MANCHESTER
1824

The University
of Manchester

# A Generic Method for Evaluating Appearance Models and Assessing the Accuracy of NRR

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

# Overview

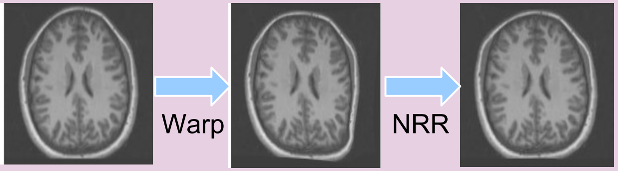
- Motivation
- Assessment methods
  - overlap-based
  - model-based
- Experiments
  - validation
  - comparison of methods
  - practical application
- Conclusions

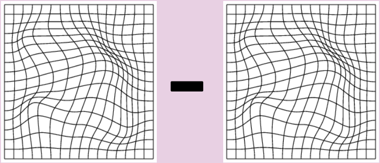
# Motivation

- Competing approaches to NRR
  - representation of warp (including regularisation)
  - similarity measure
  - optimisation
  - pair-wise vs group-wise
- Different results for same images
- Need for objective method of comparison
- QA in real applications (how well has it worked?)

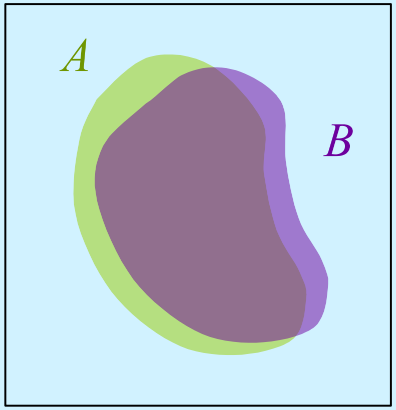
# Existing Methods of Assessment

- Artificial warps
  - recovering known warps
  - may not be representative
  - algorithm testing but not QA



- Overlap measures
  - ground truth tissue labels
  - overlap after registration
  - subjective
  - too expensive for routine QA




