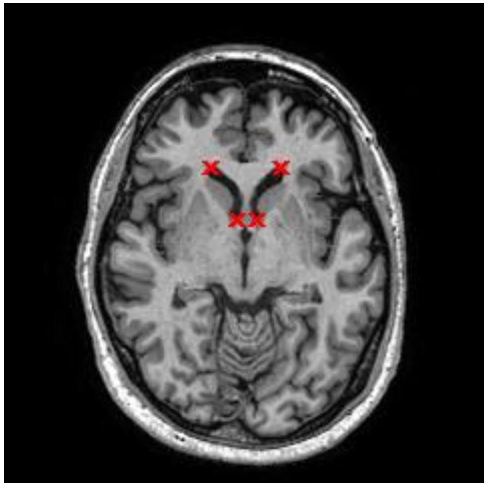
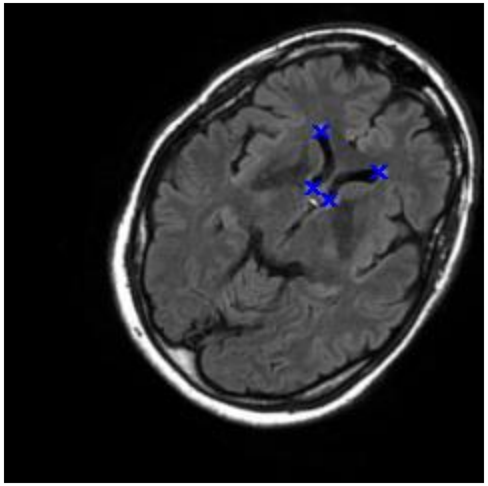
- Point-based registration requires some manual input
- Can it be used with automatic keypoint selection?



Fixed



Moving

Classification of image registration:

- Image dimensionality: 2D, 3D, 3D + time…
- Registration basis: point sets, intensity…
- Geometrical transformations: rigid, affine, nonlinear…
- Degree of interaction: automatic, manual, semi-automatic
- Optimization procedure: closed-form solution, iterative
- Modalities: multi-modal, intra-modal
- Subject: inter-patient, intra-patient, atlas
- Object: brain, head, vertebra, liver…

Intensity-based similarity metrics

Maureen van Eijnatten

Technische Universiteit Eindhoven University of Technology

Where innovation starts

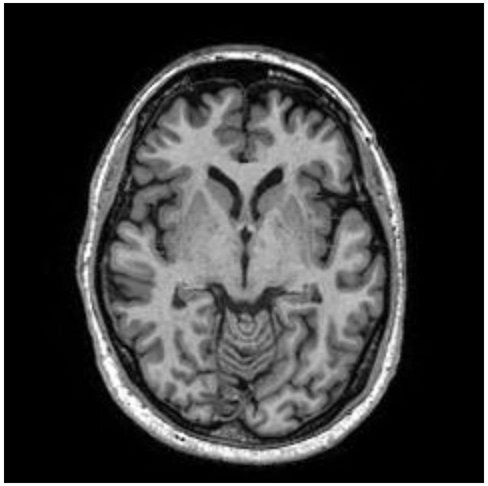
Lecture schedule

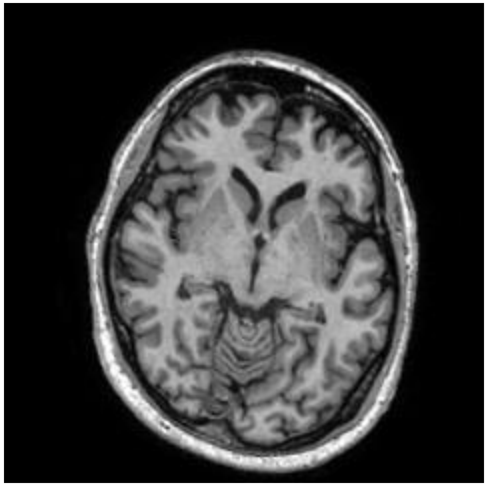
Week	Date	Lecturer	Topics
1	1 Sept.	Maureen	Course introduction; Software demo; Image registration (1)
	3 Sept.	Maureen	Image registration (2); Geometrical transformations
2	8 Sept.	Maureen	Point-based registration
	10 Sept.	**Maureen**	**Intersity-based registration**; Evaluation metrics
3	15 Sept.	*Catch-up day (no lecture)*	
	17 Sept.	Cornel Zachiu (UMCU)	Guest lecture 1: Image analysis for adaptive radiotherapy
4	22 Sept.	Mitko	Introduction to CAD; k-NN; Decision trees
	24 Sept.	Mitko	Generalization and overfitting
5	29 Sept.	Mitko	Logistic regression; Neural networks
	1 Oct.	Friso *(no on-campus lecture!)*	Convolutional neural networks *(pre-recorded)*
6	6 Oct.	Friso	Deep learning frameworks and applications
	8 Oct.	Friso	Unsupervised machine learning
7	13 Oct.	Maureen	Deep learning for deformable image registration
	15 Oct.	Geert-Jan Rutten (ETZ)	Guest lecture 2: Image analysis in neurosurgery applications
8	20 Oct	*Self-study (no lecture)*	Active shape models
	22 Oct	*Self-study (no lecture)*	Active shape models

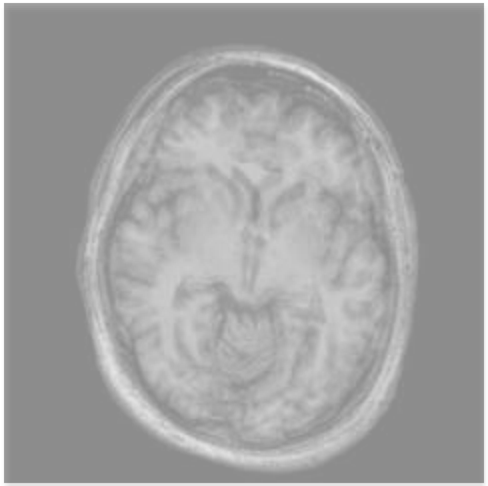
Image intensity is an alternative registration basis to points.

It is the most widely used registration basis.

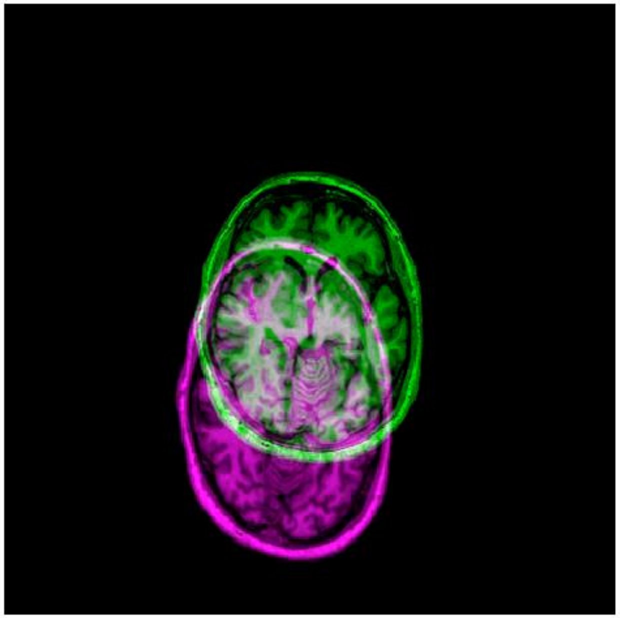
Compared to point-based registration, requires less user interaction.

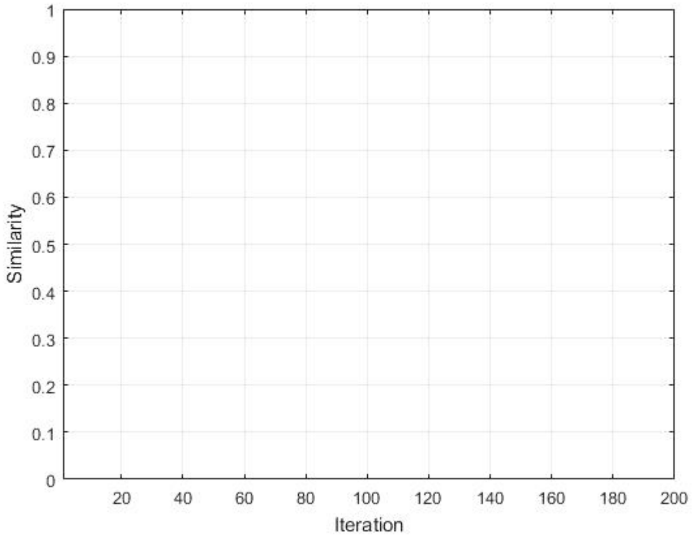






Intensity-based image registration works by iterative optimization of an intensity-based similarity measure.





Outline:

Intensity-based similarity measures:

- Sum of square differences
- Cross-correlation
- Mutual information

Optimization for intensity-based registration:

- Gradient ascent (descent)

Let I and J be two images and i the pixel locations.

A simple and intuitive intensity-based measure of the similarity of I and J is the sum of squared differences (SSD):

$$\mathrm{SSD}(I,J) = \sum_{i=1}^{n} (I(i) - J(i))^2$$

10	**12**	**16**
8	**12**	**18**
4	**8**	**10**

8	**14**	**16**
12	**12**	**20**
2	**6**	**12**

$(10-8)^2$	$(12-14)^2$	$(16-16)^2$
$(8-12)^2$	$(12-12)^2$	$(18-20)^2$
$(4-2)^2$	$(8-6)^2$	$(10-12)^2$

If I is the fixed image in a registration problem, and J is the moving image transformed with a transformation $\mathbf{T}$ the similarity measure will be a function of the transformation :

$$\mathrm{SSD}(I, J, \mathbf{T}) = \sum_{i=1}^{n} (I(i) - J_{\mathbf{T}}(i))^2$$

The SSD will be lowest when the images are perfectly aligned and will increase with misalignment.

When lowest == zero?

It can be shown that this measure is optimal when two images differ only by Gaussian noise. This is an implicit assumption of this measure.

- Not true for inter-modality registration
- Rarely true for intra-modality registration (e.g. MRI noise in not Gaussian, there will be changes between acquisition etc.)

Nevertheless, SSD can be still used with success in intra-modality registration.

Drawback: It can be very sensitive to a few “outlier” intensity differences.

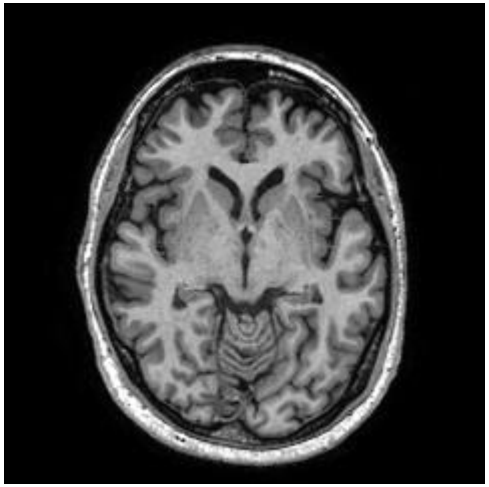
Another measure that makes slightly less assumptions is normalized cross correlation C:

$$C(I,J) = \frac{\sum_{i=1}^{n}(I(i)-\bar{I})(J(i)-\bar{J})}{\sqrt{\sum_{i=1}^{n}(I(i)-\bar{I})^2\sum_{j=1}^{n}(J(i)-\bar{J})^2}}$$

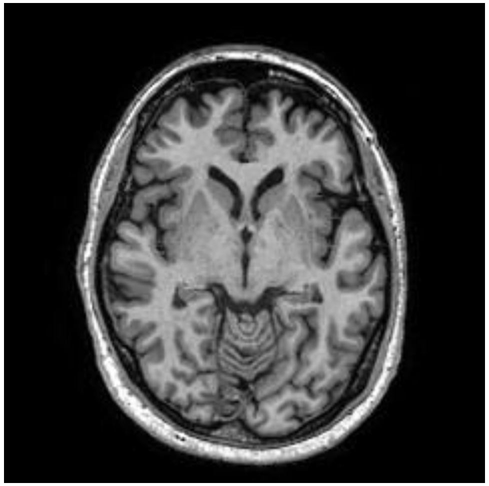
The main assumption of normalized cross-correlation is that there is a linear relationship between the pixel intensities in the two images.

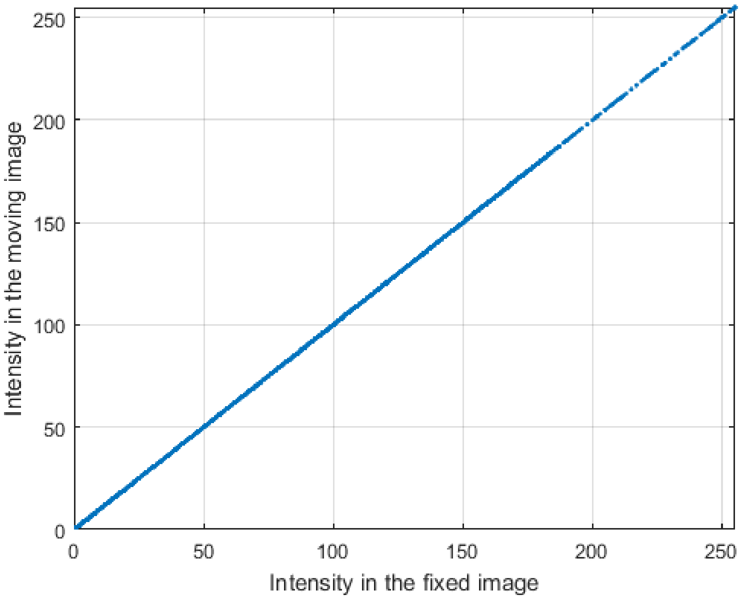
- Mostly true for intra-modality registration
- Usually not true for inter-modality registration

T1

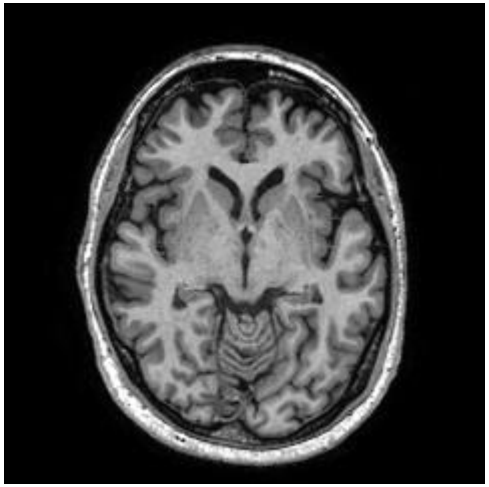


T1

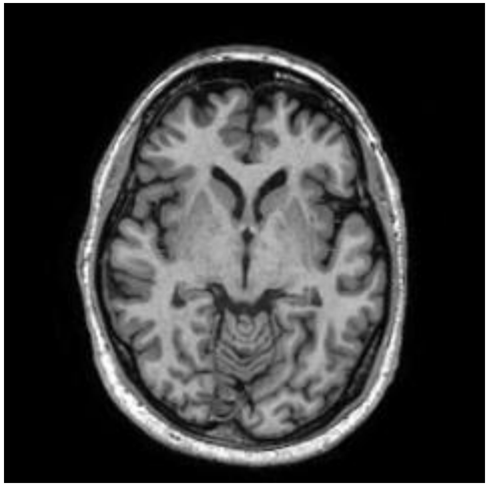


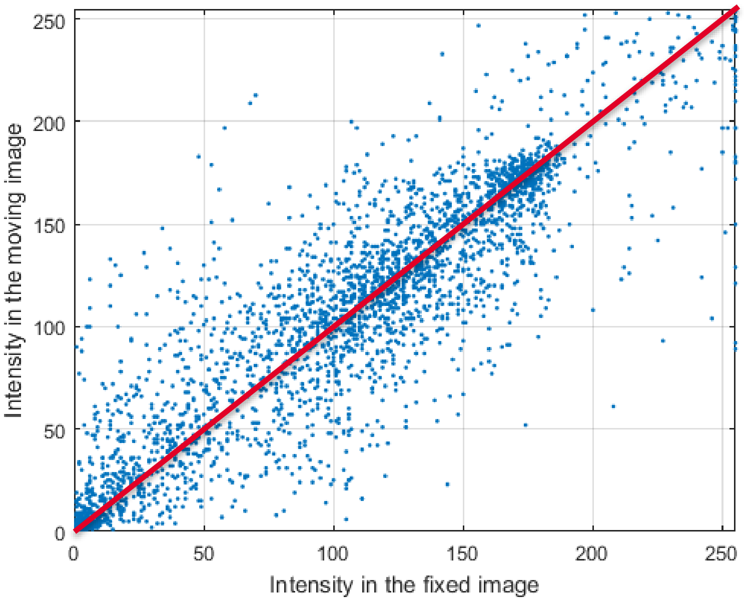

Intensity in the moving image
Intensity in the fixed image
0
50
100
150
200
250
0
50
100
150
200
250

T1



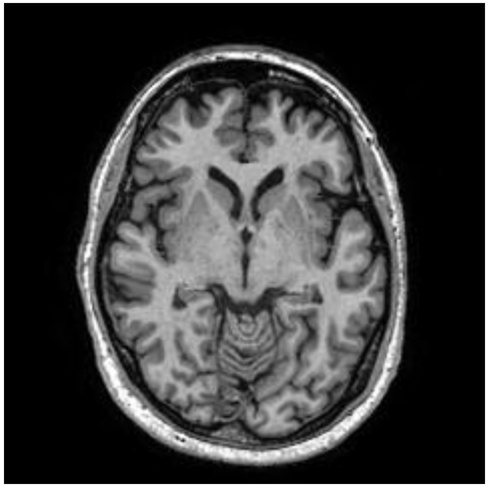
T1



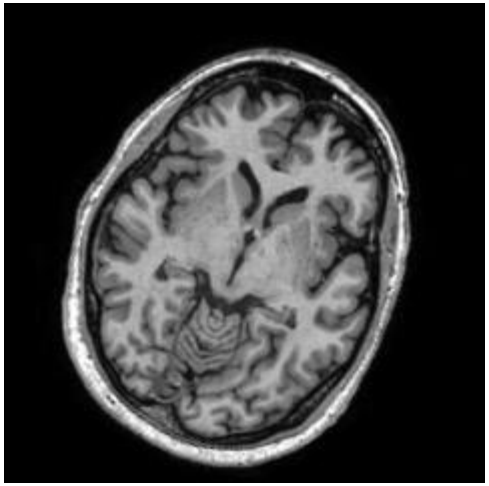


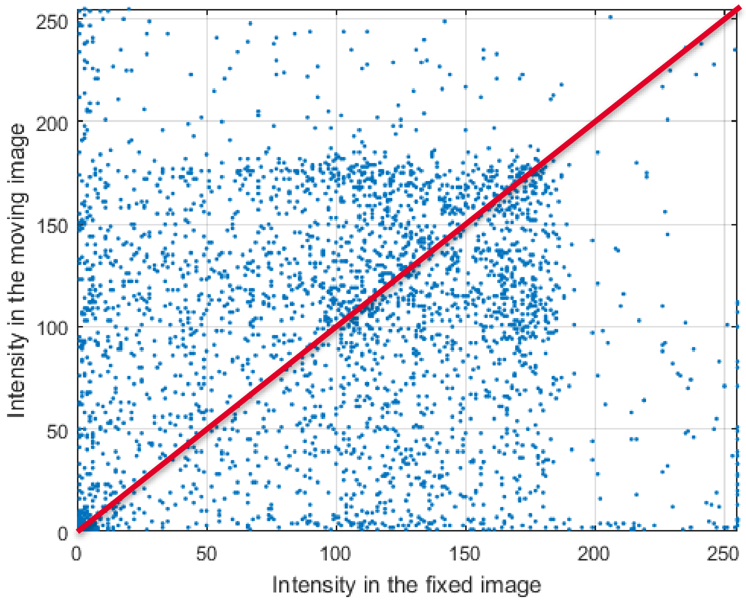
Intensity in the moving image
Intensity in the fixed image
0
50
100
150
200
250

