MANCHESTER 1824

The University of Manchester

Assessing the Accuracy of NRR with and without Ground Truth

Roy Schestowitz, Bill Crum, Vlad Petrovic, Carole Twining, Tim Cootes, Chris Taylor

… including

Generalised Overlap Measures

William R. Crum, Oscar Camara, Daniel Rueckert,
Kanwal K. Bhatia, Mark Jenkinson, and Derek L.G. Hill

Overview

- Background and motivation
- Assessment methods
  - overlap-based
  - model-based
- Experiments
  - validation
  - comparison of methods
- Conclusions

Non-rigid Registration (NRR)

- Alignment of image sets
  - dense correspondence
  - alignment of anatomical structures
- Alignment established by
  - image warping
  - comparison with other image(s)
  - maximising similarity
- Competing NRR algorithms produce different results

Motivation for Assessment

- Different methods for NRR
  - representation of warp (including regularisation)
  - similarity measure
  - optimisation
  - pair-wise vs group-wise
- Limitations of current methods of assessment
  - ground-truth deformations
  - binary overlap measures

Two New Approaches

- Generalised overlap
  - multiple labels
  - label interpolation
  - multiple images
- Model-based
  - NRR $\Rightarrow$ combined appearance model
  - good registration $\Rightarrow$ good model

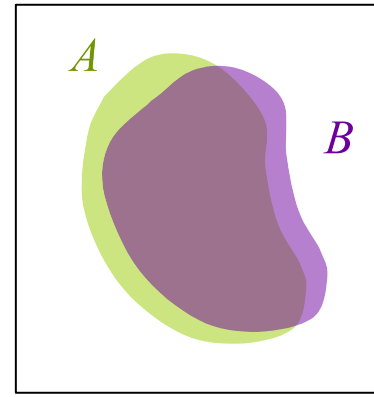
Generalised Overlap

Overlap Measures

- Existing overlap measures
  - assume binary labels
  - evaluate one label at a time
  - cannot easily be applied to groupwise registration
- In practice
  - labels may be interpolated (pv) or fuzzy
  - there may be lots of labels
  - there may be lots of images
- Generalise existing overlap measures

Binary Overlap Measures

- Consider label regions A and B
- Tanimoto/Jacaard overlap



$$O_P = \frac{N(A \quad B)}{N(A \quad B)} = \frac{\text{Number of voxels in } A \text{ AND } B}{\text{Number of voxels in } A \text{ OR } B}$$

- Dice overlap

$$O_Q = \frac{2N(A \quad B)}{N(A) + N(B)} = \frac{\text{Number of voxels in } A \text{ AND } B}{\text{Mean number of voxels in } A \text{ and } B}$$

Alternate Form

- Binary value at each voxel A_i and B_i

```
00000000000000000000
00000000000111110000
00000000011111100000
00000011111111000000
00011111111110000000
00001111111111000000

00000111111111100000
00000111111111000000
00000111111110000000
00000011111100000000
00000000000000000000
```

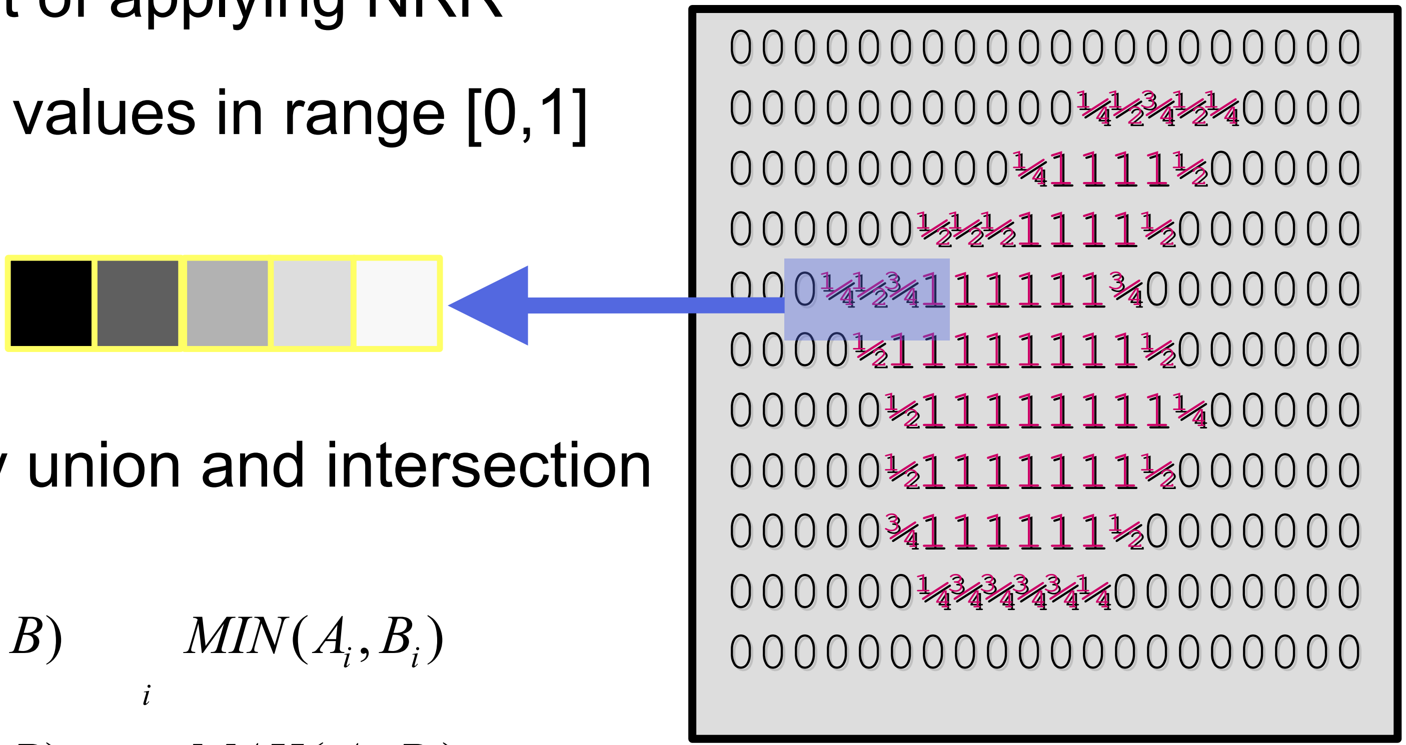
Region A

$$N(A = B) \qquad {}_{i} \; MIN(A_i, B_i)$$

$$N(A = B) \qquad {}_{i} \; MAX(A_i, B_i)$$

Interpolated Label Images

- Result of applying NRR
- Label values in range [0,1]



- Fuzzy union and intersection

$$N(A= B) \quad {}_{i} \quad MIN(A_i, B_i)$$

$$N(A= B) \quad {}_{i} \quad MAX(A_i, B_i)$$

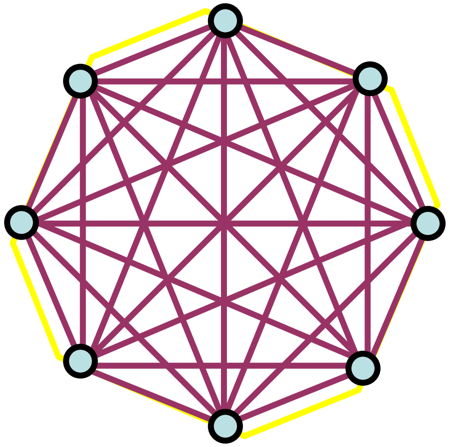
Generalised Overlap

- Fractional overlap

$$O_F = \frac{\underset{voxels,i}{} MIN(A_i, B_i)}{\underset{voxels,i}{} MAX(A_i, B_i)}$$

- Accumulated over labels and image pairs

$$O_{PMF} = \frac{\underset{pairs,k}{} \; \underset{labels,l}{} \alpha_l \; \underset{voxels,i}{} MIN(A_{kli}, B_{kli})}{\underset{pairs,k}{} \; \underset{labels,l}{} \alpha_l \; \underset{voxels,i}{} MAX(A_{kli}, B_{kli})}$$



Label Weighting

- Implicit volume weighting

$$\alpha = 1$$

- Equal weighting

$$\alpha = \frac{1}{V_l}$$

- Inverse volume weighting

$$\alpha = \frac{1}{\left(V_l\right)^2}$$

- Label complexity

$$\alpha = \underset{label\,voxels}{\quad} \left| \quad (\text{Intensity}) \right|$$