- Need for new approach

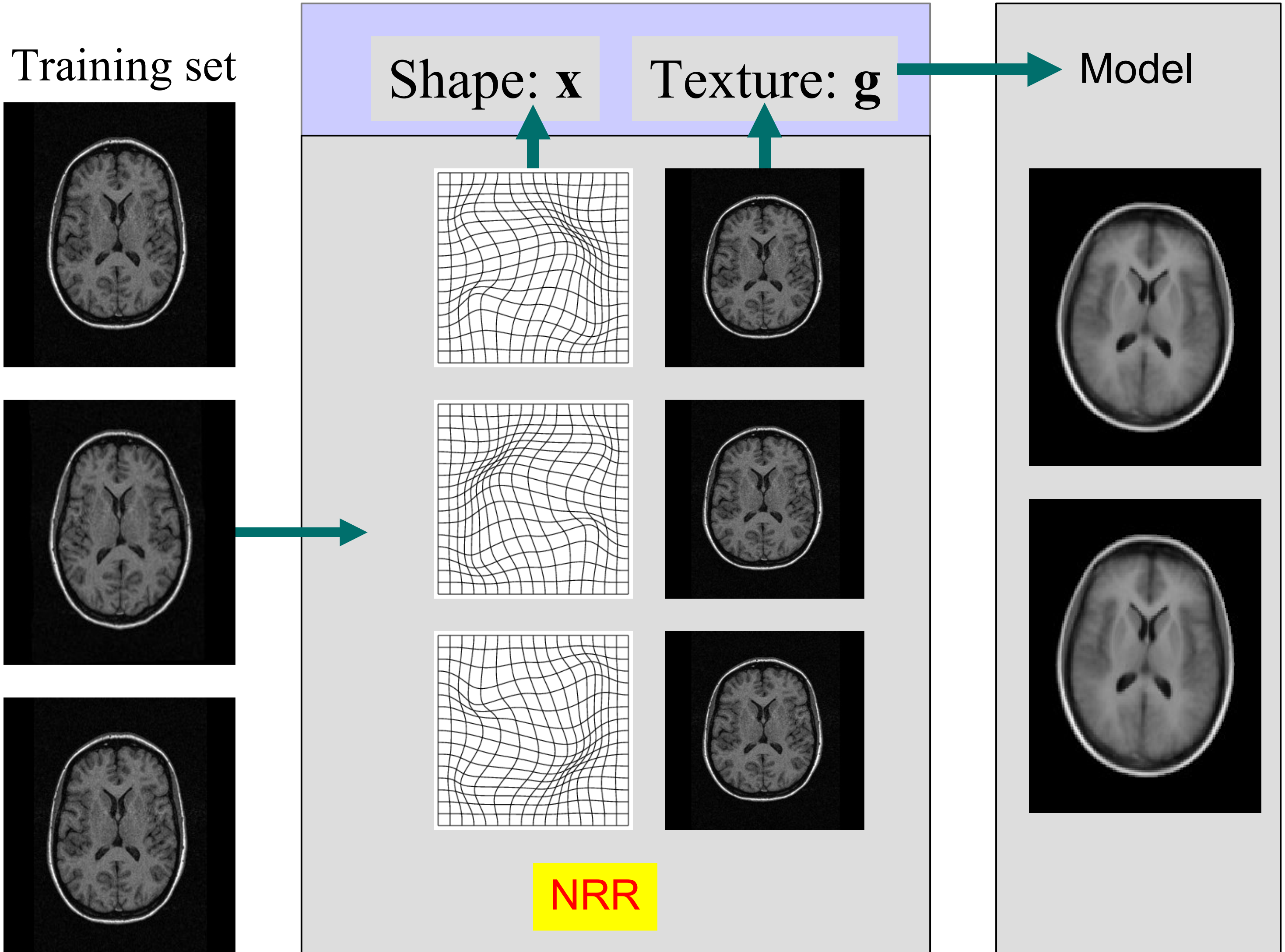
The University of Manchester

# Model-Based Assessment

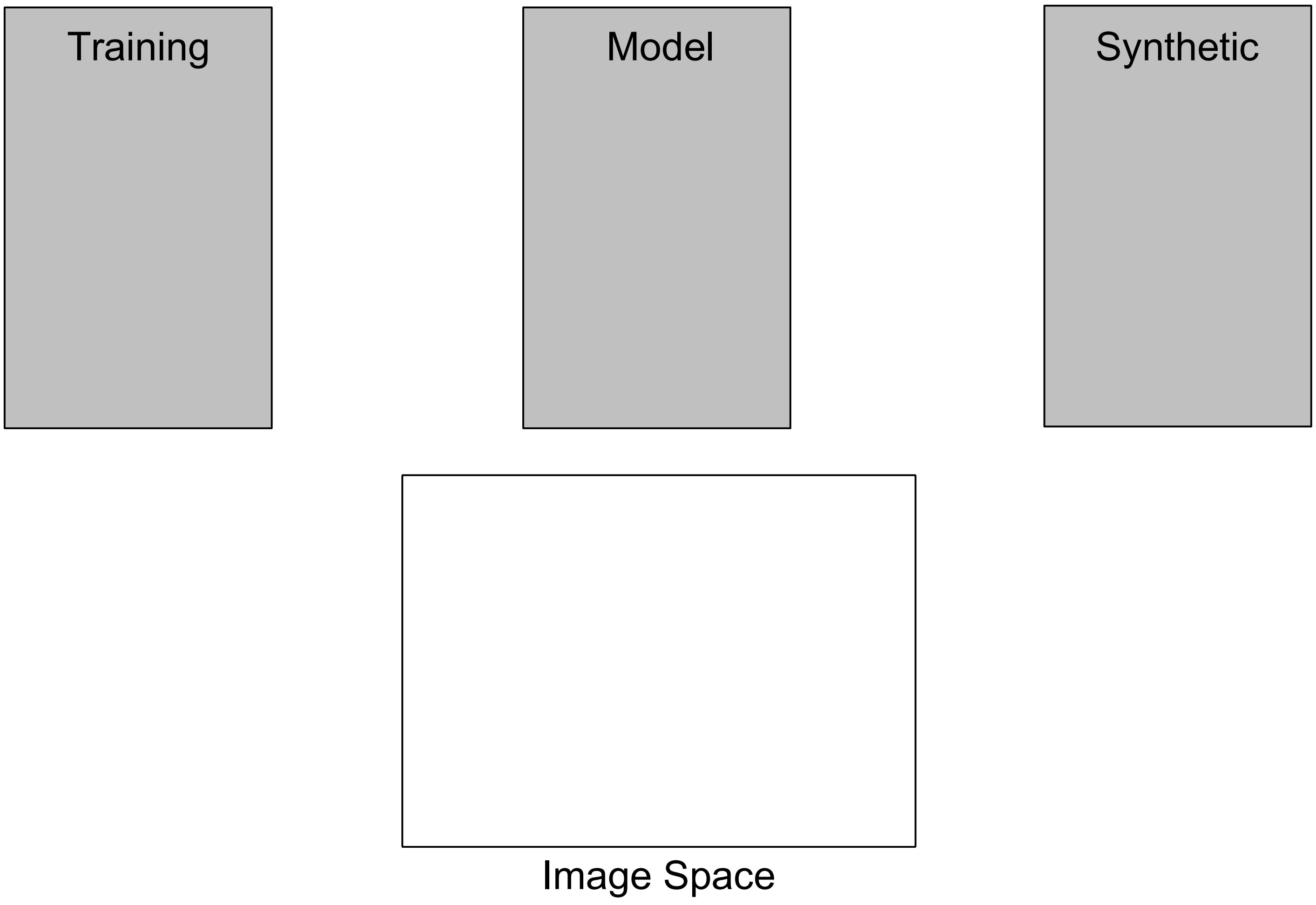
# Model-based Framework

- 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

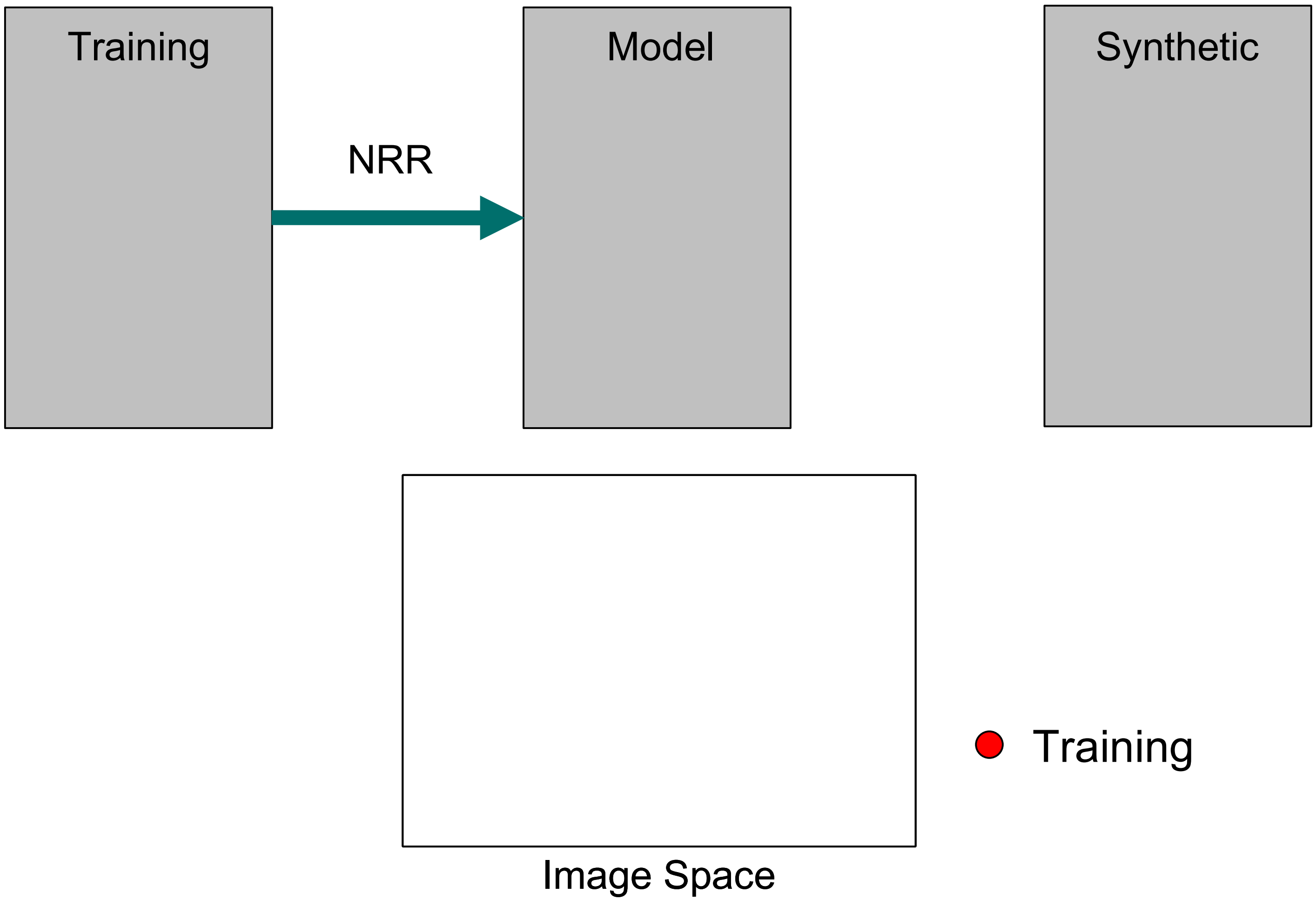
# Building an Appearance Model



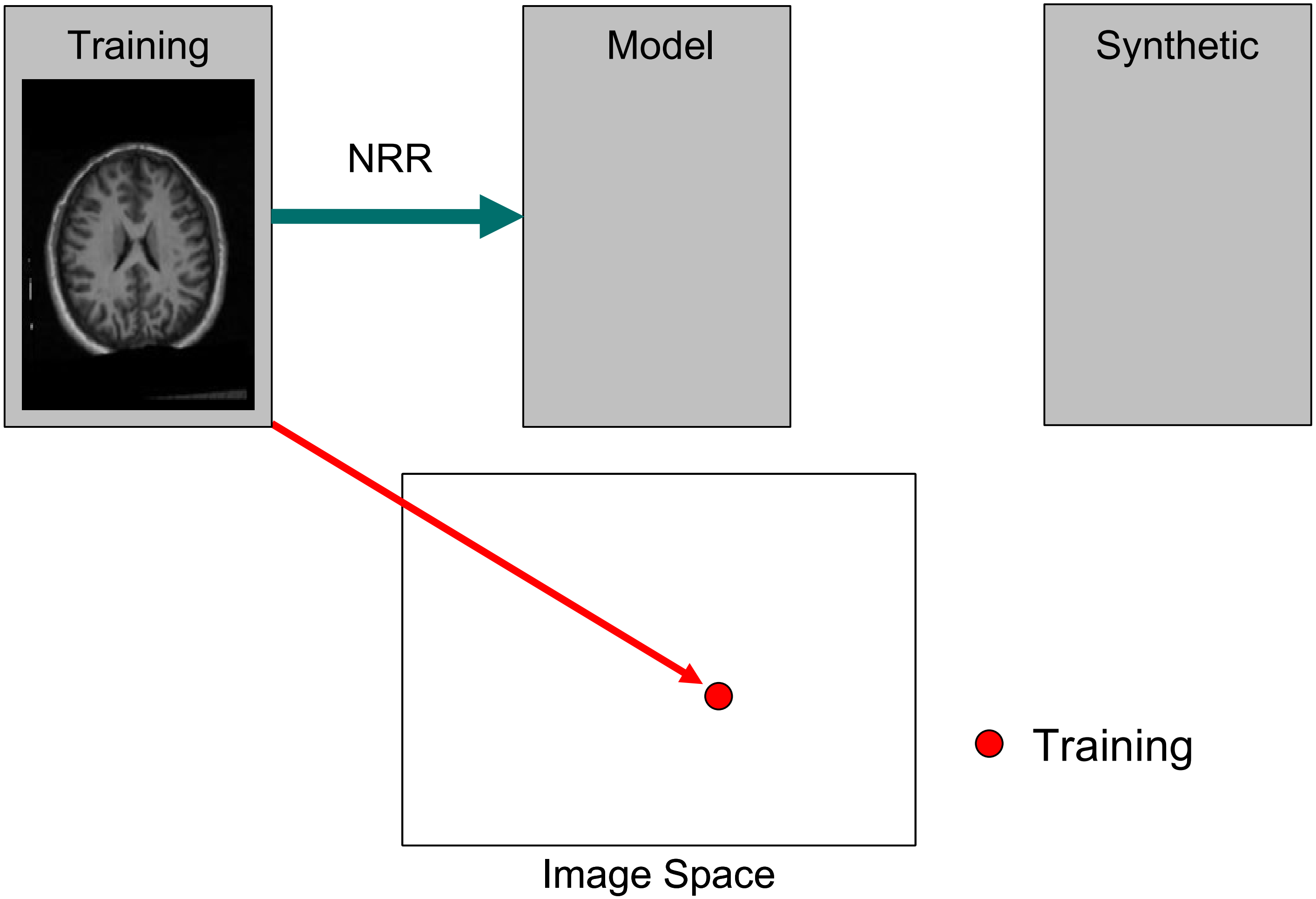
