DARQ: Deep learning of quality control for stereotaxic registration of human brain MRI

Neuro-iX

Introduction

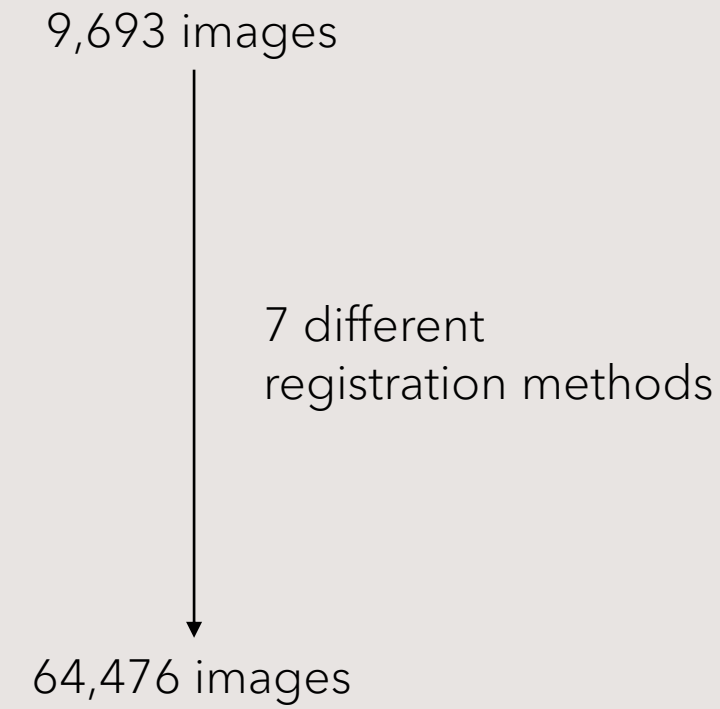
- Registration is a crucial step in MRI preprocessing
  - Preprocessing pipeline's success is often highly dependent on registration.
- Registration has a low success rate with abnormal brains (Atrophy or space-occupying lesions)
  - Tools often developed and tested on young healthy subjects
- Expert rating of registration quality is long
- Method focused on quality control of T1 weighted MRI scans with real data
  - Comparison with a method using artificially generated data
  - Aim to minimize false positive rate

Datasets

TRAIN / VALIDATION / TEST	GENERALIZATION TEST
• Alzheimer's Disease Neuroimaging Initiative (ADNI) • Parkinson Progression Marker Initiative (PPMI) • Pre-symptomatic Evaluation of Novel or Experimental Treatments for Alzheimer's Disease (PREVENT-AD) • Human Connectome Project (HCP)	• International Progressive Multiple Sclerosis Alliance (IPMSA)