T1

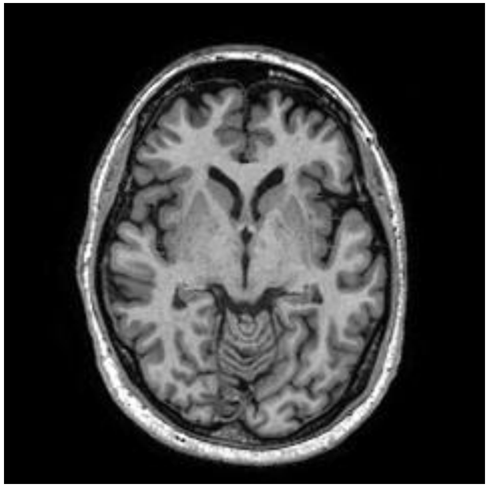


T1

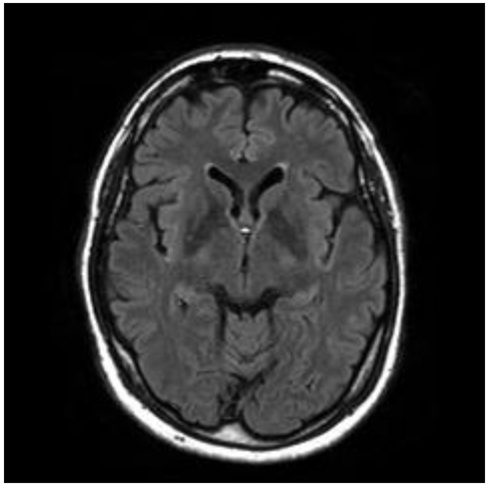


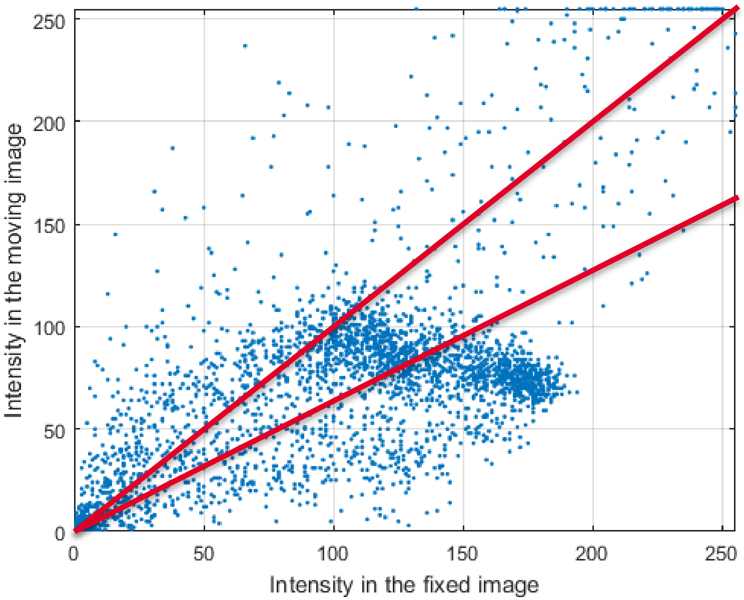

250
200
150
100
50
0
0
50
100
150
200
250
Intensity in the moving image
Intensity in the fixed image

T1



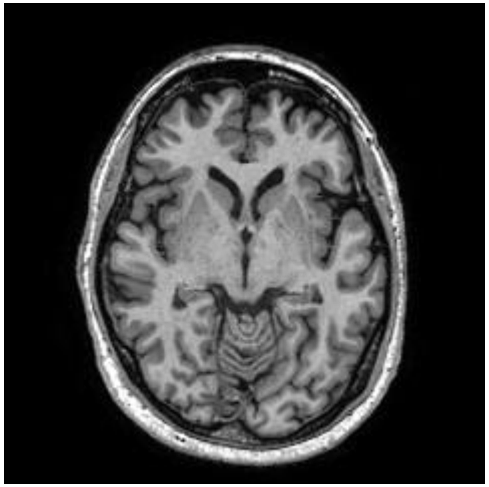
T2



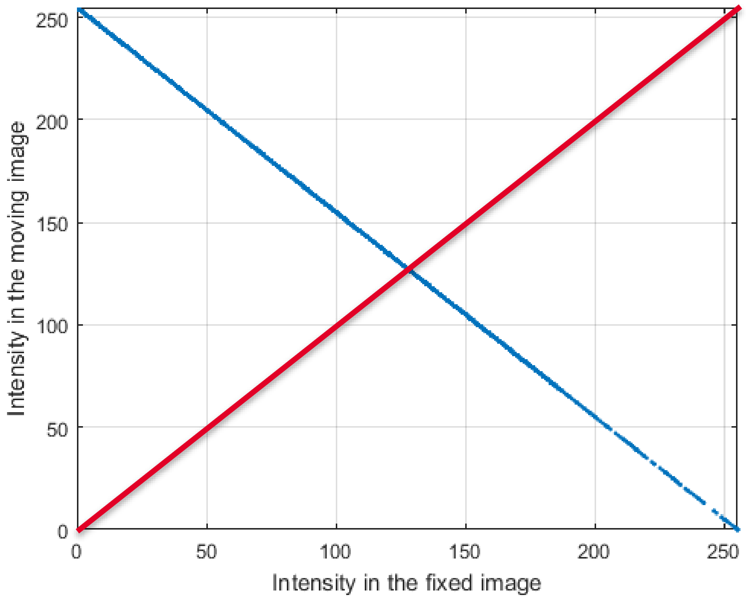


Violates the assumption of linear relationship between the intensities – not optimal

T1



Simulated modality



Completely opposite of the assumption (inverse linear relationship)

Probability theory

Random variables map the outcomes of random phenomena to numbers.

Example random phenomenon: coin toss.

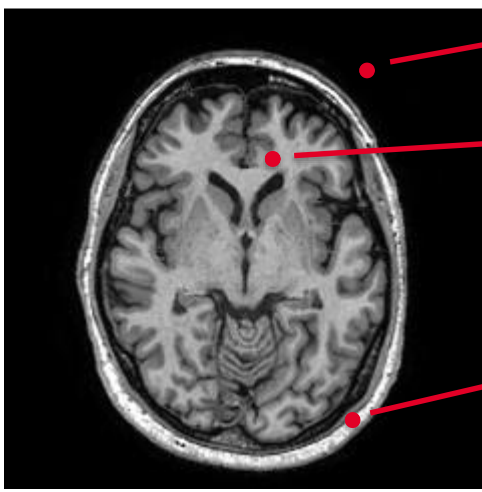
Random variable X: the outcome of the coin toss.

Another random variable Y: the number of heads in a series of 3 tosses.

$$X = \begin{cases} 1, & \text{if heads} \\ 0, & \text{if tails} \end{cases}$$

Image intensities as random variables:

Random phenomenon: pick a random pixel location.
In this case, the pixel intensity can be treated as a random variable.



$$X = 0$$

$$X = 134$$

$$X = 241$$

Each outcome from the random phenomenon we are studying can be associated with a probability.

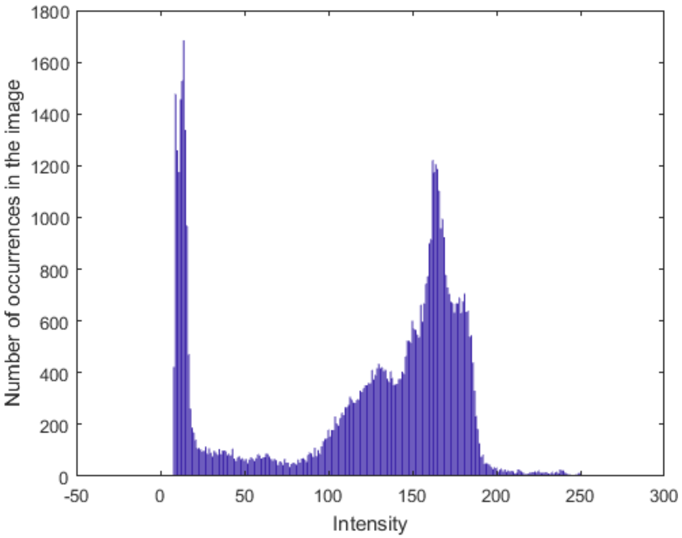
If a random variable X can have a finite set of possible values, we can define a function that maps each possible value to a probability. This function is called <u>probability mass function</u> (p.m.f).

Probability mass function:

$$p_X(x) = P(X = x)$$

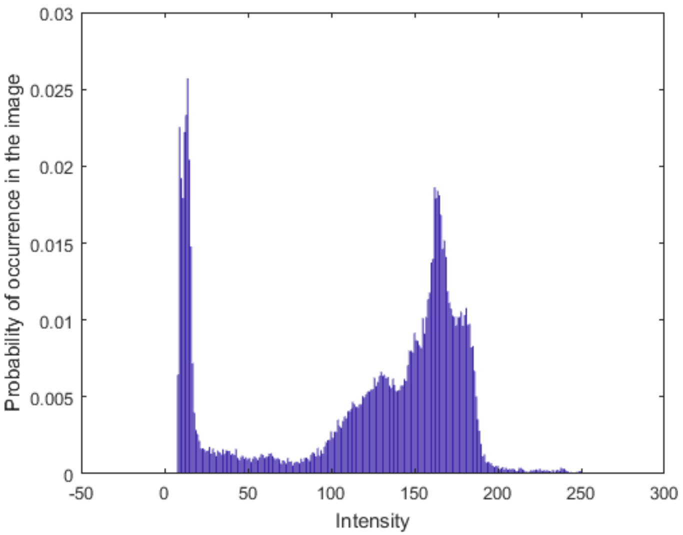
How can we define the probability mass function for the image intensities?

Image histogram – count of the number of occurrences of each intensity value in the image.

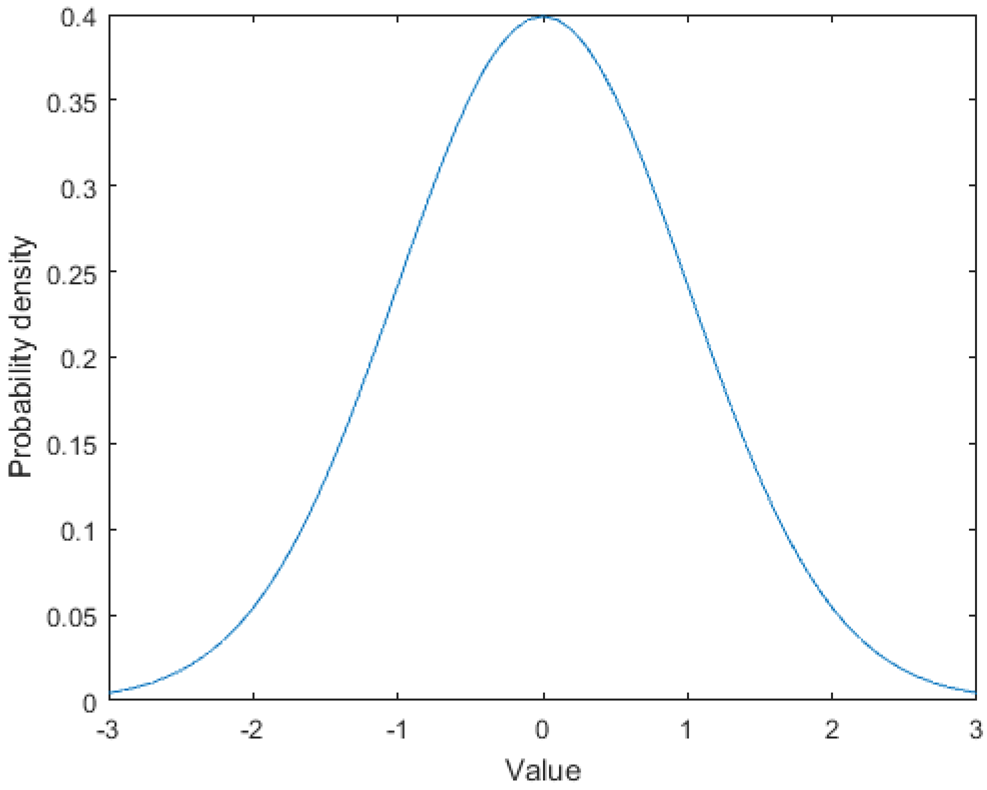


In order to treat the counts of the histogram as probability values, we must normalize the histogram in such a way that all values sum to 1.

This is the probability mass function for the pixel intensity as a random variable.



For continuous random variables (can take infinite number of possible values), we can define the probability density function (p.d.f):

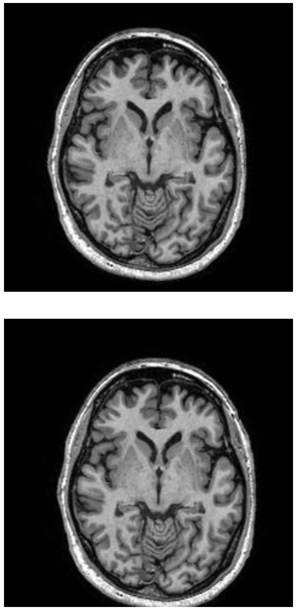


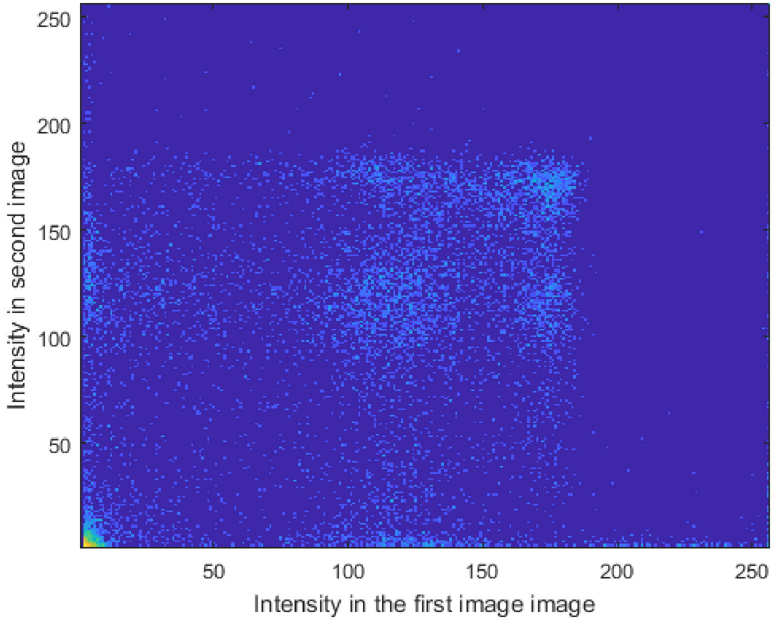
What if we have two random variables? For example, the pixel intensity in two images.

In this case we can define a joint probability mass function:

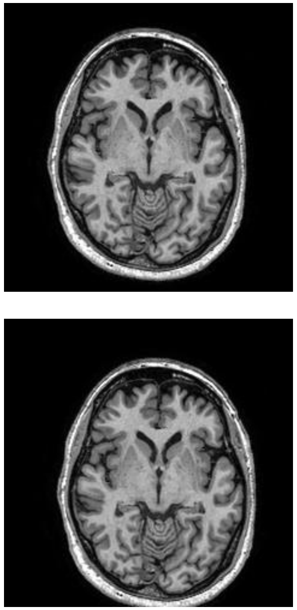
$$p_{X,Y}(x, y) = P(X = x, Y = y)$$

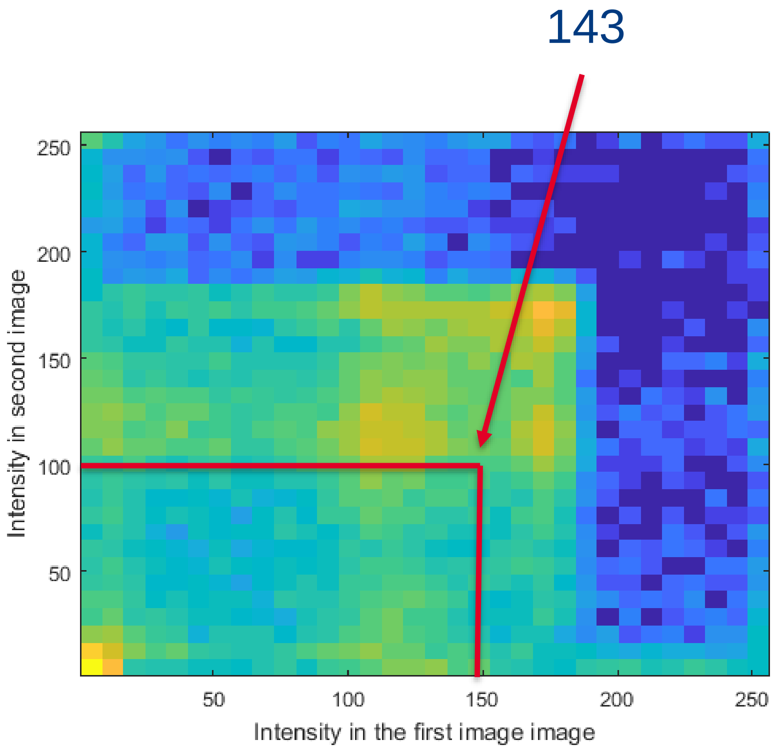
Example: the pixel intensity in two images. We can estimate this joint p.m.f. from the joint histogram of the two images.