MANCHESTER
1824

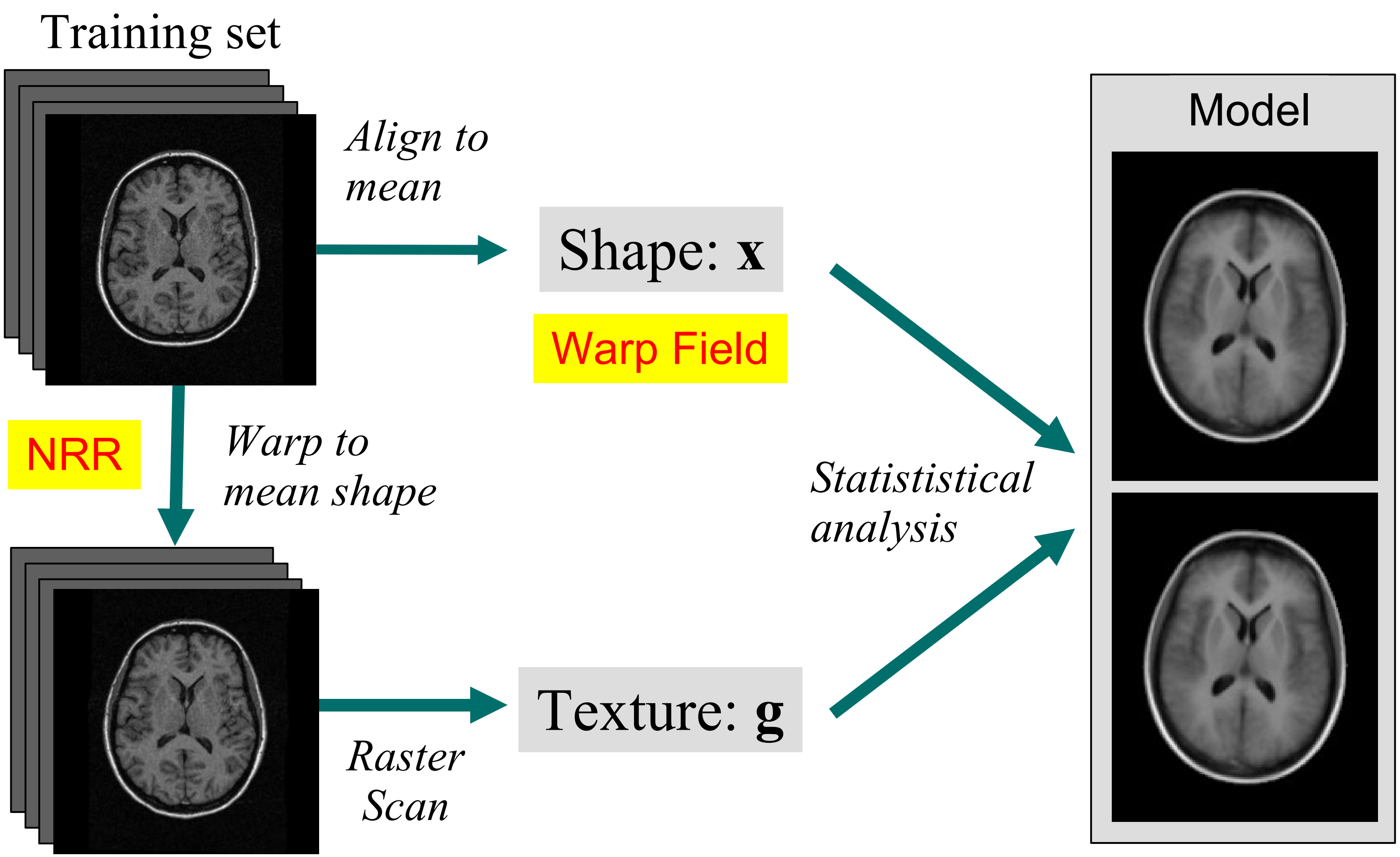
The University
of Manchester

Model-Based Assessment

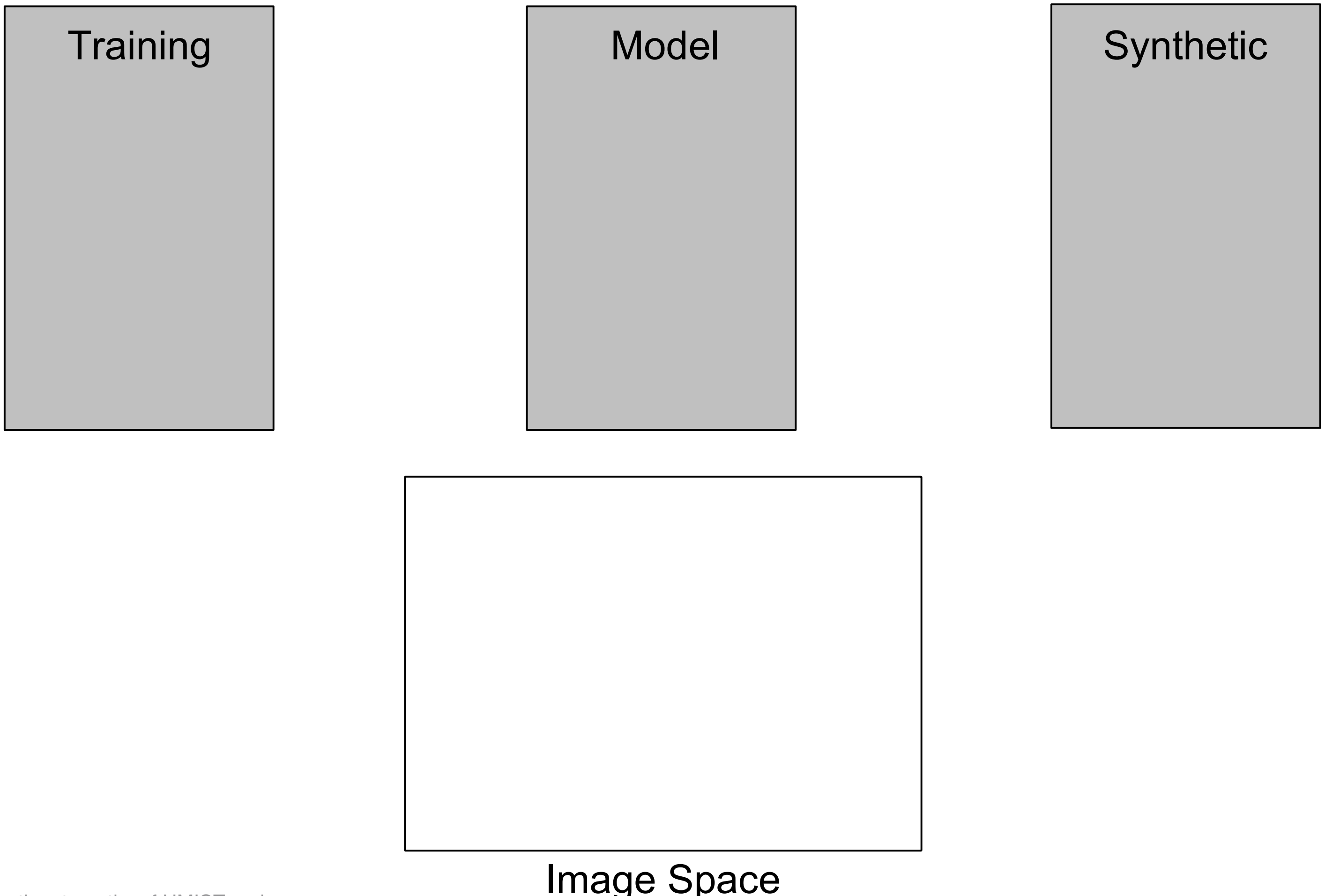
Model-based Framework

- Registered image set $\Rightarrow$ statistical appearance model
- Good registration $\Rightarrow$ good model
  - generalises well to new examples
  - specific to class of images
- Registration quality $\Leftrightarrow$ Model quality
  - problem transformed to defining model quality
  - ground-truth-free assessment of NRR

