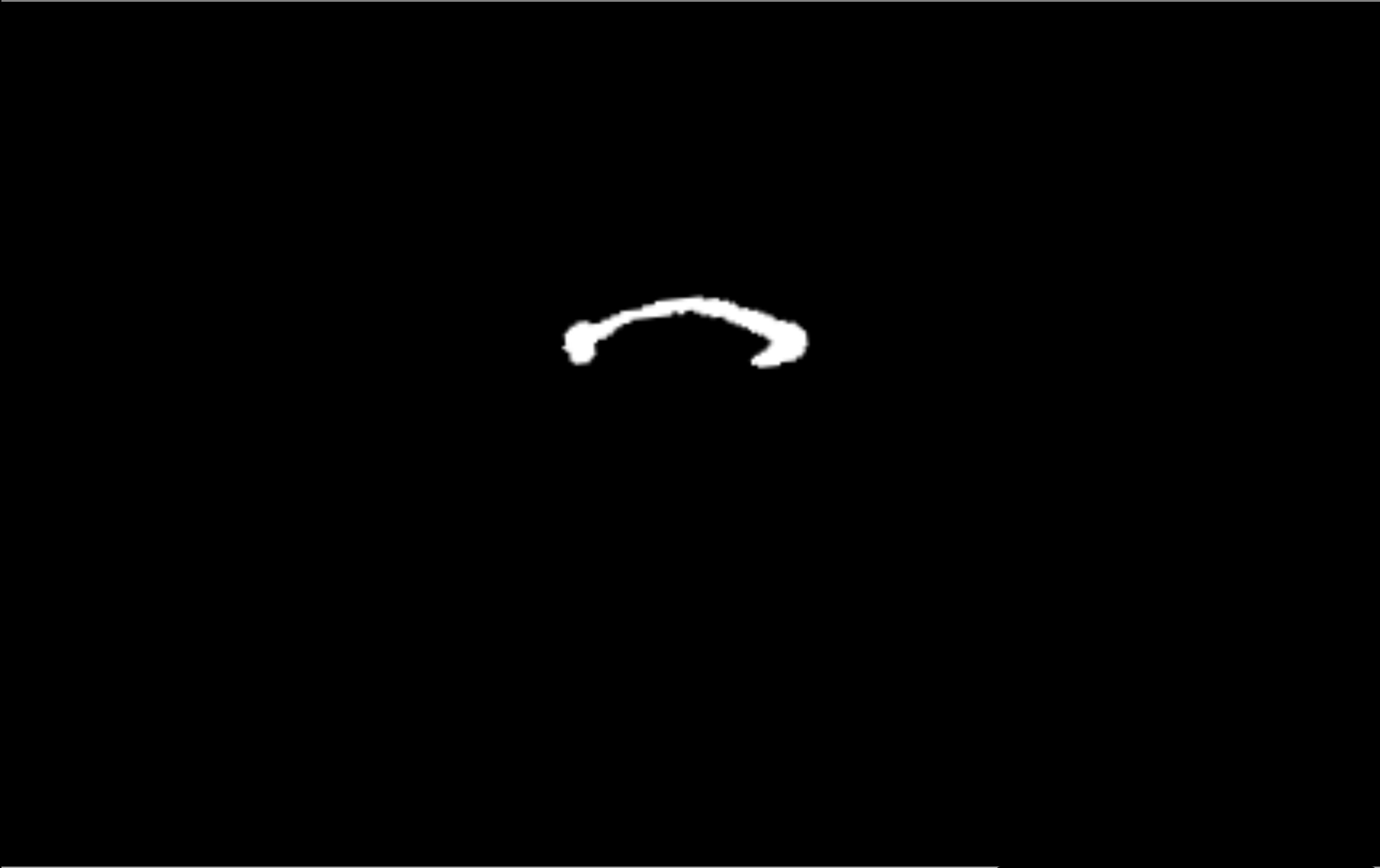
chris.mcintosh@rmp.uhn.on.ca

Diffusion Tensor MRI



chris.mcintosh@rmp.uhn.on.ca

Common Tasks

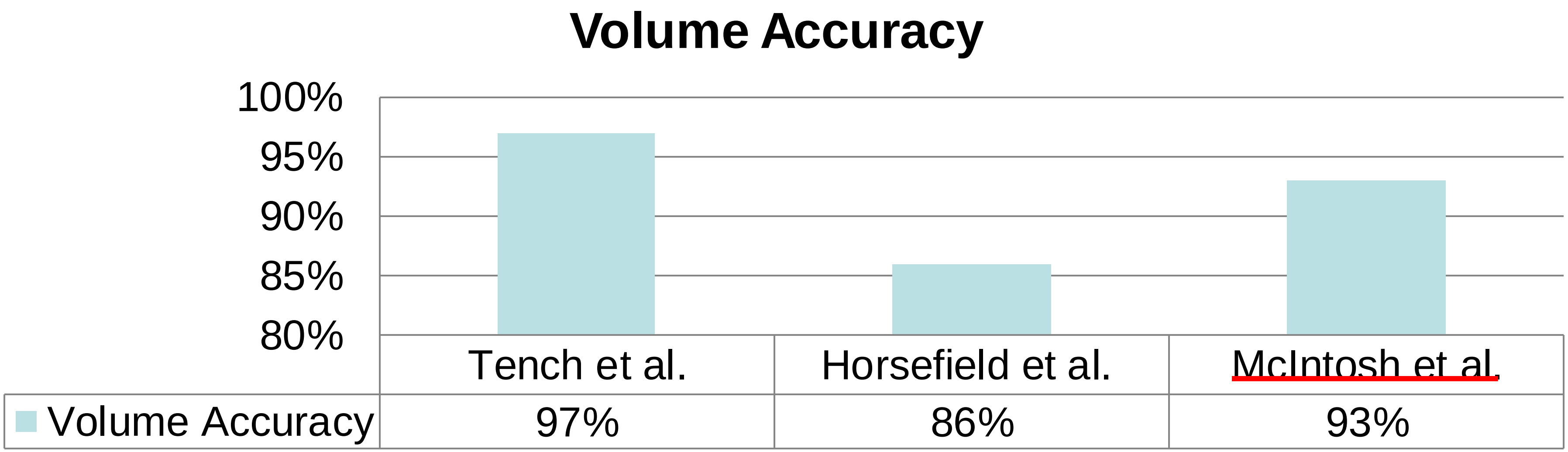


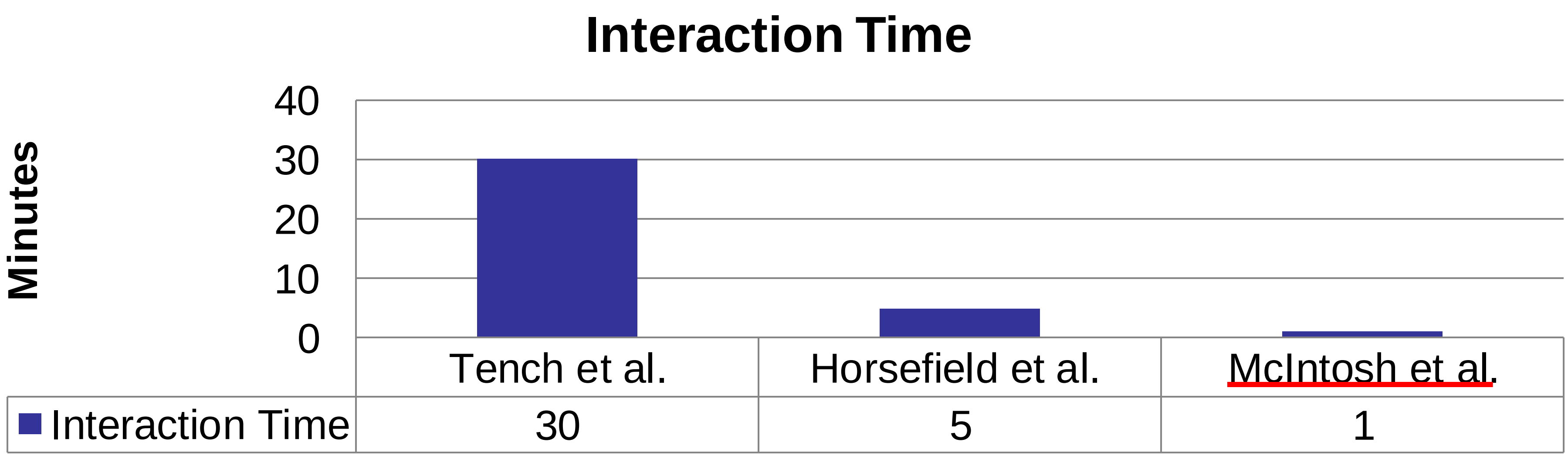
chris.mcintosh@rmp.uhn.on.ca

Disease Understanding

chris.mcintosh@rmp.uhn.on.ca

Disease Understanding





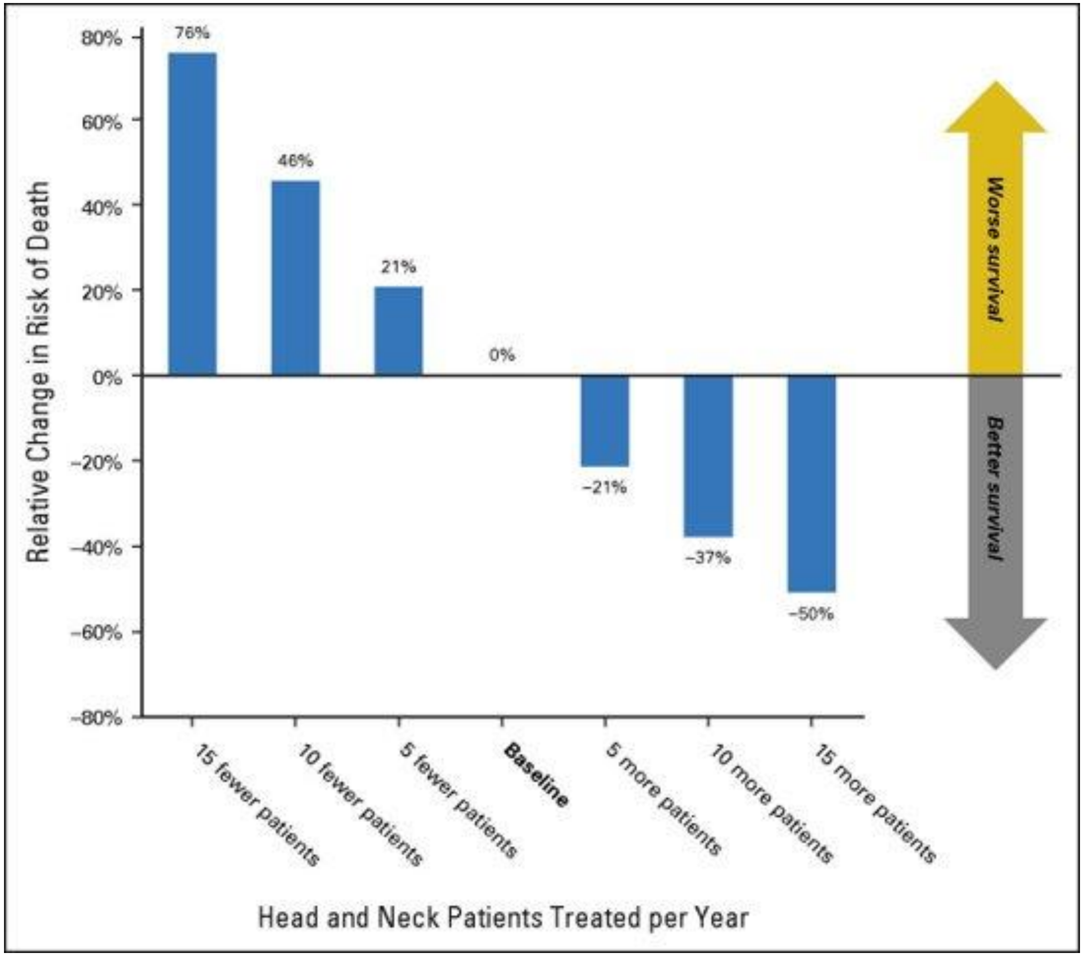
chris.mcintosh@rmp.uhn.on.ca

Summary

- Understanding the end-user
  - Not trying to replace medical experts
  - Augmenting ability
- It's about standardization and time
  - Make the region of interest more obvious (visualization)
  - Point me to the interesting data (Detection and CAD)
  - Extract the interesting data faster (segmentation, and shape analysis for disease understanding)
- A patient should receive the same diagnosis and treatment on a Monday morning in Toronto, or Friday night in Whitehorse

chris.mcintosh@rmp.uhn.on.ca

Consistent Medicine

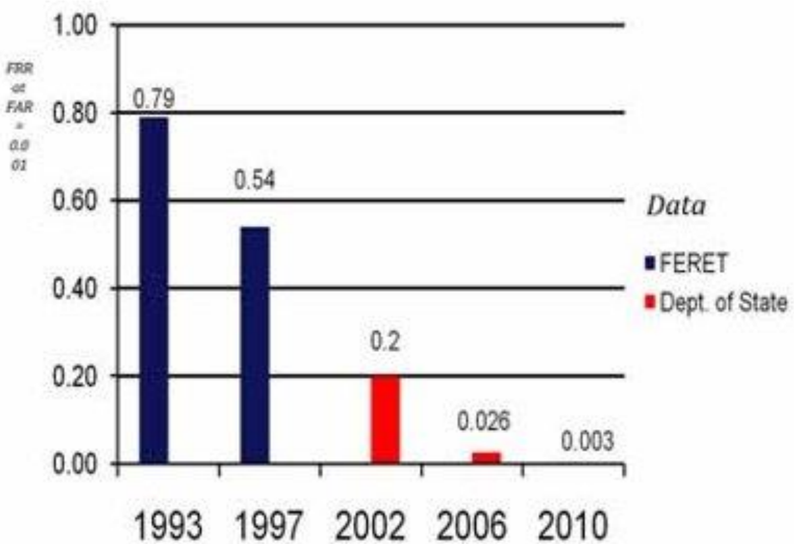


Boero, Isabel J., et al. "Importance of radiation oncologist experience among patients with head-and-neck cancer treated with intensity-modulated radiation therapy." *Journal of Clinical Oncology* 34.7 (2016): 684-690.

chris.mcintosh@rmp.uhn.on.ca

Advances in AI

- Computers can now recognize digits and faces with an accuracy that surpasses the average person.
  - Advances in computer science, mathematics, and cloud computing



DeepFace: Closing the Gap to Human-Level Performance in Face Verification

Yaniv Taigman Ming Yang Marc'Aurelio Ranzato Lior Wolf

Facebook AI Group
Menlo Park, CA, USA
{yaniv, mingyang, ranzato}@fb.com

Tel Aviv University
Tel Aviv, Israel
wolf@cs.tau.ac.il

Abstract

In modern face recognition, the conventional pipeline consists of four stages: detect ⇒ align ⇒ represent ⇒ classify. We revisit both the alignment step and the representation step by employing explicit 3D face modeling in order to apply a piecewise affine transformation, and derive a face representation from a nine-layer deep neural network. This deep network involves more than 120 million parameters using several locally connected layers without weight sharing, rather than the standard convolutional layers. Thus we trained it on the largest facial dataset to-date, an identity labeled dataset of four million facial images belonging to more than 4,000 identities. The learned representations