# Training and Synthetic Images



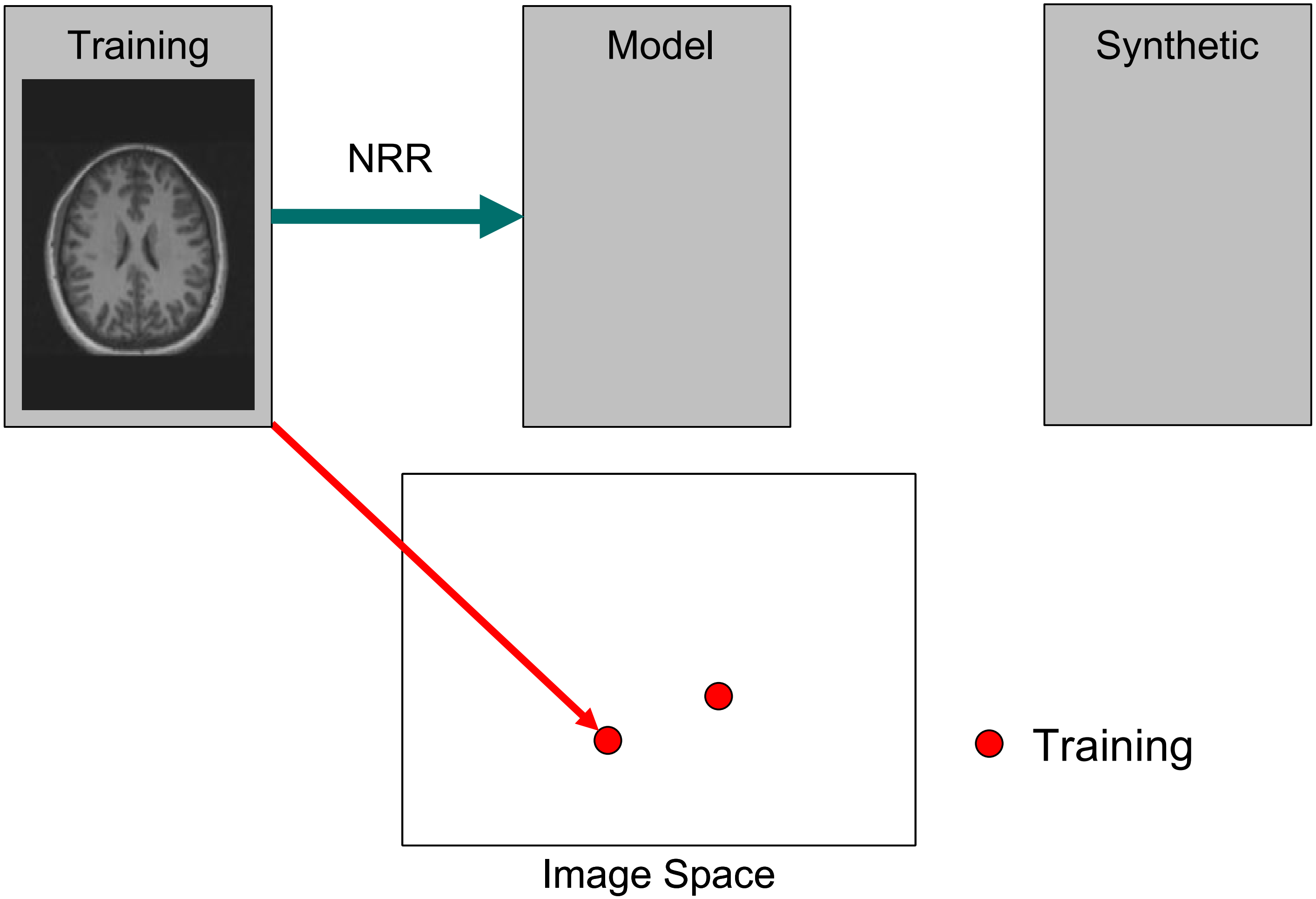
# Training and Synthetic Images



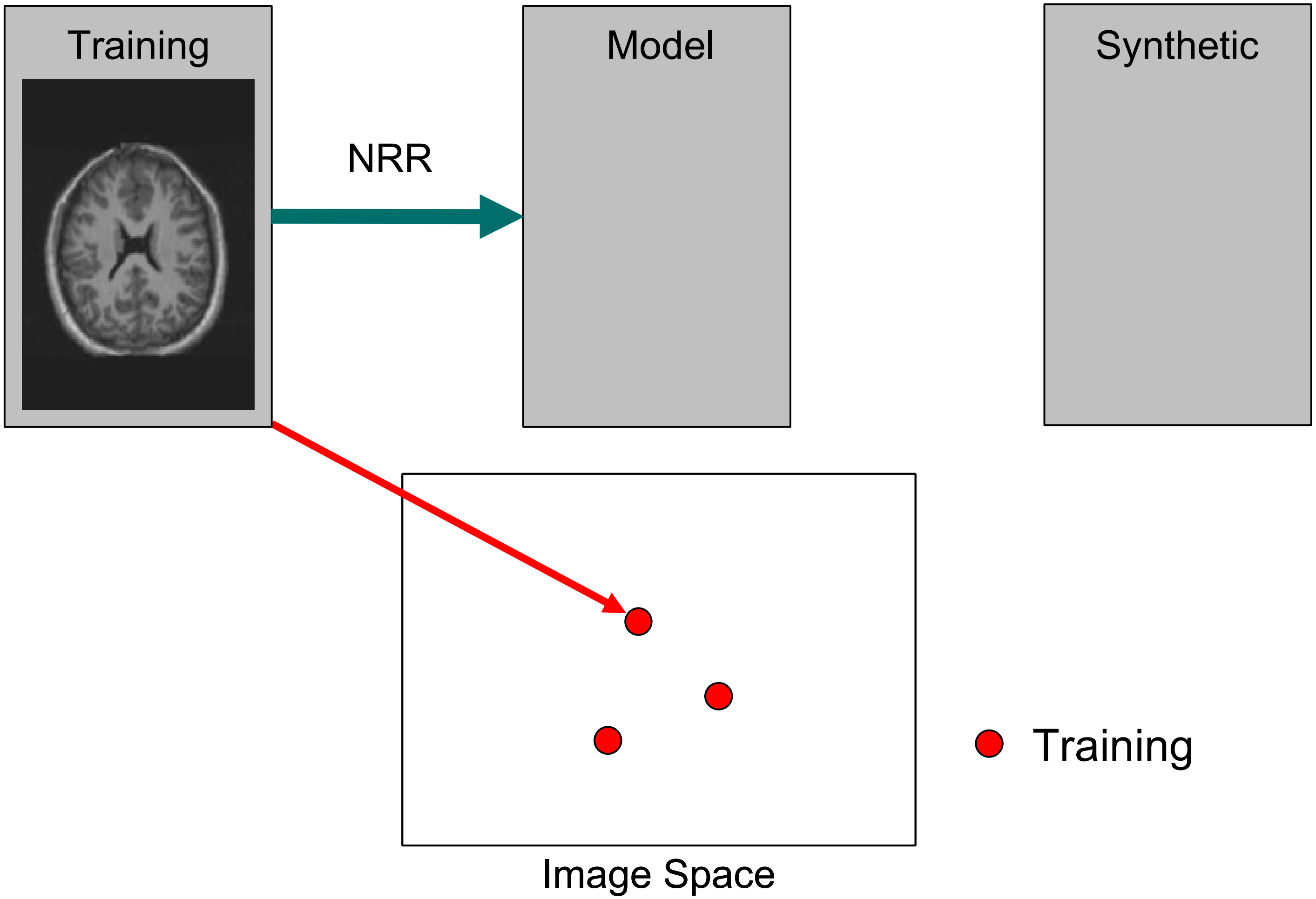
# Training and Synthetic Images



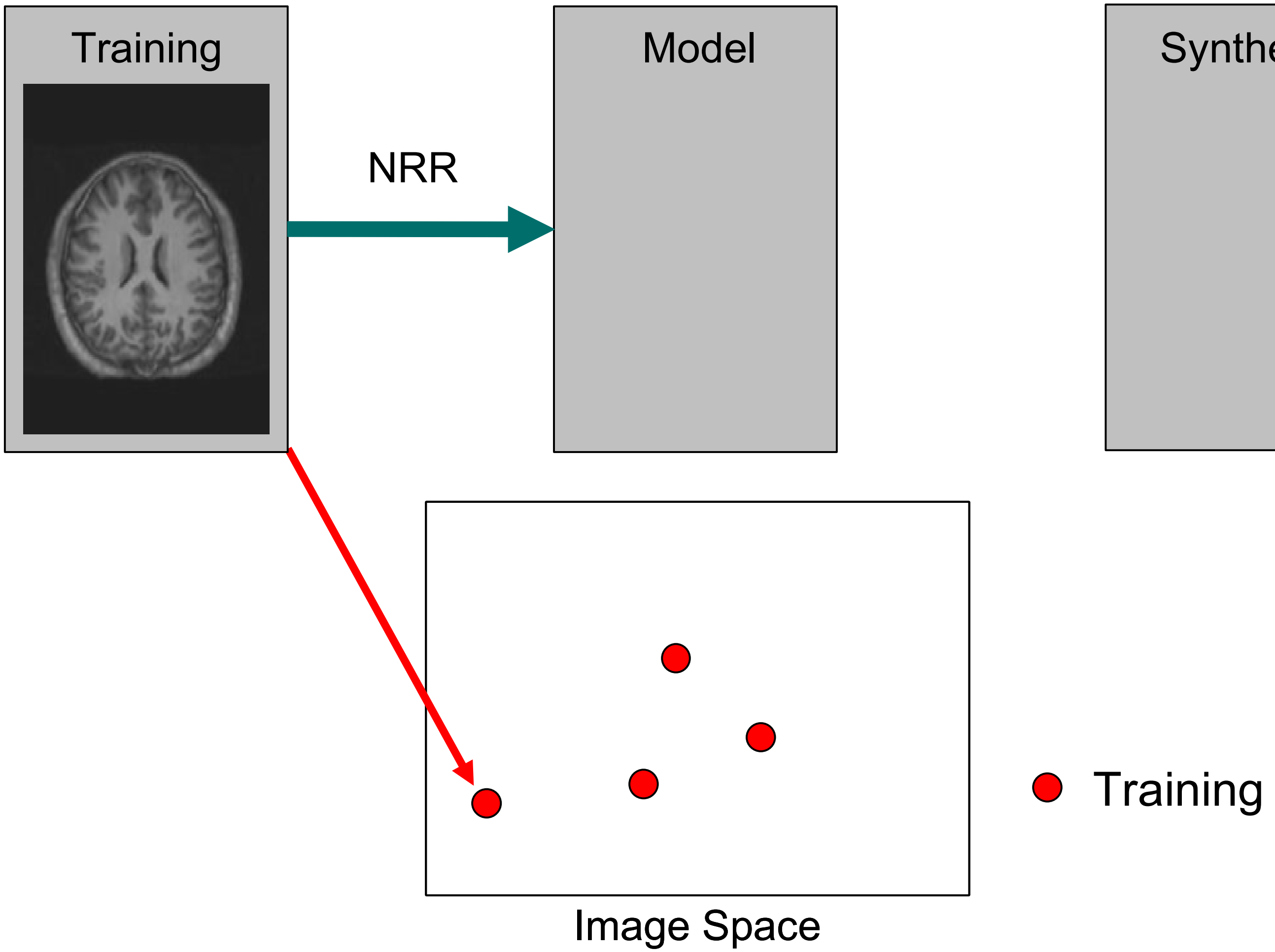
# Training and Synthetic Images



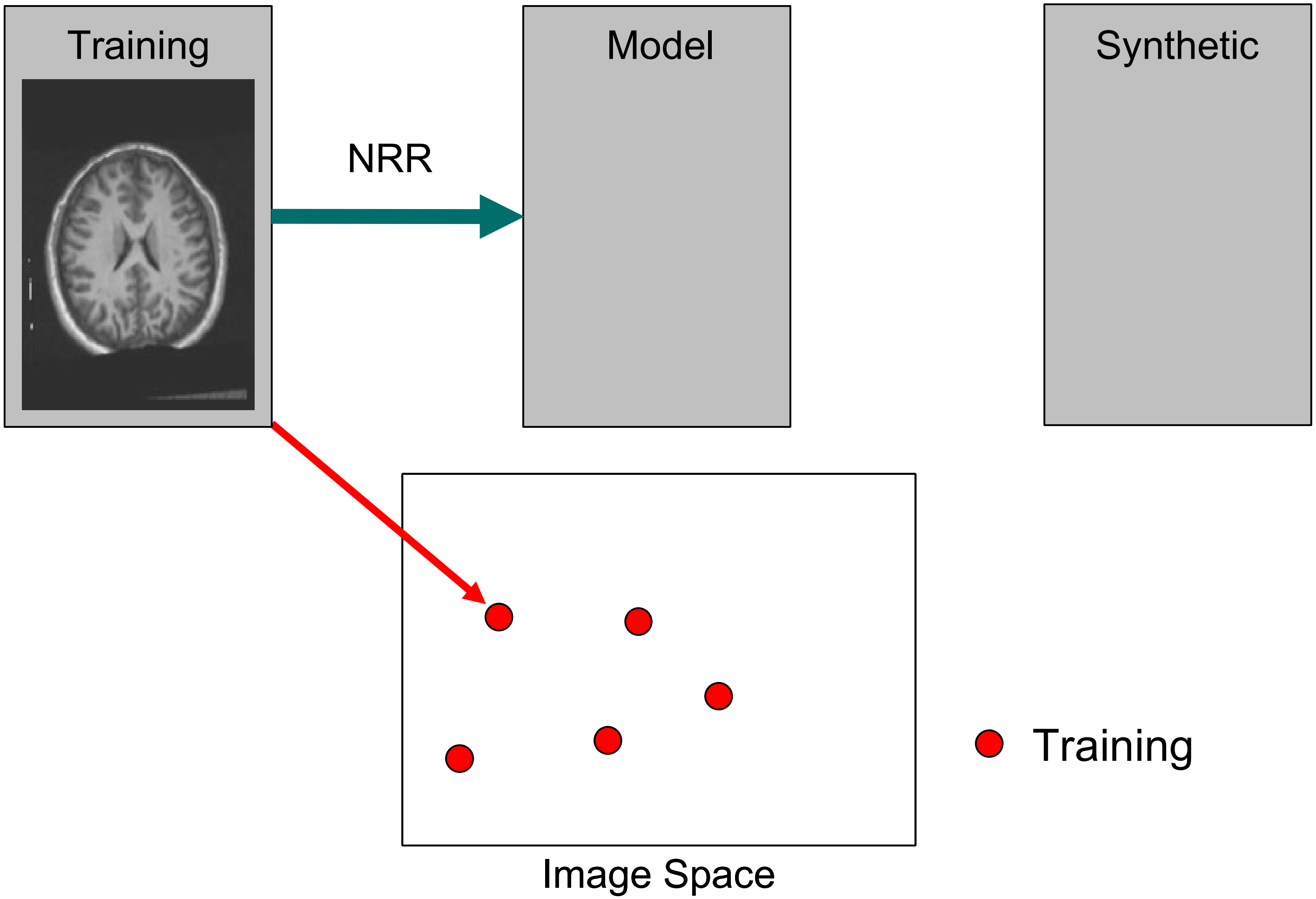