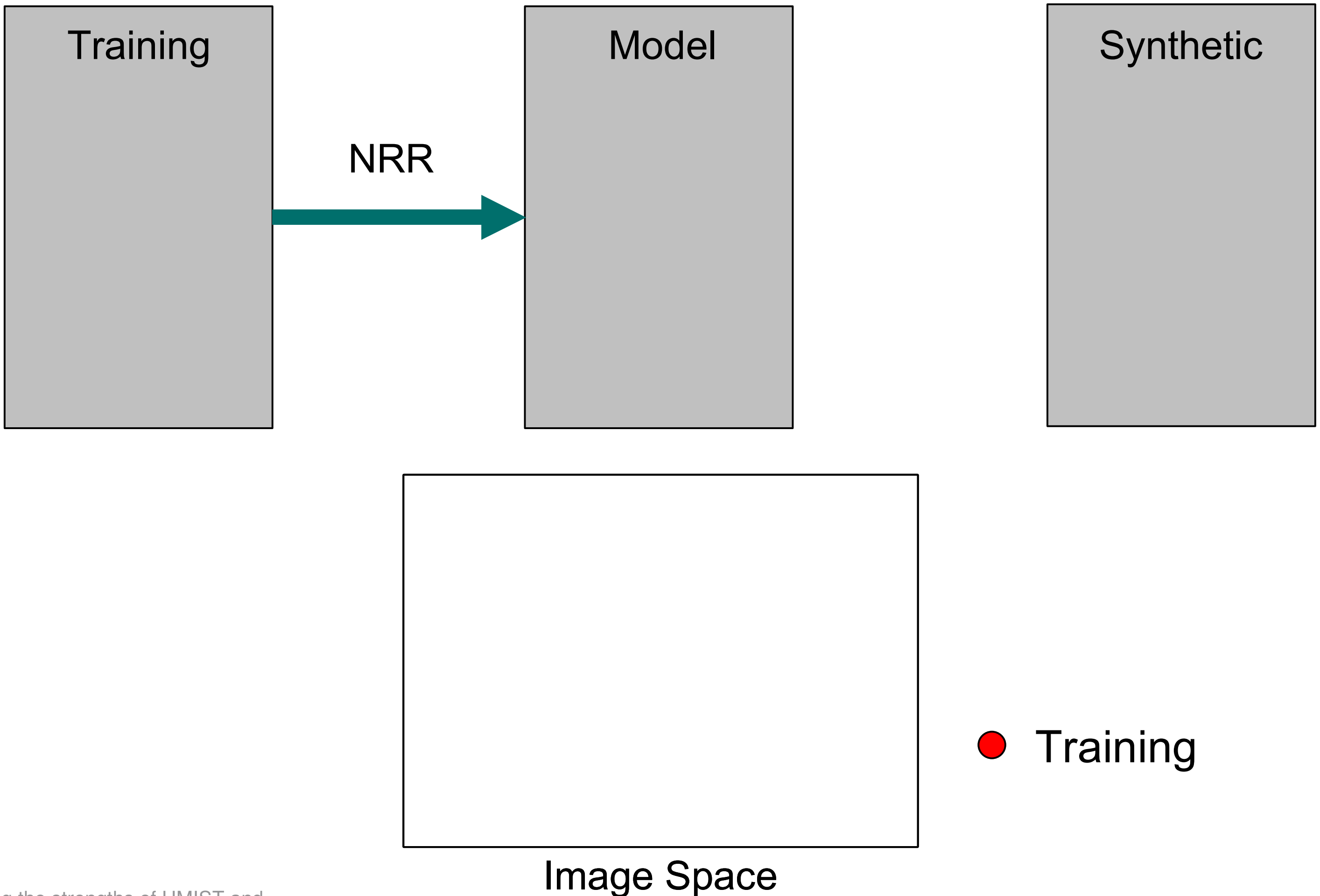
Building an Appearance Model



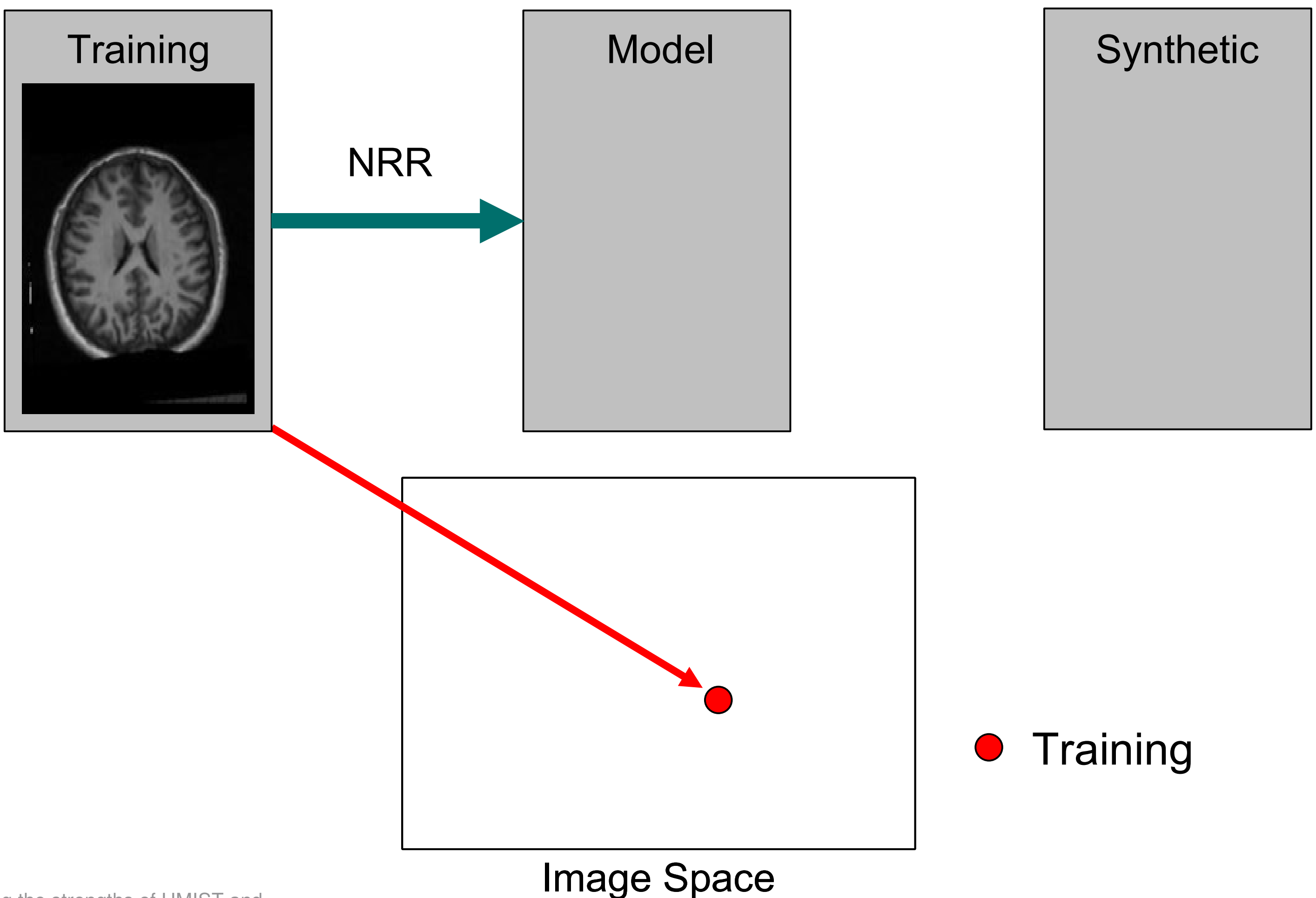
Training and Synthetic Images



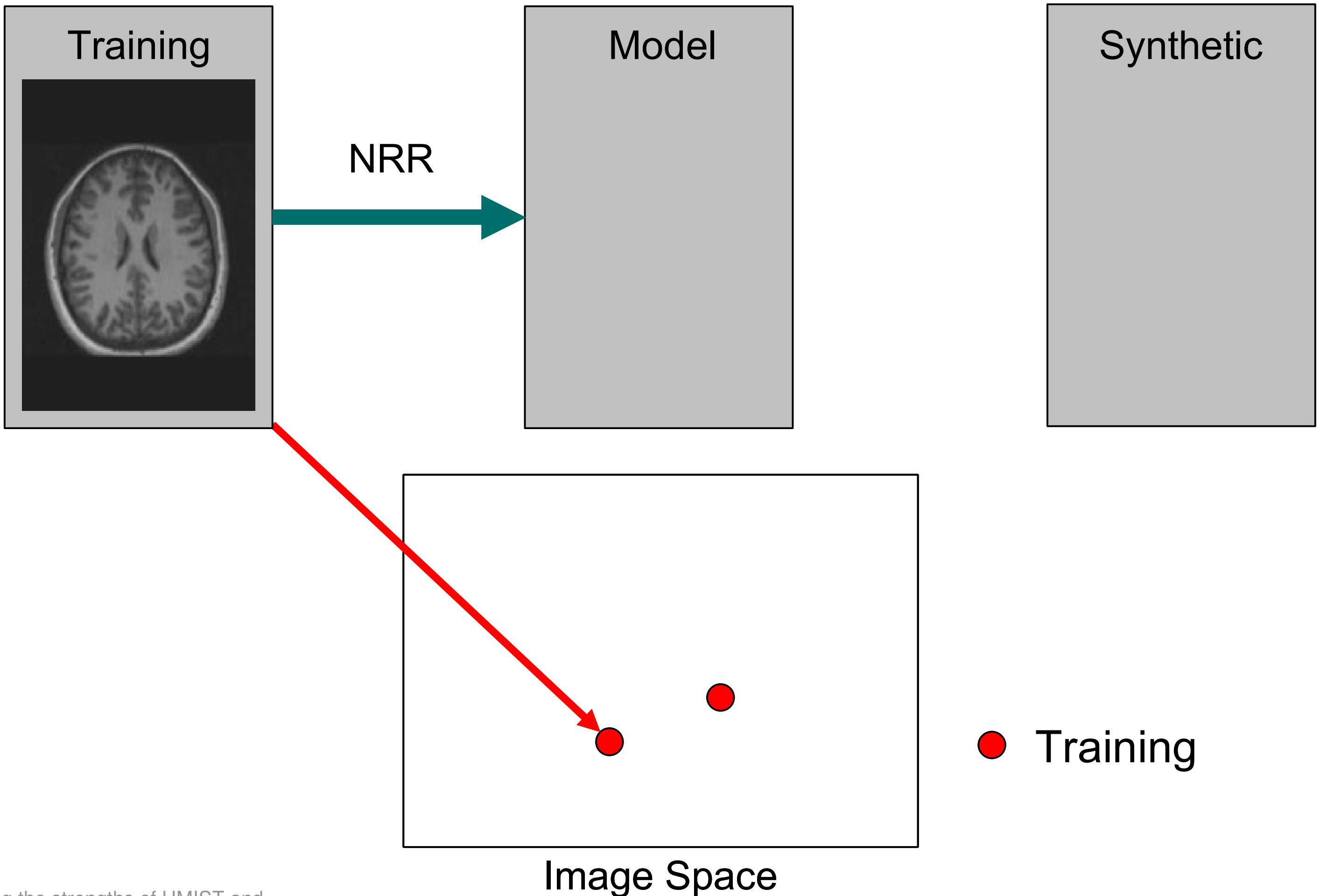
Training and Synthetic Images



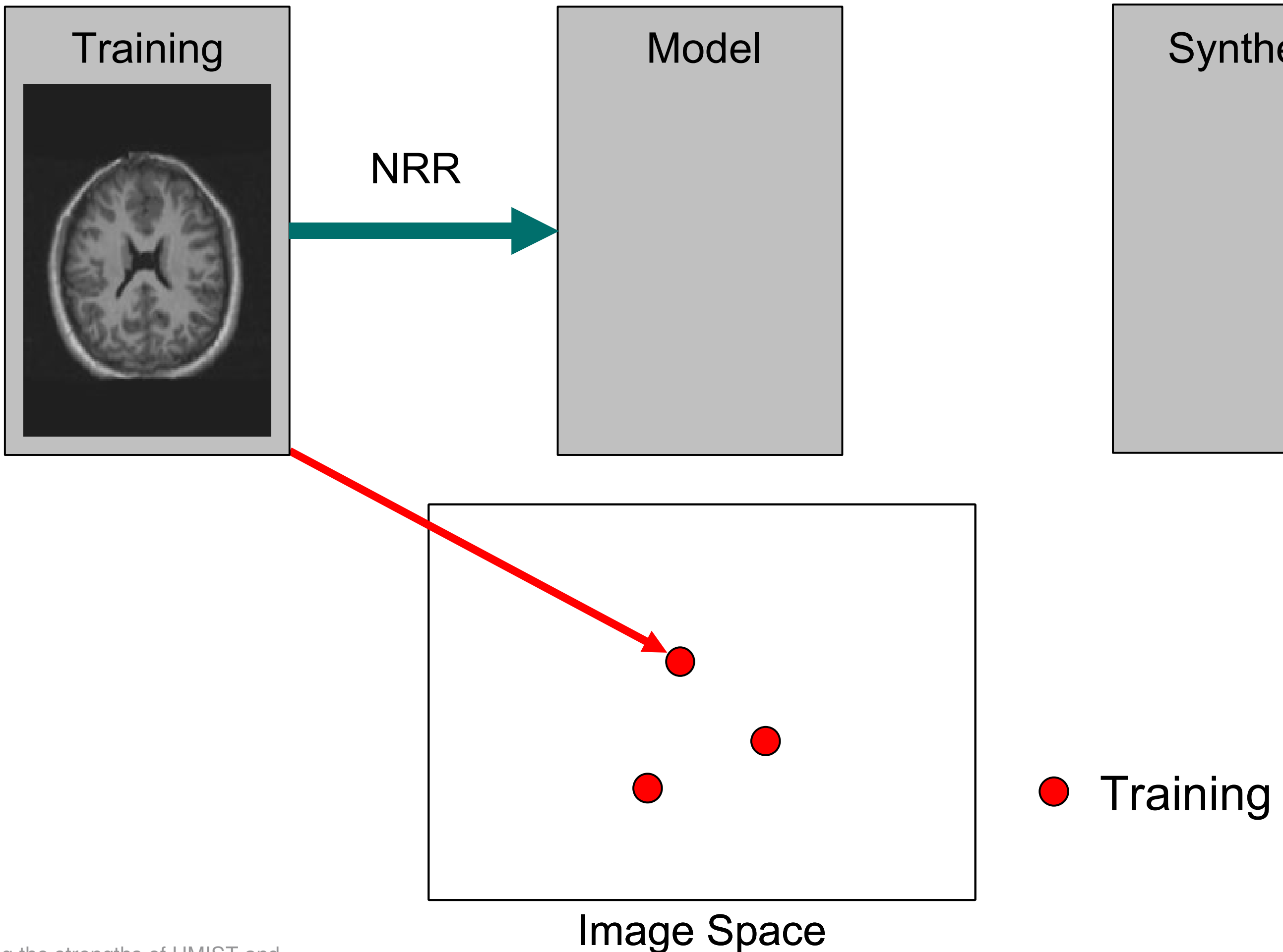
Training and Synthetic Images



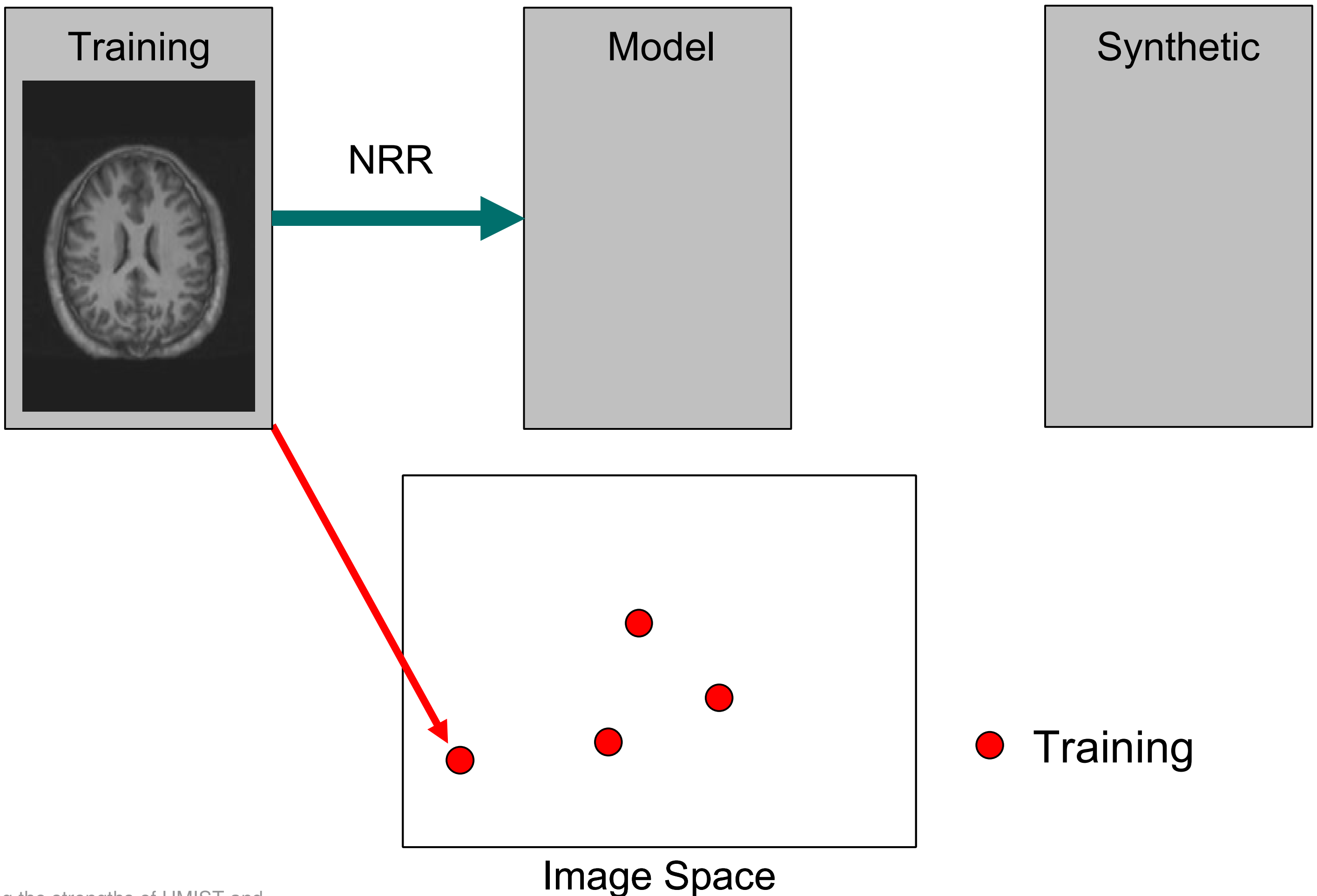
Training and Synthetic Images