coupling the accurate model-based alignment with the large facial database generalize remarkably well to faces in unconstrained environments, even with a simple classifier. Our method reaches an accuracy of 97.25% on the Labeled Faces in the Wild (LFW) dataset, reducing the error of the current state of the art by more than 25%, closely approaching human-level performance.

toward tens of thousands of appearance features in other recent systems [5, 7, 2].

The proposed system differs from the majority of contributions in the field in that it uses the deep learning (DL) framework [3, 21] in lieu of well engineered features. DL is especially suitable for dealing with large training sets, with many recent successes in diverse domains such as vision, speech and language modeling. Specifically with faces, the success of the learned net in capturing facial appearance in a robust manner is highly dependent on a very rapid 3D alignment step. The network architecture is based on the assumption that once the alignment is completed, the location of each facial region is fixed at the pixel level. It is therefore possible to learn from the raw pixel RGB values, without any need to apply several layers of convolutions as is done in many other networks [19, 21].

In summary, we make the following contributions : (i) The development of an effective deep neural net (DNN) architecture and learning method that leverage a very large labeled dataset of faces in order to obtain a face representation that generalizes well to other datasets; (ii) An effective facial alignment system based on explicit modeling of 3D

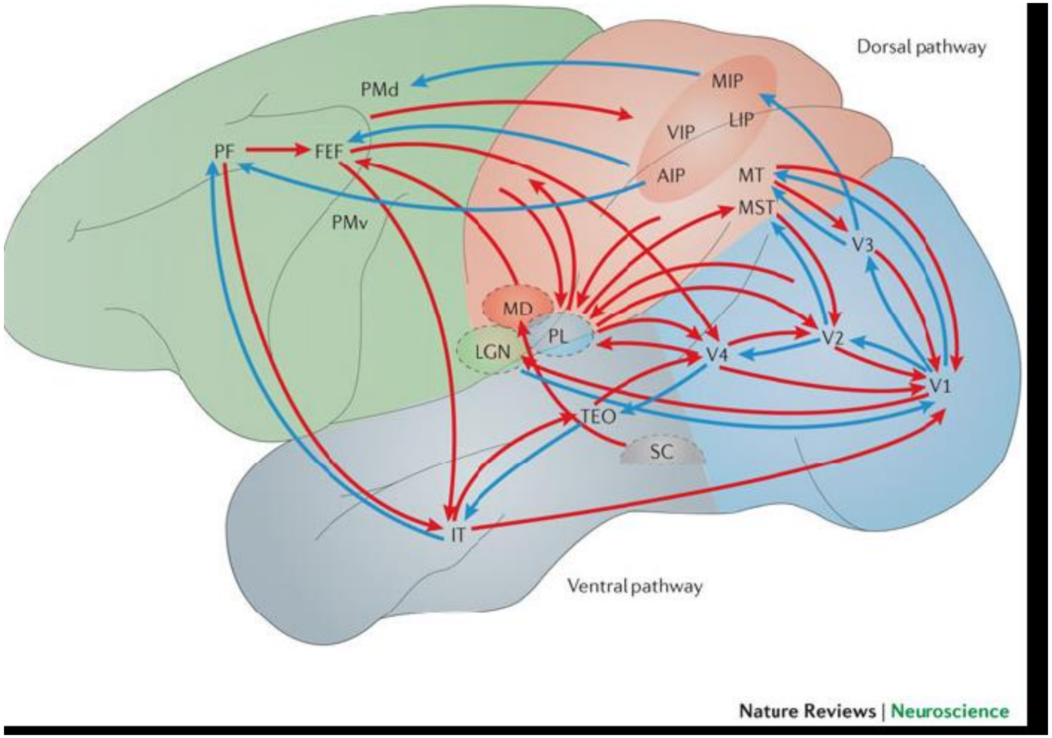
What is Vision?

- If the goal is observe or highlight something a person can see in an image, we must first understand how a person can see

chris.mcintosh@rmp.uhn.on.ca

Winning a Nobel Prize

- Hubel & Wiesel won the Nobel Prize in Physiology or Medicine in 1981
- Inserted microelectrodes into cats, and monkeys, and studied the response of different areas of the brain under different stimuli

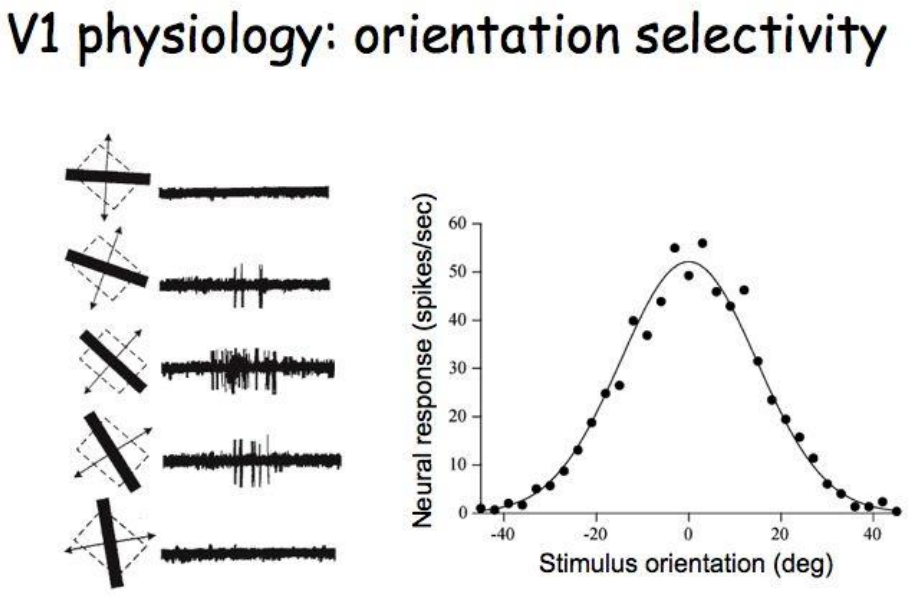


Gilbert and Li, Nature Reviews, 2013

chris.mcintosh@rmp.uhn.on.ca

The First Layer

- Primary visual cortex (V1)
- Subjects were shown different orientations of black and white bars
- Measure how neurons respond under different visual stimulus
- We call this edge response