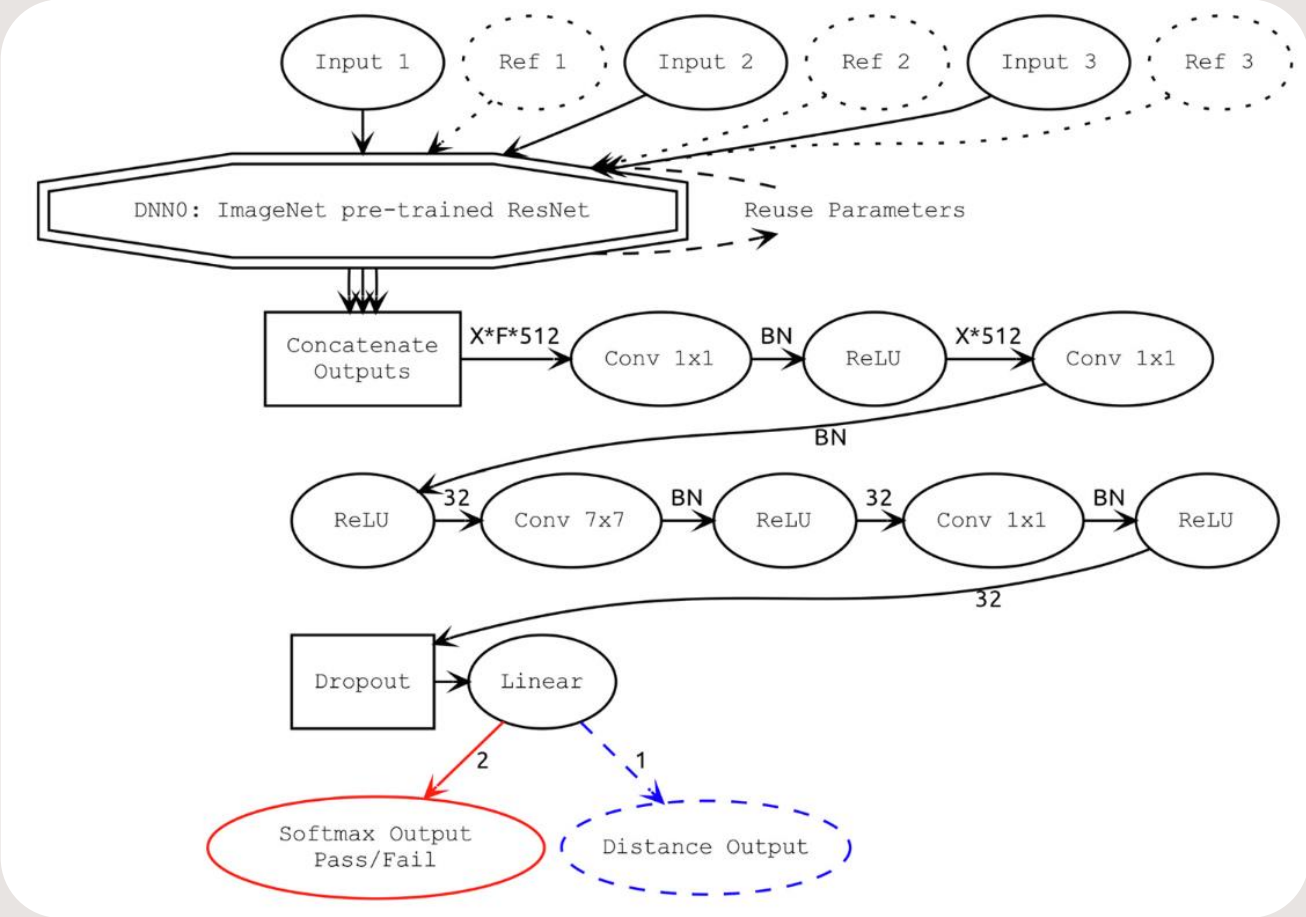
Registration Technics

- Different registrations applied
  - MRITOTAL (two versions)
  - BestLinReg
  - Revised BestLinReg (two versions)
  - Elastix
  - ANTS
- Manual QC from previous study
- 54,458 Pass (84.5%) / 10,018 Fail (15.5%)



Models

- Simple set of 3 images (with references)
  - Axial
  - Sagittal
  - Coronal
- Standard ResNET as a feature extractor
  - Model with 18, 34, 50, 101, 152 layers
  - Collapsed Input layer to accept grayscale images
  - Removed the last layers
  - Processed each image individually and concat outputs for final layers
- Two tasks: Classification or distance estimation



Into The Unclear

SILVER STANDARD	SYNTHETIC DATASET
• Average of all transformations that passed manual QC for a subject • Compute the distance between silver standard and individual transform for labelisation	• Create a synthetic dataset for distance estimation • Threshold distance to determine Pass / Fail labels • Not clear how samples where generated

Data Augmentations

CLASSIFIER

- Downsample Z by random factor (1-3)
  - Simulate thick slices
- Randomly crop top or bottom 20%
  - Simulate restricted FOV
- Random rotation on x, y, and z of [-0.1,0.1] degrees
  - Simulate small imperfections of registration parameters
- Random flip along x (left-right)
- Generate 5 for 1 => 322,460 samples

DISTANCE ESTIMATION

- Only use samples where manual QC was passed
- Generated random transformations with uniform distribution of distances from the “silver standard”
- Random flips along x
- Random cropping in z direction
- Generate 65,257 samples