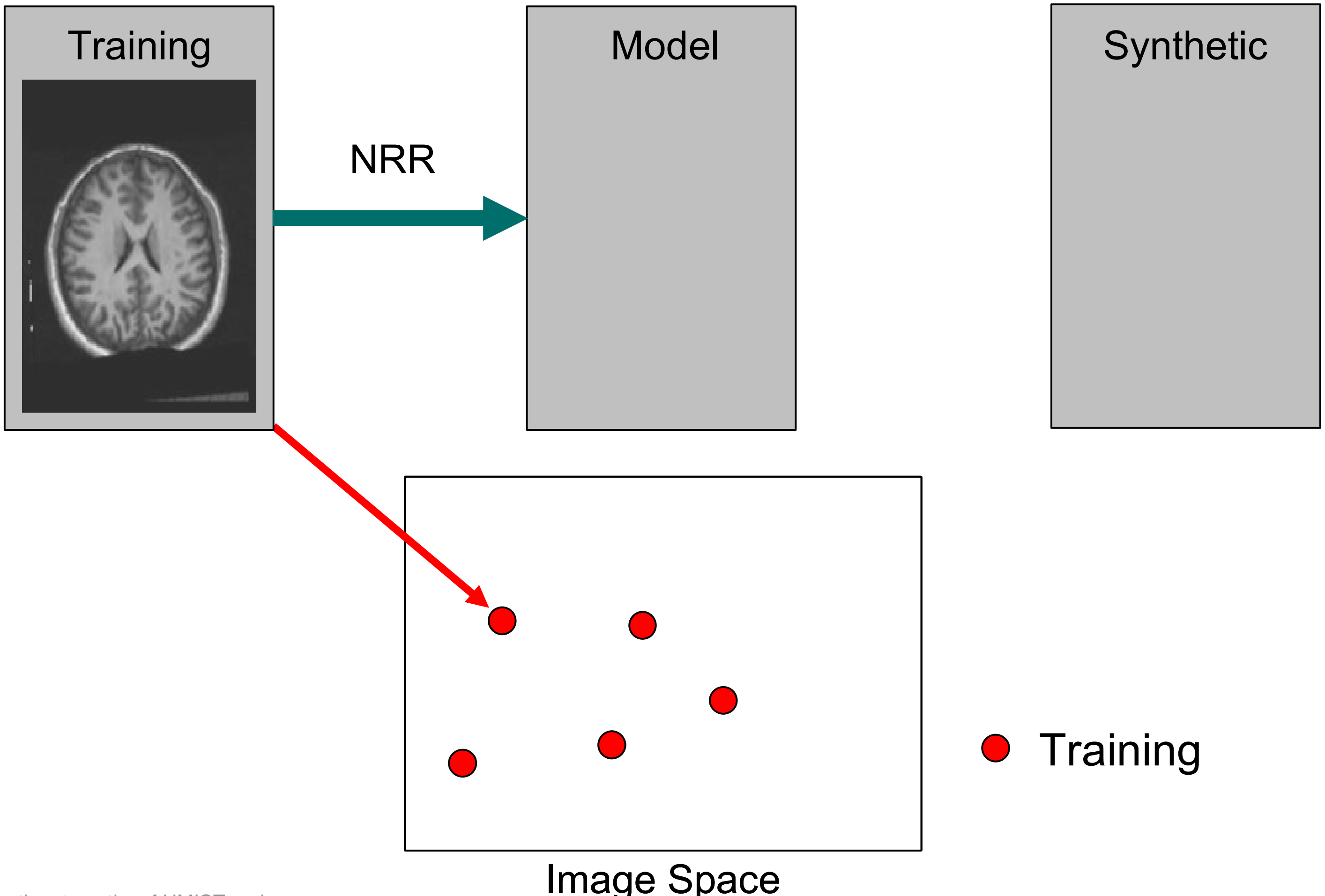
Training and Synthetic Images



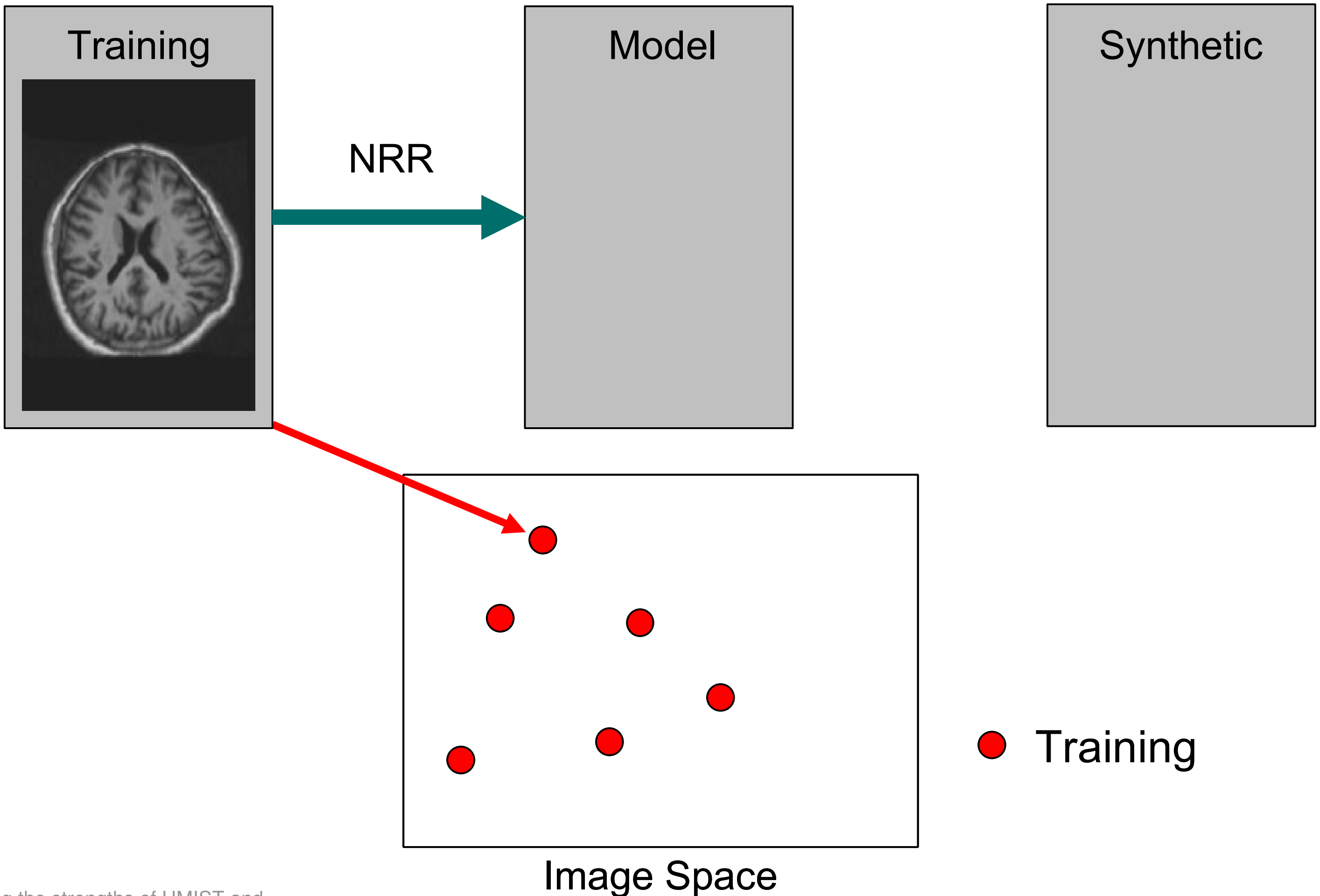
Training and Synthetic Images



Training and Synthetic Images



Training and Synthetic Images



Training and Synthetic Images

Training

Model

Synthetic

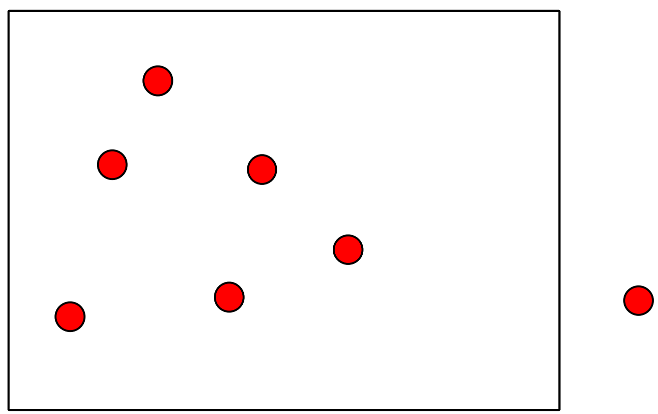
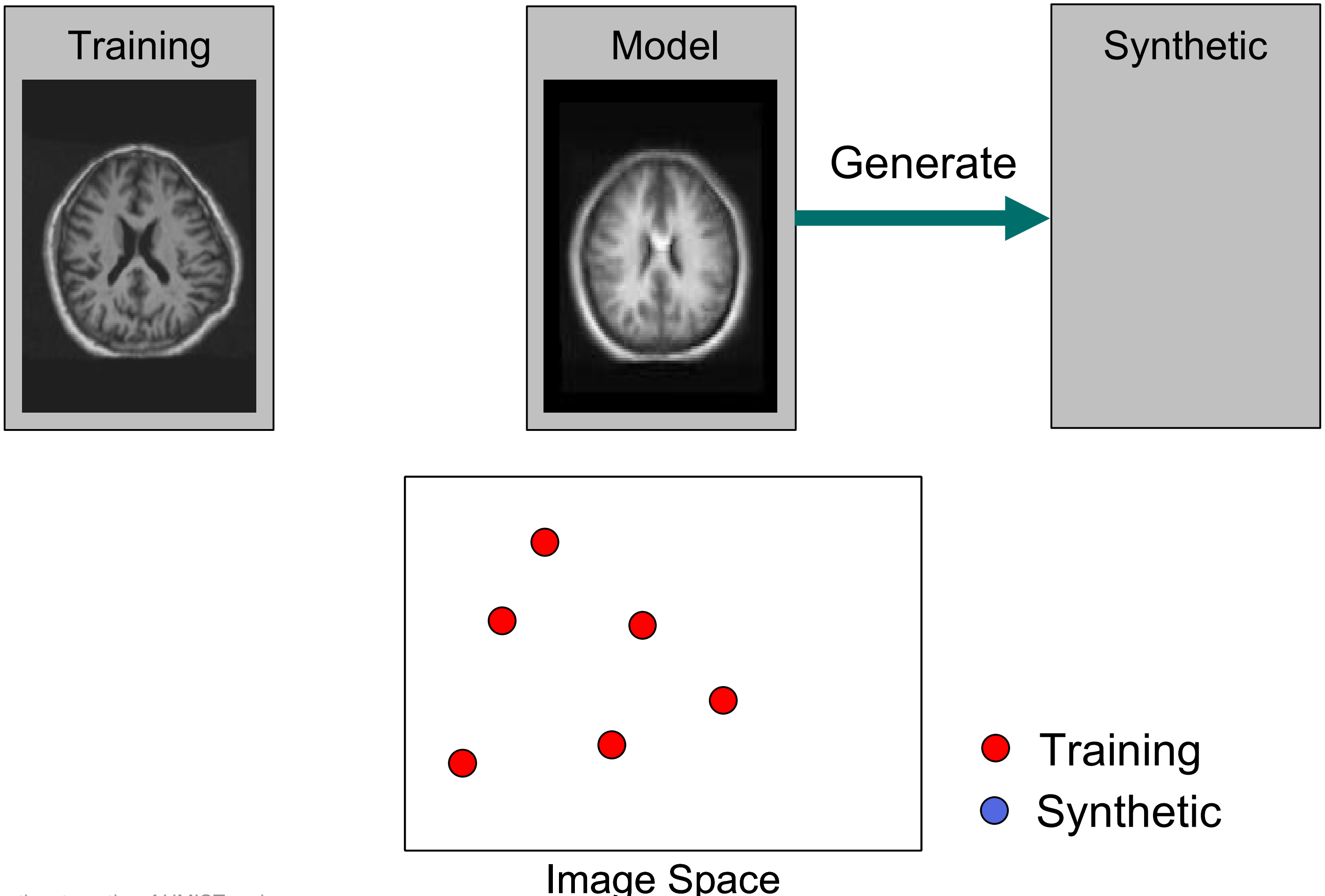
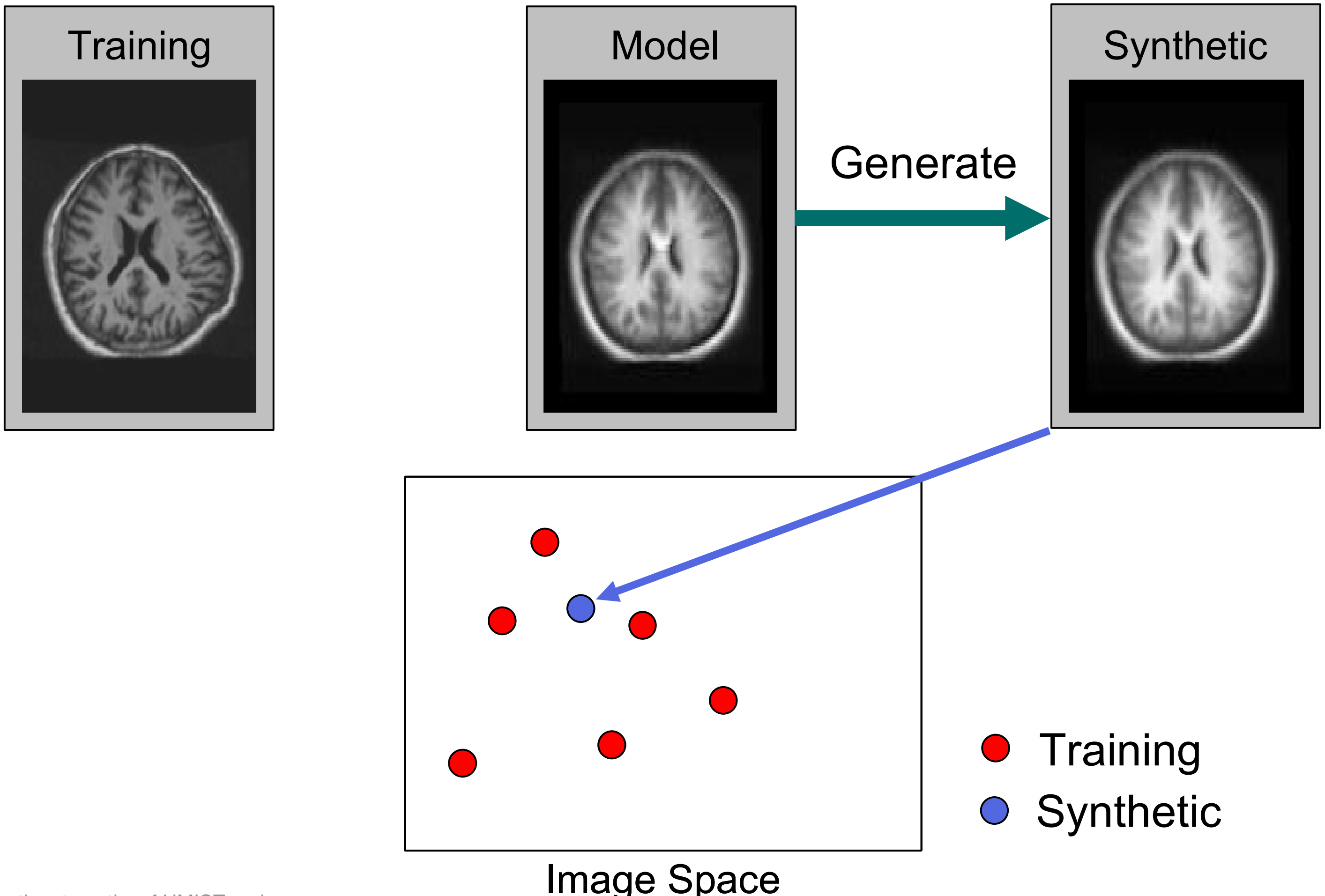