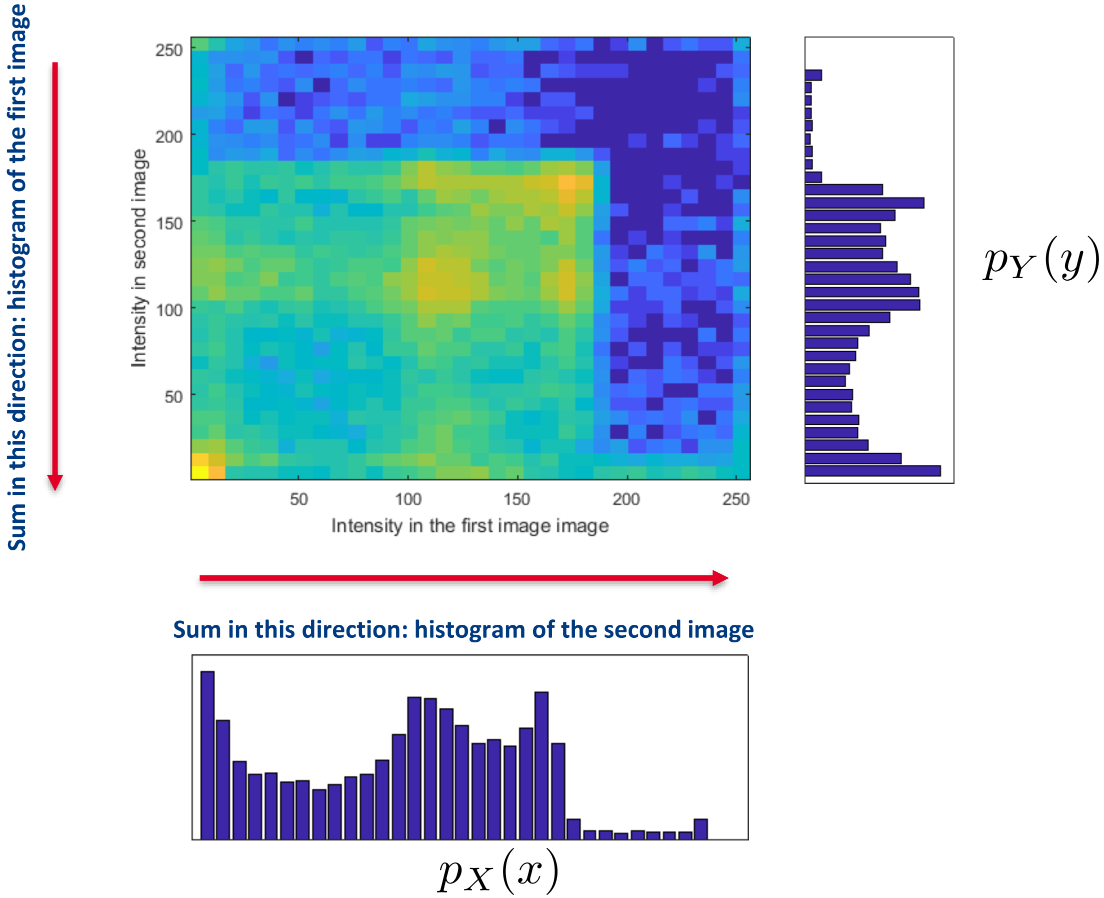


Example: the pixel intensity in two images. We can estimate this joint p.m.f. from the joint histogram of the two images.





From the joint p.m.f. we can compute the p.m.f.'s of the individual variables (called marginal p.m.f.'s):



Conditional distributions answer the question: what is the probability of distribution over Y when we know that X takes a certain value?

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$$

$$p_{X,Y}(x,y) = p_{Y|X}(y|x)(x)p_X(x)$$

Example: if we pick a random image location, and the first image has intensity value of 124 at that location, what is the probability distribution for the intensity values in the second image at that location?

The random variables X and Y are independent if:

$$p_{X,Y}(x, y) = p_X(x)p_Y(y)$$

Bayes's rule is a very useful formula that we will use later in the computer-aided diagnosis sections of this course:

$$p_{Y|X}(y|x) = \frac{p_{Y|X}(x|y)p_Y(y)}{p_X(x)}$$

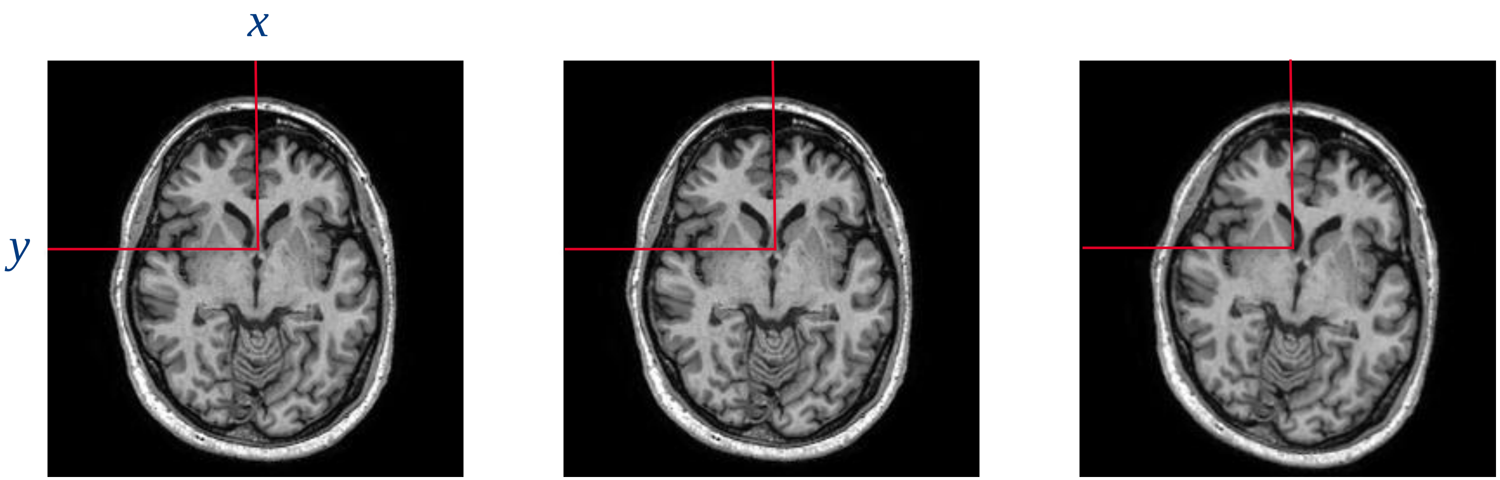
Intensity-based similarity metrics

Continued.

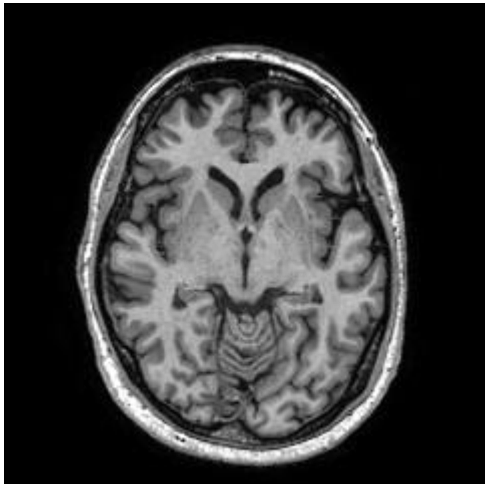
Another measure with even less assumptions: mutual information (MI).

An intuitive interpretation of the MI between two images:

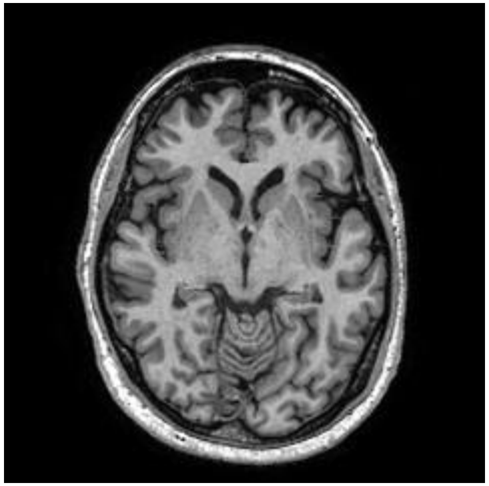
If we know the pixel intensity value at some location in the fixed image, how much information do we have about the pixel intensity at the same location in the other image?



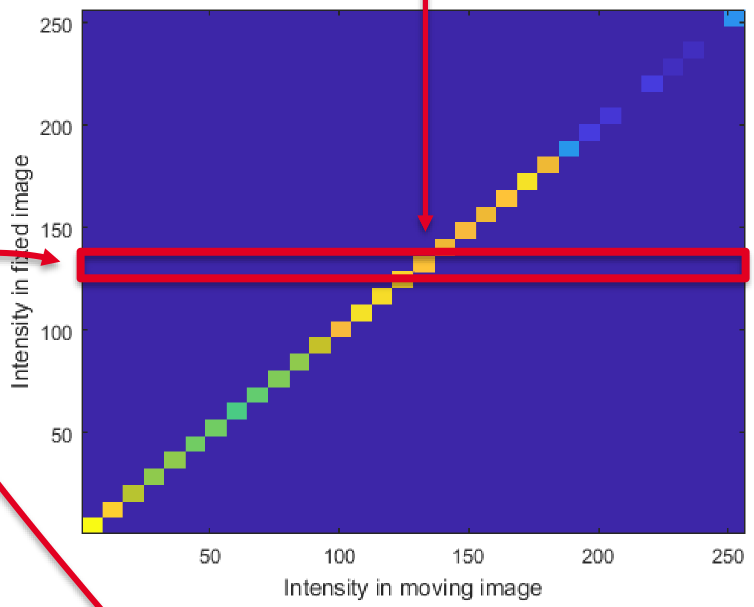
T1



T1

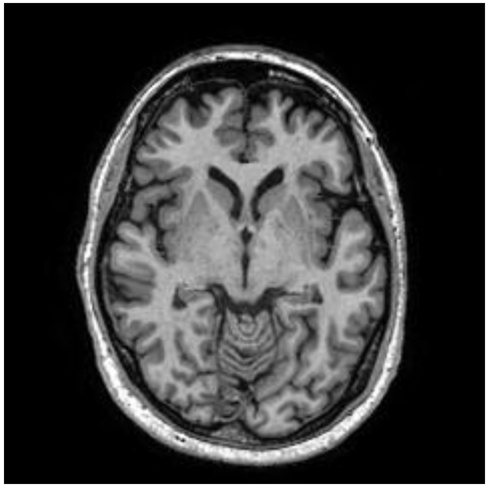


There is only one option for the other value (all other values have zero probability).

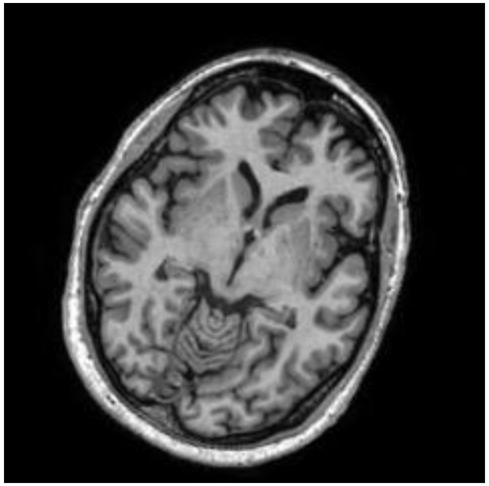


One value is “fixed”.

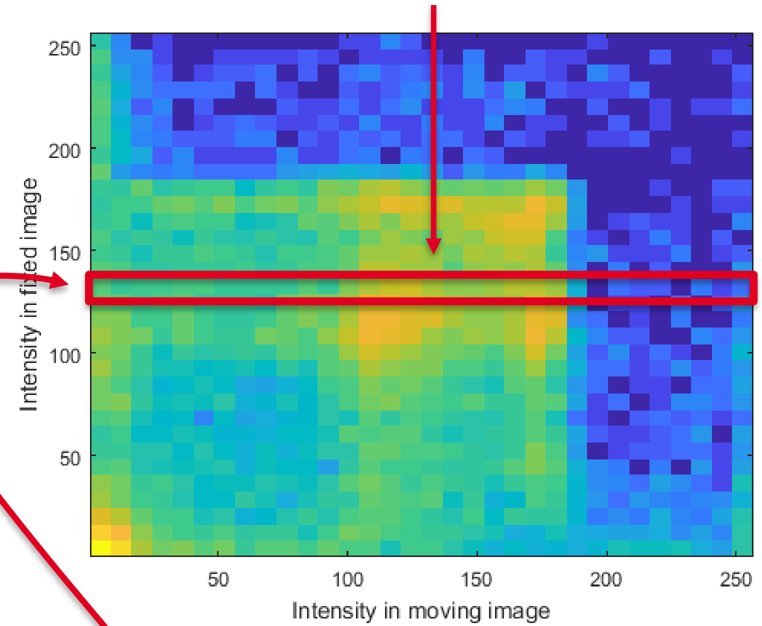
T1



T1

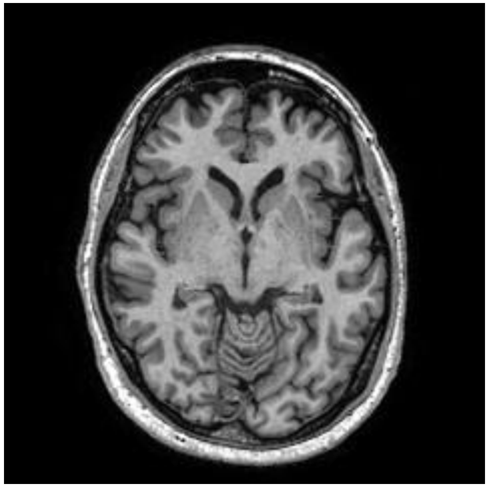


There are many probable values.

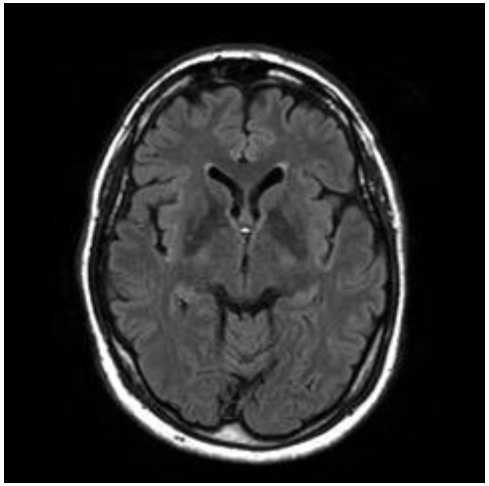


One value is “fixed”.

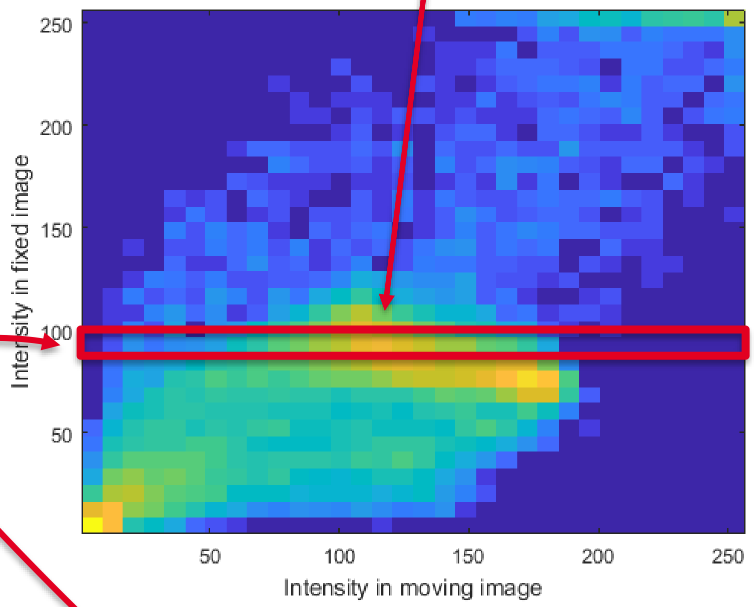
T1



T2

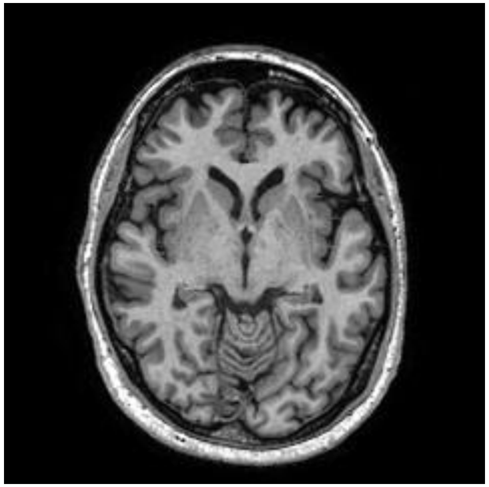


There are a few values with very high probability.

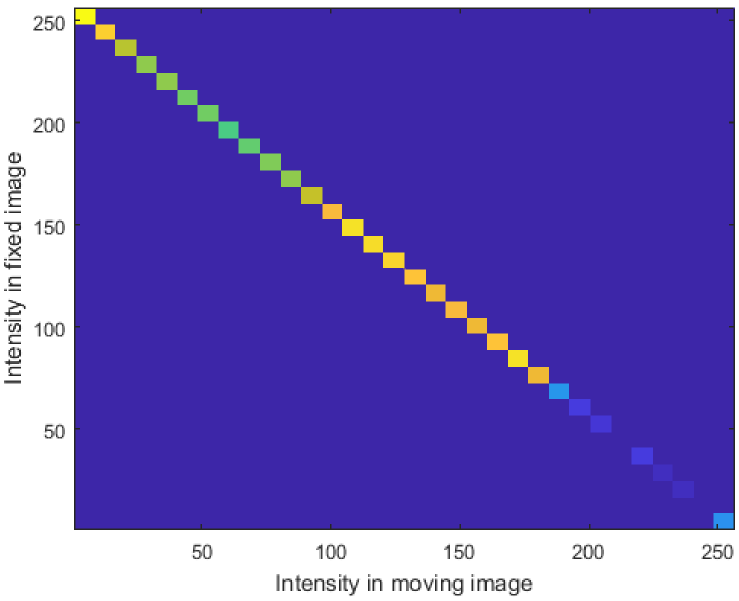


One value is “fixed”.