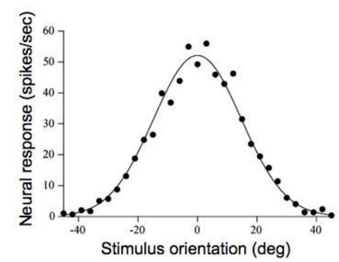


Hubel & Wiesel, 1968

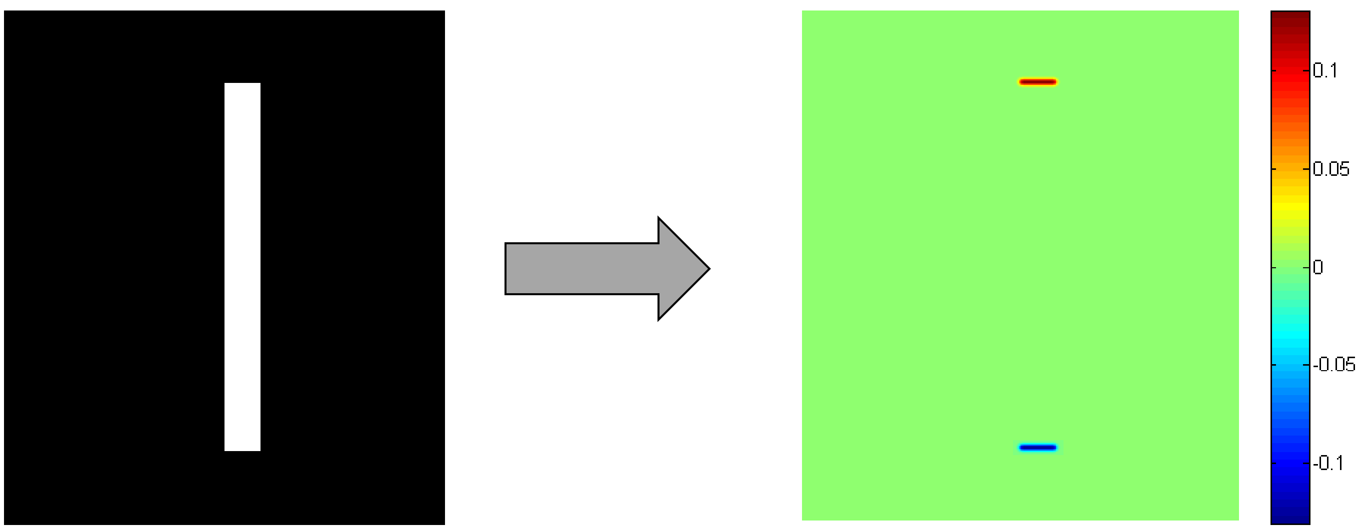
chris.mcintosh@rmp.uhn.on.ca

Example

- Examine response for a single neuron



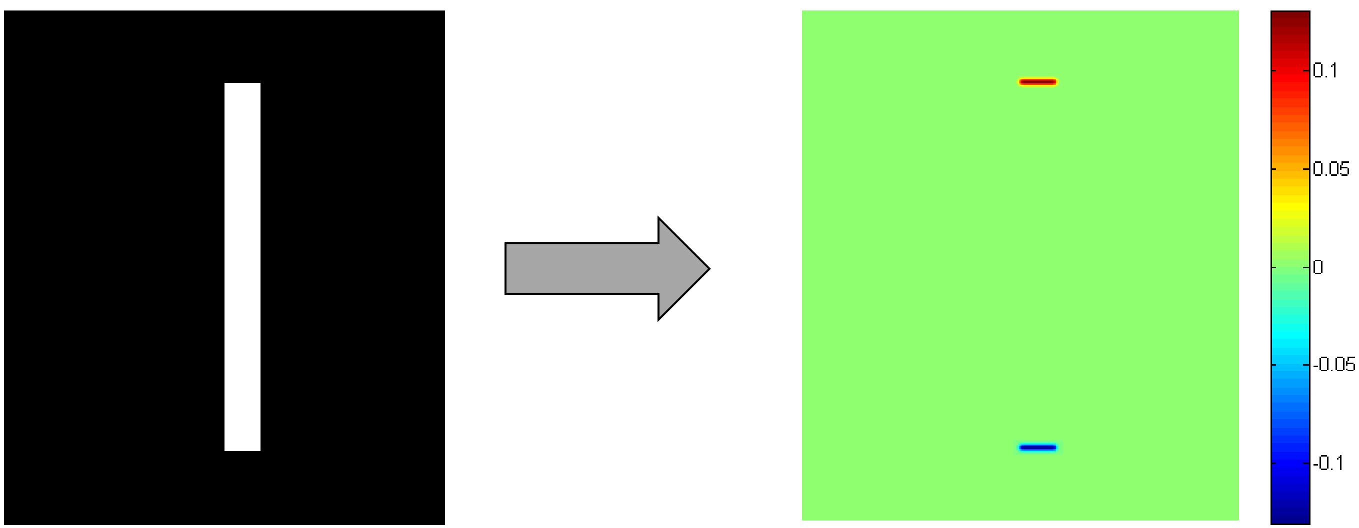
Hubel & Wiesel, 1968



chris.mcintosh@rmp.uhn.on.ca

Filtering

- We need examine every pixel of an MxN image and compare it to the pattern's discovered by Hubel and Weise
- We call the pattern a filter, and it will be a [2*k+1,2*k+1] array.
  - Odd sizes ensure a clearly defined centre



chris.mcintosh@rmp.uhn.on.ca

Example

Filter

$k = 1$

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

Input Image

$F(x, y)$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Output Result

$G(x, y)$

	0								

Source: S. Seitz

chris.mcintosh@rmp.uhn.on.ca

Example

Filter

$k = 1$

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

Input Image

$F(x, y)$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Output Result

$G(x, y)$

	0	10							

Source: S. Seitz

chris.mcintosh@rmp.uhn.on.ca

Example

Filter

$k = 1$

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

Input Image

$F(x, y)$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Output Result

$G(x, y)$

	0	10	20						

Source: S. Seitz

chris.mcintosh@rmp.uhn.on.ca

Boundary Conditions

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

chris.mcintosh@rmp.uhn.on.ca

Example

Filter

$k = 1$

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

Input Image

$F(x, y)$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Output Result

$G(x, y)$

	0	10	20	30	30	30	20	10	
	0	20	40	60	60	60	40	20	
	0	30	60	90	90	90	60	30	
	0	30	50	80	80	90	60	30	
	0	30	50	80	80	90	60	30	
	0	20	30	50	50	60	40	20	
	10	20	30	30	30	30	20	10	
	10	10	10	0	0	0	0	0	

Source: S. Seitz, S. Fidler

chris.mcintosh@rmp.uhn.on.ca

Algorithms

- Three main variants
  - 2D Filtering
  - Convolution
    - Same as filtering with a filter flipped in Y and then X to gain a few important mathematical properties
  - Normalized cross-correlation or template-matching
    - Same as filtering, but the response at each pixel normalized by the magnitude of the filter times the pixel-window
    - Example later

chris.mcintosh@rmp.uhn.on.ca

Edge Response

Filter

$k = 1$

$$\begin{bmatrix} ? & ? & ? \\ ? & ? & ? \\ ? & ? & ? \end{bmatrix}$$

Input Image

$F(x, y)$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Output Result

$G(x, y)$

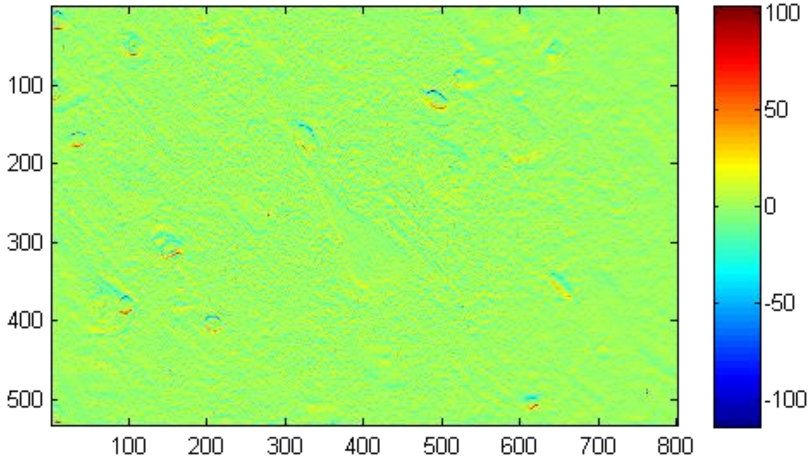
0	0	-90	-90	-90	-90	-90	-90	0	0

Source: S. Seitz

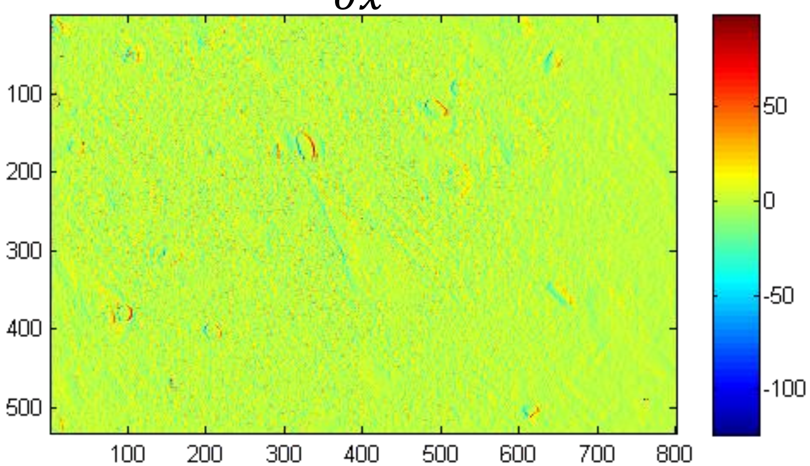
chris.mcintosh@rmp.uhn.on.ca

Edge Response (Gradient)