Image Space

Training and Synthetic Images



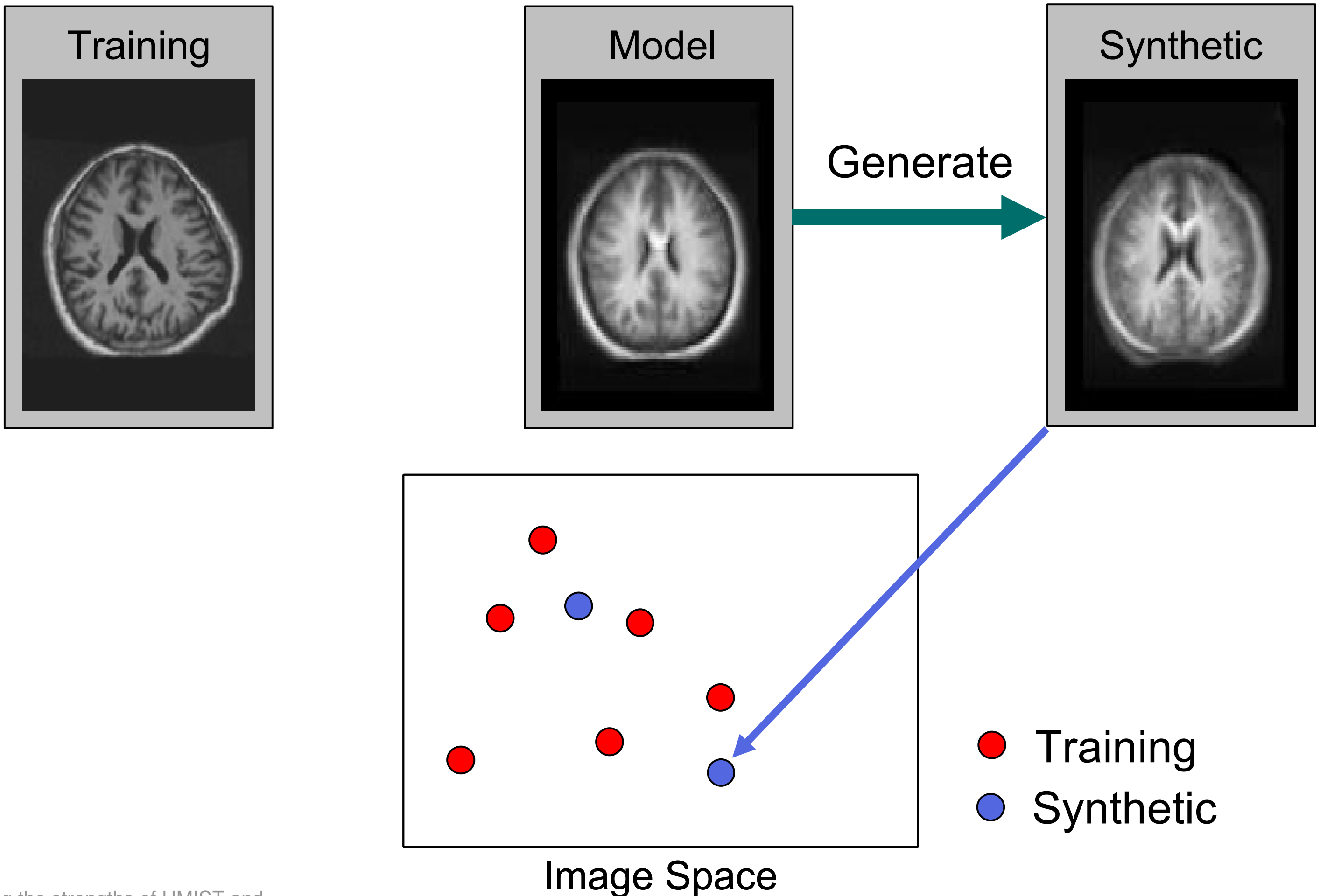
Training and Synthetic Images



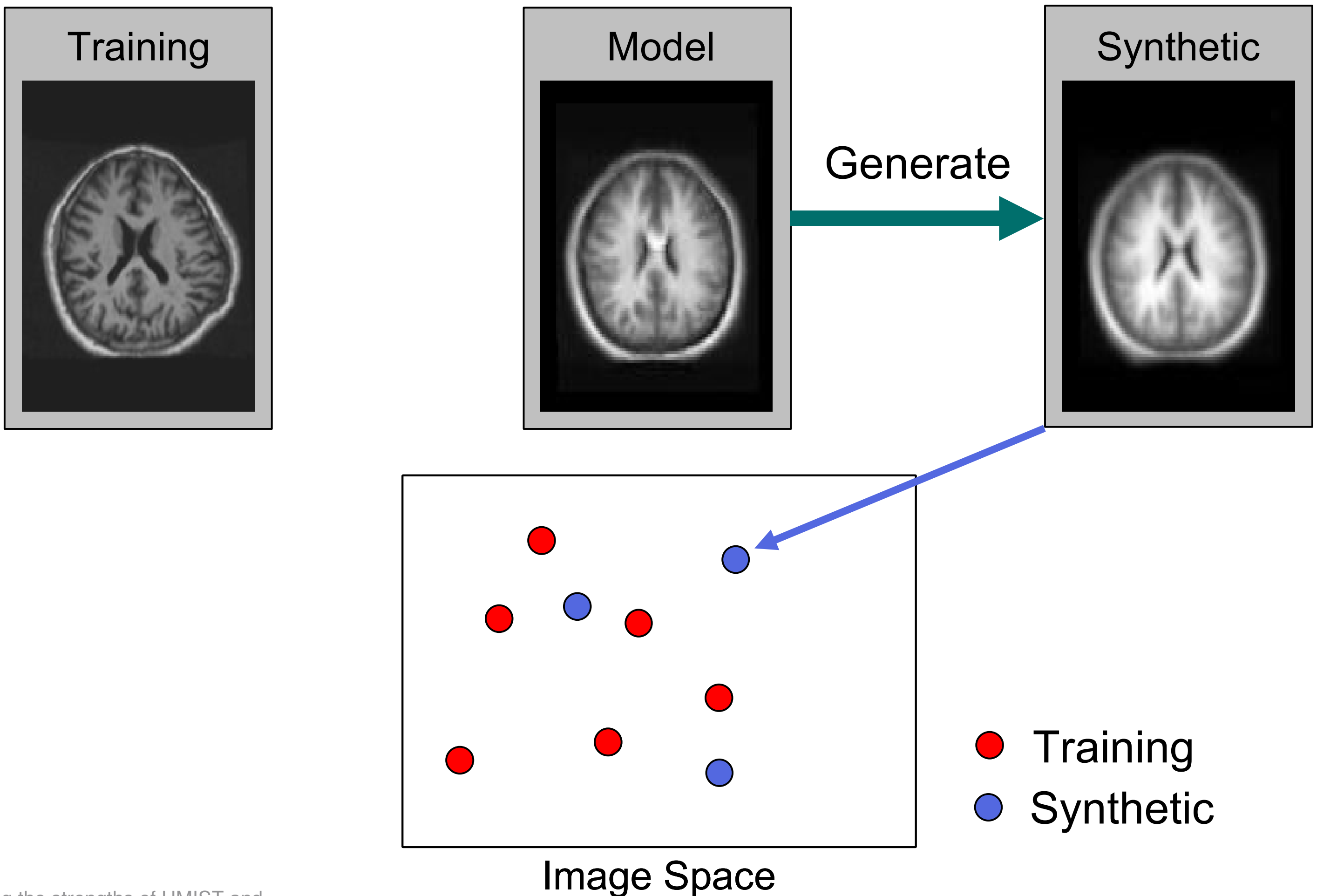
MANCHESTER
1824

The University of Manchester

Training and Synthetic Images



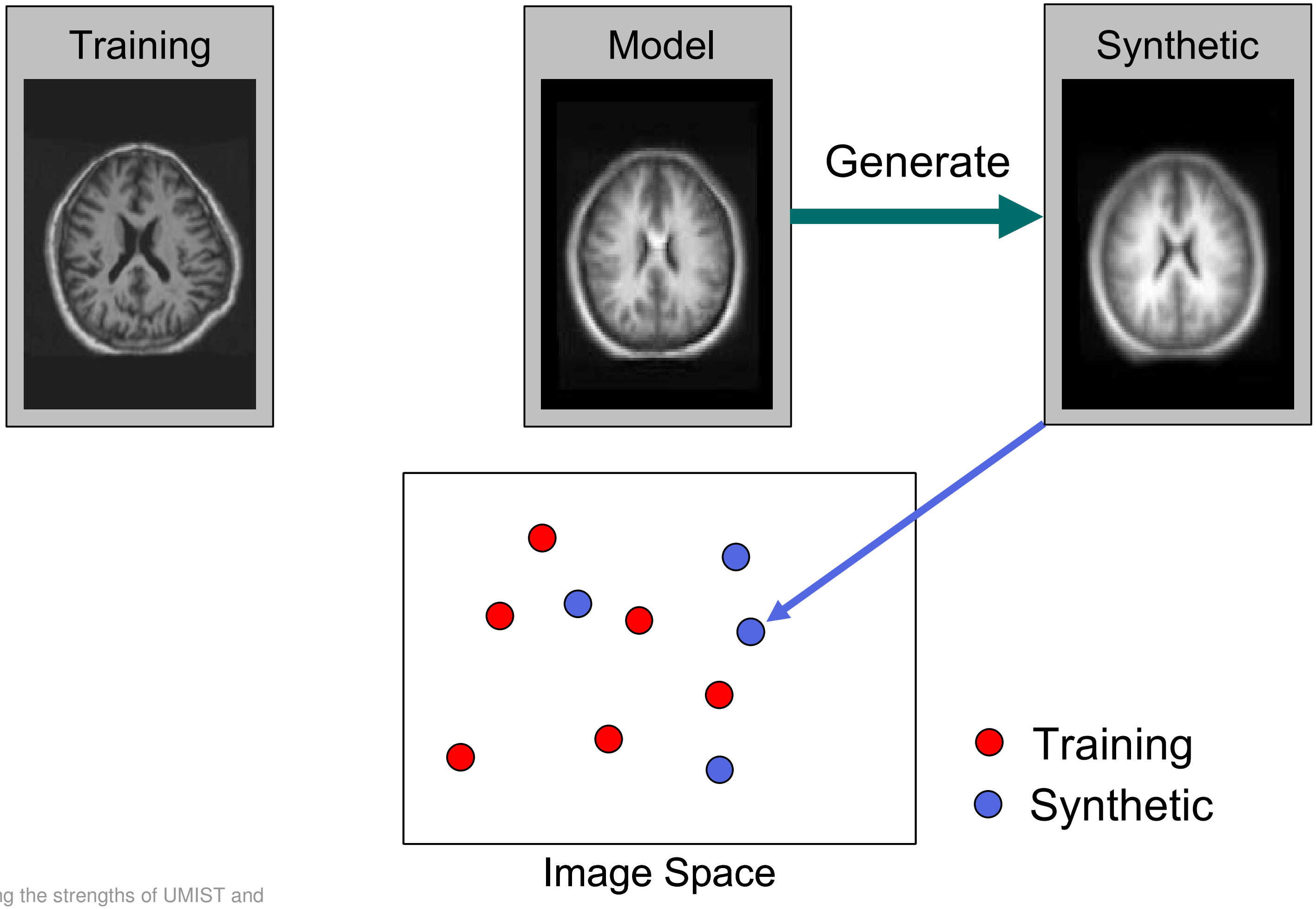
Training and Synthetic Images



MANCHESTER
1824

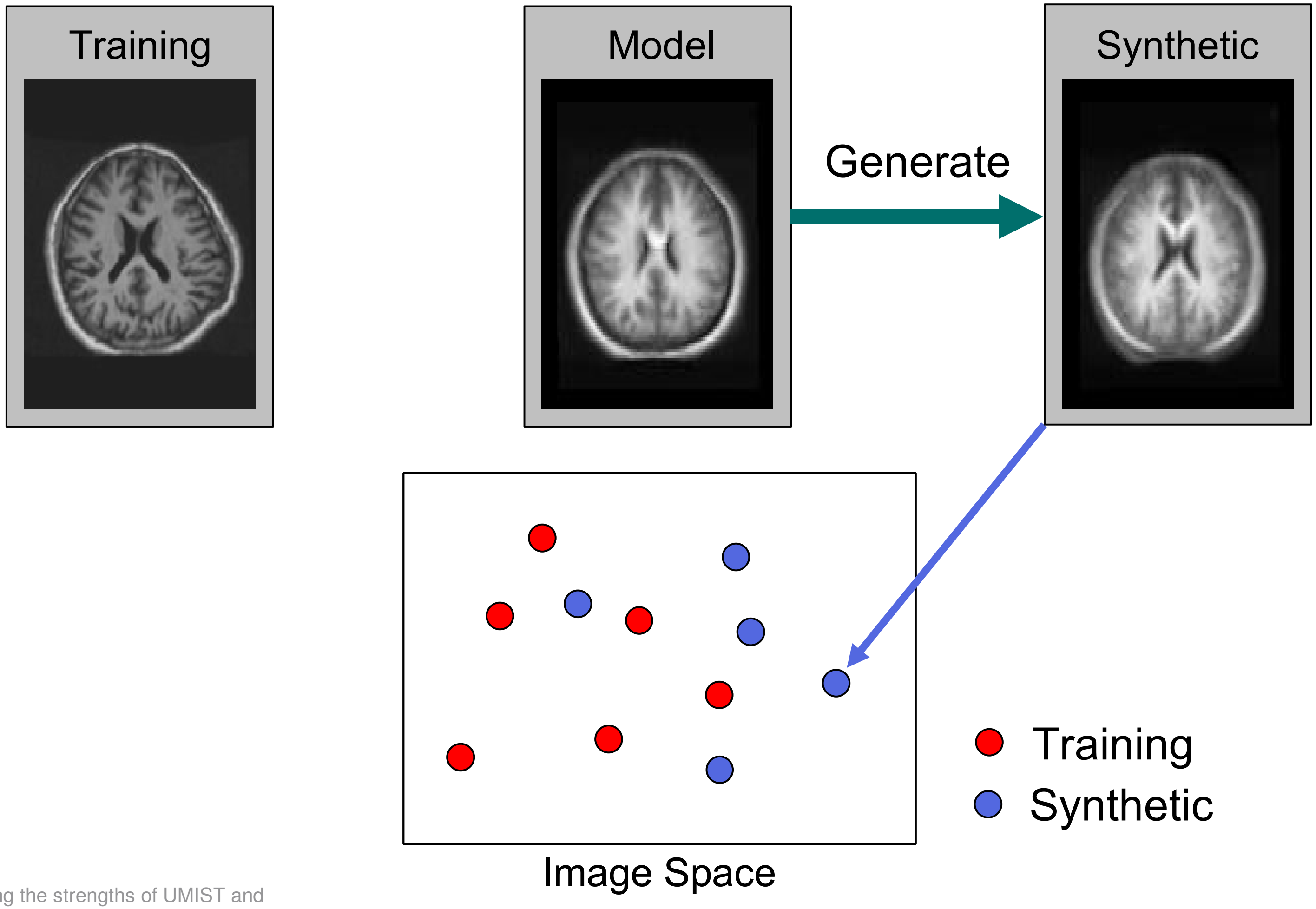
The University of Manchester