# Training and Synthetic Images



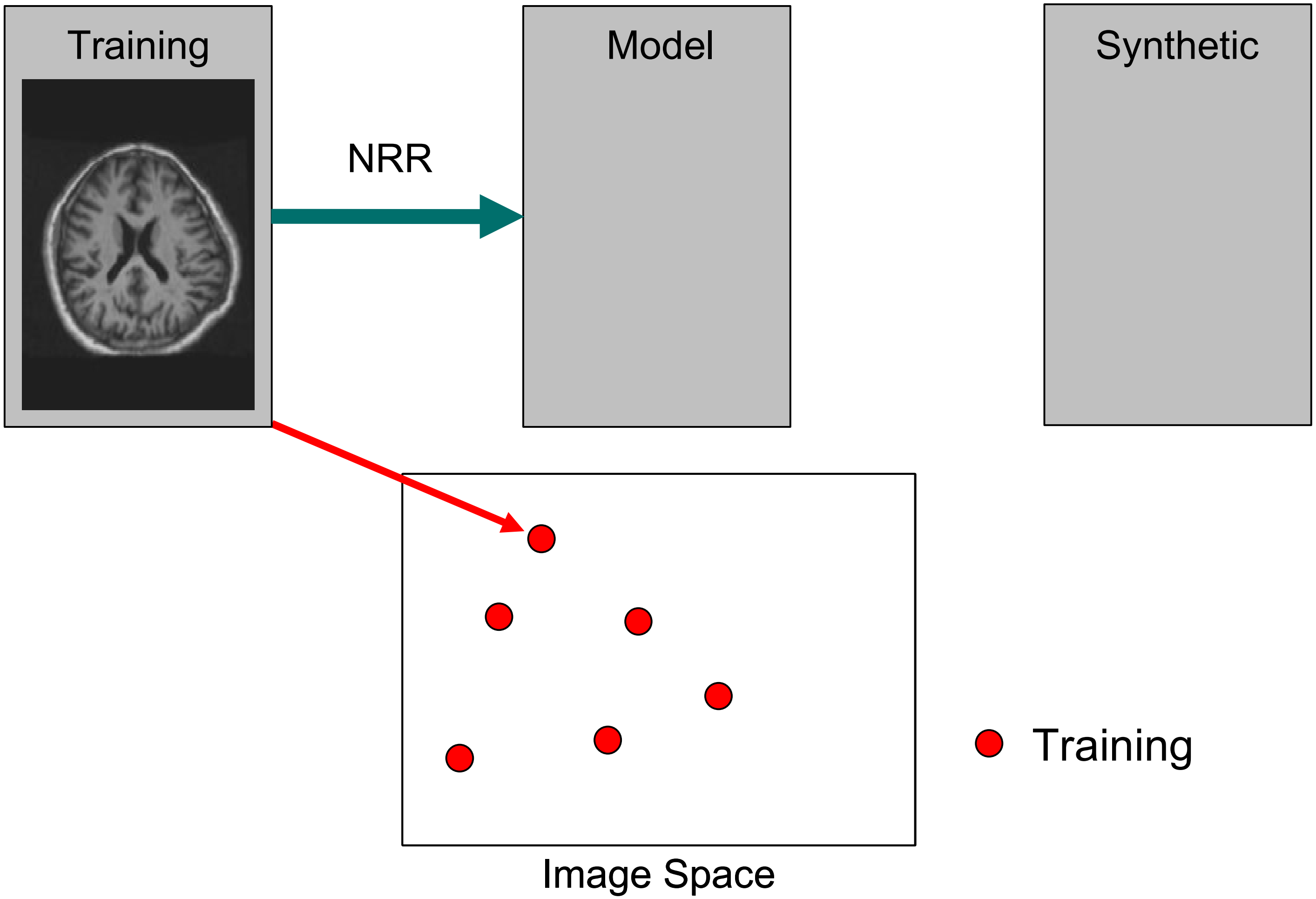
# Training and Synthetic Images



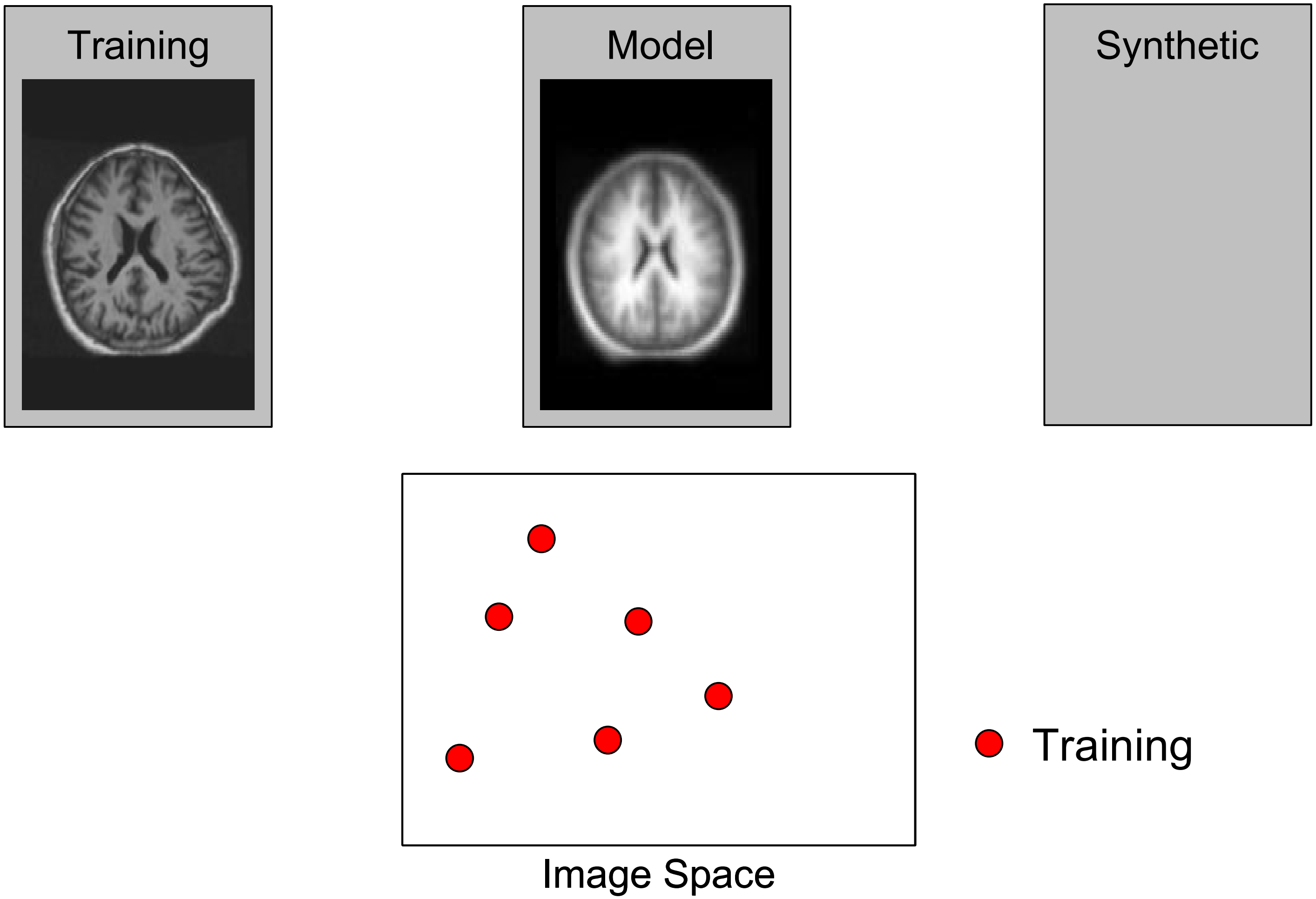
# Training and Synthetic Images



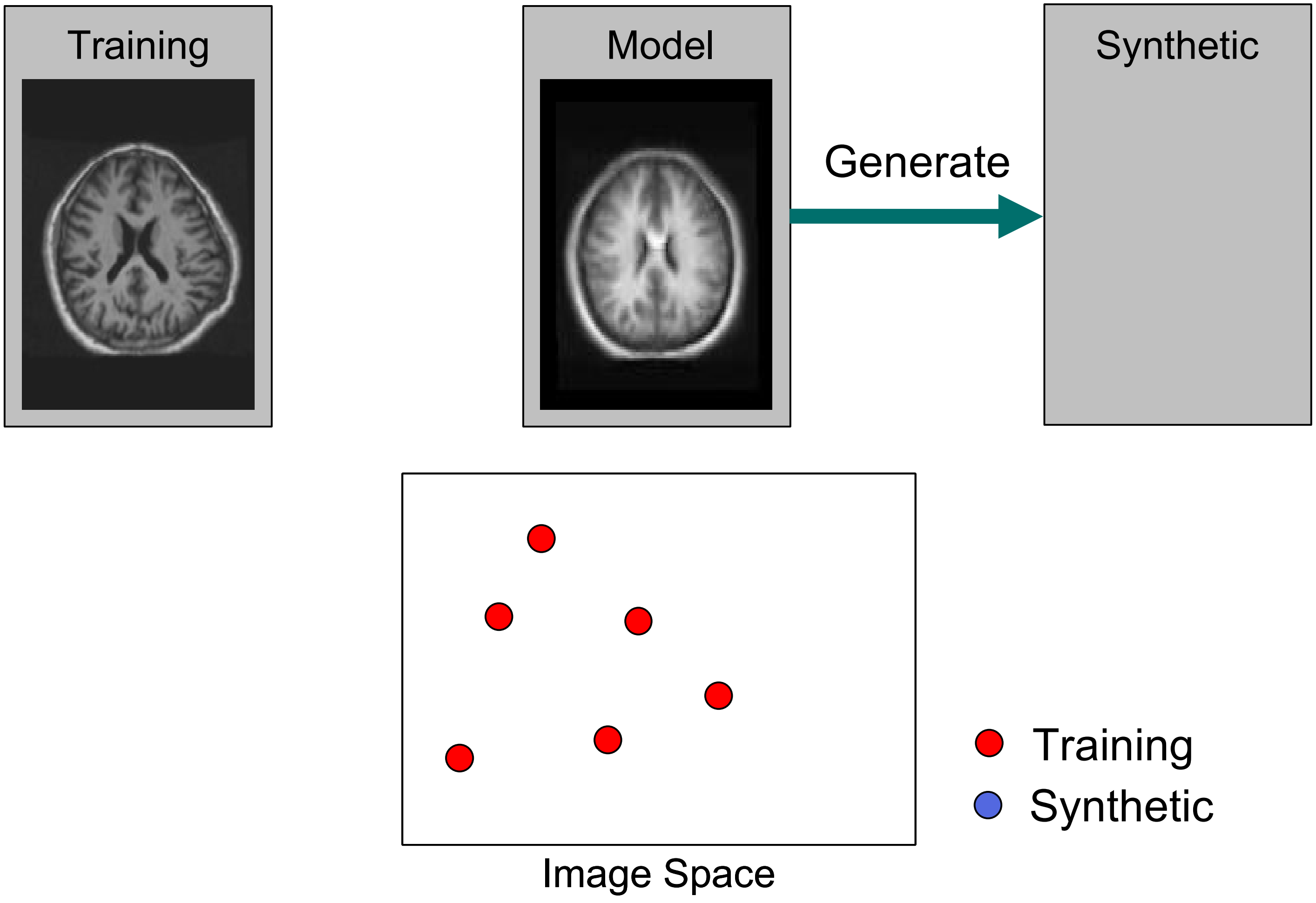
# Training and Synthetic Images



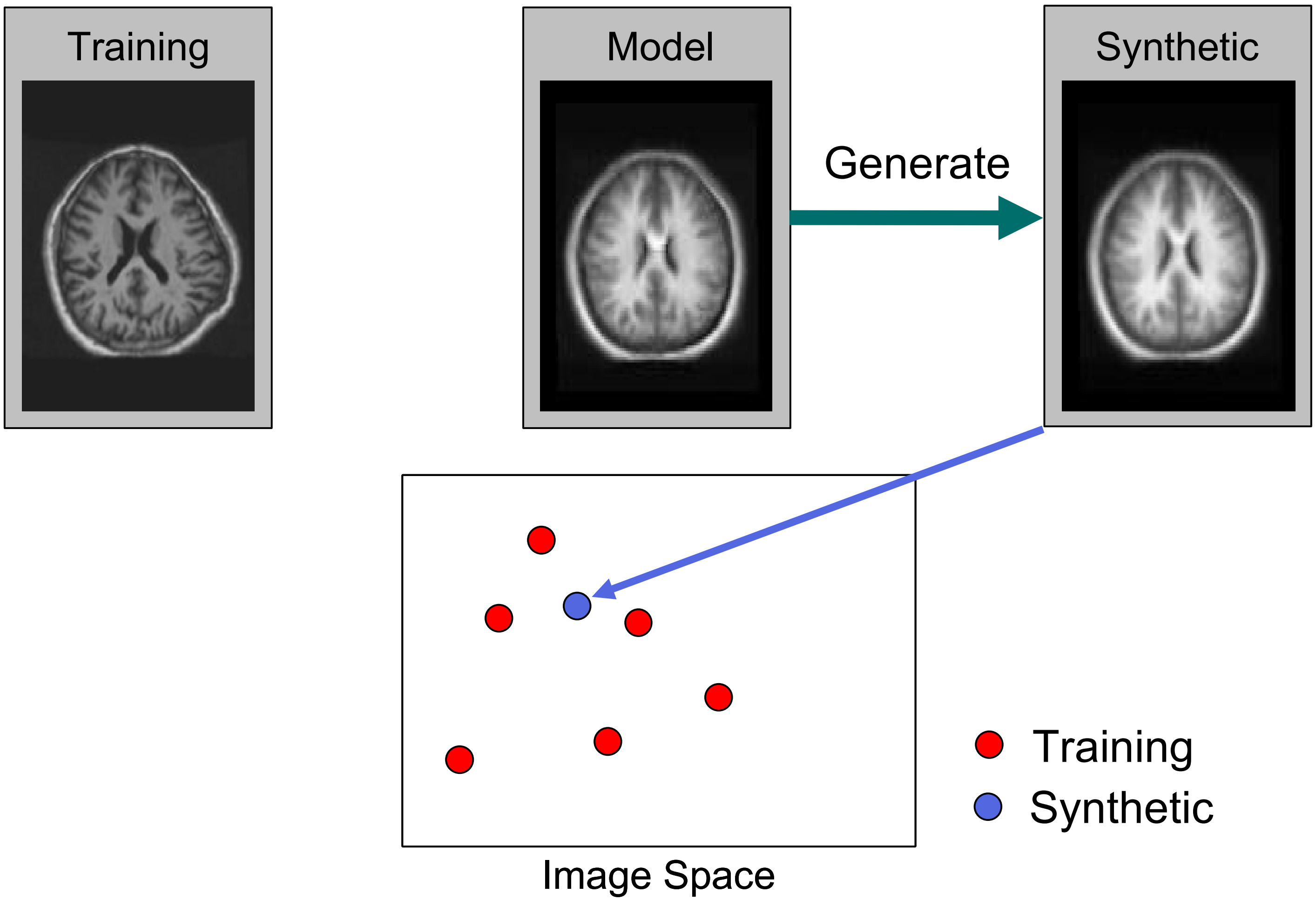
# Training and Synthetic Images



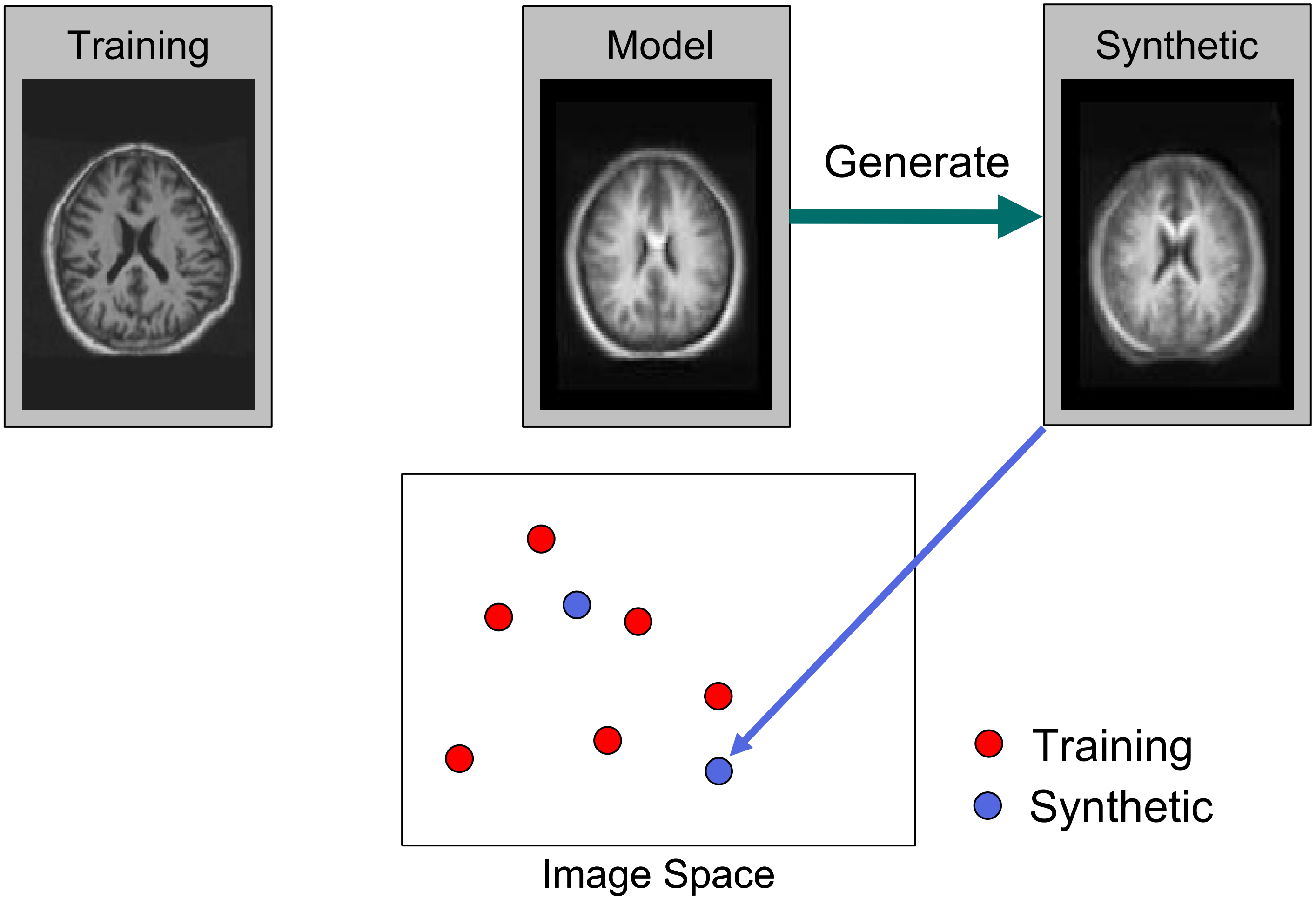