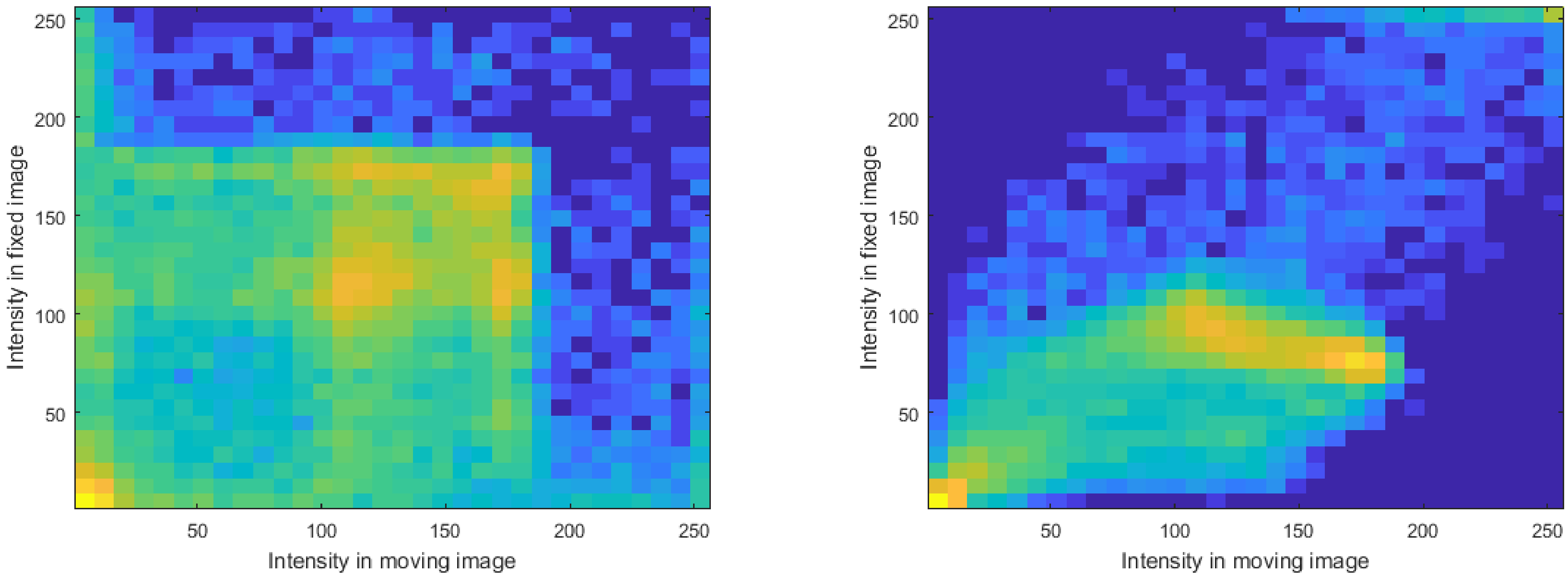
T1



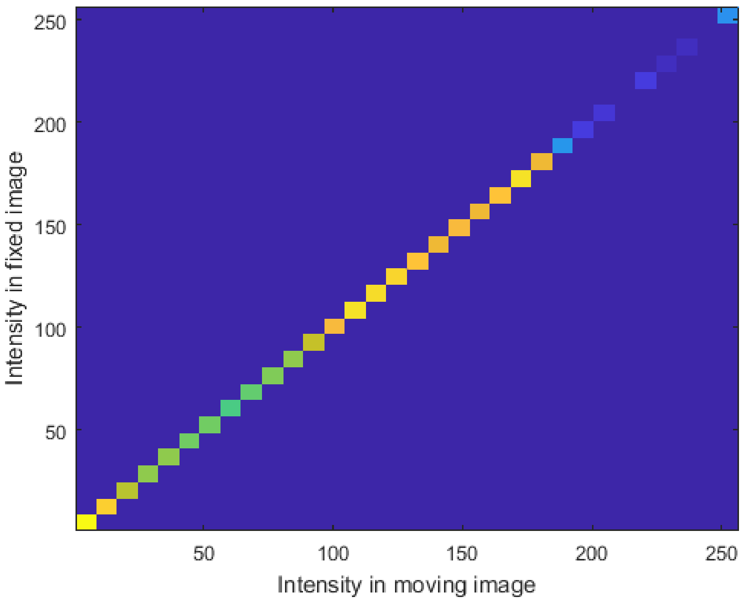
Simulated modality

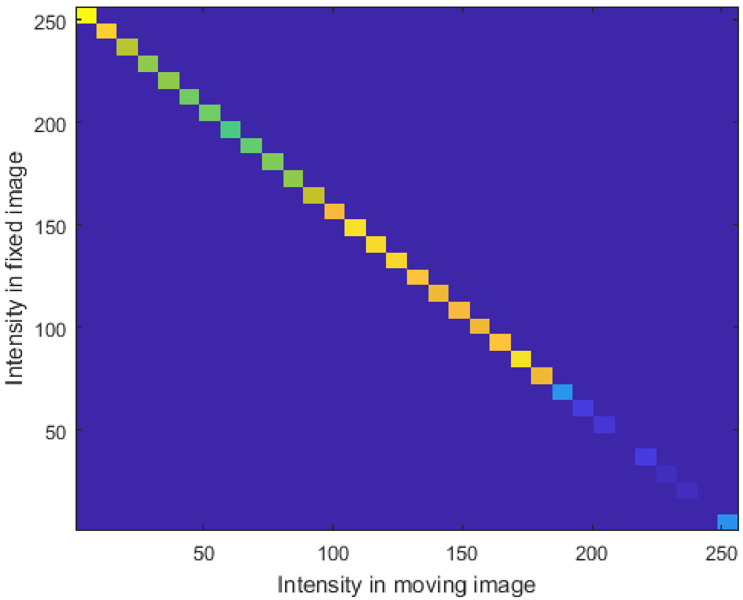


Which pair of images is better aligned according to the join histogram? Which histogram is more "compact"?



Which pair of images is better aligned according to the join histogram? Which histogram is more “compact”?





Given the joint p.m.f. of two images and the two marginal p.m.f.'s, the mutual information between the two images can be computed with the following formula:

$$MI(I,J) = \sum_{i=1}^{n}\sum_{j=1}^{n} p_{I,J}(i,j)\log\frac{p_{I,J}(i,j)}{p_I(i)p_J(j)}$$

The unit of MI depends on the particular *log* function: when using the natural logarithm the unit is *nats*, when using base 2 logarithm the unit is *bits*.

MI in essence is a measure of the “compactness” of the joint p.m.f. of two images.

When the two images are well registered the joint p.m.f. is compact.

When the two images are not well aligned the joint p.m.f. is “spread out”.

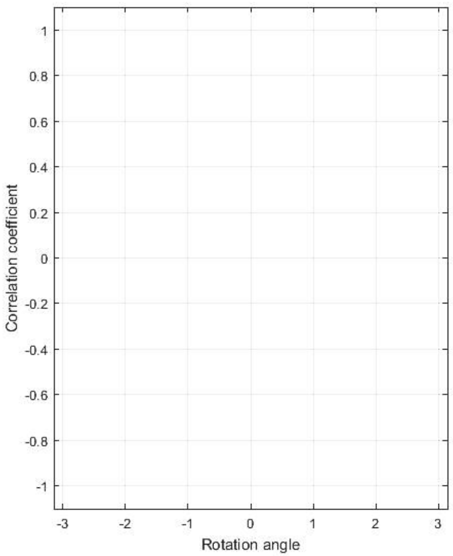
We have now defined several intensity-based similarity measures.

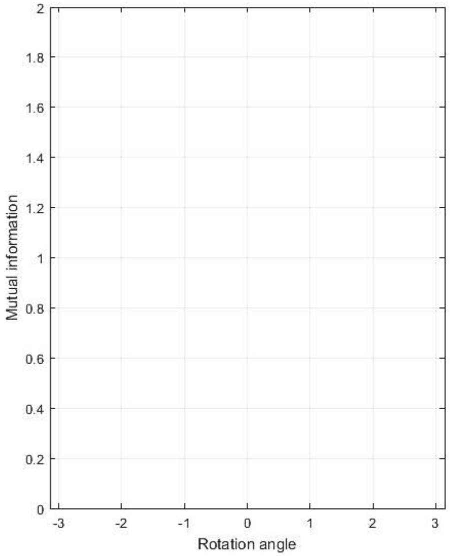
When one of the images is being transformed, the similarity measures are a function of the image transformation.

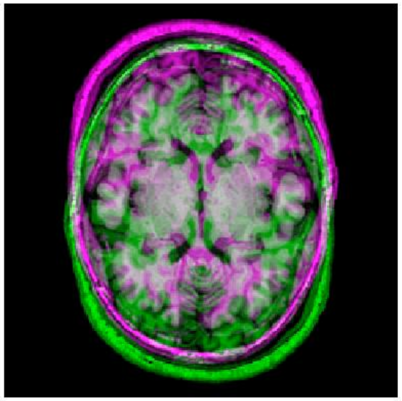
This is “step 1” in our general approach to registering two images.

“Step 2” is finding the parameters that find the transformation that maximizes the similarity between two images.

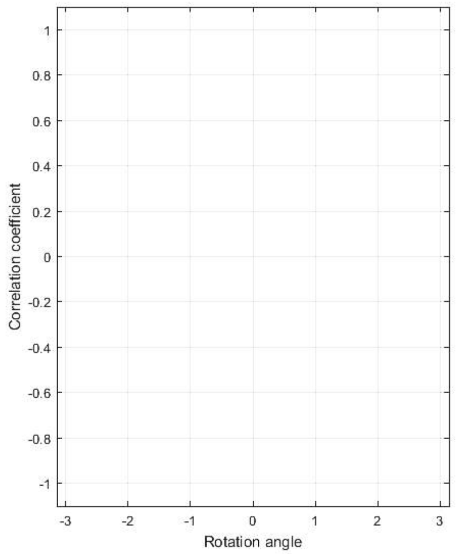
Similarity as a function of transformation (T1 to T1):

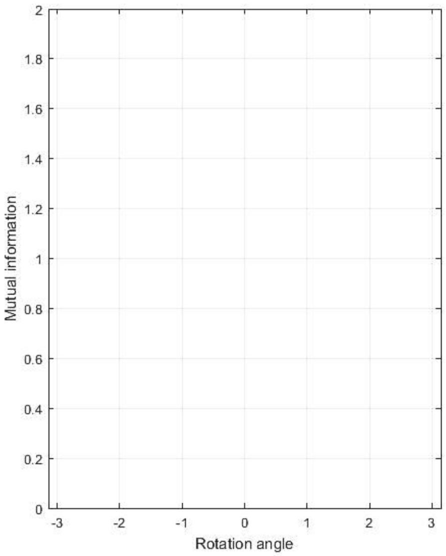


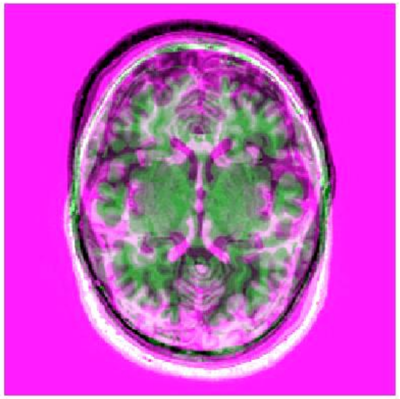




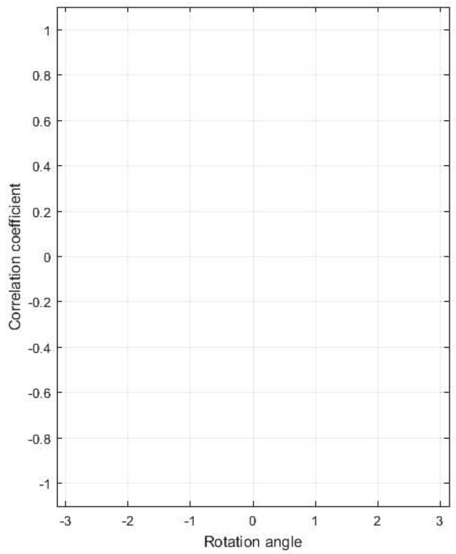
Similarity as a function of transformation (T1 to sim. modality):

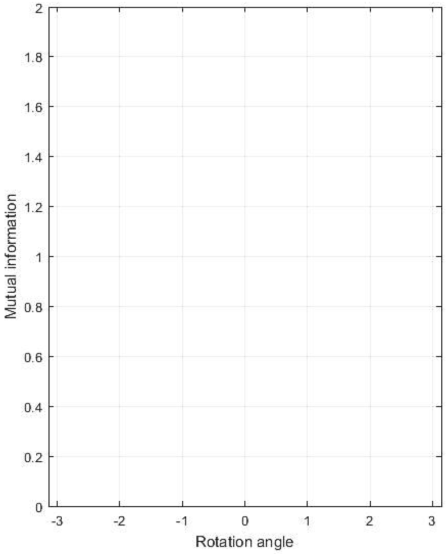


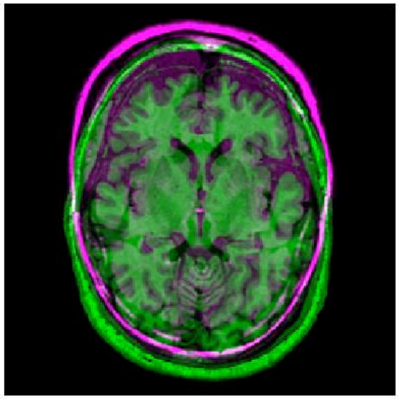




Similarity as a function of transformation (T1 to T2):



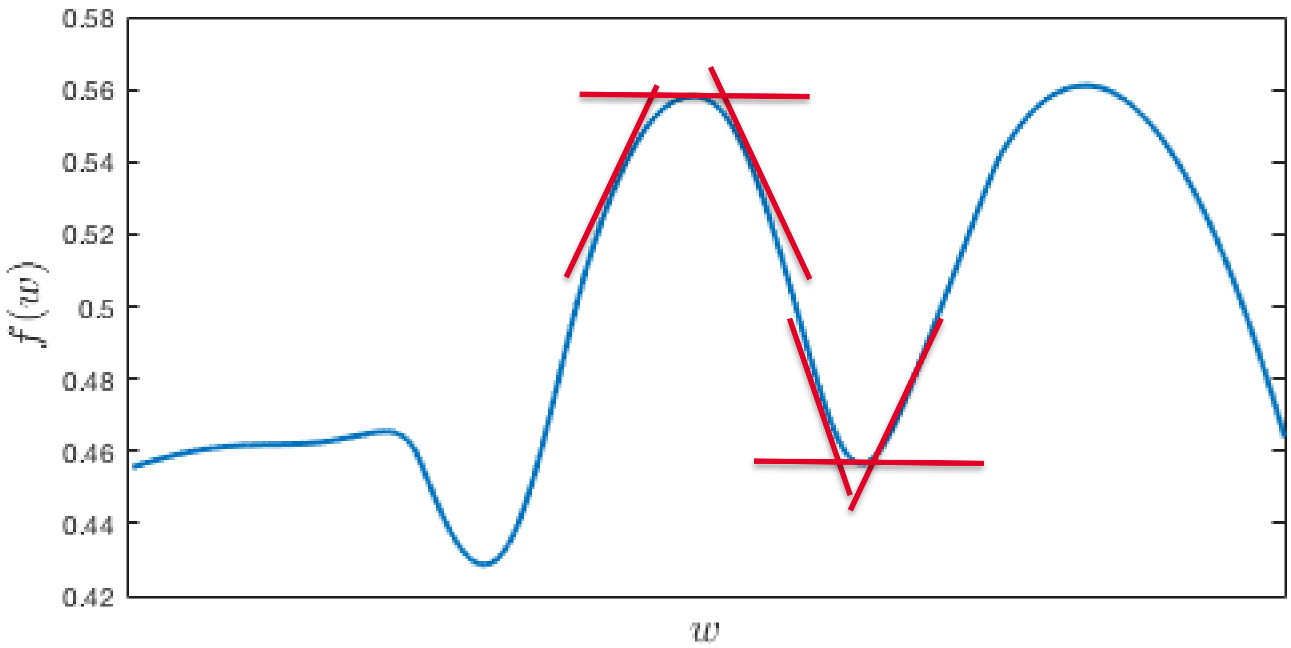




Optimization

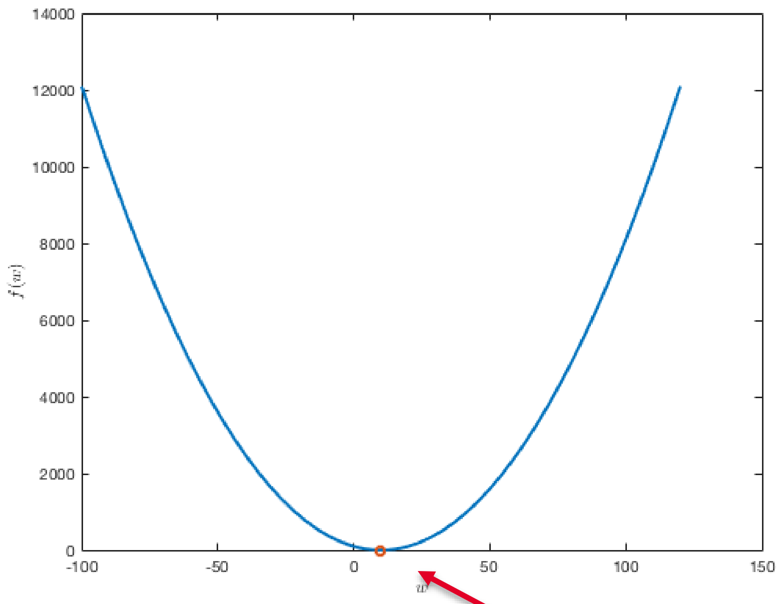
Continued.

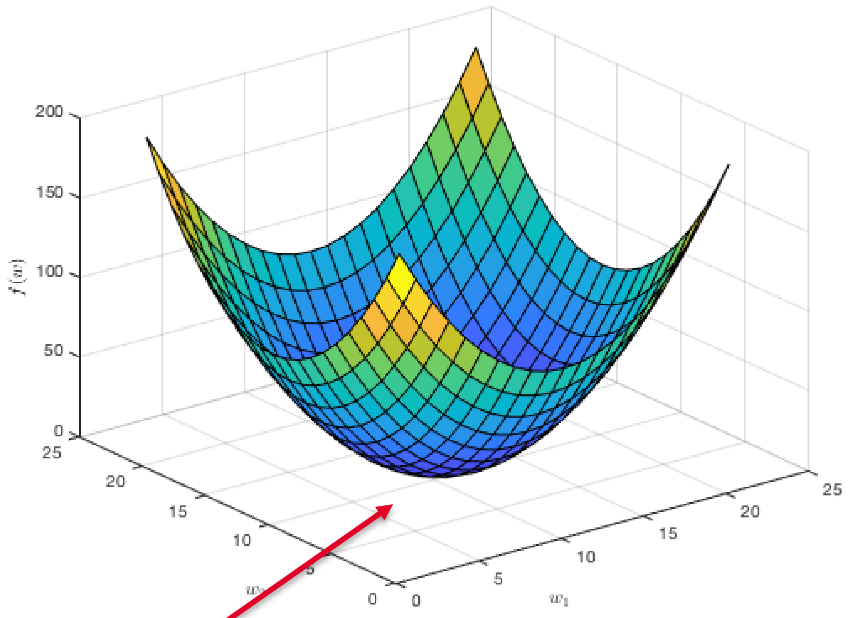
How to find the min. and max. of this function analytically?