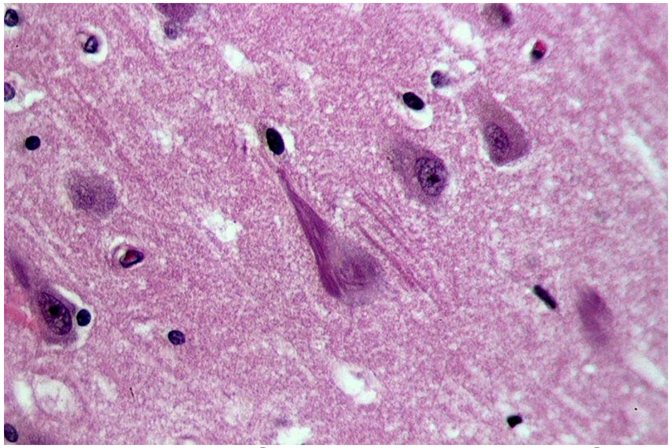
$$\frac{\partial}{\partial y} I(x,y)$$



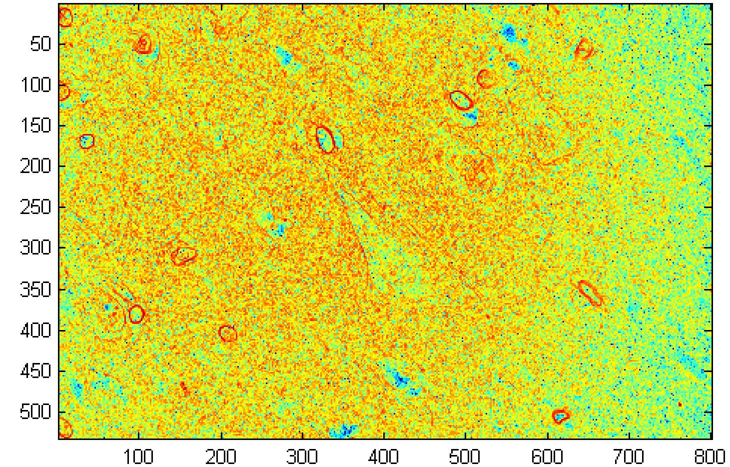
$$\frac{\partial}{\partial x} I(x,y)$$



$$I(x,y)$$



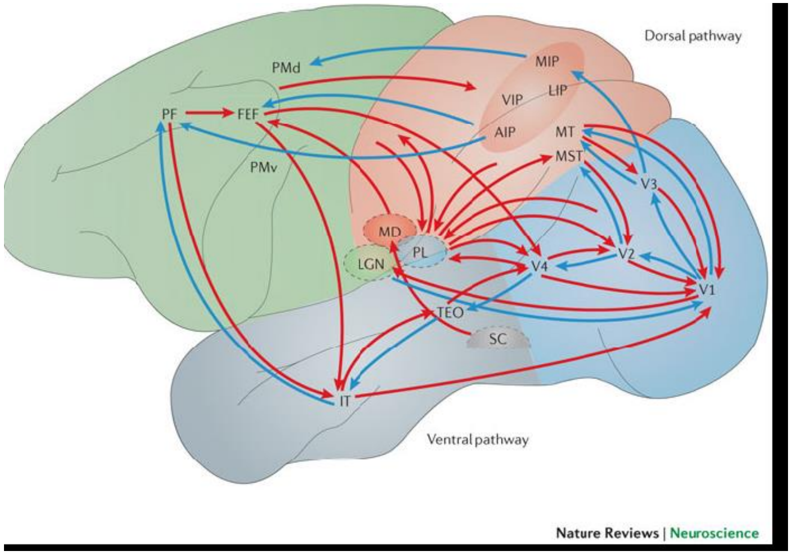
$$\left(\frac{\partial}{\partial x} I(x,y)\right)^2 + \left(\frac{\partial}{\partial y} I(x,y)\right)^2$$



chris.mcintosh@rmp.uhn.on.ca

Higher levels of vision

- As we progress through the visual cortex we begin grouping responses from lower levels to create more complex representations
- This grouping gives rise to contours and shapes (*Gilbert and Li, Nature Reviews, 2013*)
- V4 has shown strong response to texture (*Kastner et al., J Neurophysiology, 2000*)



Gilbert and Li, Nature Reviews, 2013

chris.mcintosh@rmp.uhn.on.ca

Vision is challenging

- Do both blocks have a gradient?

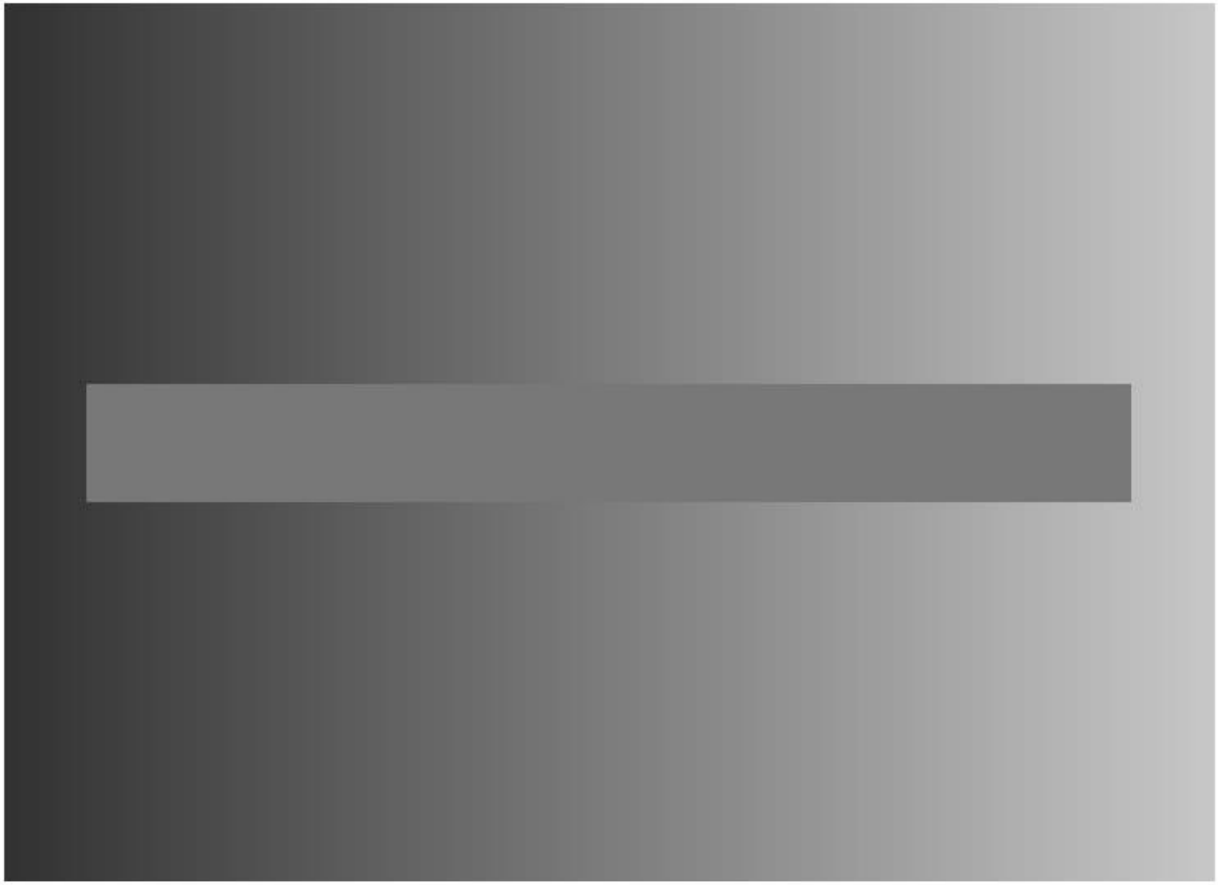


Photo by Dodek, CC BY-SA 3.0

chris.mcintosh@rmp.uhn.on.ca

Vision is challenging

- Do both blocks have a gradient?

Photo by Dodek, CC BY-SA 3.0

chris.mcintosh@rmp.uhn.on.ca

Vision is challenging

- The same dog or different? (Gold dress or blue)

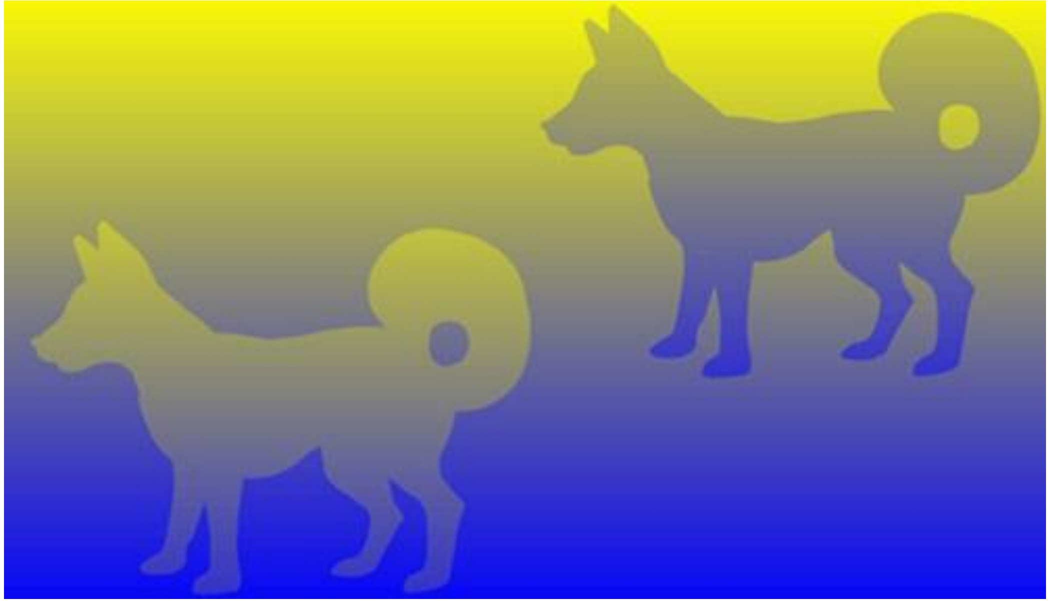


Photo by (WP:NFCC#4), Fair use

chris.mcintosh@rmp.uhn.on.ca

Vision is challenging

- Let's watch a short video and perform a basic vision task

chris.mcintosh@rmp.uhn.on.ca

Psychological

- Vision is a psychological and perceptual phenomena, not a physical measurement
  - Why did we evolve this way?
- Computer vision can emulate human vision
- It can also build measured responses to assist human vision
- Measurements lead to quantitative markers that can help decision making
- We call these features (some are perceptual, some are physical, some are both)