Training and Synthetic Images

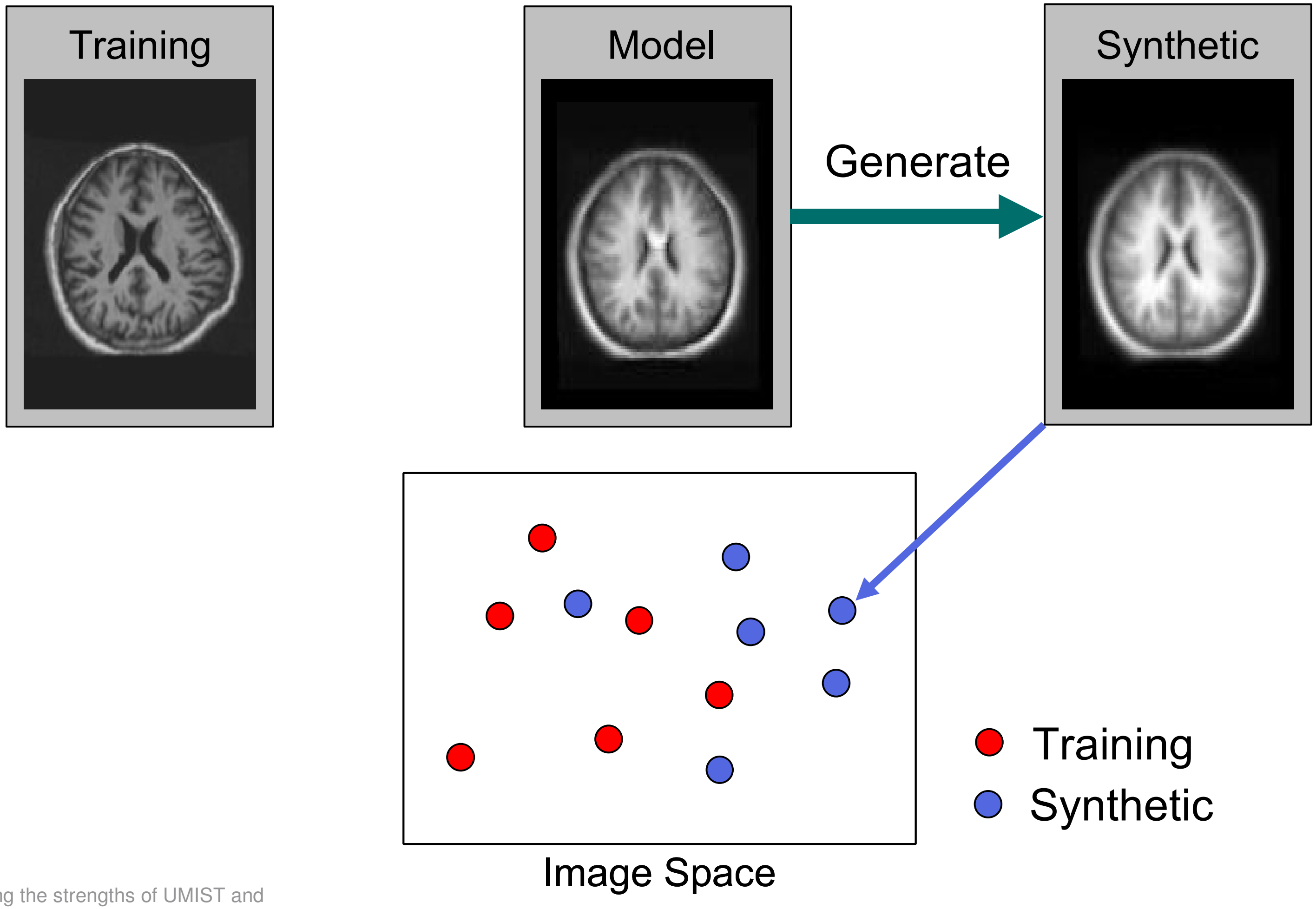


Combining the strengths of UMIST and
The Victoria University of Manchester

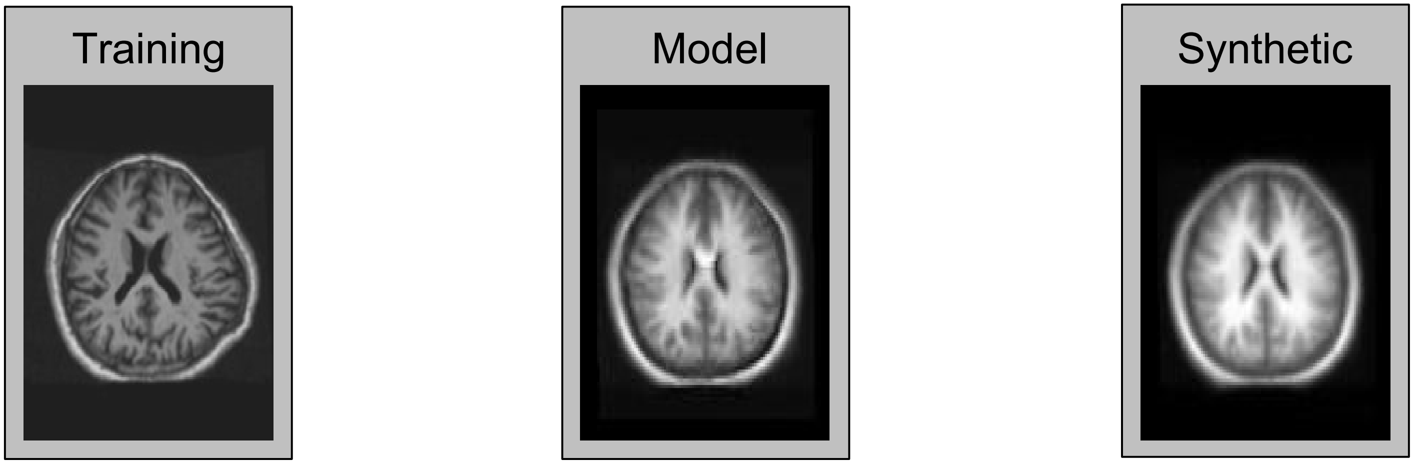
Training and Synthetic Images



Training and Synthetic Images



Training and Synthetic Images



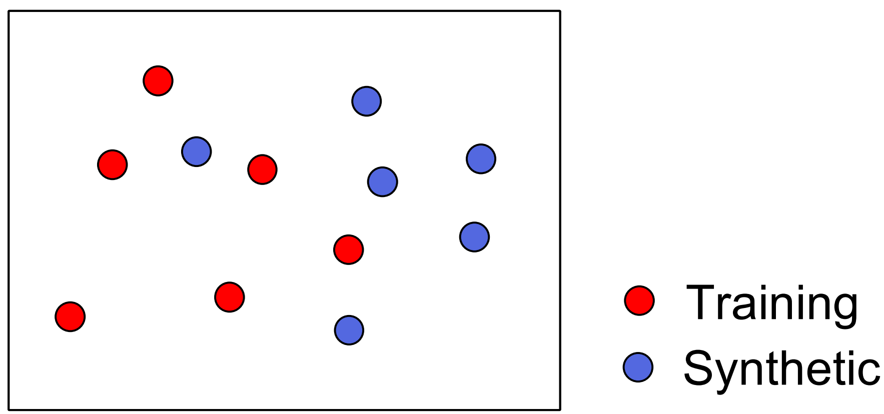
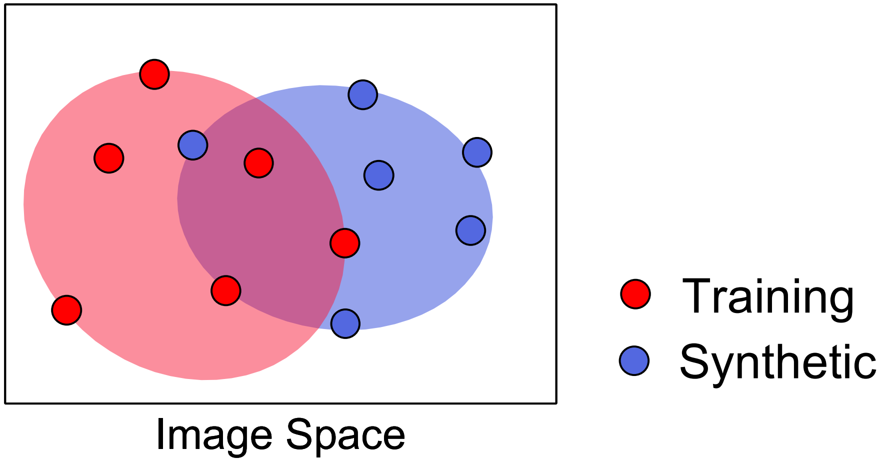
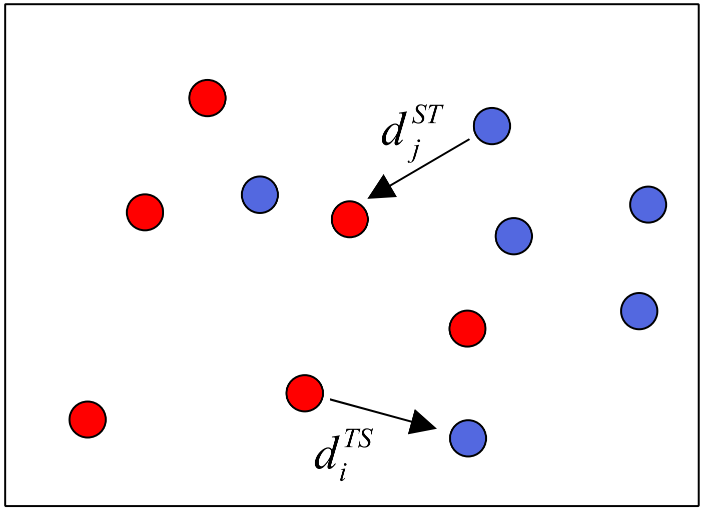