Compute the derivative and set it to zero. If the function has more than one variable, set the partial derivatives (or gradient vector) to zero.

Convex functions:





Single minimum

For point based affine registration we had:

$$E(\mathbf{T}) = ||\mathbf{T}\mathbf{X}' - \mathbf{X}||_F^2$$

$$\nabla_{\mathbf{T}} E(\mathbf{T}) = 0$$

Find the expression of the gradient and set it to zero. This will result to a system of equations.

$$\mathbf{T} = \mathbf{X}'\mathbf{X}^{\intercal}(\mathbf{X}\mathbf{X}^{\intercal})^{-1}$$

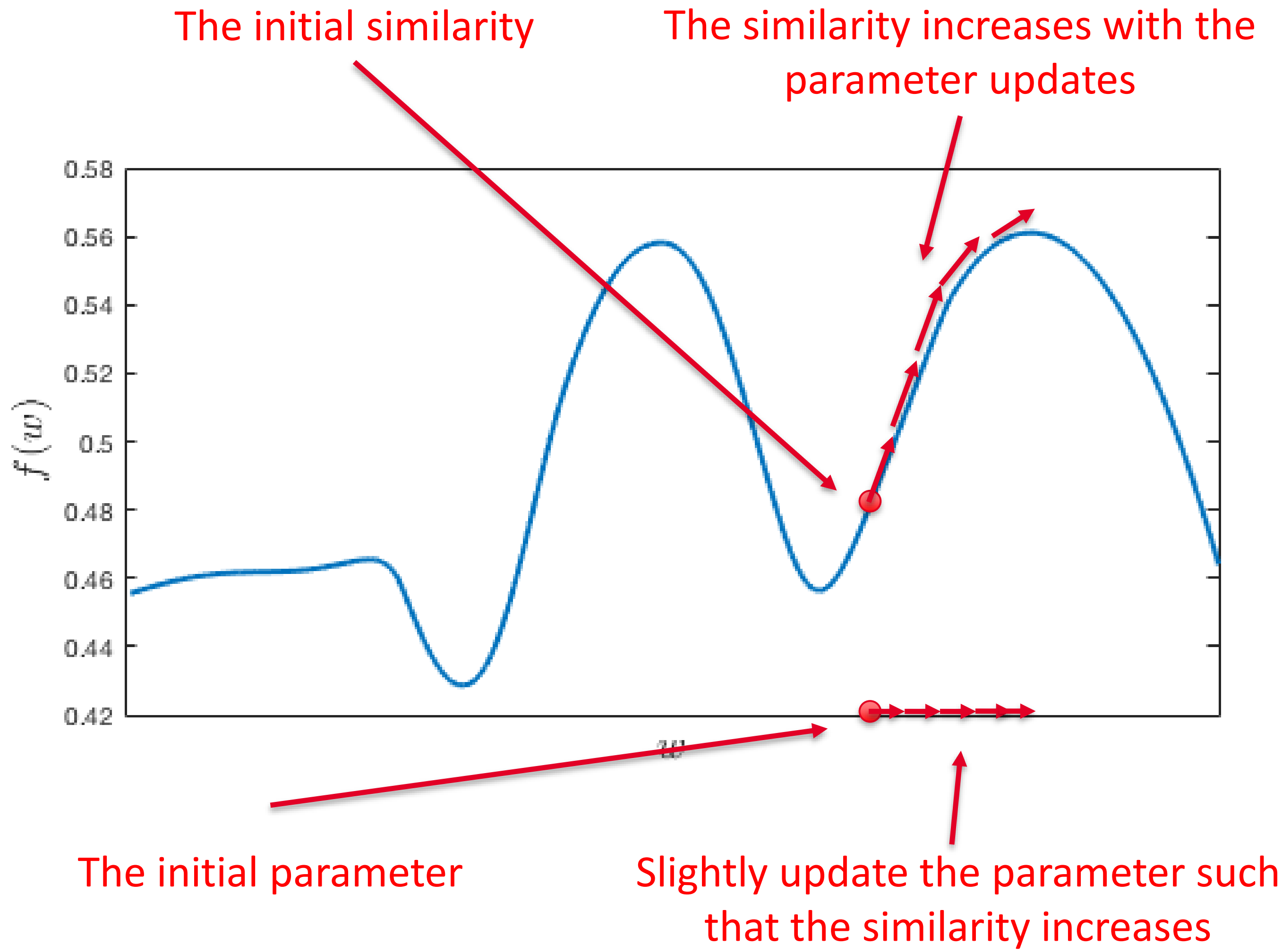
The solution of this system is the optimal value of $\mathbf{T}$.

However, it might be the case that the system of equations produced by setting the gradient to zero is not solvable.

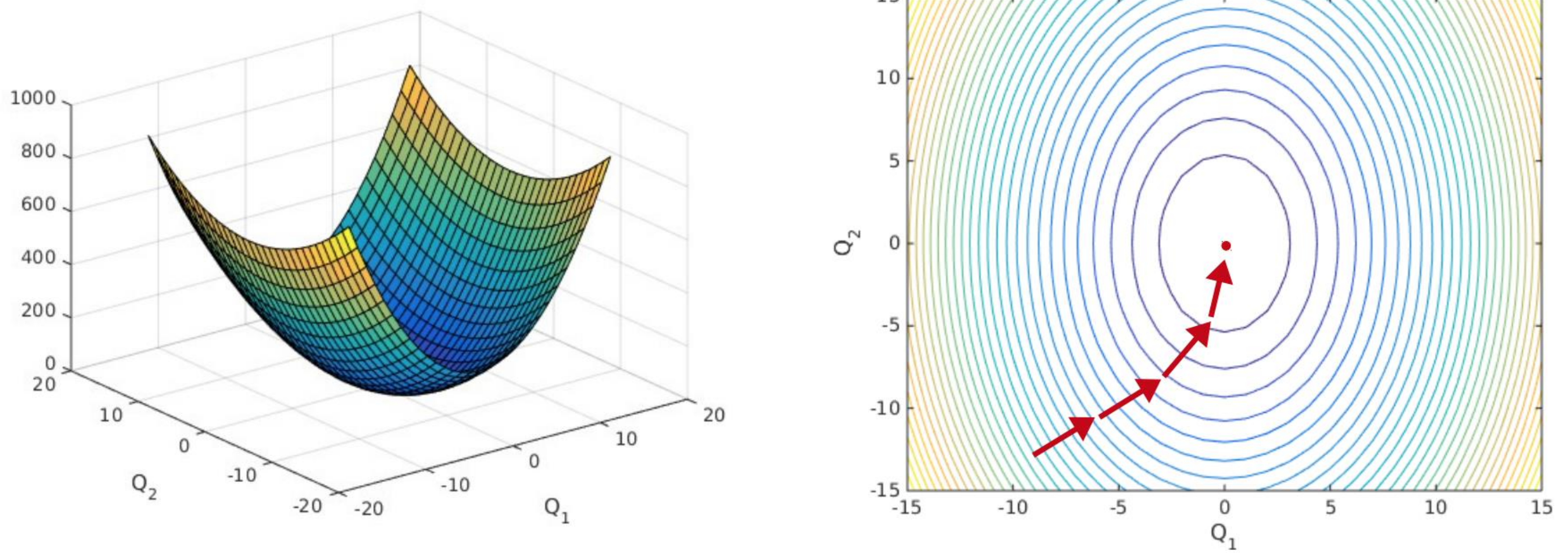
In that case we have to result to numerical methods for finding the minimum of the error (or the maximum of the similarity).

General procedure (for maximization of a similarity function):

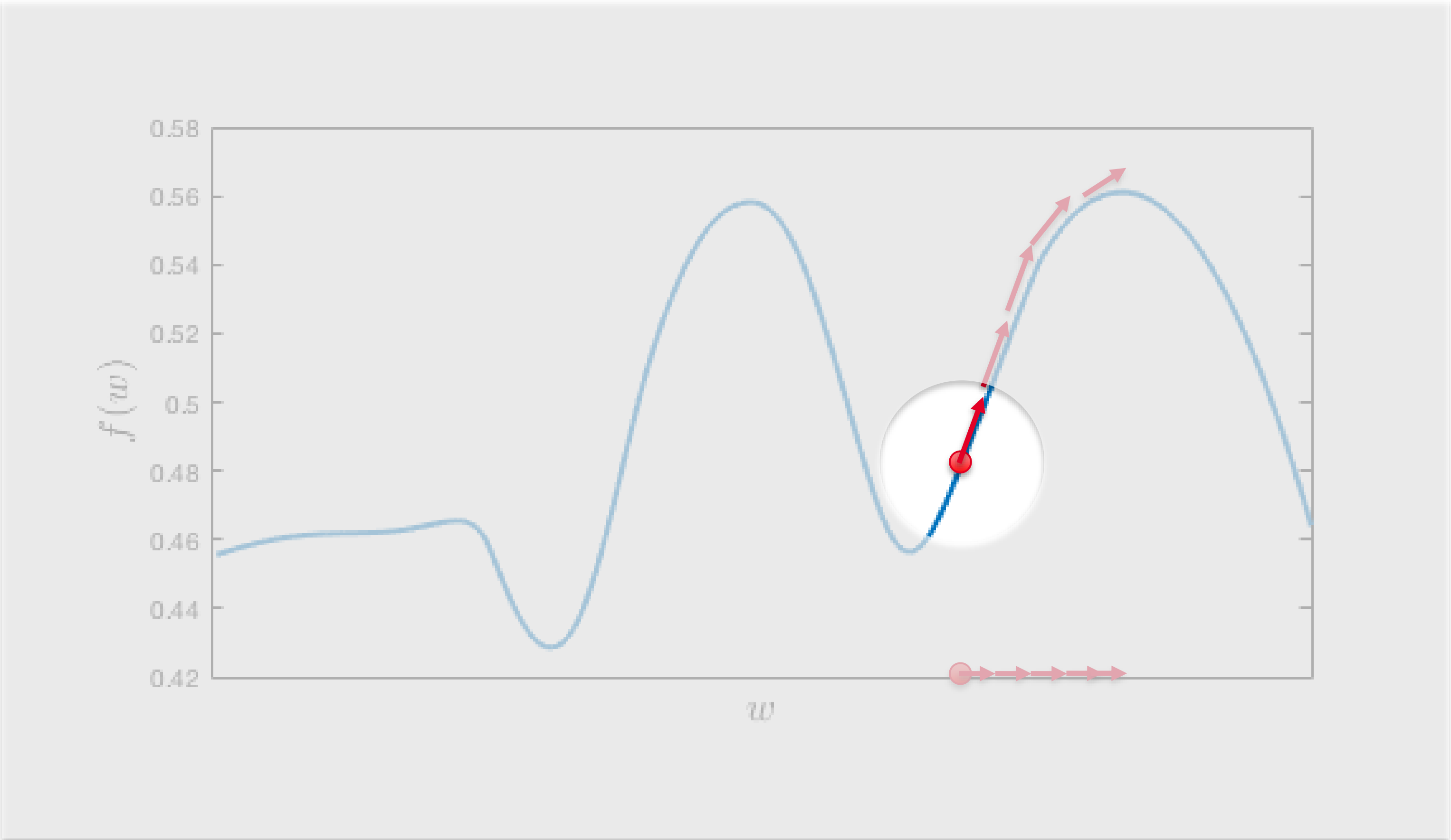
1. Start with some initial values for the parameters (in this case the transformation $\mathbf{T}$).
2. Slightly update the parameters in such a wat that the similarity will slightly increase.
3. Repeat until the similarity stops increasing.


The initial similarity
The similarity increases with the parameter updates
0.58
0.56
0.54
0.52
0.5
0.48
0.46
0.44
0.42
$f(w)$
w
The initial parameter
Slightly update the parameter such that the similarity increases

Minimizing a function with two parameters:

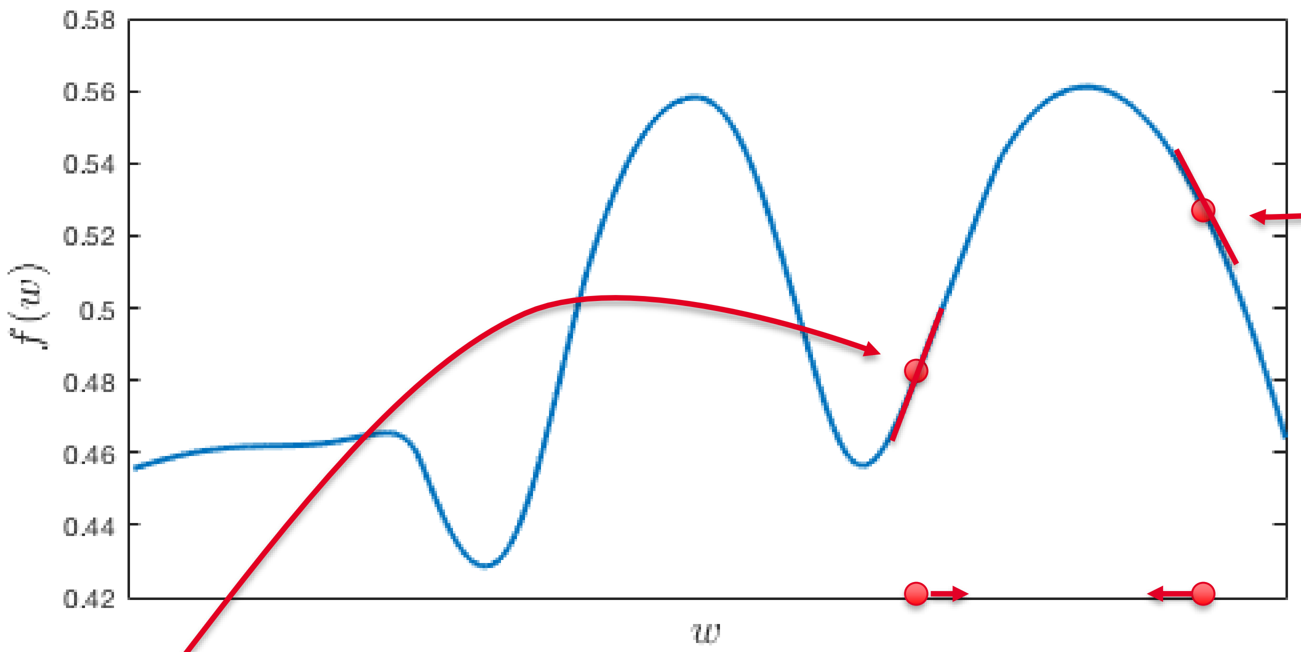


How to find out in which direction to do the parameter update?



How to find out in which direction to do the parameter update?

How to find out in which direction to do the parameter update?



If we are here → positive gradient → increase the parameter (move to the right)

If we are here → negative gradient → decrease the parameter (move to the left)

Gradient ascent algorithm for maximizing a function $f(\mathbf{w})$:

1. Choose some initial values of the parameters $\mathbf{w}$
2. Calculate the value for the gradient of $f(\mathbf{w})$ for the current parameters
3. Update the parameters in the direction of the gradient:

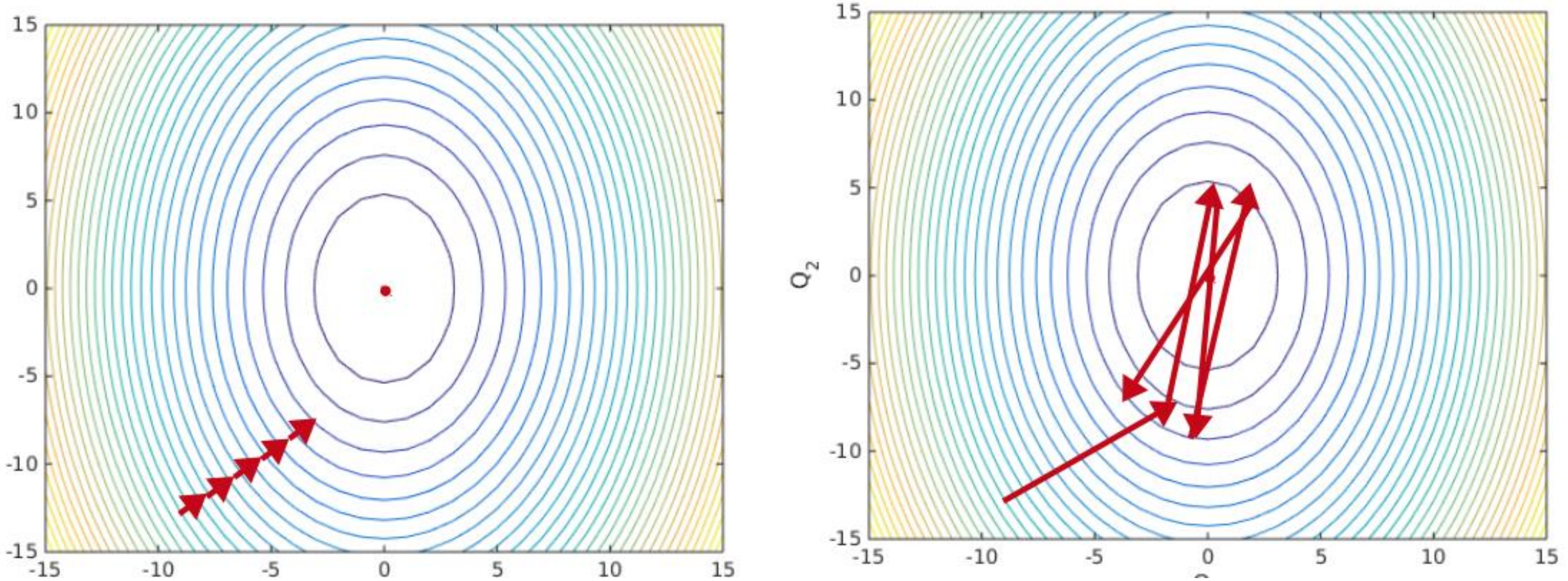
$$\mathbf{w} \leftarrow \mathbf{w} + \mu \nabla_{\mathbf{w}} f(\mathbf{w})$$