chris.mcintosh@rmp.uhn.on.ca

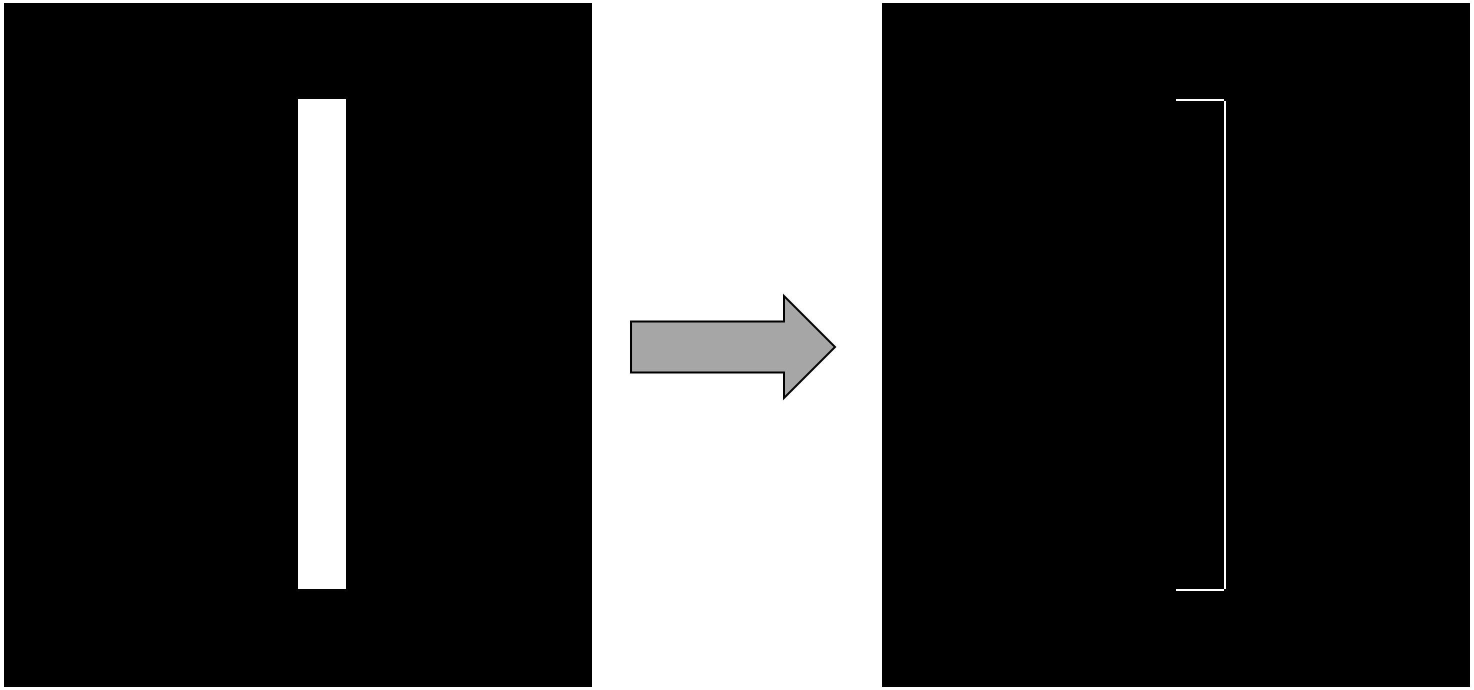
Features

- In computer vision we call the lower level responses/structures used to define an object features
- Features can be anything
- Represent a voxel or group of voxels by a number
  - Try and describe the local structure
  - Bright vs dark
  - Wavy vs smooth
  - Round vs square
- The features will enable us to build our applications (e.g. image segmentation)

chris.mcintosh@rmp.uhn.on.ca

Edge Detection

- Our edge filter is local (per pixel)
- Edges are connections of strong responses that group into a logical contour



chris.mcintosh@rmp.uhn.on.ca

Edges Are Both

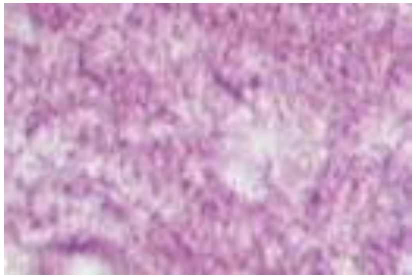
- Humans implicitly ignore edges that are not relevant to their perception of an image

Martin et al., A Database of Human Segmented Natural Images…, Computer Vision and Pattern Recognition, 2001

Texture

- Repeating intensity patterns in the data

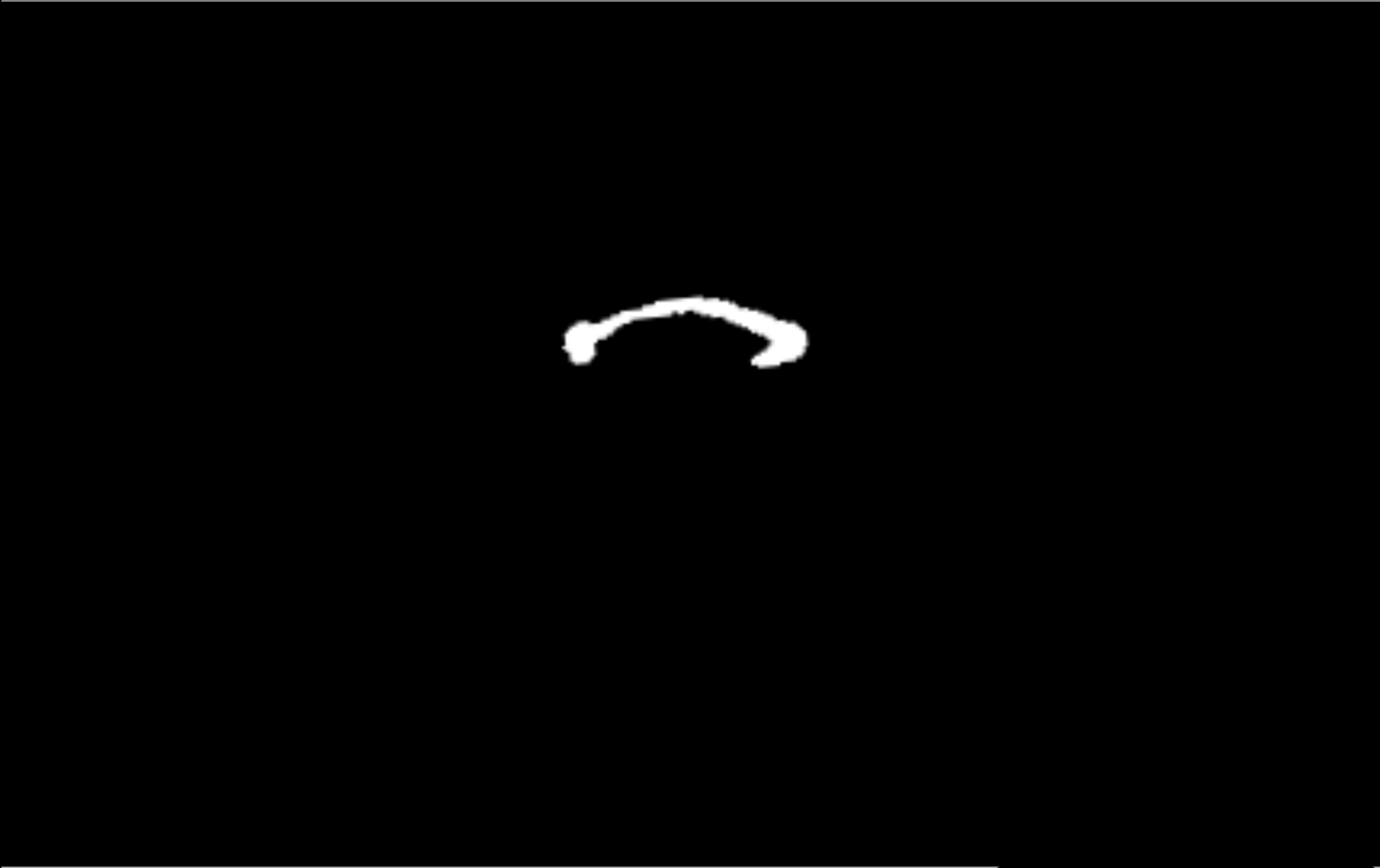




- Texture analysis, called radiomics, in CT has been shown to correlate with genomic features and cancer outcomes (*Aerts et al., Nature Methods, 2014*)