# Training and Synthetic Images



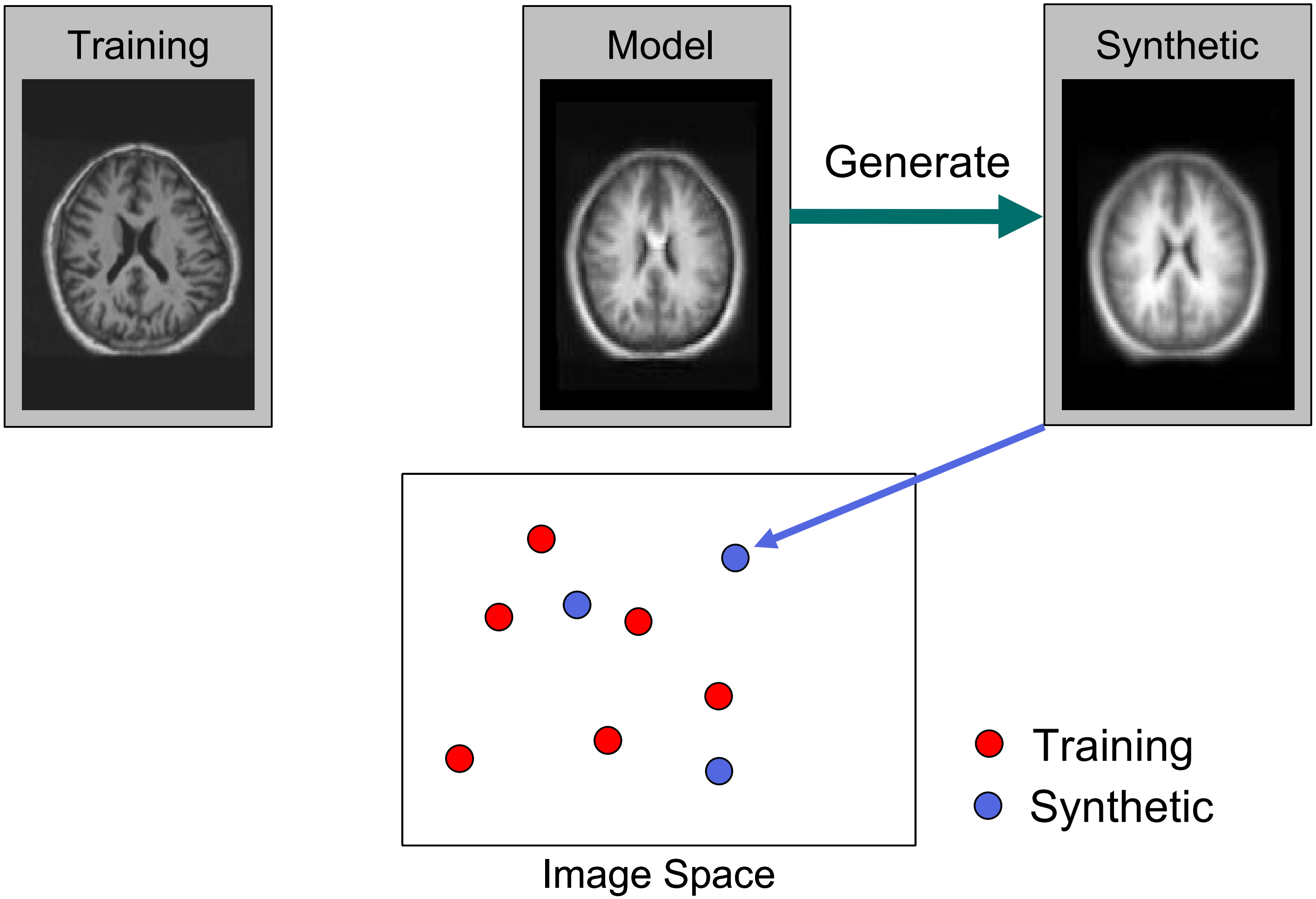
# Training and Synthetic Images



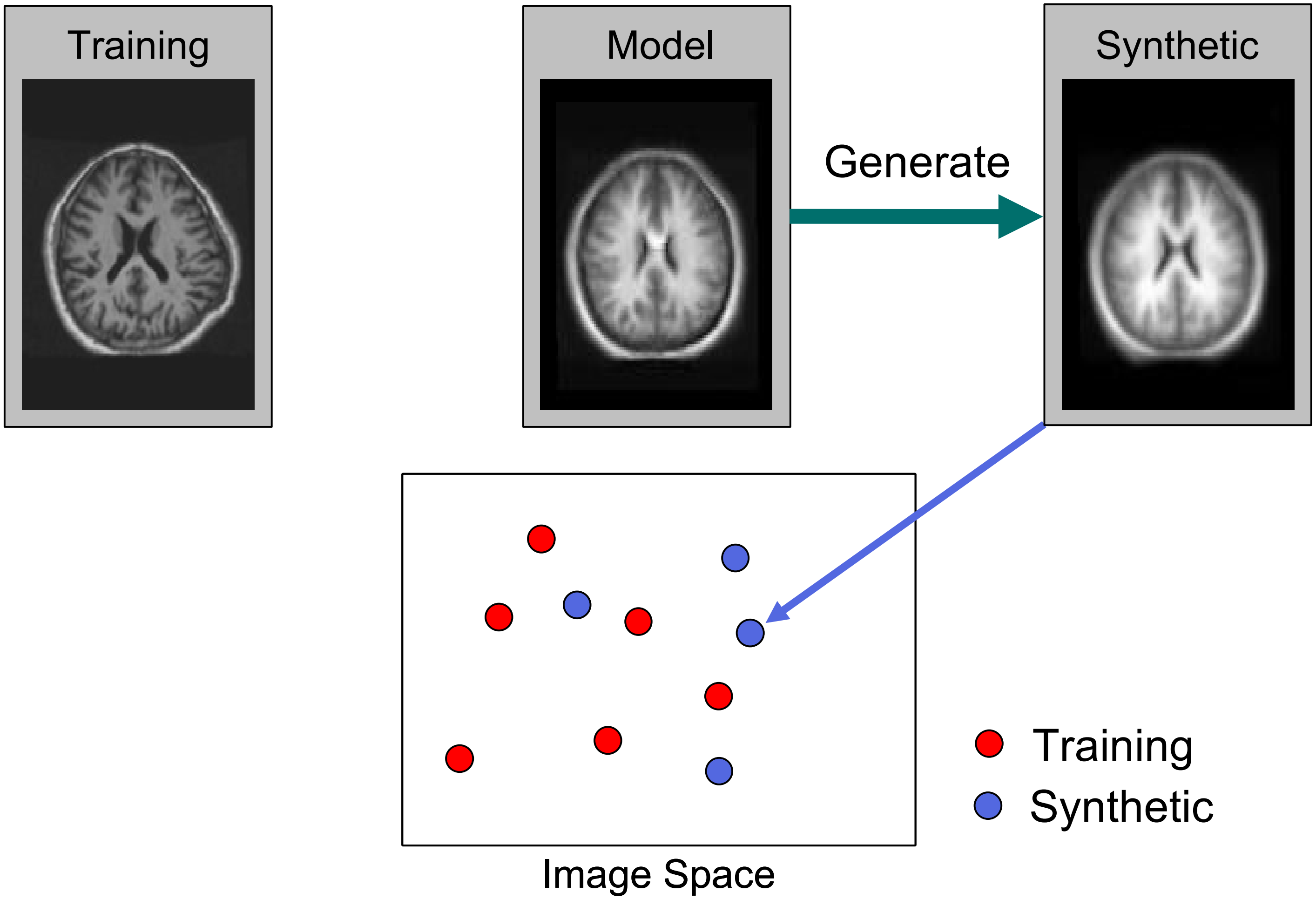
# Training and Synthetic Images



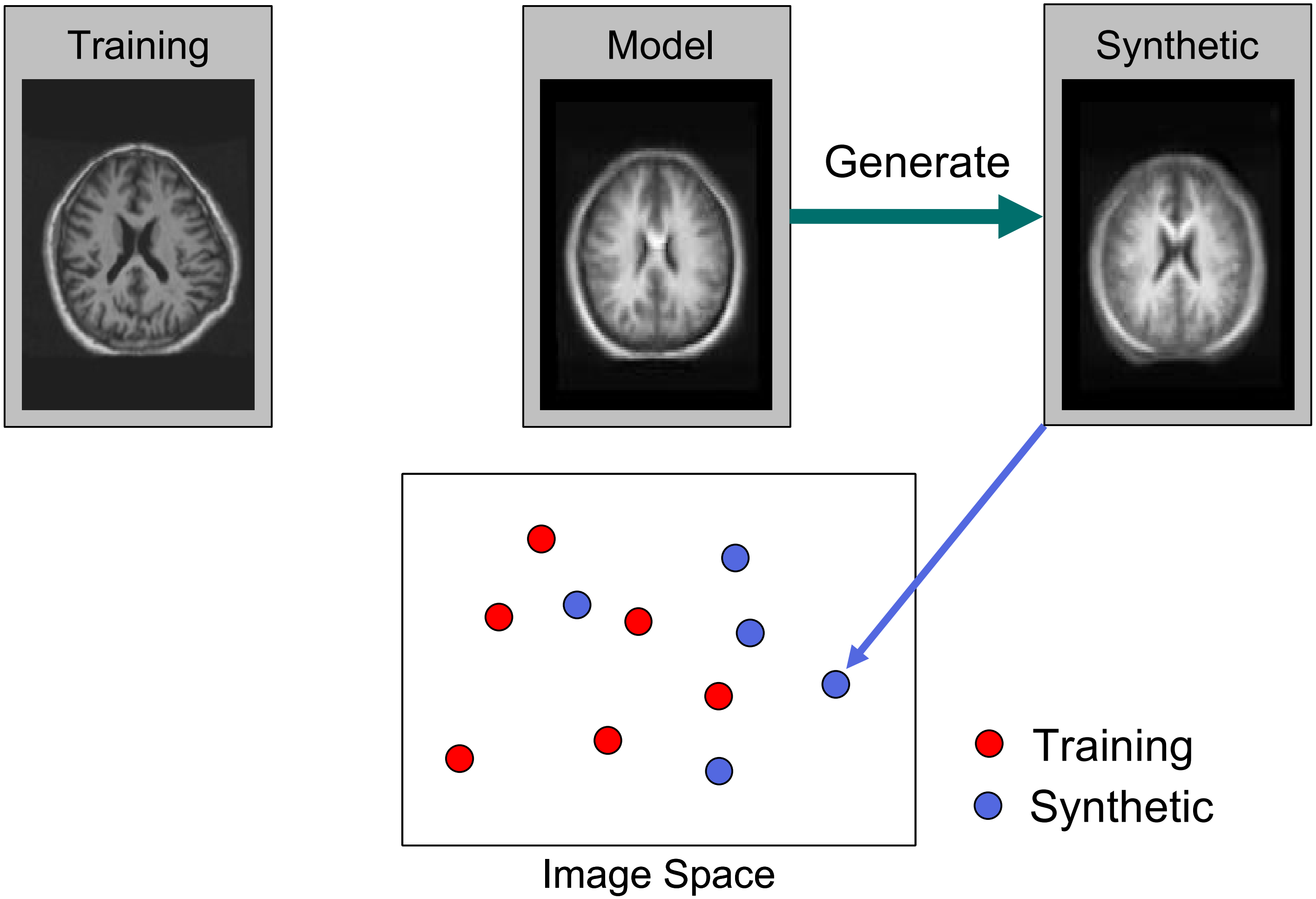
# Training and Synthetic Images



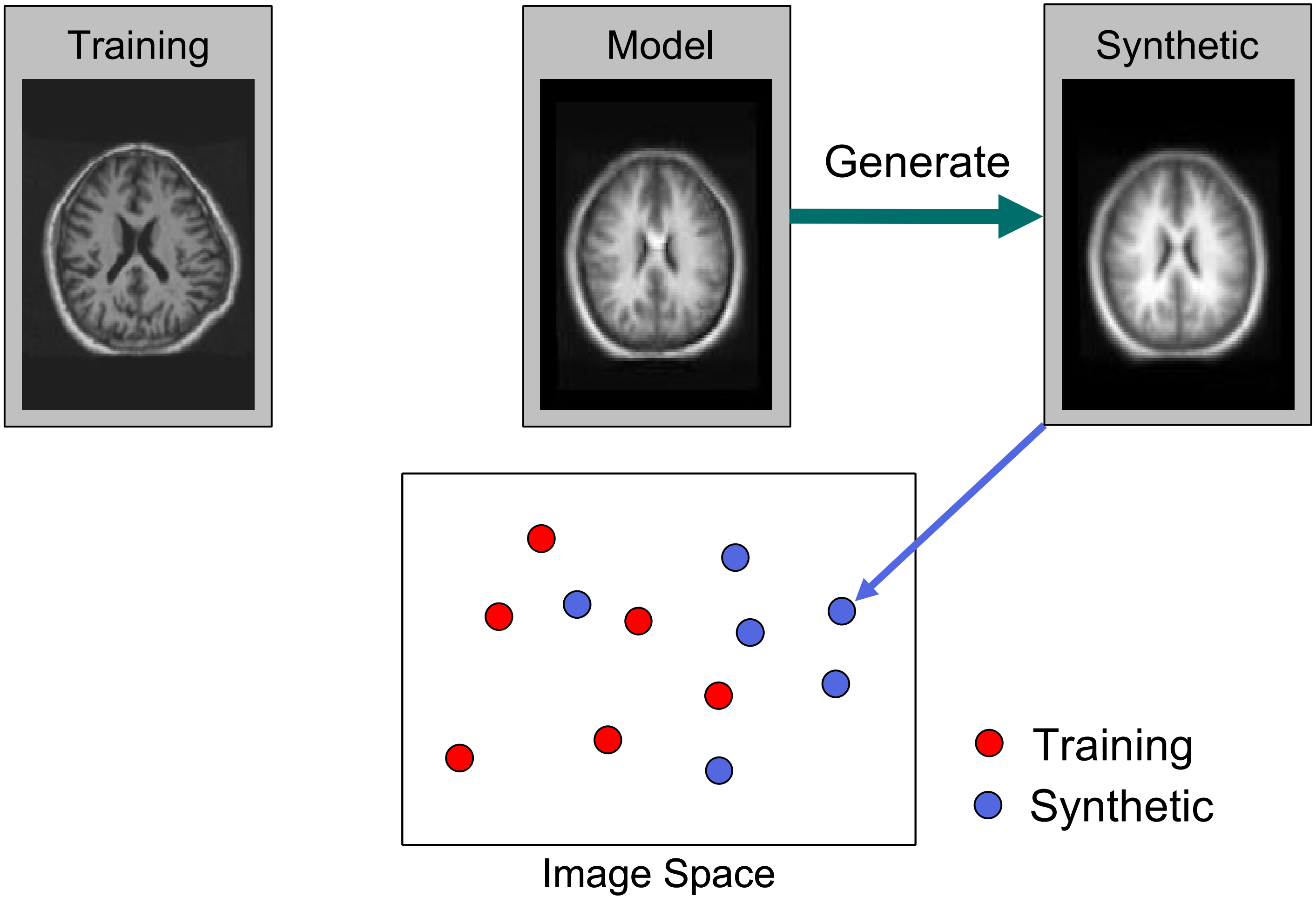
# Training and Synthetic Images



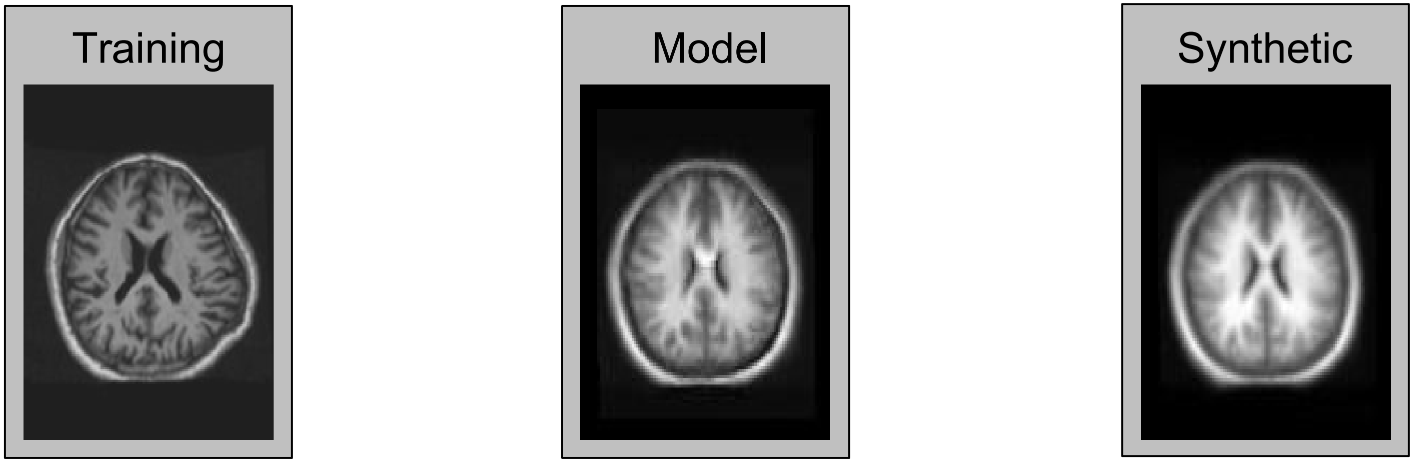