If we want to minimize the function we move in the direction opposite of the gradient (gradient descent):

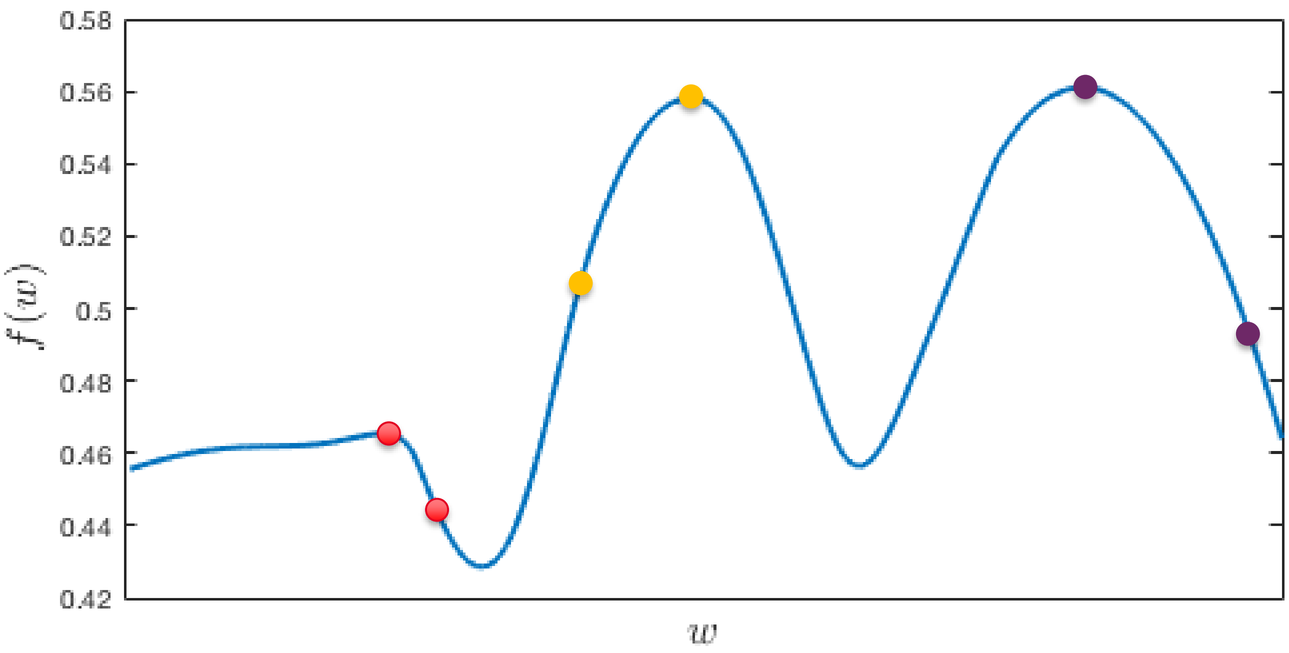
$$\mathbf{w} \leftarrow \mathbf{w} - \mu \nabla_{\mathbf{w}} f(\mathbf{w})$$

The parameter μ is called learning rate. It controls how fast we move towards the maximum (minimum).

If μ is too small, the maximum (or minimum) might not be reached in reasonable time.
If μ is too large, the maximum (minimum) might be missed.



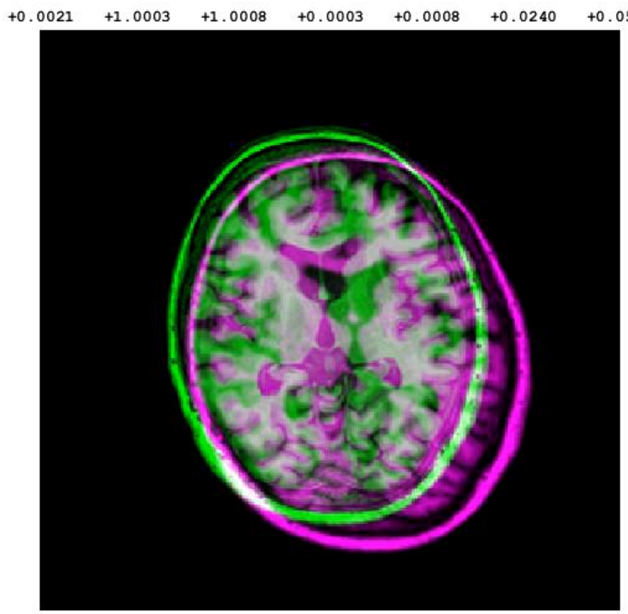
Initialization is important. Different starting points will result in different found maxima (and not always global).

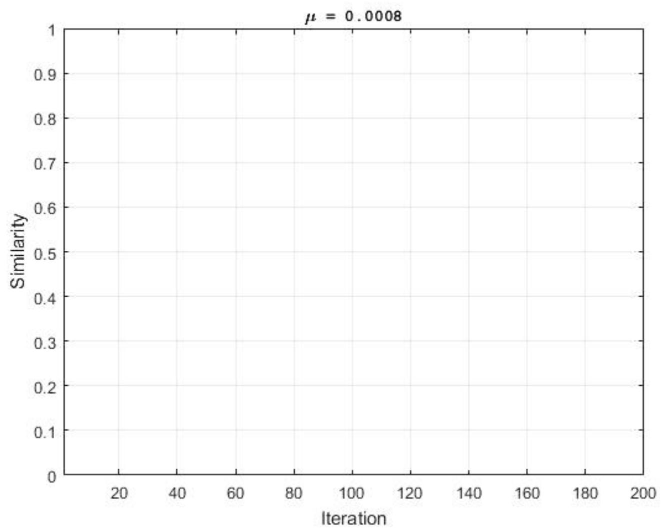


Intensity-based image registration

Continued.

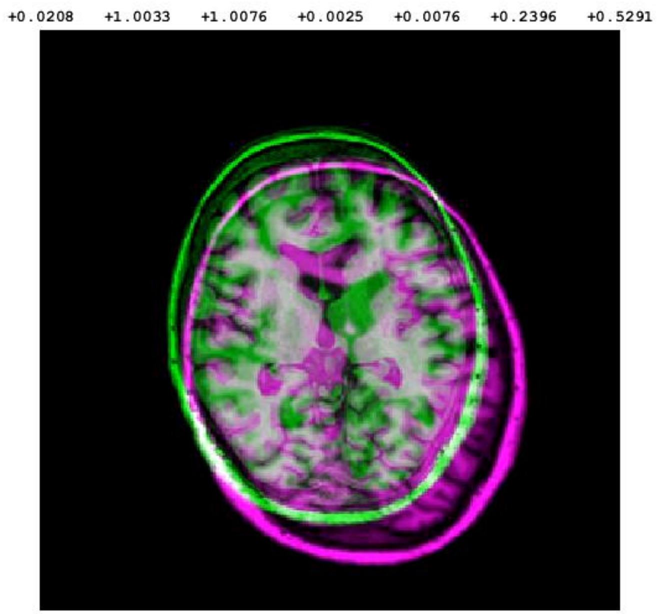
parameters of affine transformation

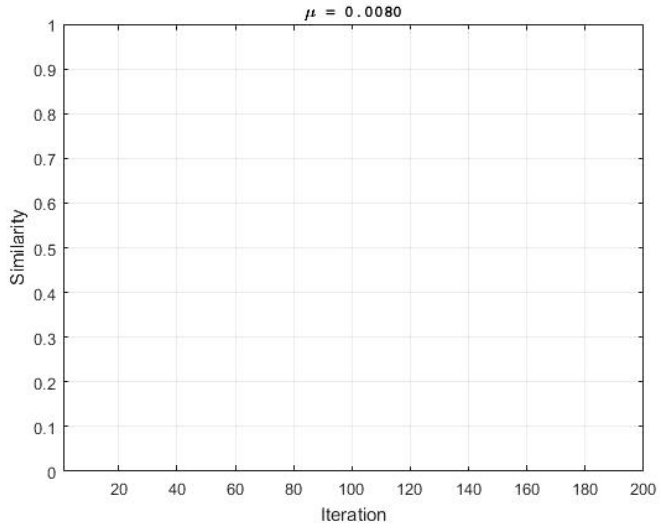




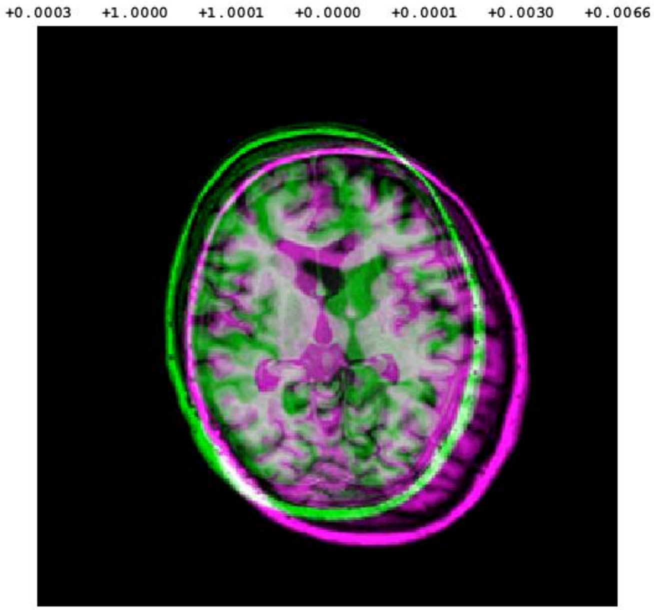
$$\mathbf{w} \leftarrow \mathbf{w} + \mu \nabla_{\mathbf{w}} f(\mathbf{w})$$

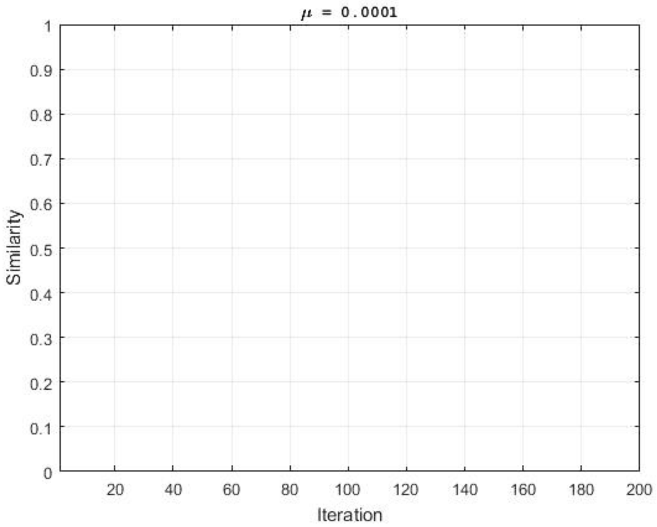
gradient of the similarity: SSD, CC or MI





$$\mathbf{w} \leftarrow \mathbf{w} + \mu \nabla_{\mathbf{w}} f(\mathbf{w})$$





$$\mathbf{w} \leftarrow \mathbf{w} + \mu \nabla_{\mathbf{w}} f(\mathbf{w})$$

Validation (evaluation metrics)

Lecture schedule

Week	Date	Lecturer	Topics
1	1 Sept.	Maureen	Course introduction; Software demo; Image registration (1)
	3 Sept.	Maureen	Image registration (2); Geometrical transformations
2	8 Sept.	Maureen	Point-based registration
	10 Sept.	**Maureen**	Intersity-based registration; **Evaluation metrics**
3	15 Sept.	*Catch-up day (no lecture)*	
	17 Sept.	Cornel Zachiu (UMCU)	Guest lecture 1: Image analysis for adaptive radiotherapy
4	22 Sept.	Mitko	Introduction to CAD; k-NN; Decision trees
	24 Sept.	Mitko	Generalization and overfitting
5	29 Sept.	Mitko	Logistic regression; Neural networks
	1 Oct.	Friso *(no on-campus lecture!)*	Convolutional neural networks *(pre-recorded)*
6	6 Oct.	Friso	Deep learning frameworks and applications
	8 Oct.	Friso	Unsupervised machine learning
7	13 Oct.	Maureen	Deep learning for deformable image registration
	15 Oct.	Geert-Jan Rutten (ETZ)	Guest lecture 2: Image analysis in neurosurgery applications
8	20 Oct	*Self-study (no lecture)*	Active shape models
	22 Oct	*Self-study (no lecture)*	Active shape models

Important characteristics to consider when evaluating medical image analysis methods:

- **Accuracy** = deviation of results from known ground truth.
- **Precision**, **reproducibility, reliability** = extent to which equal or similar input produces equal or similar results.
- **Robustness** characterizes the change of analysis quality if conditions deviate from assumptions made for analysis (e.g., when noise level increases or if object appearance deviates from prior assumptions).
- **Efficiency** = effort necessary to achieve an analysis result.

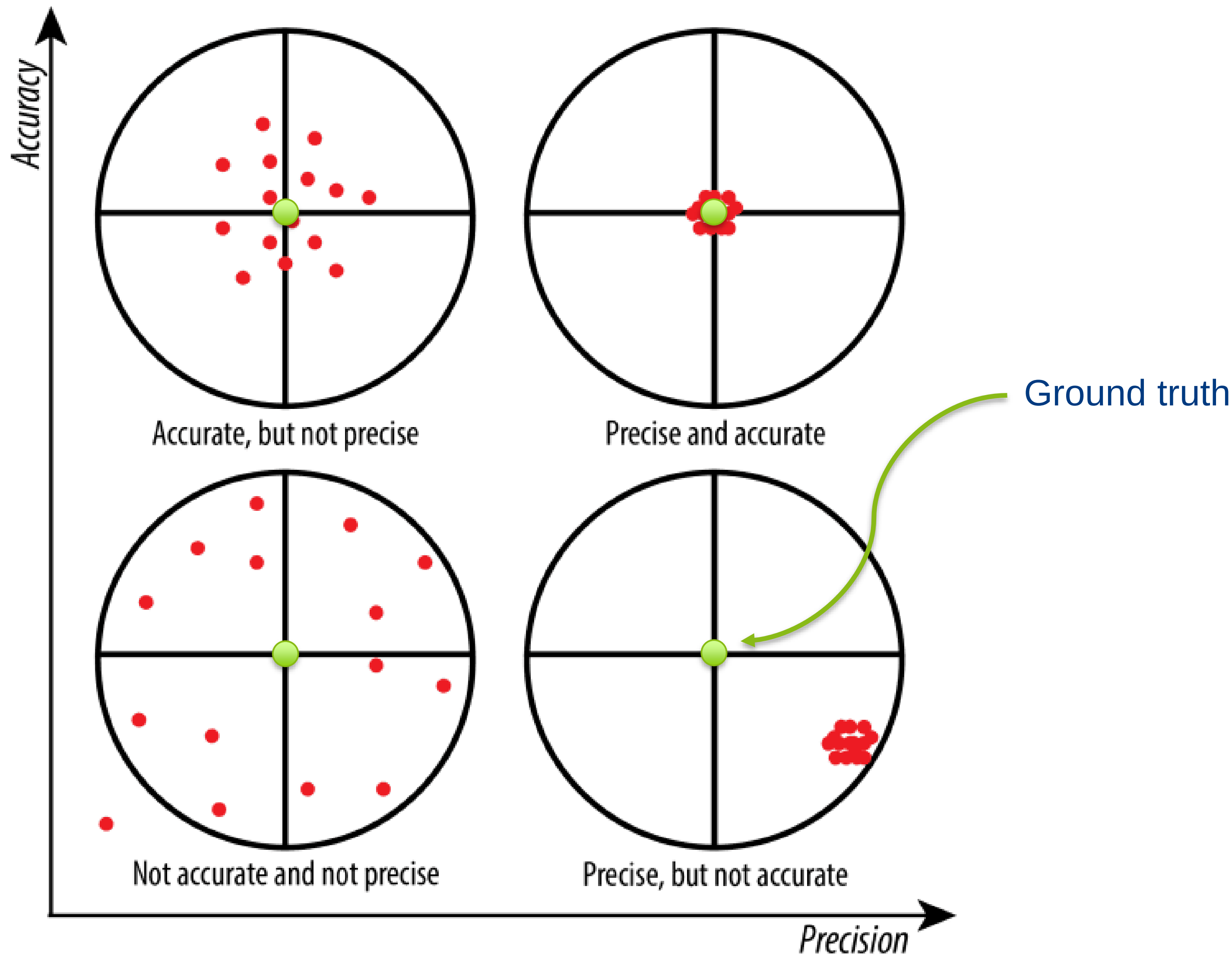


Image source: https://wp.stolaf.edu/it/gis-precision-accuracy/

ground truth = a conceptual term relative to the knowledge of the **truth** concerning a specific question (the “ideal expected result”)

But the goal of medical imaging itself poses an inherent challenge…

“In medical image analysis, the truth is difficult to come by, since the reason for producing images in the first place was to gather information about the human body that cannot be accessed otherwise.”

Technische Universiteit Eindhoven University of Technology

So, how can we get a **ground truth**?

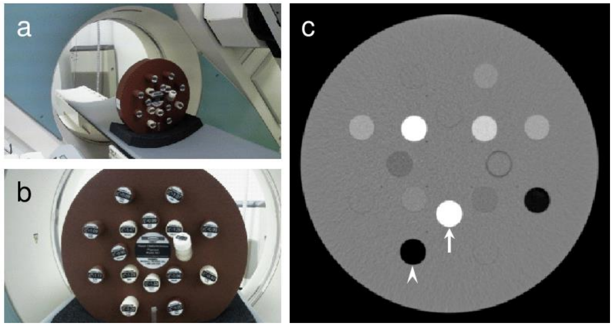
A. Based on *real data*

- Artificial hardware (imaging) phantoms
- Cadaveric material

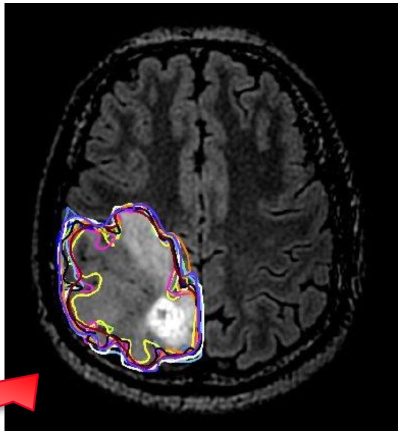
- Other imaging modality
- Other analysis method
- Expert annotations
  - e.g., radiologists, pathologists

(intra- & inter-observer variability?)