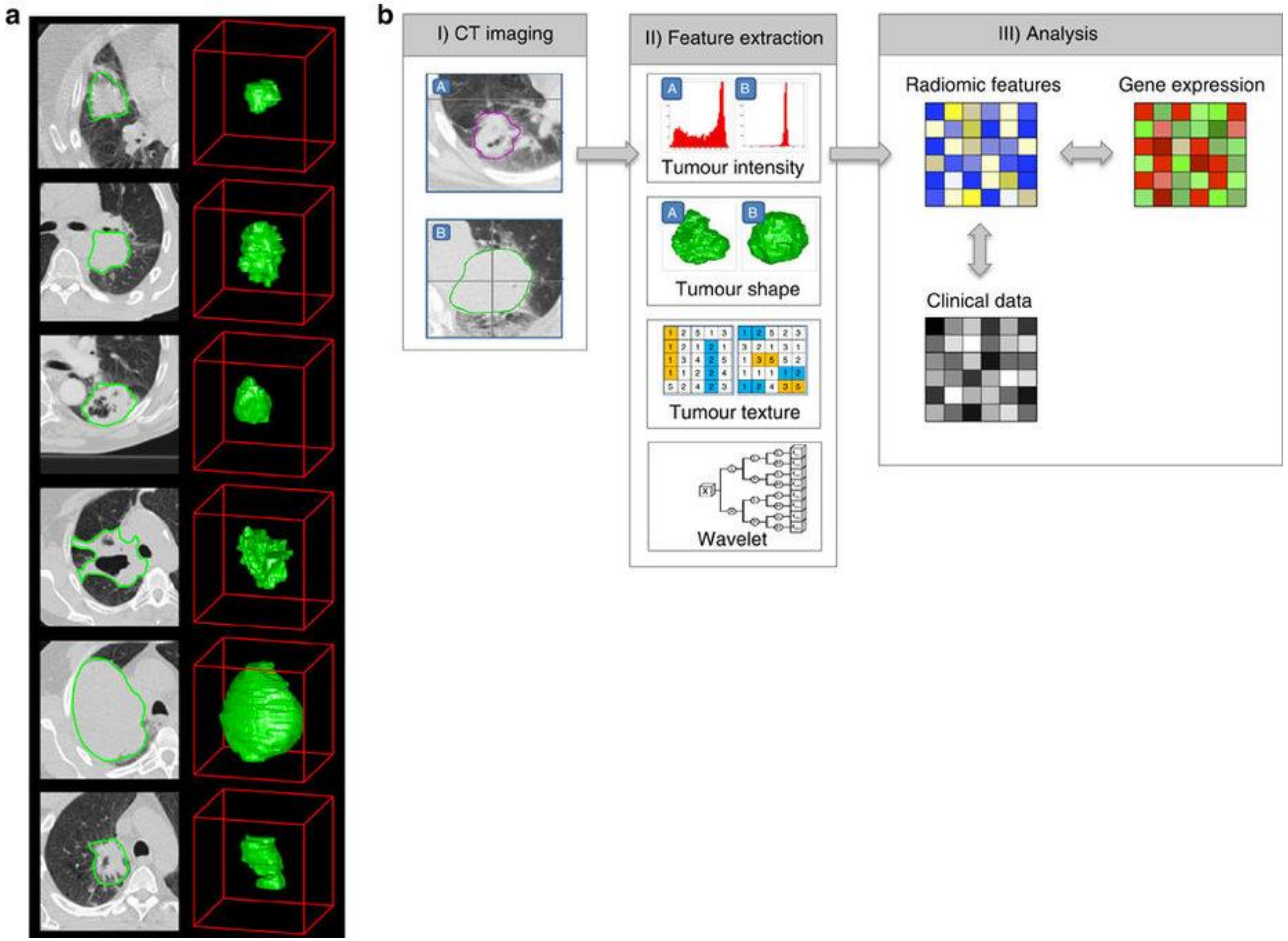
chris.mcintosh@rmp.uhn.on.ca

Shape



chris.mcintosh@rmp.uhn.on.ca

Radiomics



Aerts et al., Nature Methods, 2014

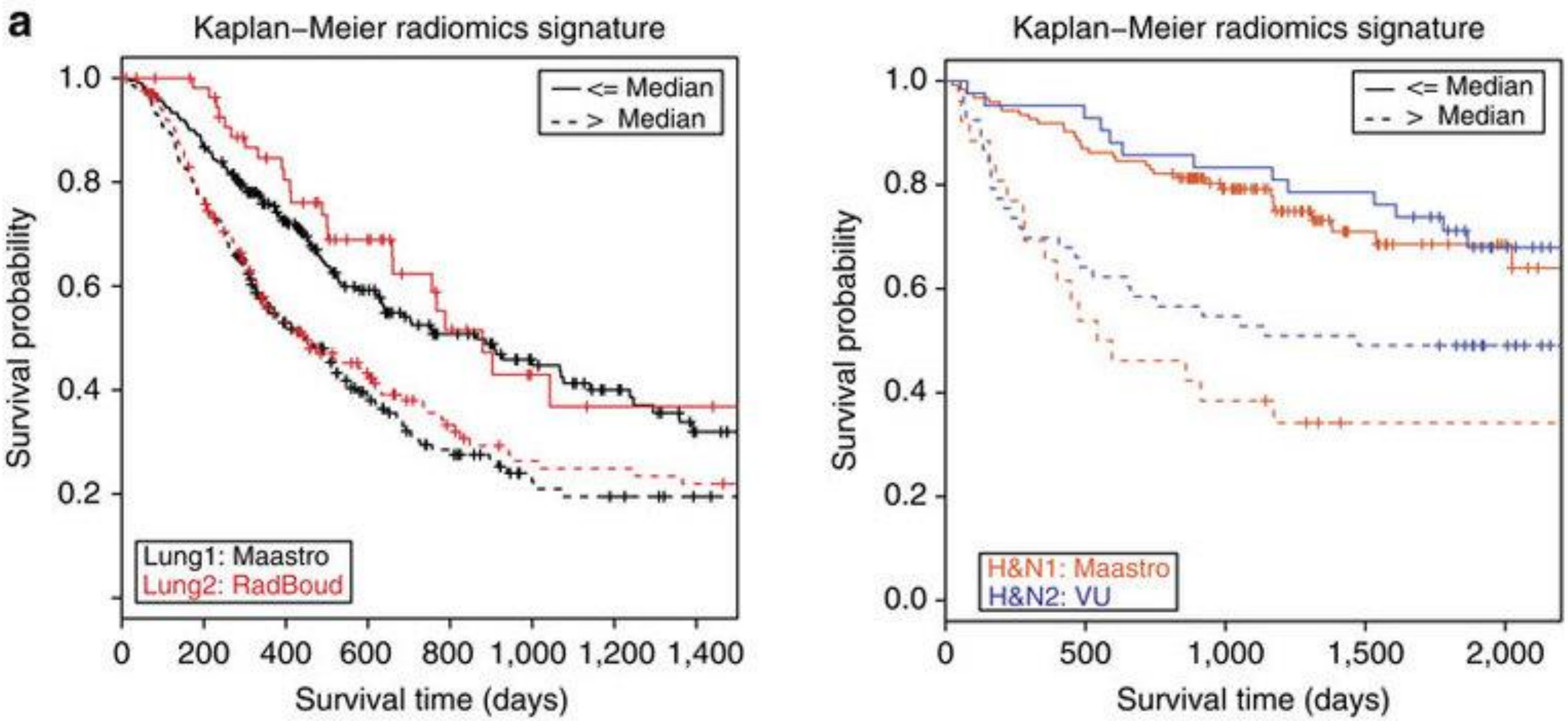
chris.mcintosh@rmp.uhn.on.ca

Building a Model

- Once features are extract the next step is to build a model to make predictions based on the features
  - $F(features) = Outcome$
- Model can be coded based on prior knowledge or learned via machine learning

chris.mcintosh@rmp.uhn.on.ca

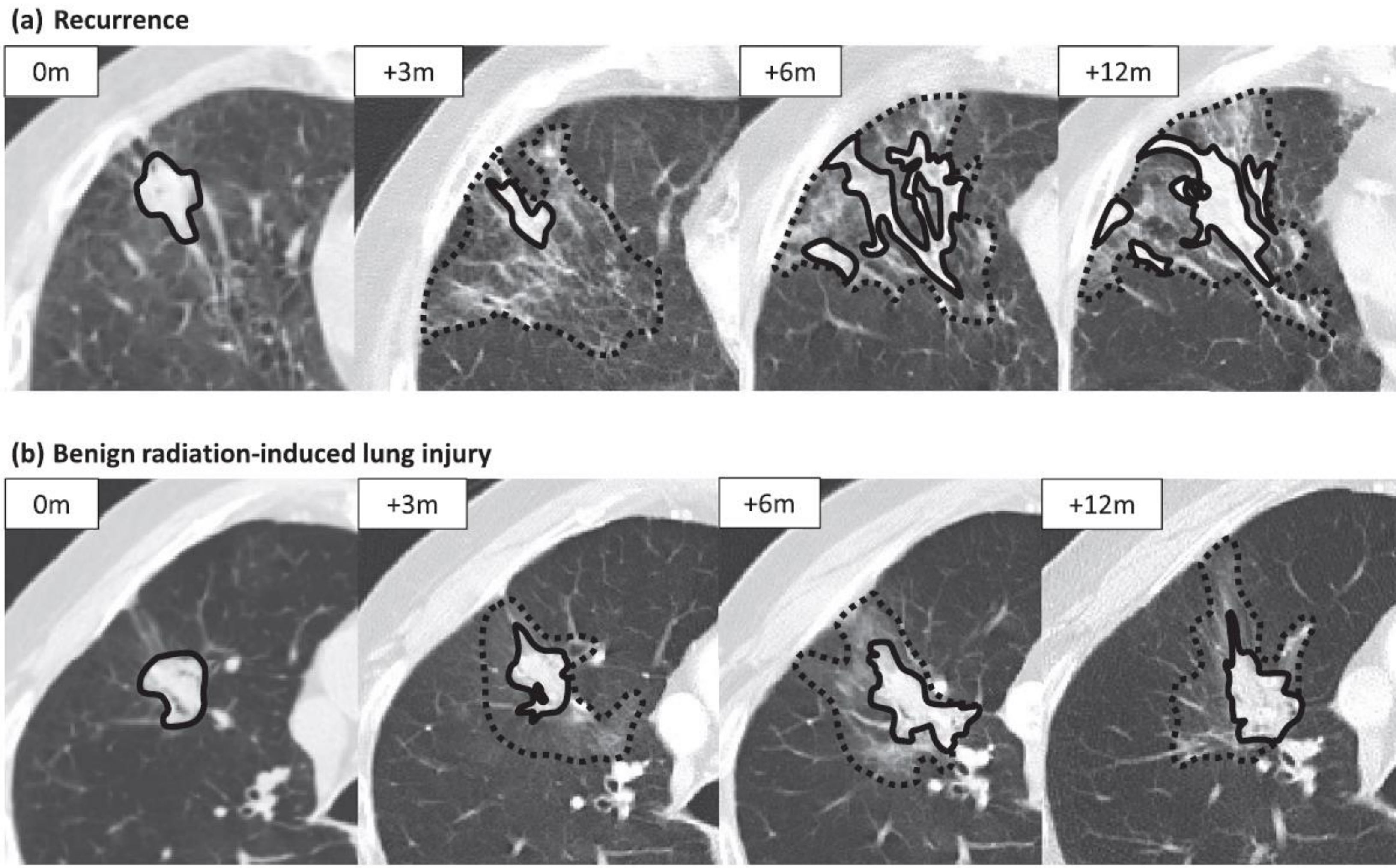
Radiomics



Aerts et al., Nature Methods, 2014

chris.mcintosh@rmp.uhn.on.ca

Lung Cancer Recurrence

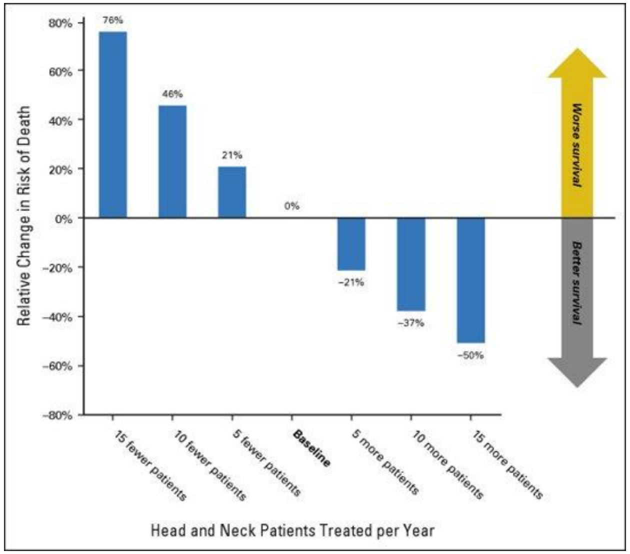


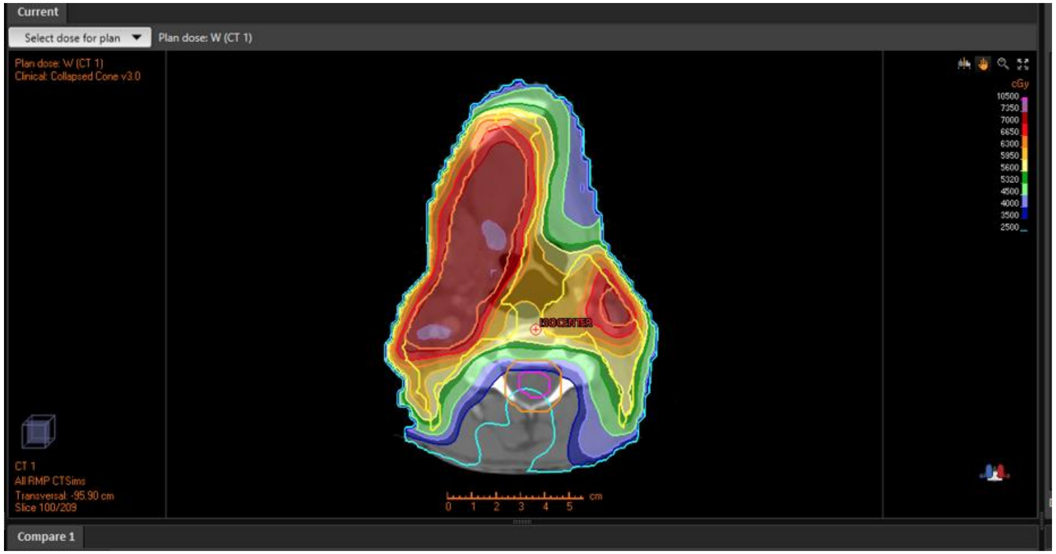
Mattonen et al., Medical Physics, 2014

chris.mcintosh@rmp.uhn.on.ca

Patients and Treatment

- Relating patient's anatomical geometry, texture, shape and appearance to radiotherapy treatment
- Given a CT image, predict treatment radiation dose to each voxel for a novel patient