# Training and Synthetic Images



# Training and Synthetic Images



# Training and Synthetic Images



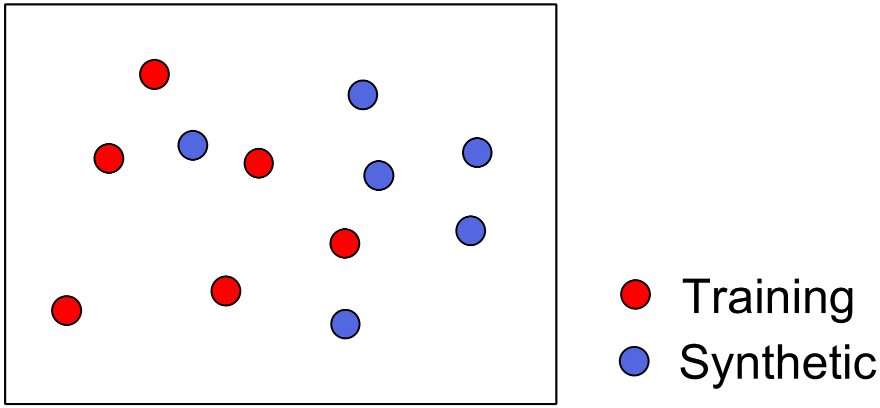
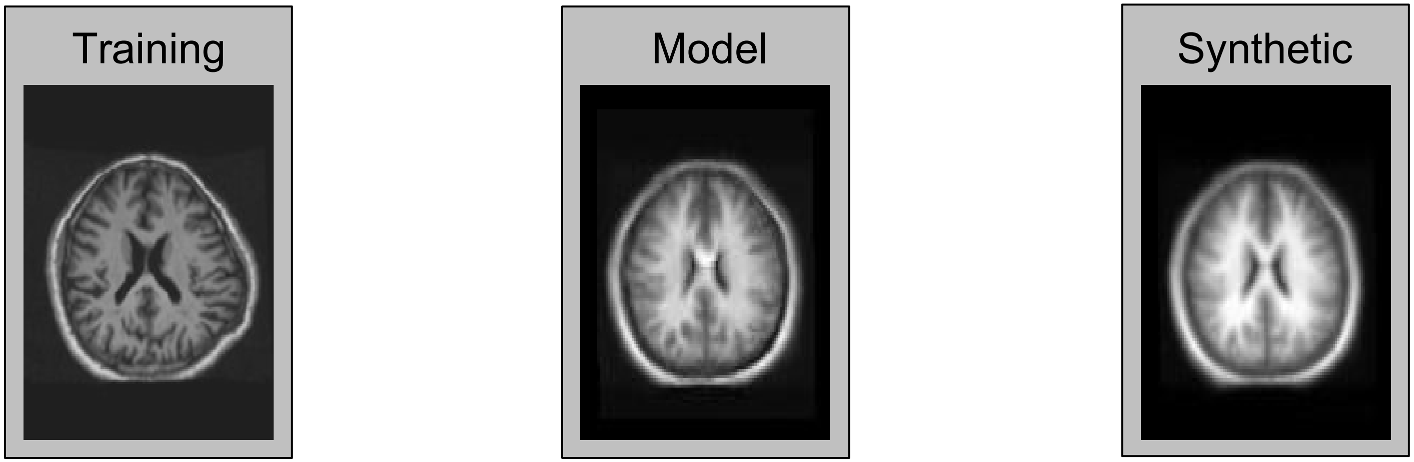




Image Space

# Training and Synthetic Images



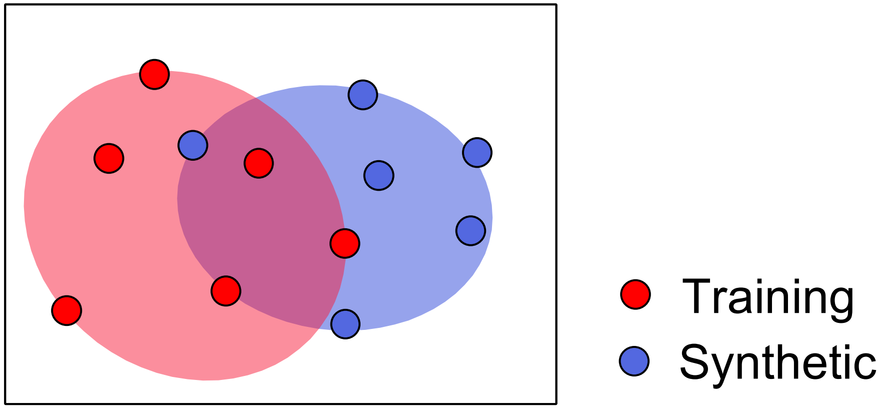
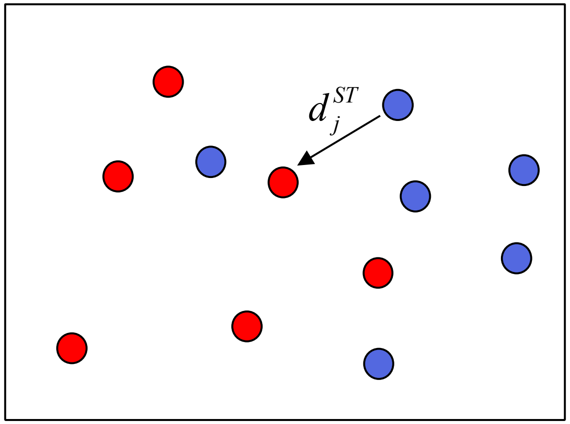