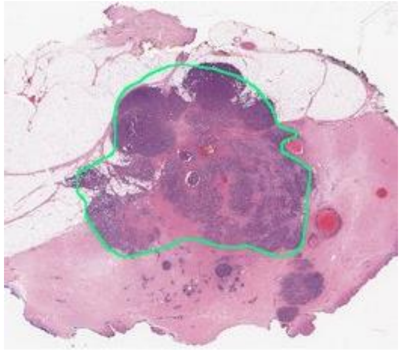
Anthropomorphic head phantom (CT)



Attenuation phantom (CT)

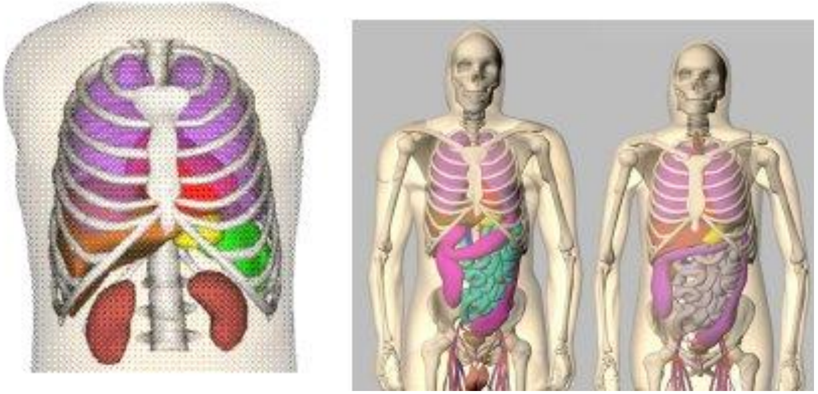


Manual glioma segmentation (MRI)



Tissue segmentation by pathologist

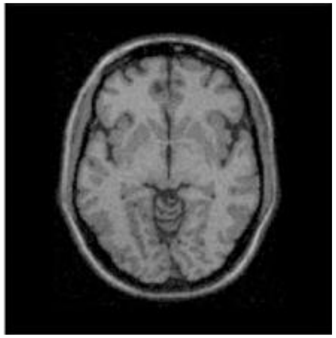
So, how can we get a **ground truth**?



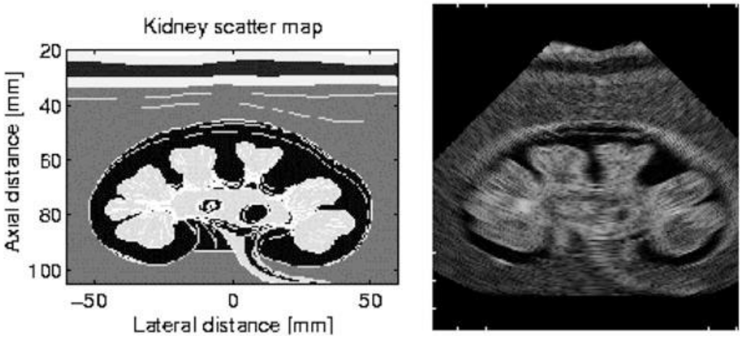
4D XCAT phantom

B. Based on *simulated data*

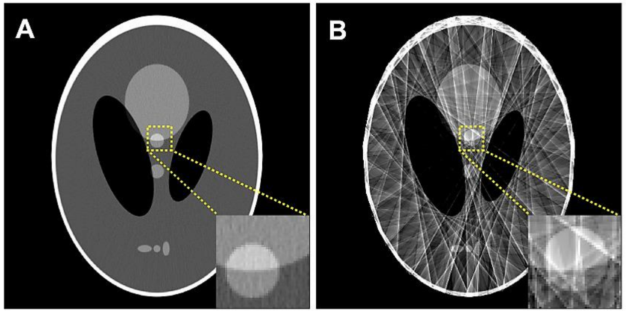
- Software phantom
  - E.g., XCAT phantom for PET validation, BrainWeb phantom, ultrasound phantom
- Mathematical simulations
  - E.g., Shepp-logan phantom



BrainWeb phantom (MRI)



Ultrasound phantom (Jensen and Svendsen 1992, 1996)



Shepp-logan (CT)

Example: evaluation of a segmentation task

- Which approach do we choose?

- Compare to ground truth:

  score = evaluate_segmentation(segmentation, ground_truth)

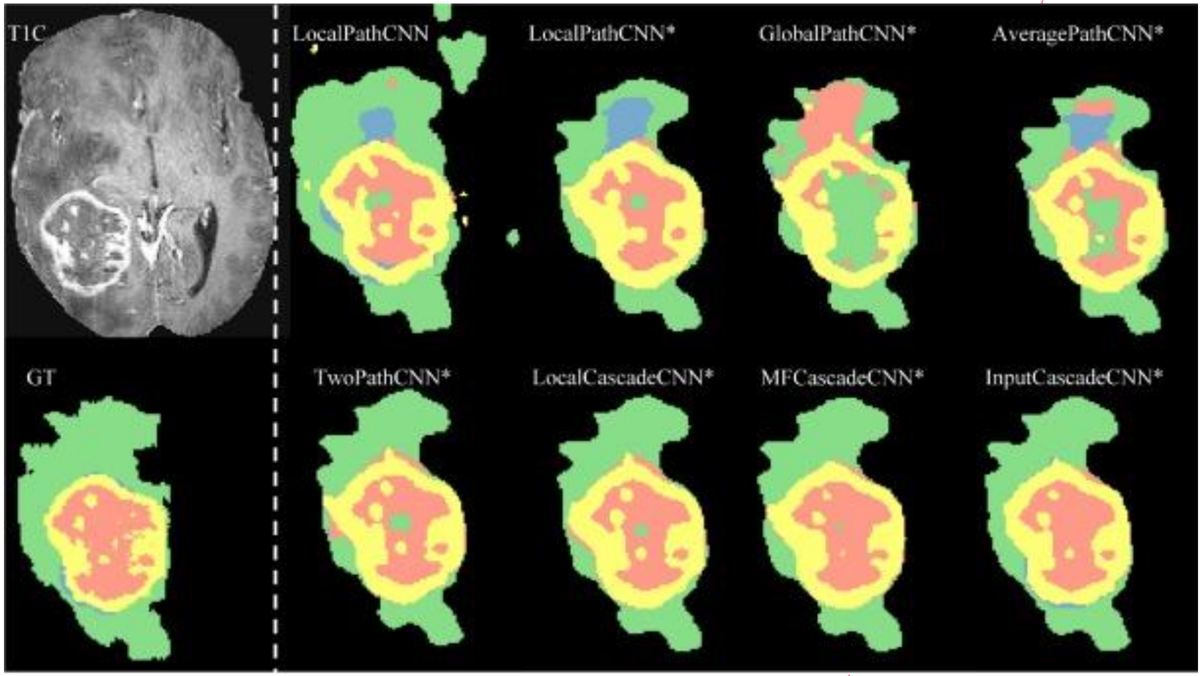
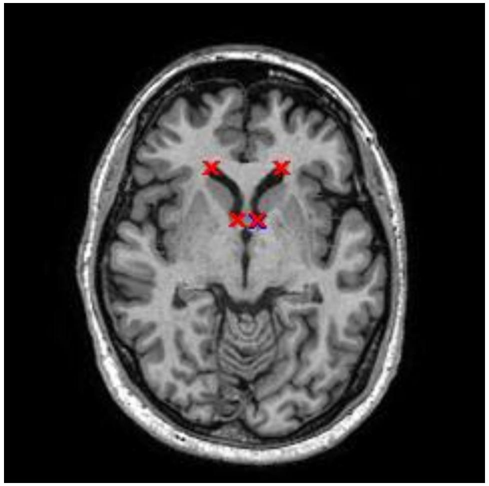


Image https://www.sciencedirect.com/science/article/pii/S1361841516300330

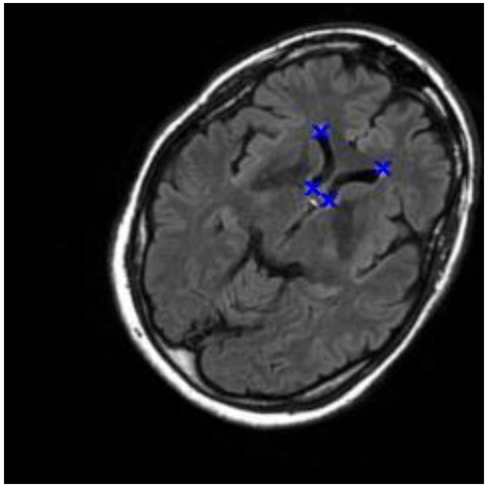
Question:

How can we measure the quality of a **registration** task?

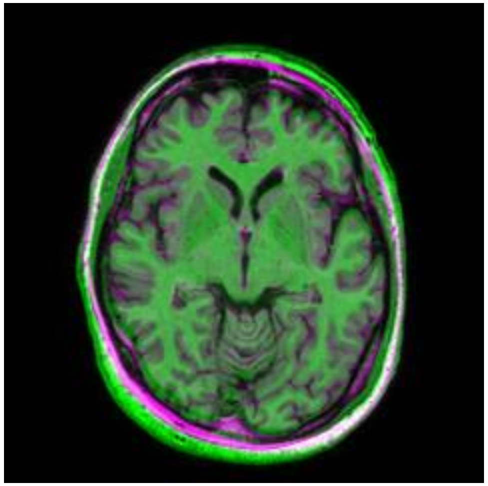
Remember this example:



Fixed



Moving

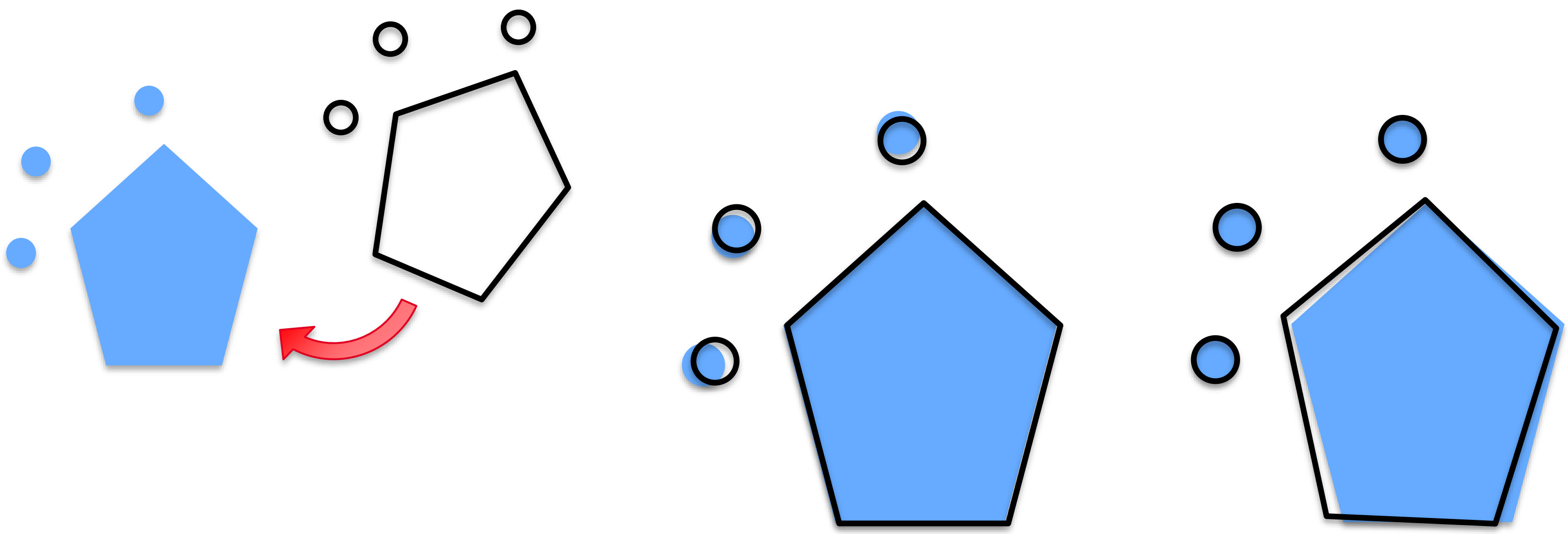


Registered

Measure registration performance by computing the registration error for some target point pairs. These target points must be different from the points used to compute the transformation! Why?

The **optimal fit** for the fiducial markers does not automatically mean that the registration itself is optimal, especially if:

- the markers are far away from the object to be registered
- too few markers are used
- it is difficult to localize the markers.

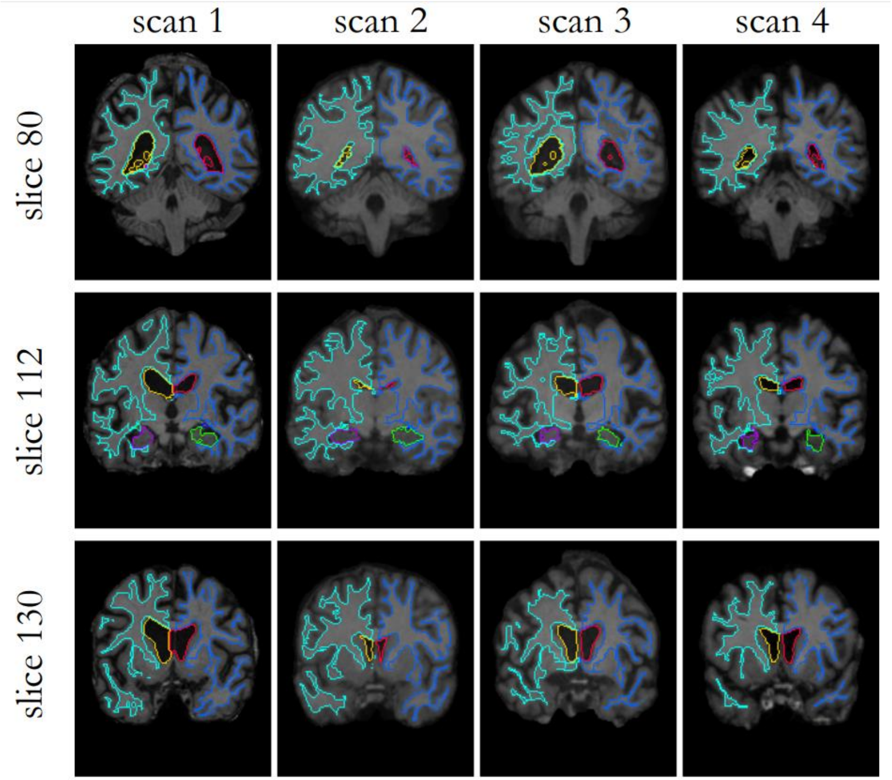


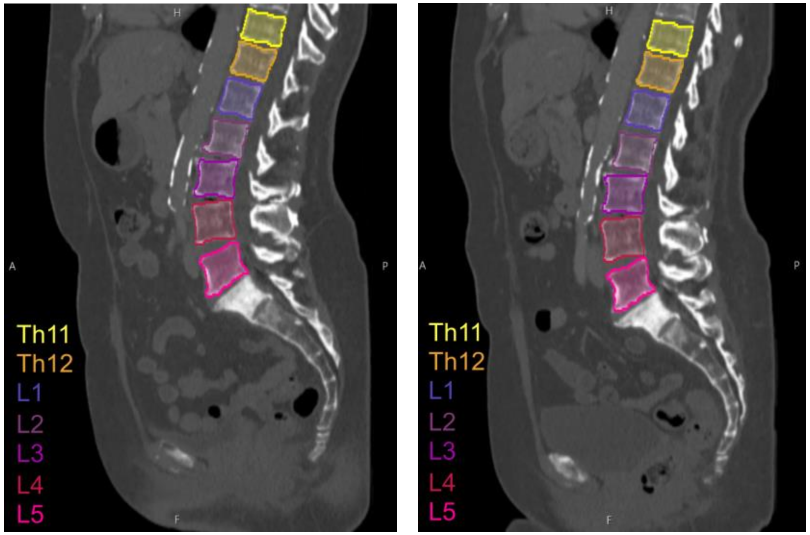
Optimal registration

Optimal fit fiducial markers

How can we measure the quality of a **registration** task?

Idea: apply the transformation to a segmentation mask that represents the anatomy of interest.





Balakrishnan, Guha, et al. "Voxelmorph: a learning framework for deformable medical image registration." *IEEE transactions on medical imaging* 38.8 (2019): 1788-1800.

Evaluation metrics

Many metrics available, we look at:

- Accuracy
- Dice score
- Hausdorff distance

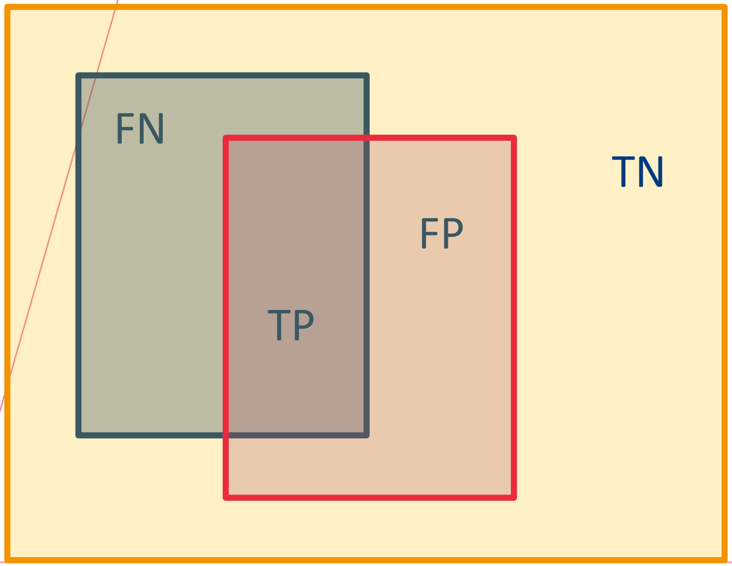
Accuracy

"How many pixels are correct?"