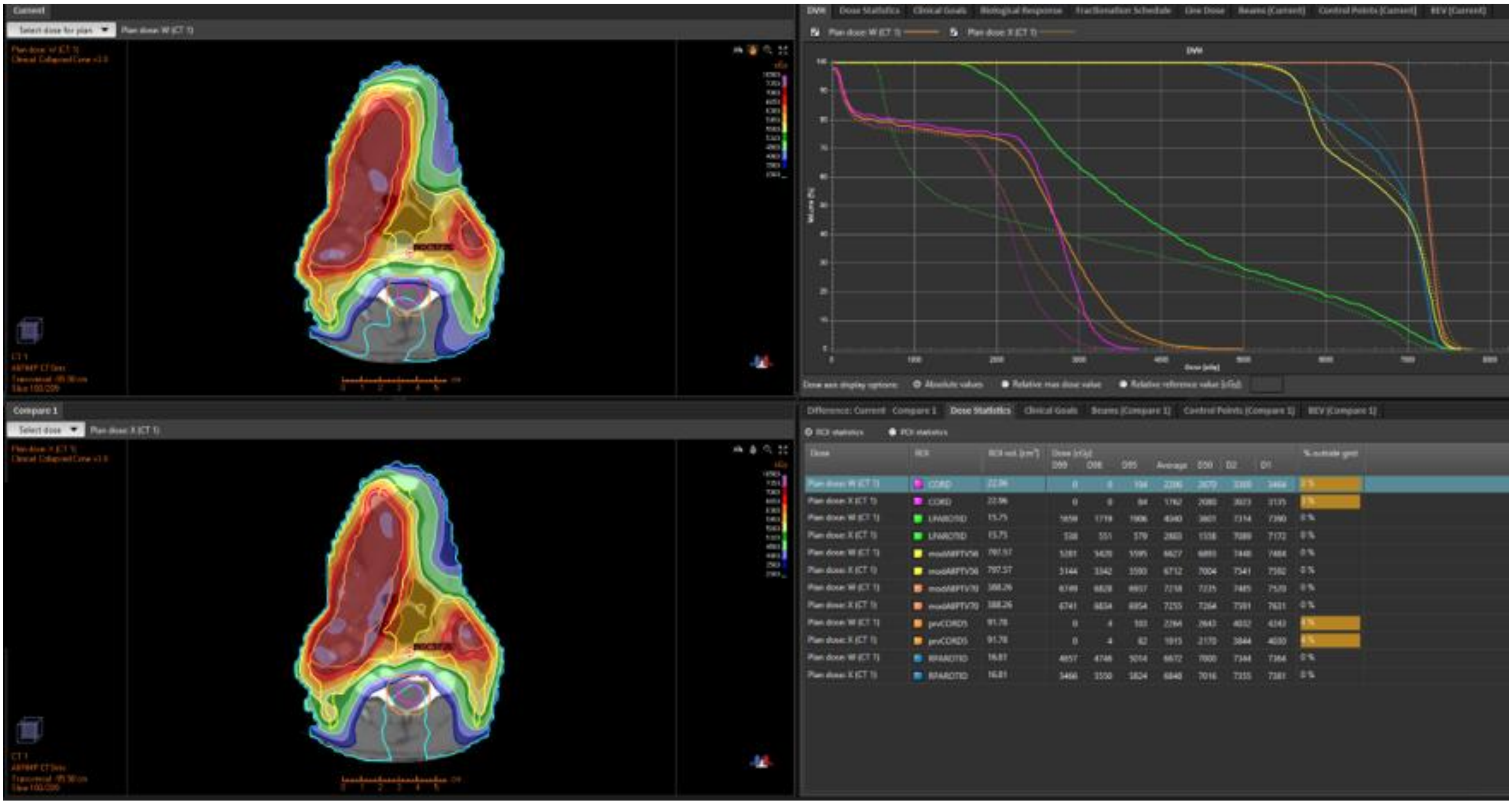




chris.mcintosh@rmp.uhn.on.ca

Automated vs Clinical

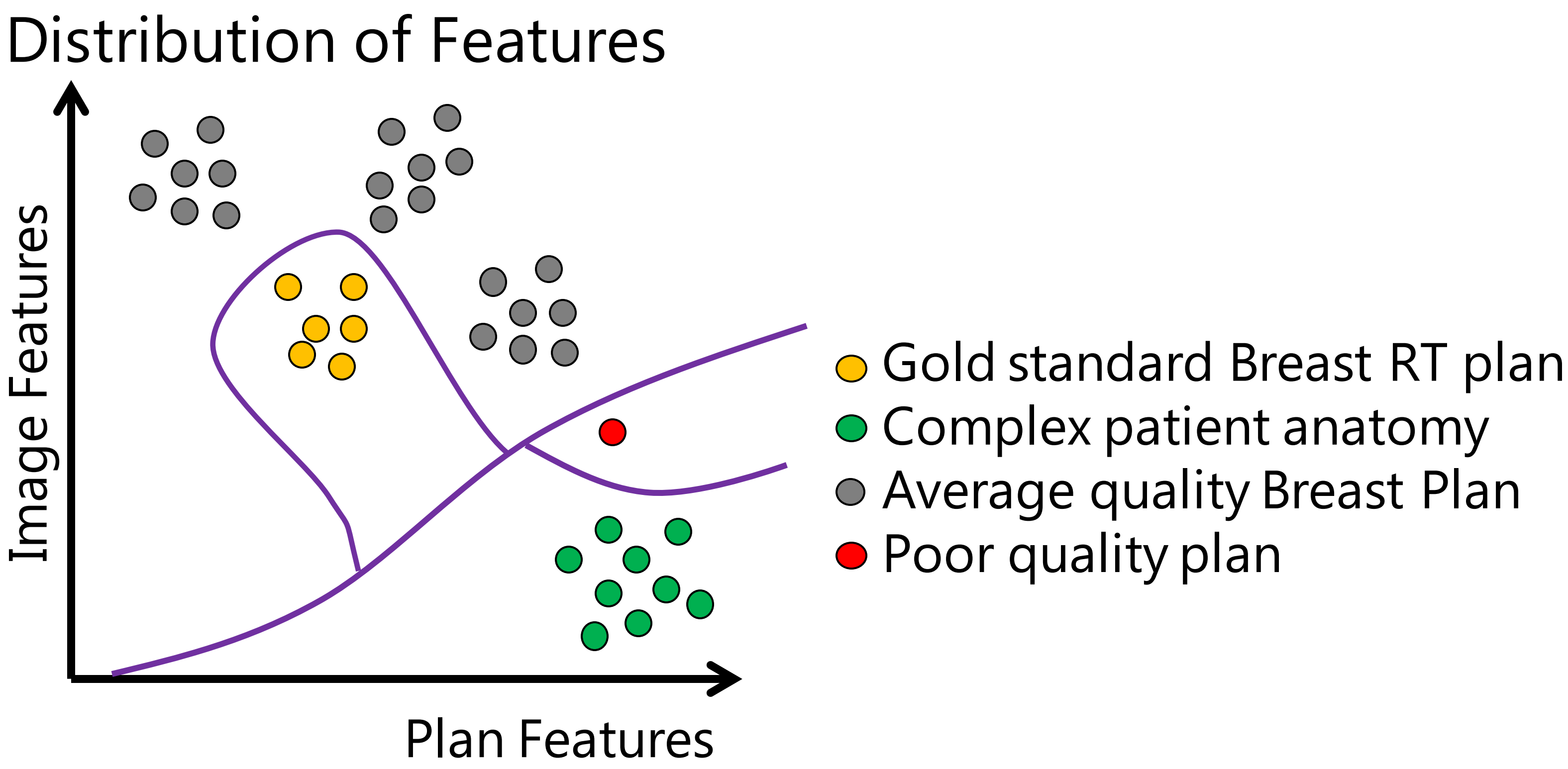
Toronto General
Toronto Western
Princess Margaret
Toronto Rehab
COURAGE LIVES HERE



chris.mcintosh@rmp.uhn.on.ca

Distribution of Features

- Different radiotherapy plans and qualities have different image and plan features
- Machine Learning learns to distinguish between the different groups
- Learns to:
  - Automatically catch low quality plans
  - Rank plans in order of least-to-most complexity for review