Image Space

# 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

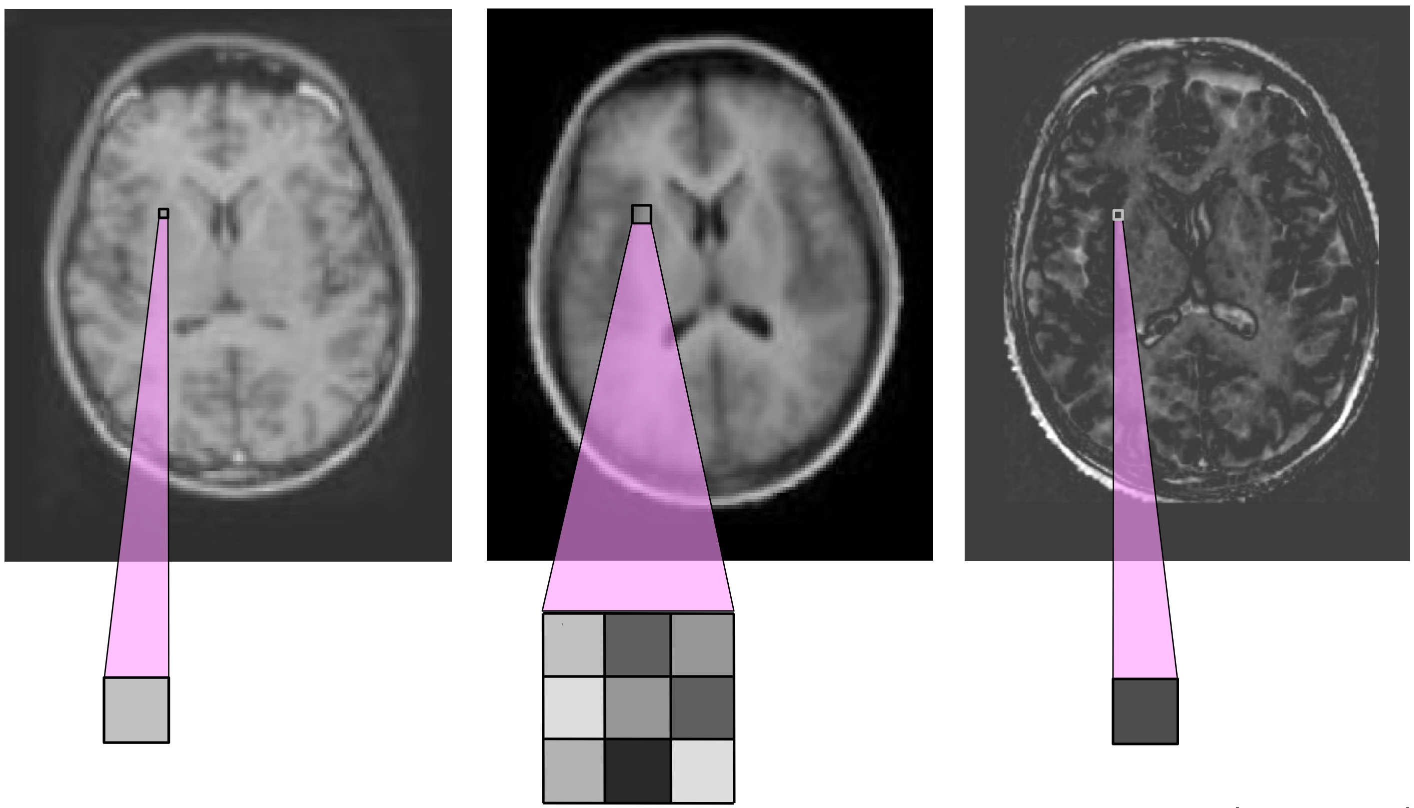
- $d$ can be the Euclidean or shuffle distance between images
- Better models have smaller distances, $d$
- We plot {-Specificity}, which decreases with misregistration

# Measuring Inter-Image Distance

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

# Shuffle Distance

Image $A$ | Image $B$ | Difference Image $\Delta S$



$A_i$ | $B_{ij}$

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

# Varying Shuffle Radius

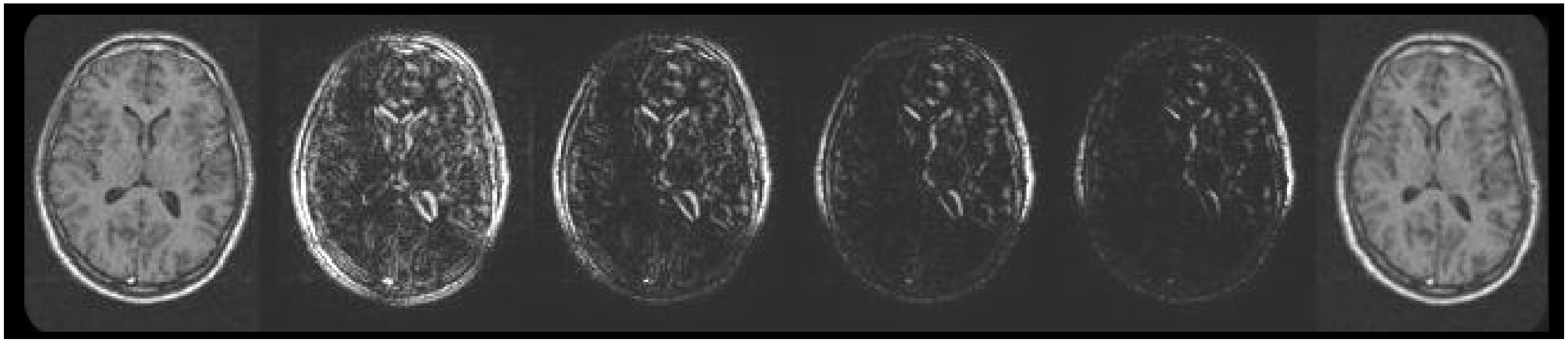


Image $A$    r = 1    r = 1.5    r = 2.1    r = 3.7    Image $B$

The University of Manchester

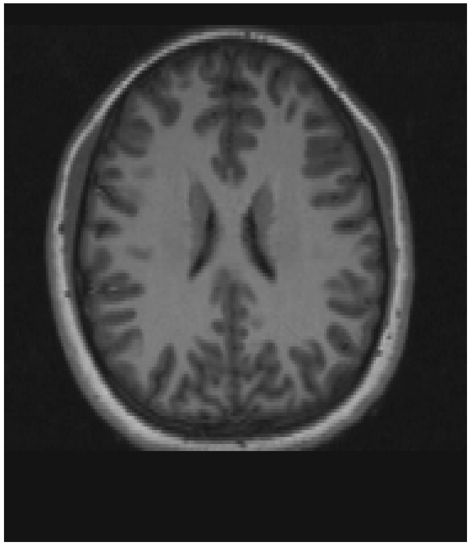
# Validation Experiments

# Experimental Design

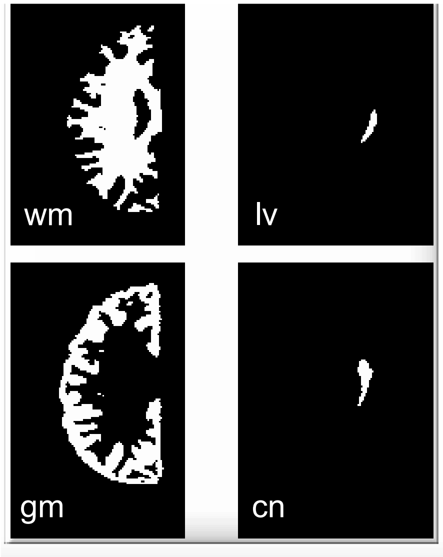
- MGH dataset (37 brains)
- Selected 2D slice
- Initial ‘correct’ NRR
- Progressive perturbation of registration
  - 10 random instantiations for each perturbation magnitude
- Comparison of the two different measures
  - overlap
  - model-based

# Brain Data

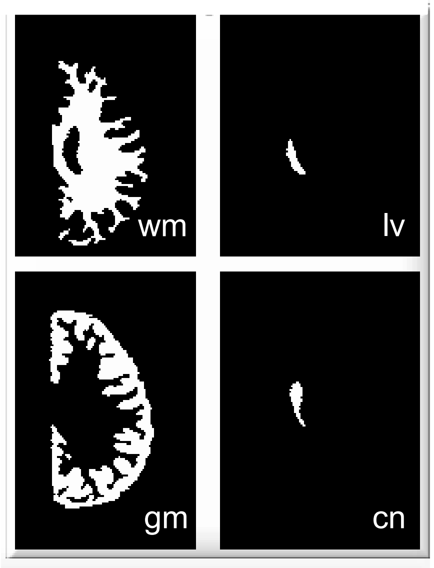
- Eight labels per image
  - L/R white/grey matter
  - L/R lateral ventricle
  - L/R caudate nucleus