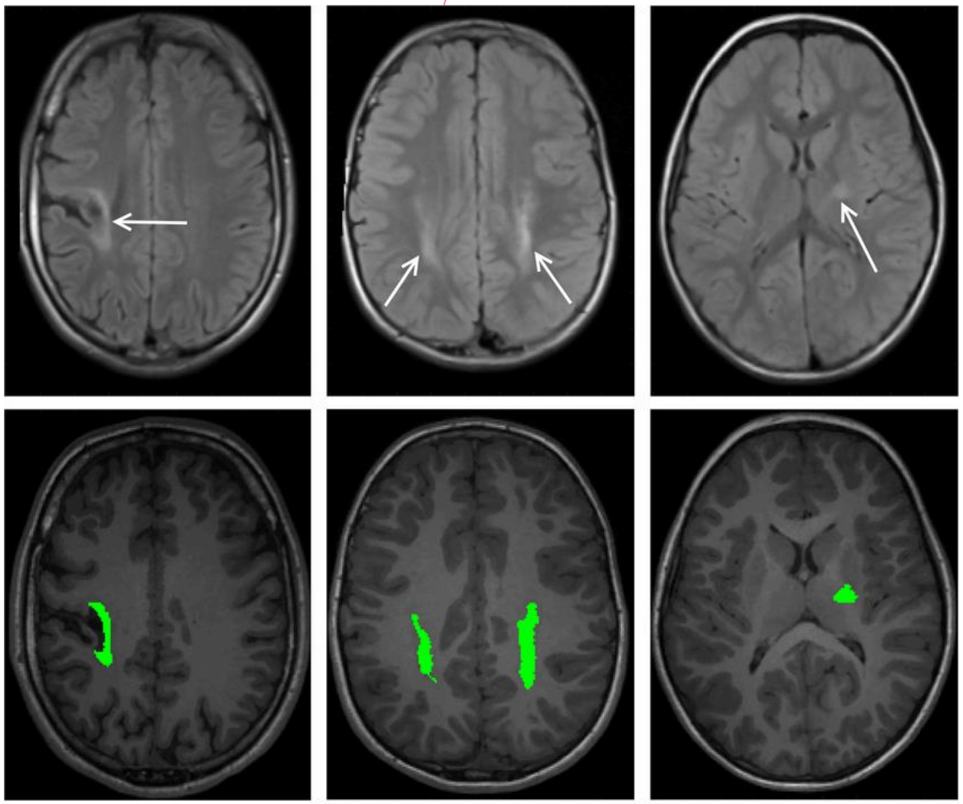
Accuracy = (TP + TN) / (TP+FP+FN+TN)

- TP = True Positive
- FP = False Positive
- FN = False Negative
- TN = True Negative

Orange = whole image
Blue = ground truth
Red= segmentation result



What if the ground truth is small?



(a) (b) (c)

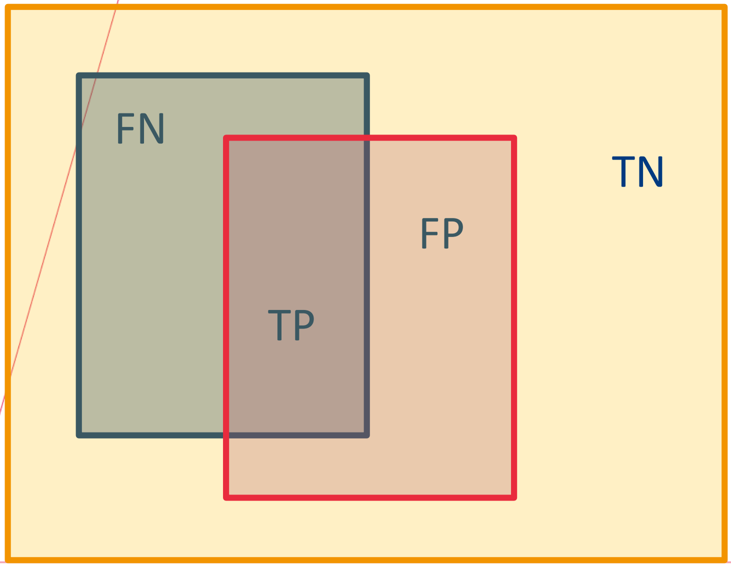
Dice score

Measure overlap excluding TN

Sørensen–Dice index a.k.a. Dice Similarity Coefficient (DSC)

DSC = 2TP / (2TP + FP + FN)

Orange = whole image
Blue = ground truth
Red= segmentation result



Dice score

Two equivalent definitions

DSC = **2**TP / (**2**TP + FP + FN)

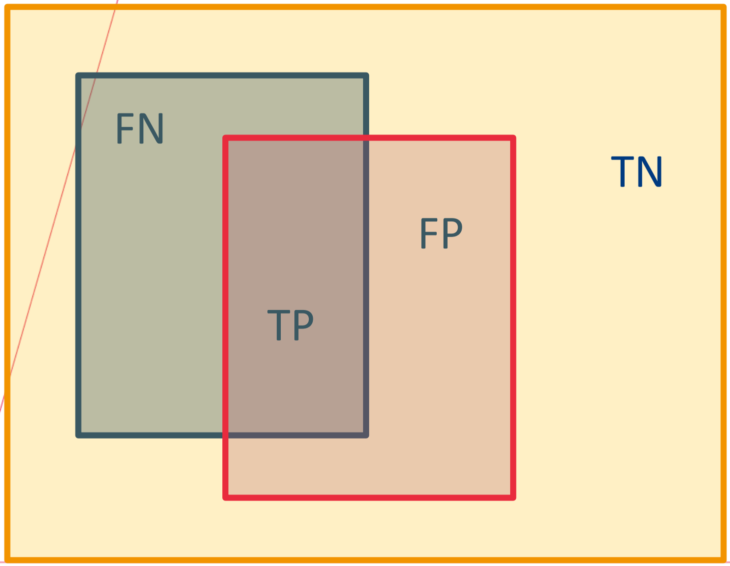
DSC = 2 | A * B | / (| A | + | B |) for binary images A and B

| A | = size of blue ground truth

| B | = size of red result

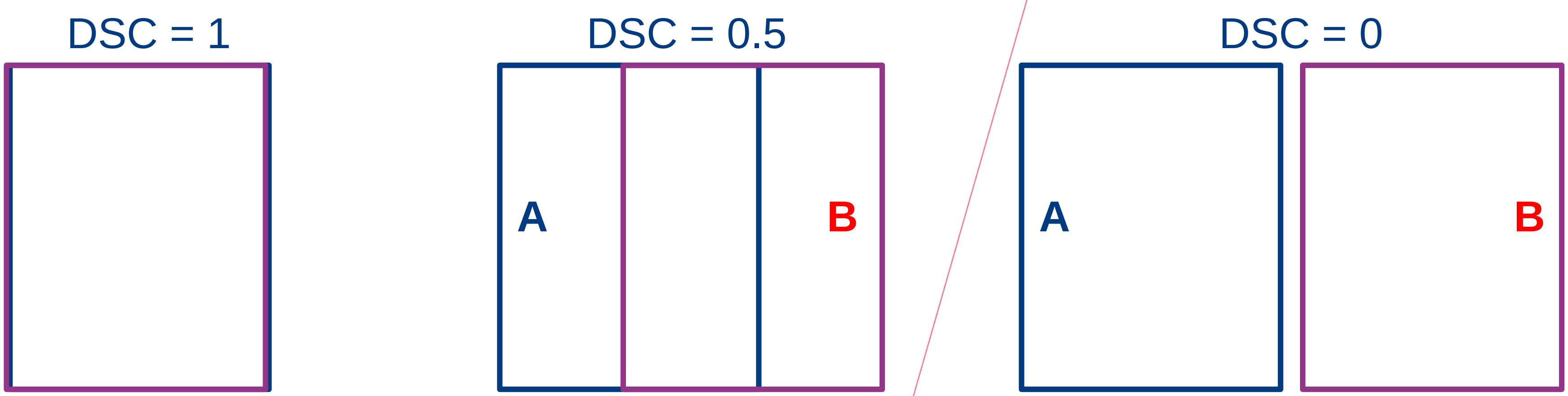
| A * B| = size of overlap

Orange = whole image
Blue = ground truth
Red= segmentation result



Dice score

- Between 1 and 0 for full / no overlap



Other relevant evaluation metrics based on TP, FP, etc. are the **sensitivity** and **specificity.**

- Sensitivity = TP / TP + FN
- Specificity = TN / TN + FP

Reality / Ground truth

Measured

	✓	✗
✓	True Positives (TP)	False Positives (FP) “Type I error”
✗	False Negatives (FN) “Type II error”	True Negatives (TN)

NB: These metrics are commonly used in **detection** tasks involving medical images. Interestingly, they are also very important when interpreting the performance of any test (e.g., airport security, breast cancer screening, quality assurance in companies, …)

Radiology

Original Research
Thoracic Imaging

Free Access

Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases

Tao Ai*, Zhenlu Yang*, Hongyan Hou, Chenao Zhan, Chong Chen, Wenzhi Lv, Qian Tao, Ziyong Sun, Liming Xia

* T.A. and Z.Y. contributed equally to this work.

Author Affiliations

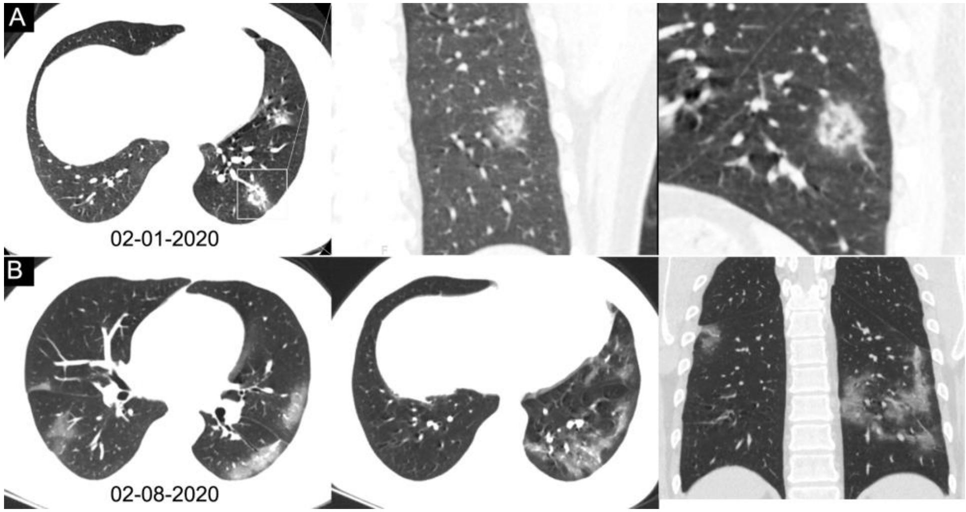
Published Online: Feb 26 2020 | https://doi.org/10.1148/radiol.2020200642

Ai T, Yang Z, Hou H, et al. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology*. 2020;296(2):E32-E40.

Results

“Of the 1014 patients, 601 of 1014 (59%) had positive RT-PCR results and 888 of 1014 (88%) had positive chest CT scans. The sensitivity of chest CT in suggesting COVID-19 was 97% (95% confidence interval: 95%, 98%; 580 of 601 patients) based on positive RT-PCR results.

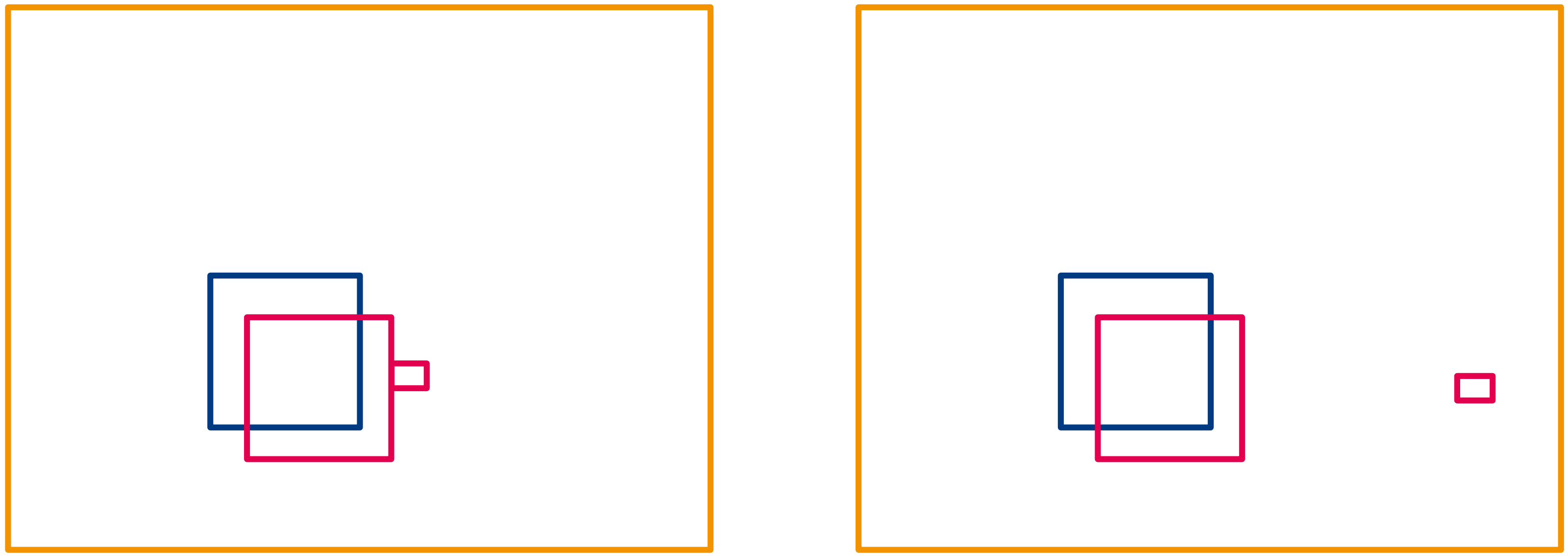
In patients with negative RT-PCR results, 75% (308/413) had positive chest CT findings; of 308, 48% were considered as highly likely cases, with 33% as probable cases.”



What about the specificity?

Ai T, Yang Z, Hou H, et al. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology*. 2020;296(2):E32-E40.

Dice (and other metrics based on TP, FP, etc.) are not sensitive to location



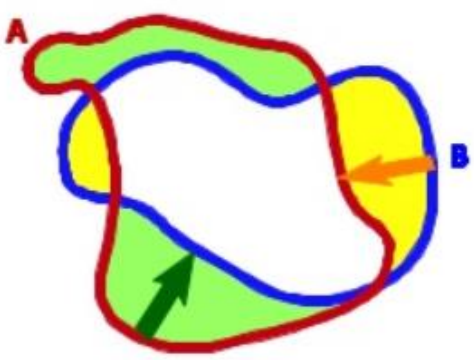
Hausdorff distance

Compare sets of points on the boundaries

Hausdorff distance = maximum shortest distance between the boundary points

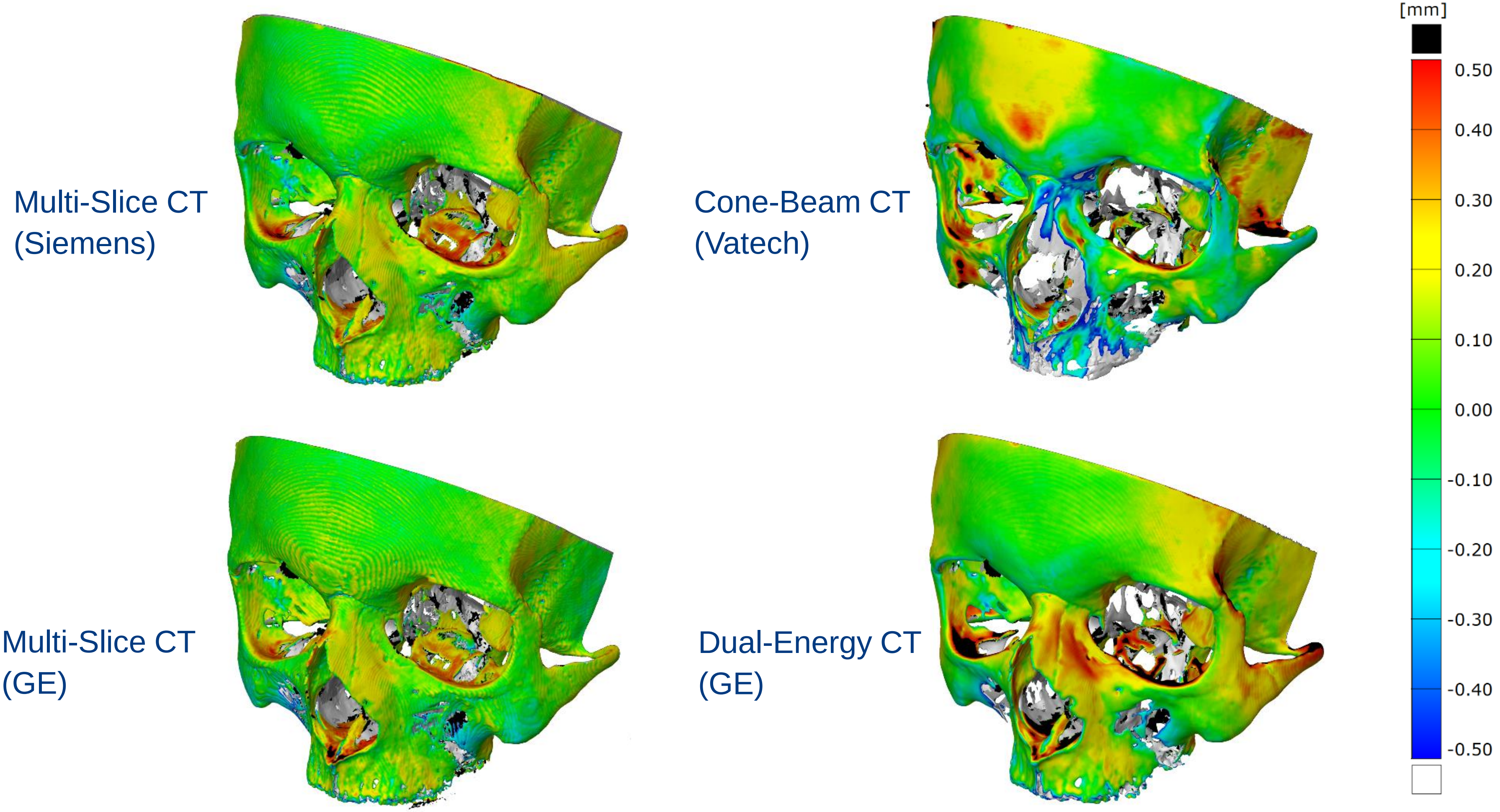
$$\mathrm{h}(\mathrm{A},\mathrm{B}) = \max_{a \in A} \min_{b \in B} d(a,b)$$

$$H(\mathrm{A},\mathrm{B}) = \max(h(A,B), h(B,A))$$



https://www.slideshare.net/UlaBac/lec14-evaluation-framework-for-medical-image-segmentation

Measuring and/or visualizing surface distances can give fascinating insights into the accuracy of a medical image analysis task:



van Eijnatten, M., Berger, F. H., de Graaf, P., Koivisto, J., Forouzanfar, T., & Wolff, J. (2017). Influence of CT parameters on STL model accuracy. Rapid Prototyping Journal, 23(4), 678-685.

Further reading:

- Guide to Medical Image Analysis - Methods and Algorithms
  https://link.springer.com/book/10.1007/978-1-4471-2751-2