Image Space

Training and Synthetic Images



Model Quality



Training

Synthetic

Given measure d of image distance

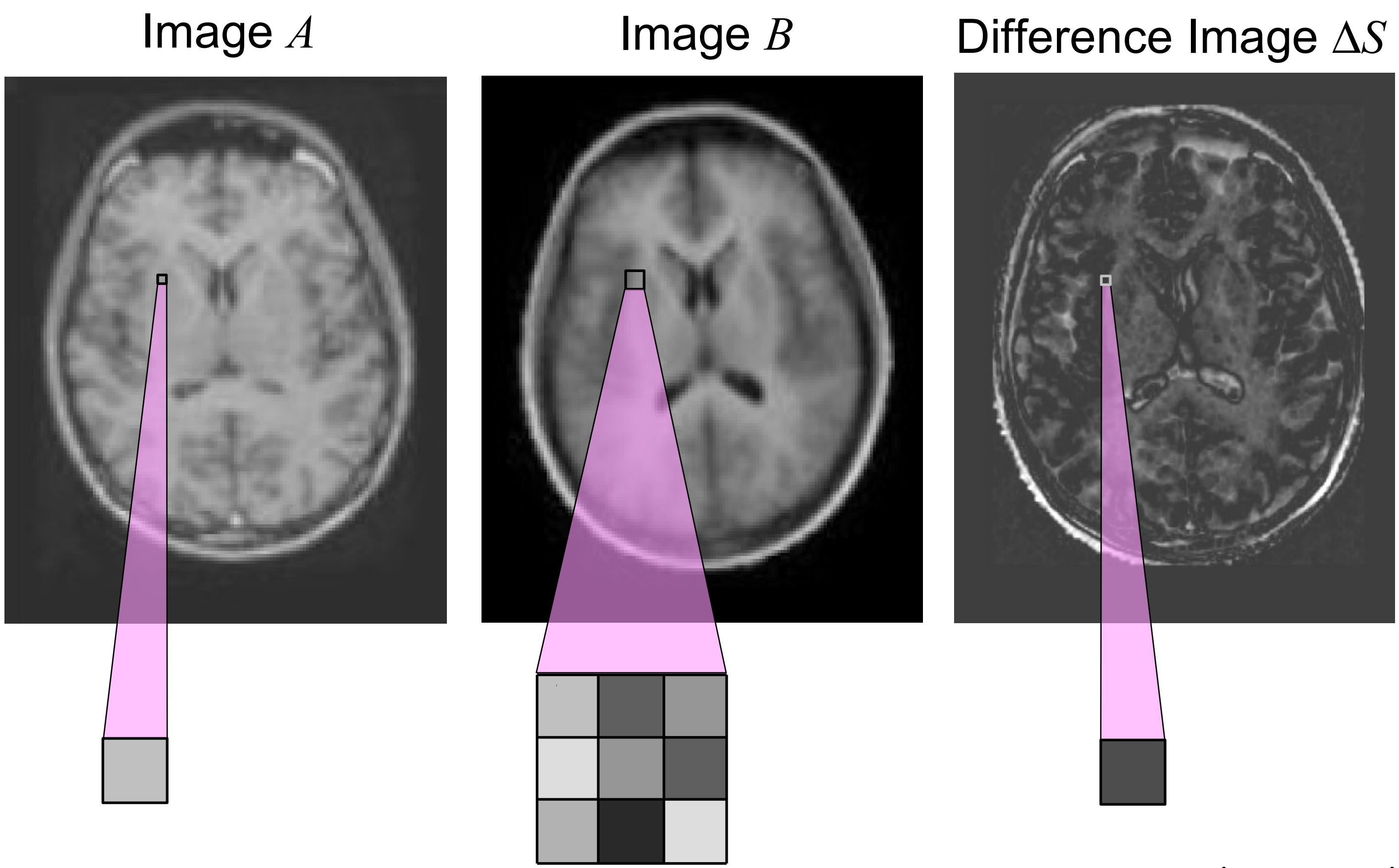
$$Specificity = \sum_{j=1}^{m} \left| d_j^{ST} \right| \Big/ m$$ Mean distance to nearest training image

$$Generalisation = \sum_{i=1}^{n} \left| d_i^{TS} \right| \Big/ n$$ Mean distance to nearest model image

Measuring Inter-Image Distance

- Euclidean
  - simple and cheap
  - sensitive to small misalignments
- Shuffle distance
  - neighbourhood-based pixel differences
  - less sensitive to misalignment

Shuffle Distance



$$A_i \qquad B_{ij} \qquad \Delta S_i = \mathrm{Min}_j \left| A_i - B_{ij} \right|$$

Varying Shuffle Radius

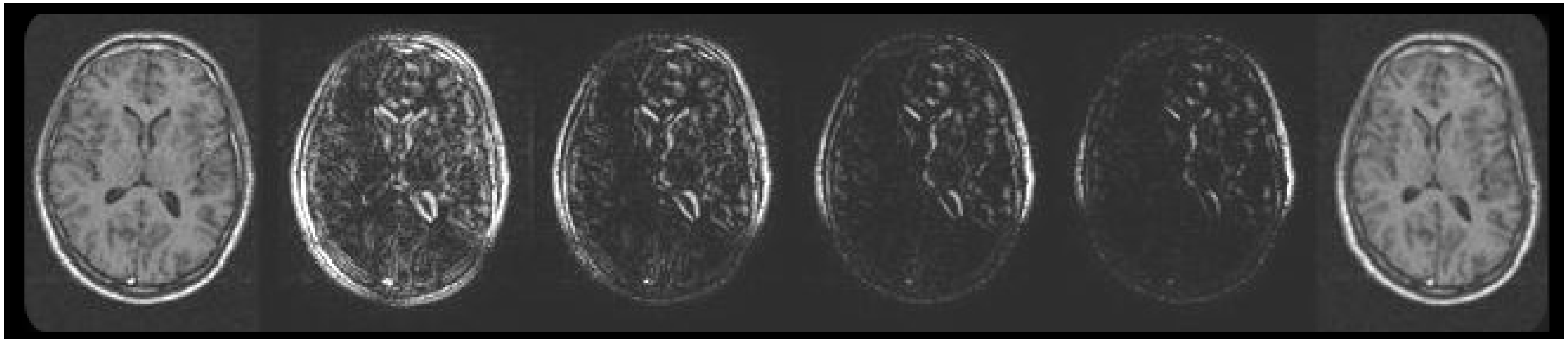


Image A r = 1 r = 1.5 r = 2.1 r = 3.7 Image B

The University of Manchester

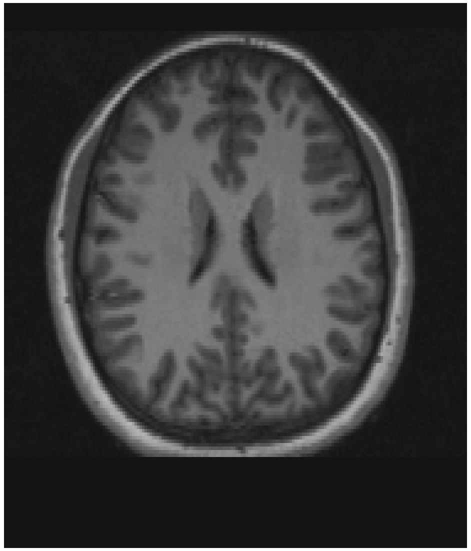
Experimental Evaluation

Experimental Design

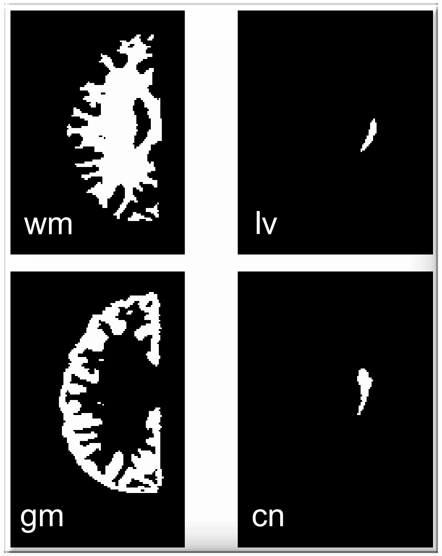
- MGH dataset (37 brains)
- Selected 2D slice
- Initial ‘correct’ NRR
- Progressive perturbation of registration
  - 10 random instantiations for each perturbation magnitude
- Comparison of the two different measures
  - overlap
  - model-based

Brain Data

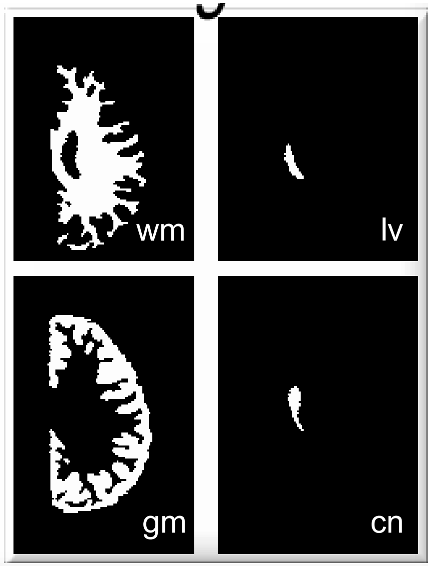
- Eight labels per image
  - L/R white/grey matter
  - L/R lateral ventricle
  - L/R caudate nucleus



Image



LH Labels



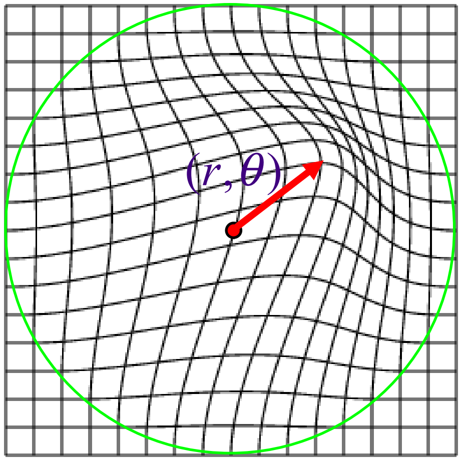
RH Labels

MANCHESTER 1824

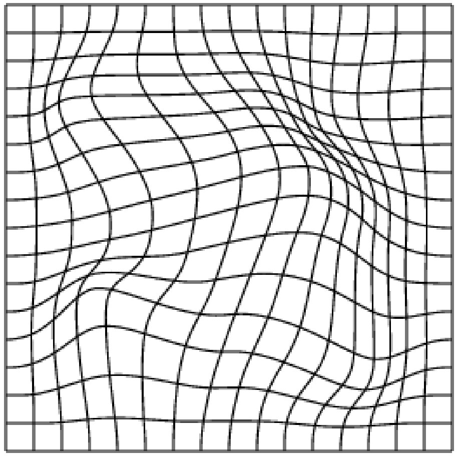
The University of Manchester

Perturbation Framework

- Alignment degraded by applying warps to data
- Clamped-plate splines (CPS) with 25 knot-points
- Random displacement (r, θ) drawn from distribution