Image



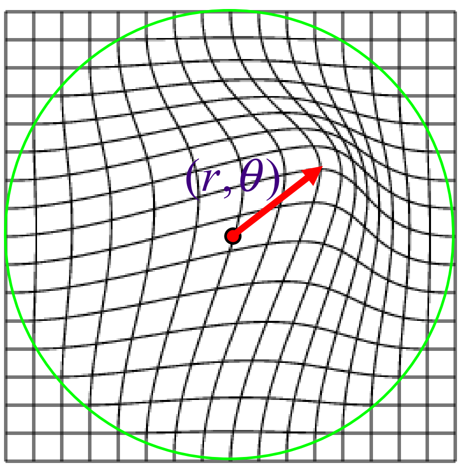
LH Labels



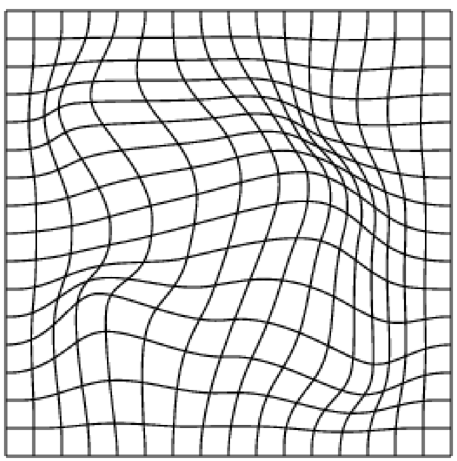
RH Labels

# Perturbation Framework

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

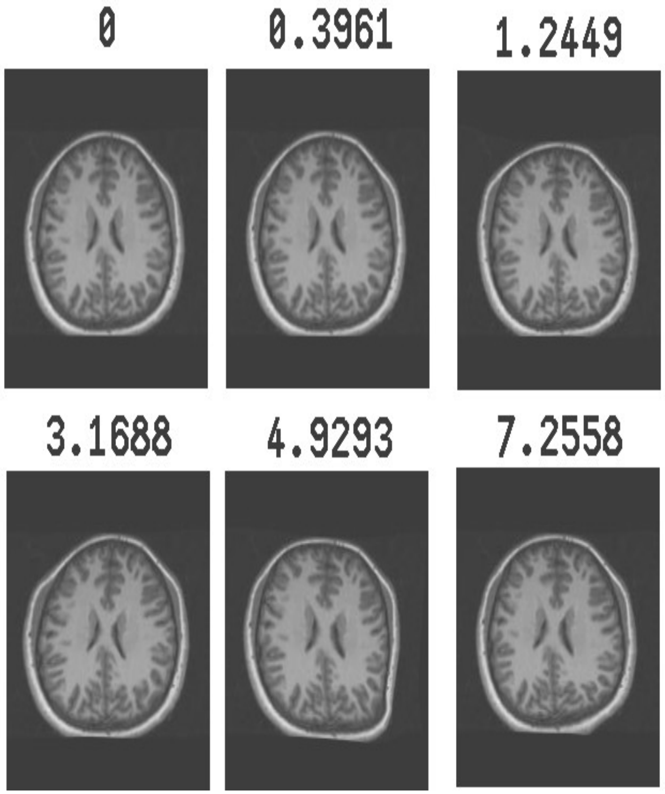



CPS with 1 knot point

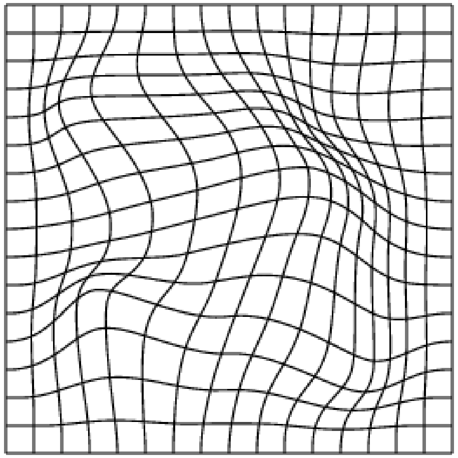


Multiple knot points

# Examples of Perturbed Images



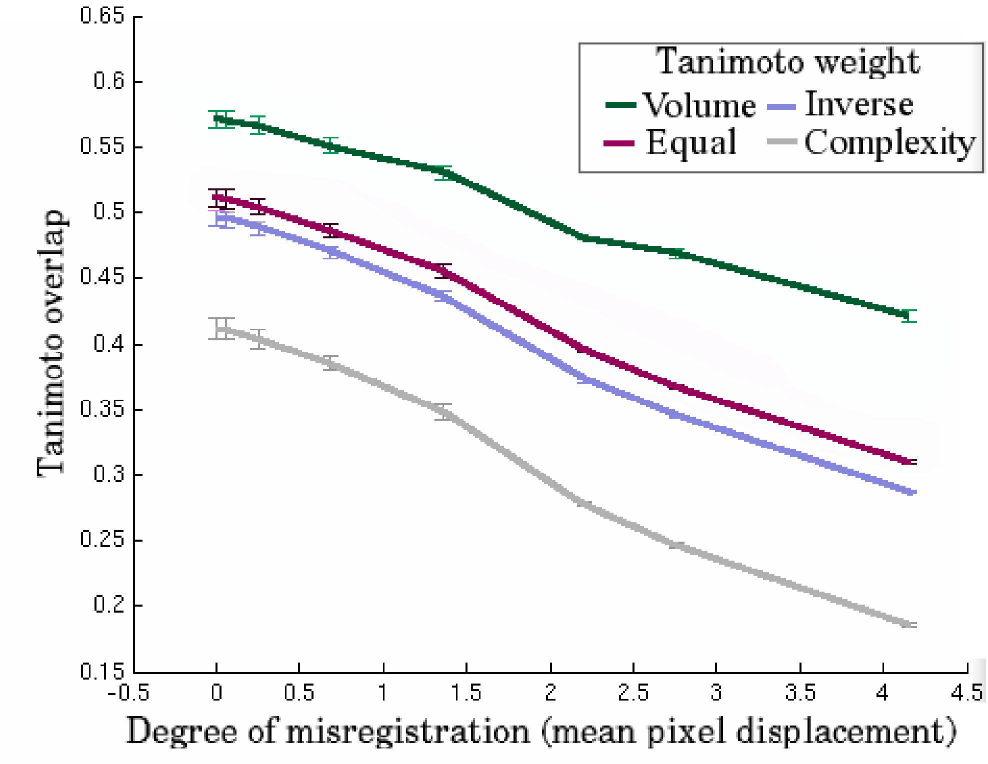

Increasing mean pixel displacement



Example warp

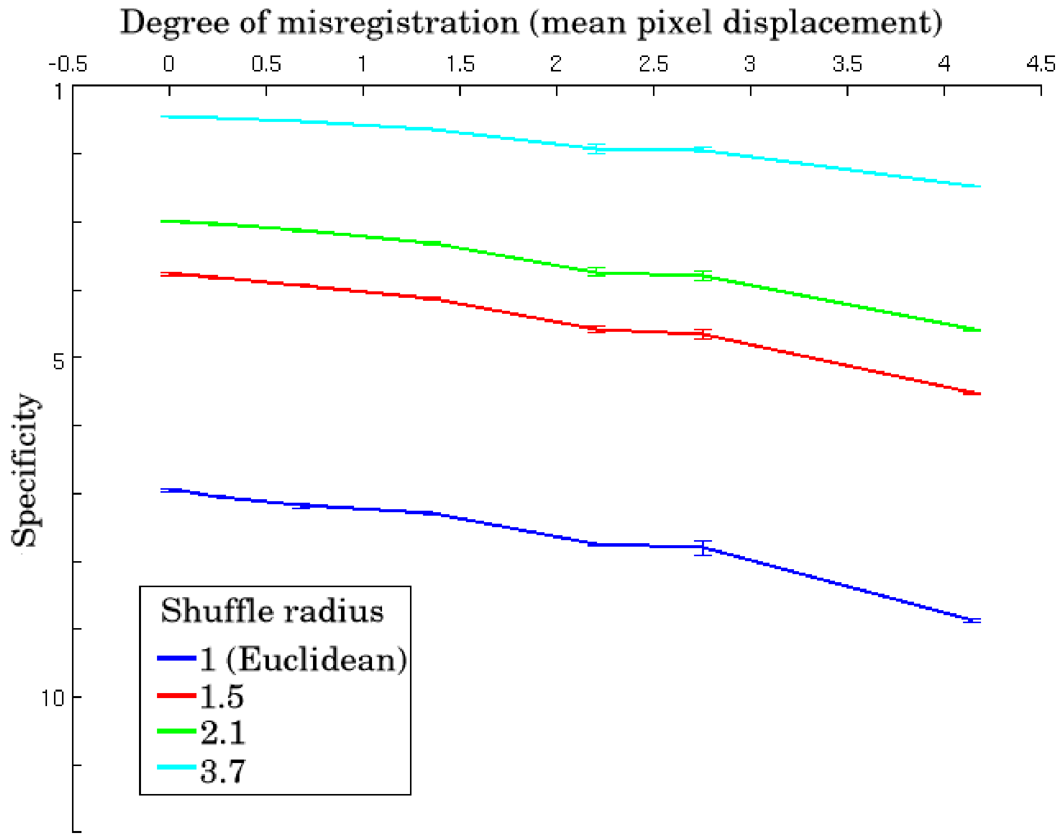
# Results – Generalised Overlap

- Overlap decreases monotonically with misregistration



# Results – Model-Based

- Measures increase monotonically with misregistration

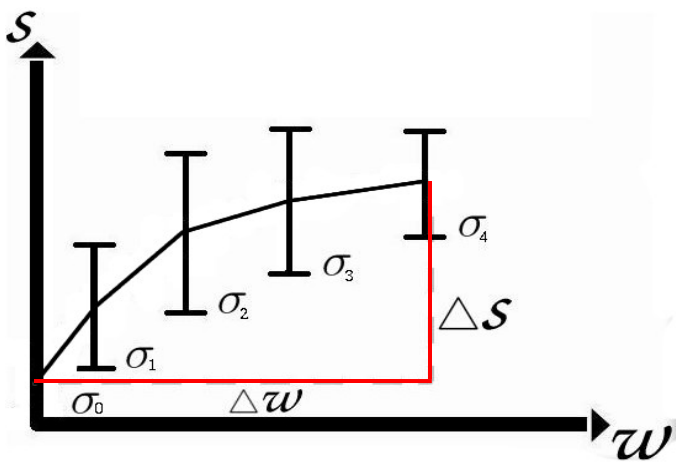


The University of Manchester

MANCHESTER 1824

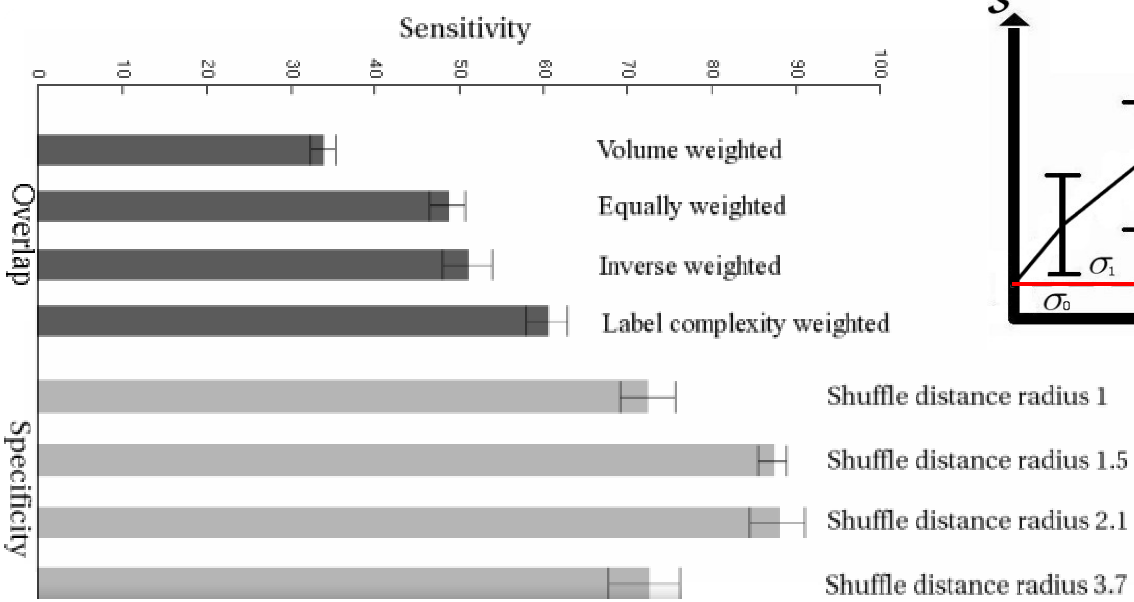
# Results – Comparison

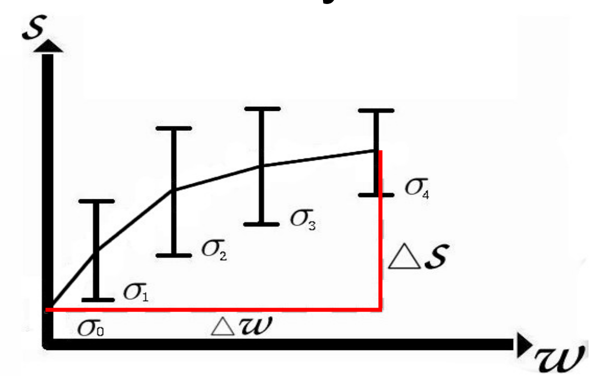
- All three measures give similar results
  - overlap-based assessment requires ground truth (labels)
  - model-based approach does not need ground truth
- Compare sensitivity of methods
  - ability to detect small changes in registration



# Results – Sensitivities

- High sensitivity = small deformations reliably detected




- Specificity most sensitive method

# Further Tests – Noise

- A measure of robustness to noise is sought
- Validation experiments repeated with noise applied
  - each image has up to 10% white noise added
  - two instantiations of set perturbation are used
- Results indicate that the model-based method is robust
  - changes in Generalisation and Specificity remain detectable
  - curves remain monotonic
  - noise can potentially exceed 10%

The University of Manchester

MANCHESTER 1824

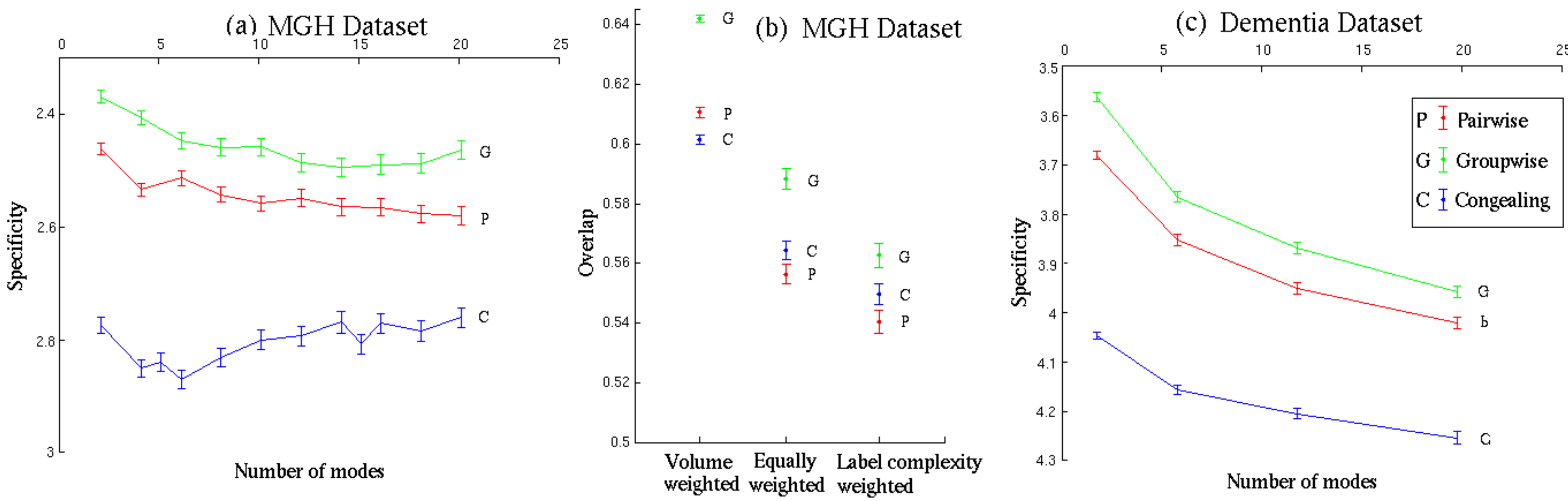
# Practical Application

# Practical Application

- 3 registration algorithms compared
  - Pair-wise registration
  - Group-wise registration
  - Congealing
- 2 brain datasets used
  - MGH dataset
  - Dementia dataset
- 2 assessment methods
  - Model-based (Specificity)
  - Overlap-based

# Practical Application - Results

- Results are consistent
- Group-wise NRR outperforms pair-wise, which outperforms congealing



MANCHESTER
1824

The University
of Manchester

# Extension to 3-D

- 3-D experiments
- Work in progress
  - validation experiments laborious to replicate
  - comparison of 4-5 NRR algorithms
- Fully-annotated IBIM data
- Results can be validated by measuring  label overlap

# Conclusions

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