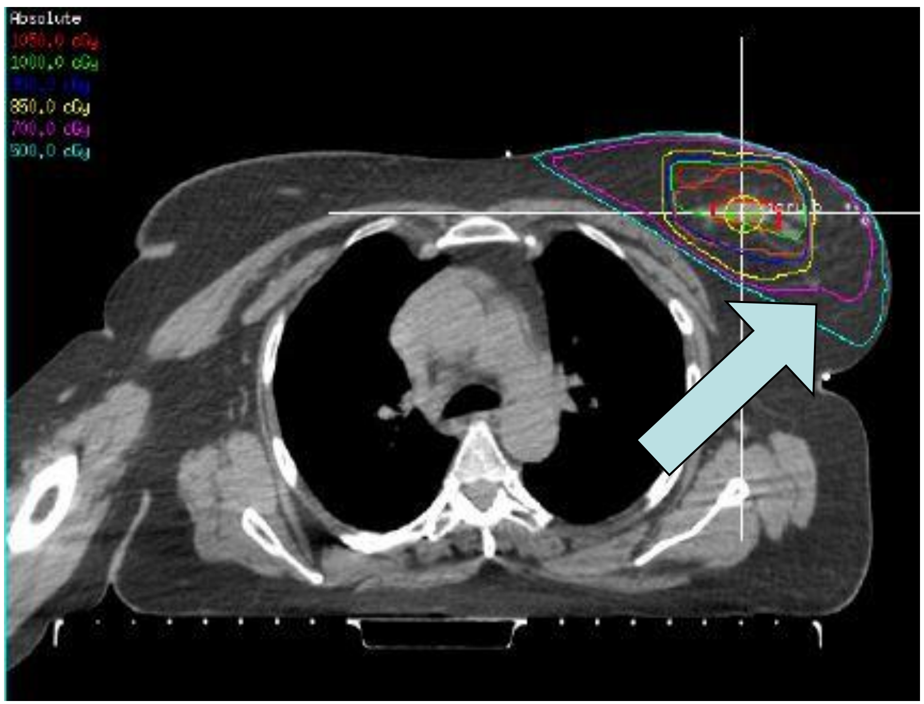


chris.mcintosh@rmp.uhn.on.ca

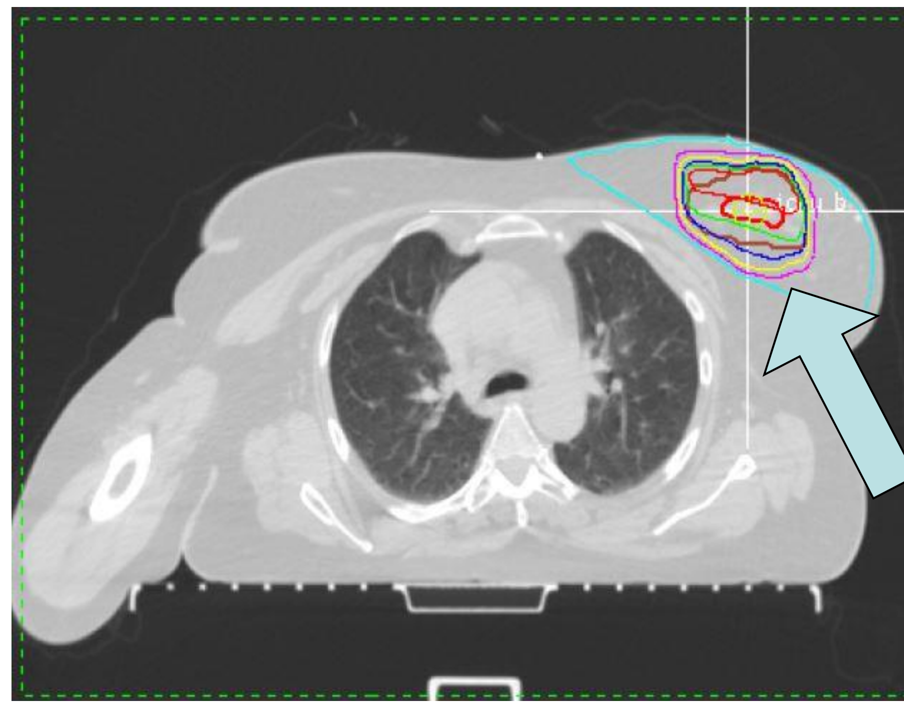
Planning Error Detection

- Preliminary breast study, detected 80% of clinically rejected plans
- Detected plan error with poor high dose conformity (700 cGy isodose)

Rejected



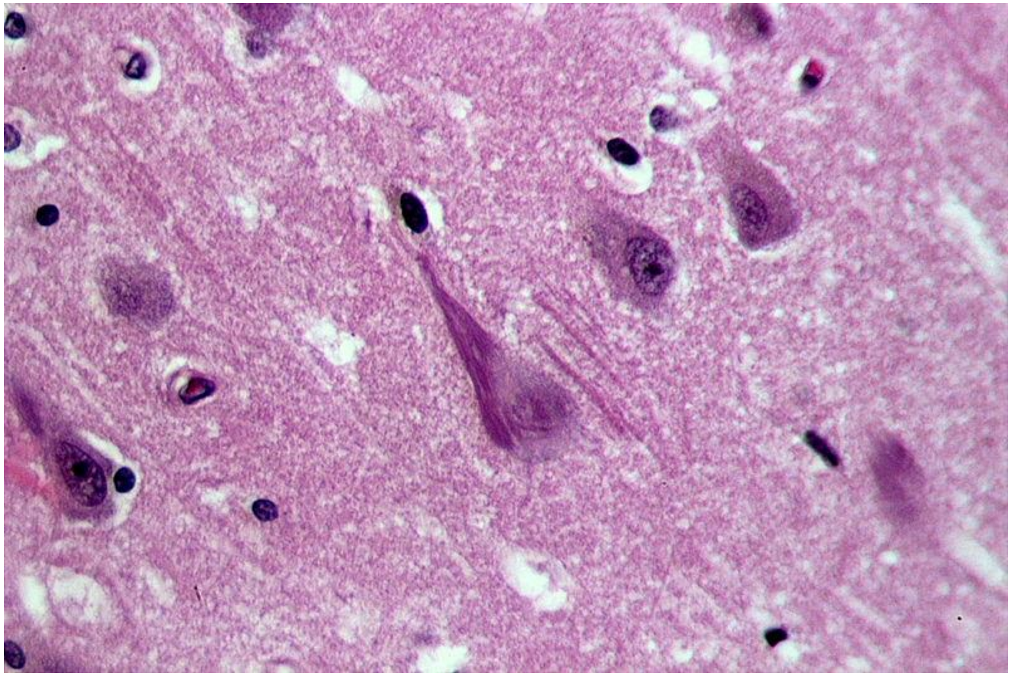
Accepted



chris.mcintosh@rmp.uhn.on.ca

Project Example

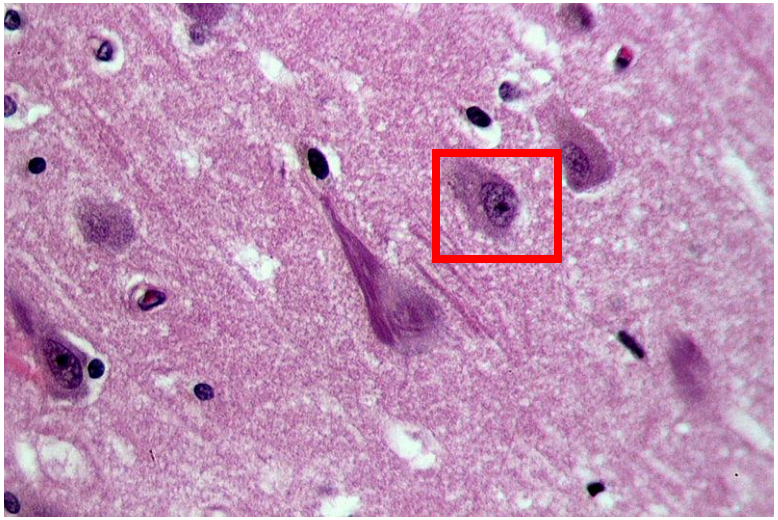
- Four main tasks:
  - Compute features
  - Train a model
  - Predict centres
  - Score

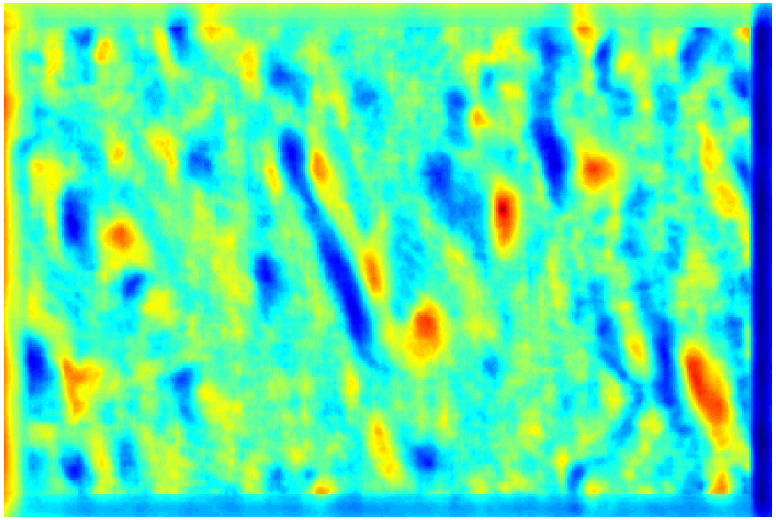


chris.mcintosh@rmp.uhn.on.ca

Template Matching

- Might be difficult to construct a good feature for nuclei
- We can use training data to build templates, and find those templates in the image
  1. Extract the template as a square around a nuclei centre
  2. Use normalized cross-correlation (match_template in python) to find similar patterns
  3. Find the (x,y) locations of all of the top responses

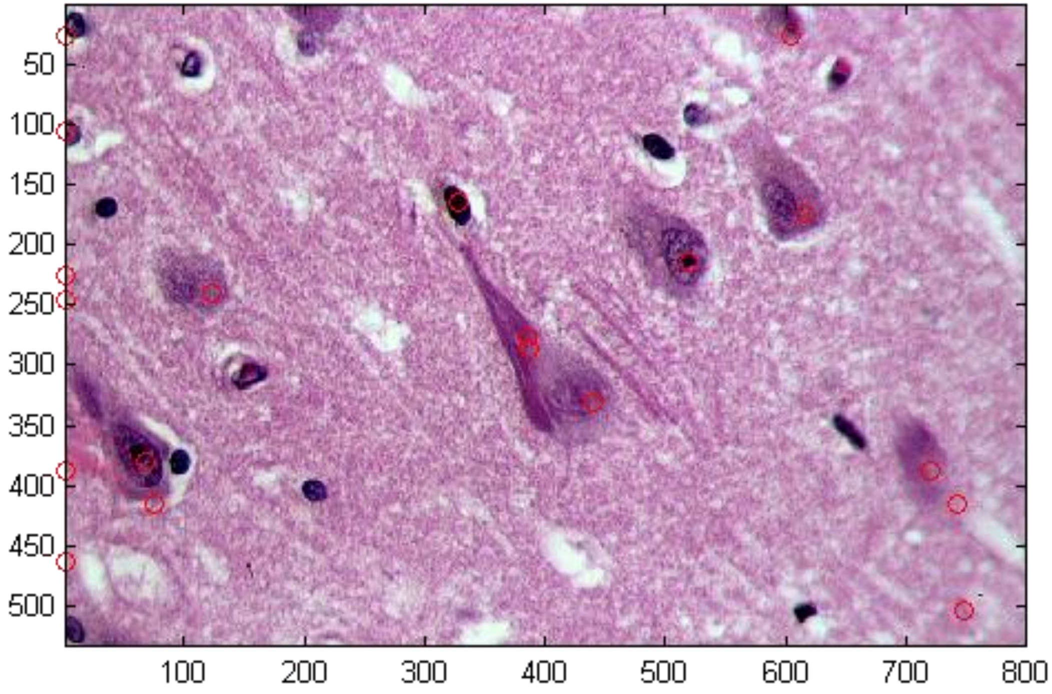




chris.mcintosh@rmp.uhn.on.ca

Scoring

- True positive: Any nuclei centre with a prediction sufficiently close by (e.g. 12 pixels)
- False negative: Any nuclei centre without a sufficiently close prediction
- False positive: Any prediction not sufficiently close to a nuclei centre



chris.mcintosh@rmp.uhn.on.ca

Summary

- Medical image analysis can aid in standardization and efficiency of measurements for outcomes, treatments, and disease understanding
- Many image features are built on filtering or convolution, emulating a similar process to the human visual system
- Human perception of colors, gradients, and edges is both psychological and physical
- The best systems will pair the strengths of medical experts (domain knowledge, compassion, understanding, interaction, high-level analysis) with the best of computers (repeatable, high throughput, quantitative)

chris.mcintosh@rmp.uhn.on.ca