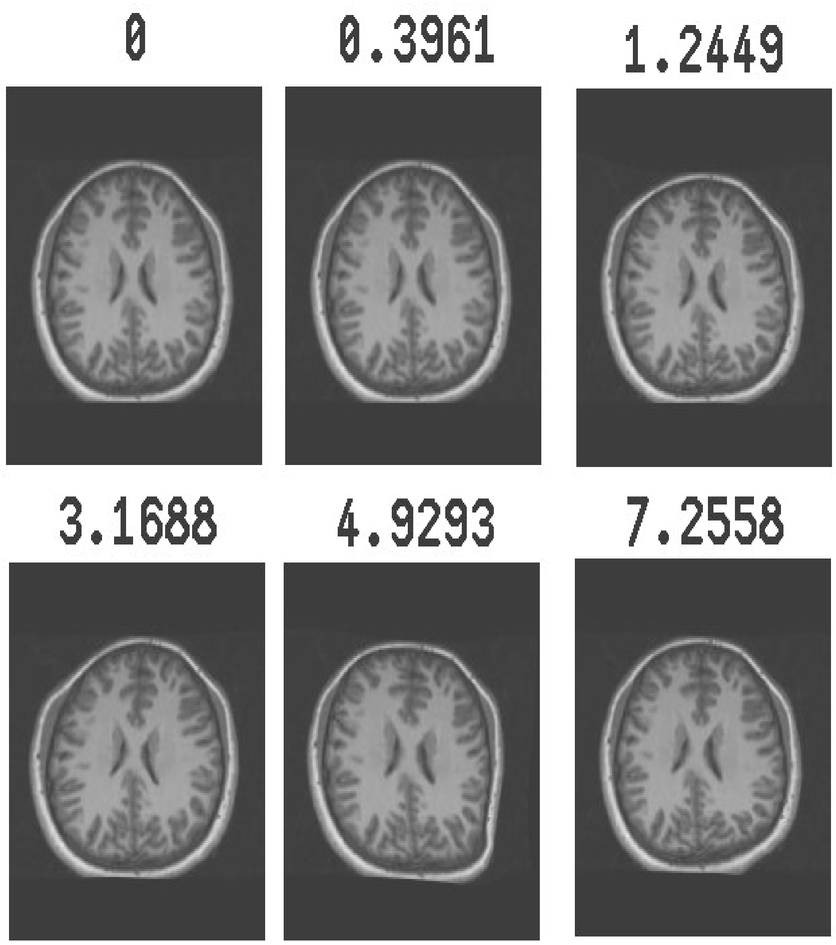


CPS with 1 knot point



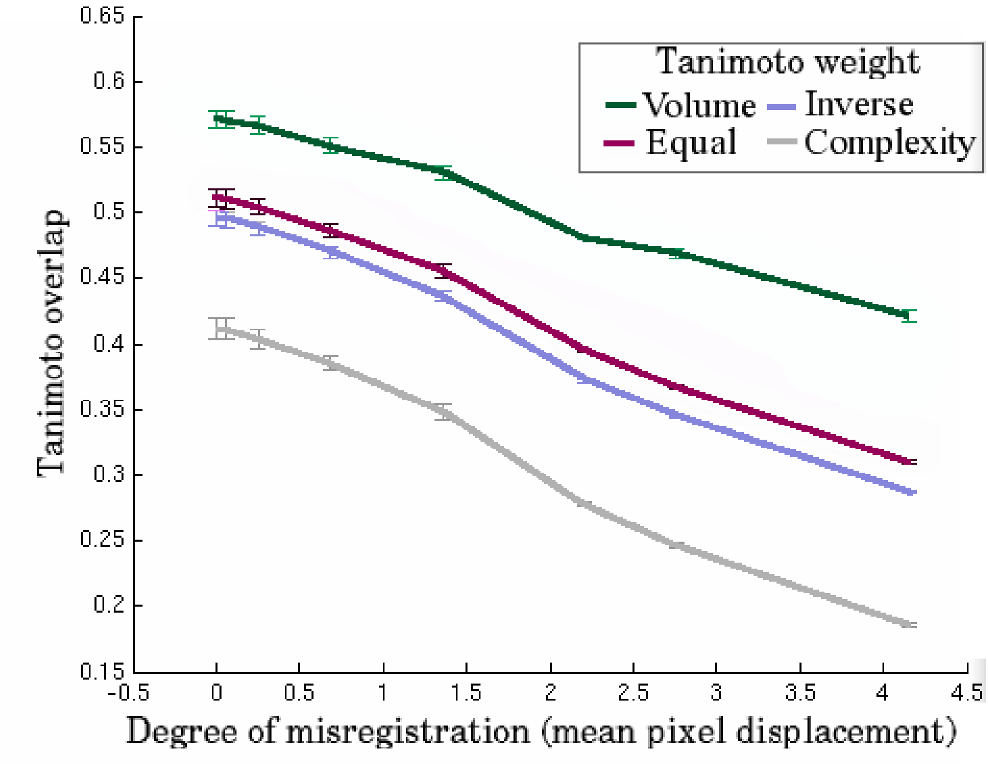
Multiple knot points

Examples of Perturbed Images



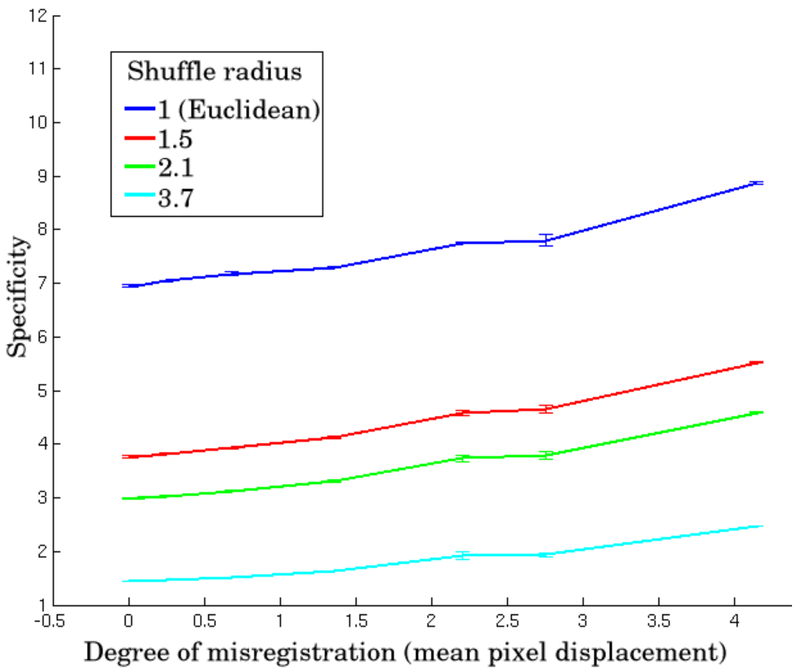
Results – Overlap

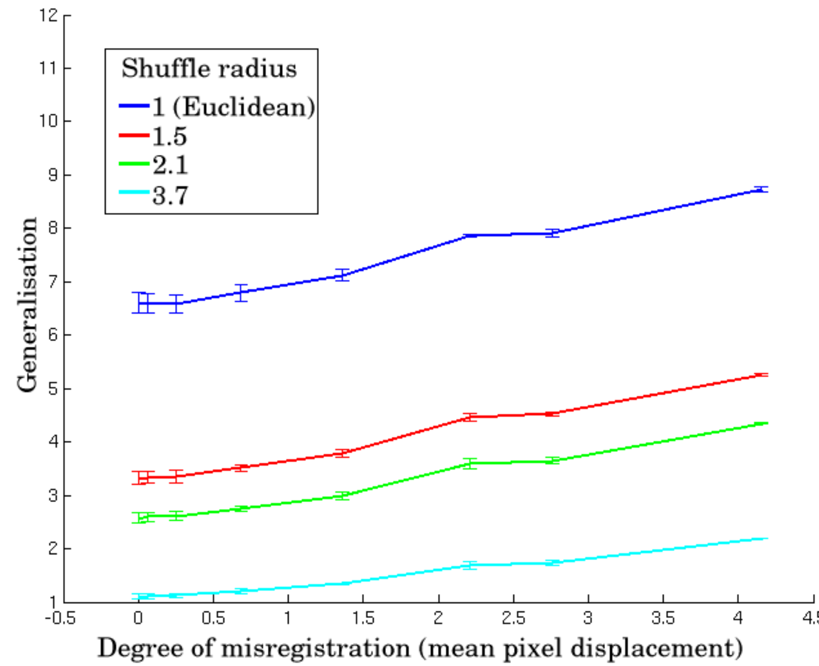
- Overlap decreases monotonically with misregistration



Results – Model-Based

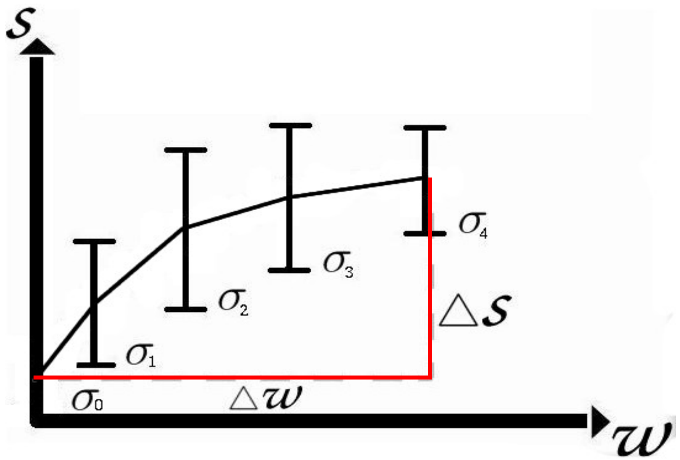
- Measures increase monotonically with misregistration





Results – Comparison

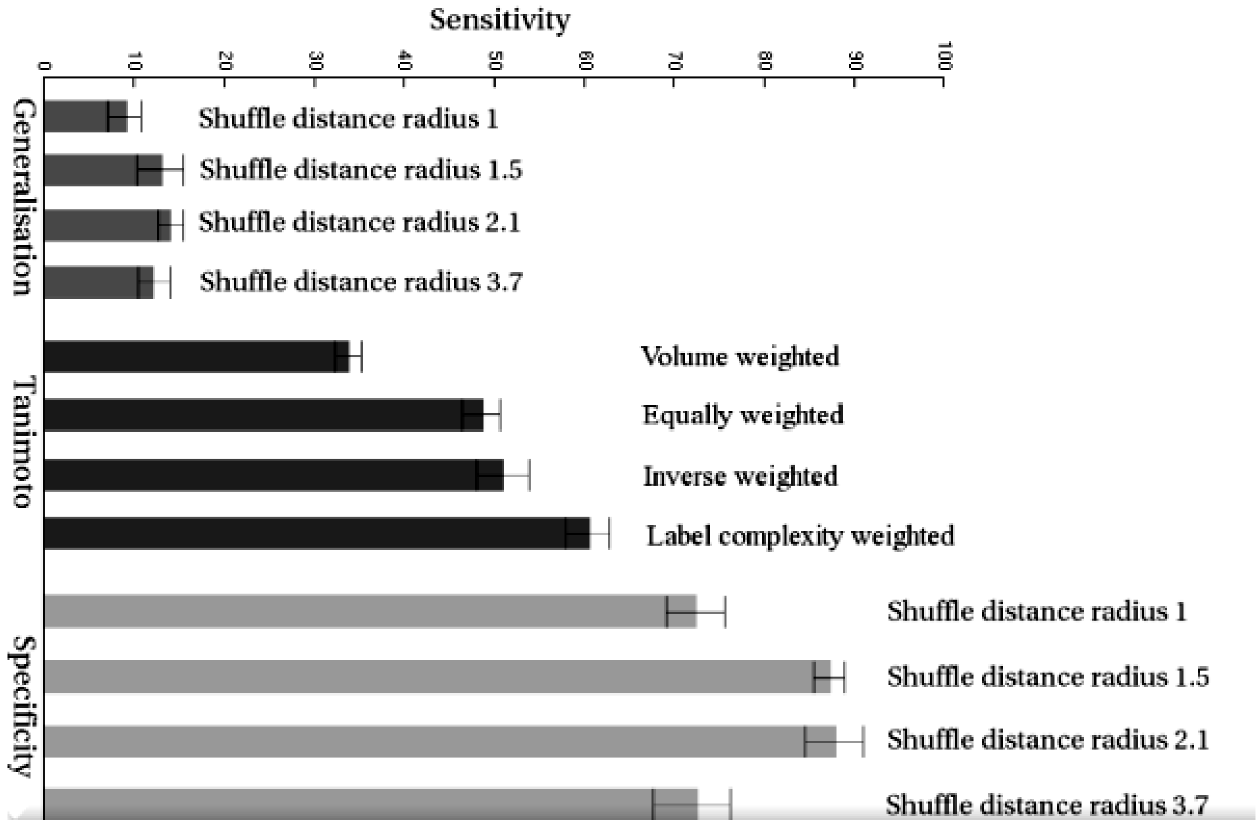
- All three measures give similar results
  - overlap-based assessment requires ground truth (labels)
  - model-based approach does not need ground truth
- Compare sensitivity of methods
  - ability to detect small changes in registration



$$Sensitivity = \frac{\Delta S}{\Delta w} \Big/ \bar{\sigma}$$

Results – Sensitivities

- Specificity most sensitive method



Further Tests – Noise

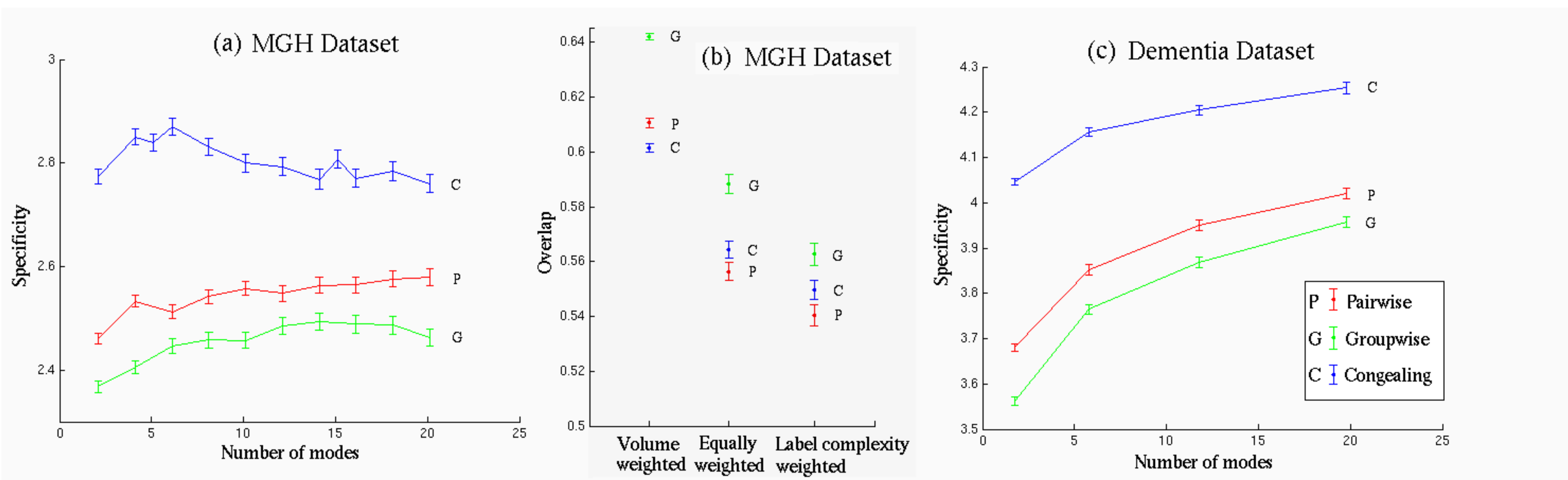
- Measure of robustness to noise is sought
- Previous experiments were repeated with noise applied
- Each image had up to 10% white noise added
- Changes in Generalisation and Sensitivity detectable
- Curves remain monotonic

Practical Application – NRR Benchmark

- 3 registration algorithm compared
  - Pairwise registration
  - Groupwise registration
  - Congealing
- 2 brain datasets used
  - MGH dataset
  - Dementia dataset
- 3 assessment methods
  - Model-based: Generalisation and Specificity
  - Overlap-based

Practical Application - Results

- Results are consistent
- Groupwise outperforms pairwise, which outperforms congealing



Extension to 3-D

- The method was implemented and tested in 3-D
- Shuffle neighbourhood to be considered can be a:
  - box
  - cube
  - plane-based comparison (slice-by-slice)
  - or sphere
- Validation experiments too laborious to replicate
- Instead, 4-5 NRR algorithms will be compared
- Ongoing work using annotated IBIM data
- Results to be compared against label overlap

Conclusions

- Both approaches sensitive to subtle misregistration
- Overlap and model-based approaches ‘equivalent’
- Overlap provides ‘gold standard’
- Specificity is a good surrogate
  - monotonically related
  - no need for ground truth
  - more sensitive
  - only applies to groups (but any NRR method)