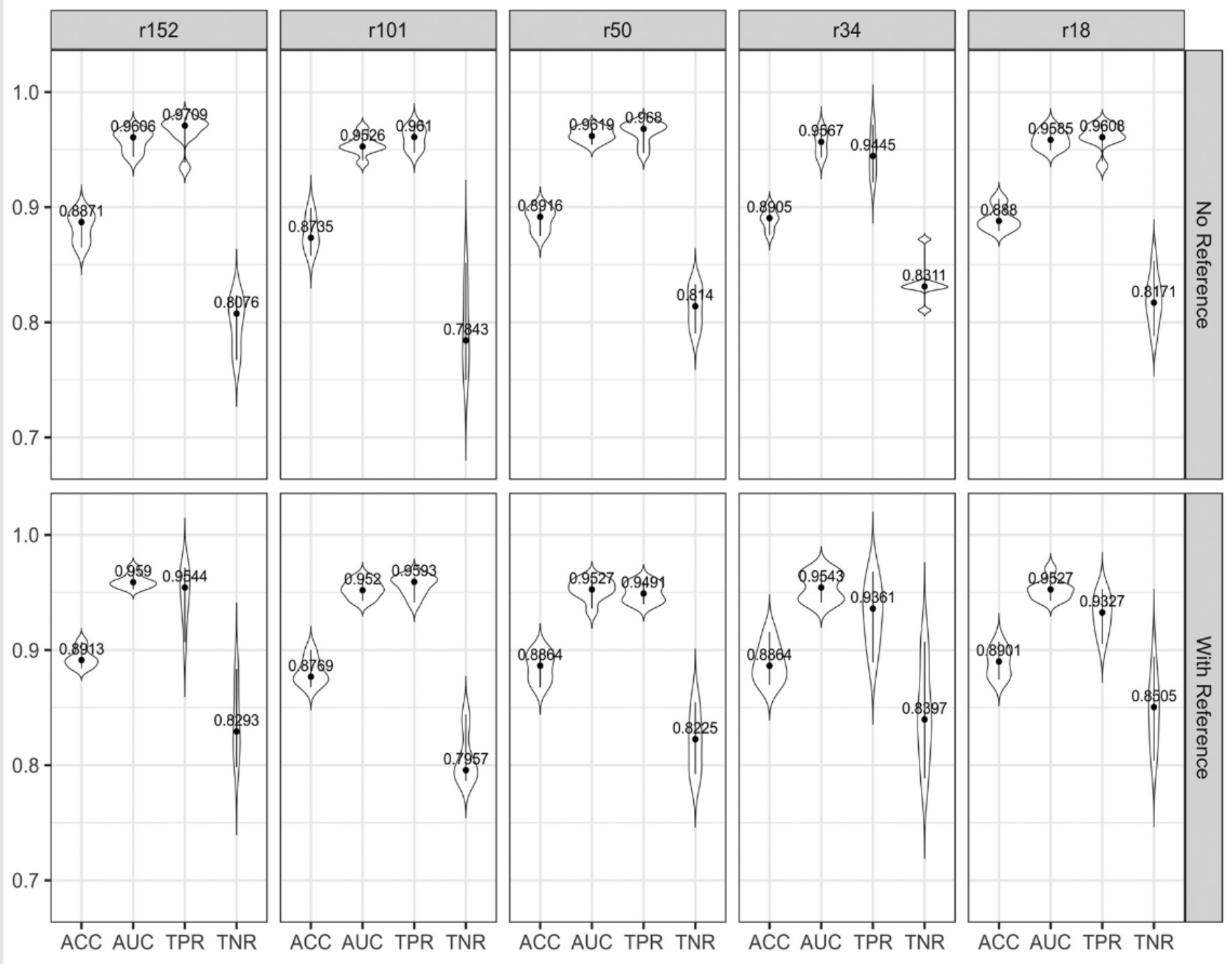
Training

- 8 fold cross validation
- Validation and testing set had 50/50 pass/fail
- Preliminary experiment noticed overfitting after 10 epochs
- Create a set of 1200 images (50/50 ratio) from IPMSA to test generalization capacity

Results

- QCResNet18 : TNR of 0,85 and AUC of 0,952
- DistResNet152 : AUC of 0,9356
- Achieve better performance with real data (compared to synthetic)
- QCResNet18 on IPMSA :

ACC	TPR	TNR	AUC
96.1%	96.7%	95.5%	98.7%


r152
r101
r50
r34
r18
No Reference
With Reference
1.0
0.9
0.8
0.7
0.8871
0.9606
0.9709
0.8076
0.8735
0.9526
0.961
0.7843
0.8916
0.9619
0.968
0.814
0.8905
0.9567
0.9445
0.8311
0.888
0.9585
0.9608
0.8171
0.8913
0.959
0.9544
0.8293
0.8769
0.952
0.9593
0.7957
0.8864
0.9527
0.9491
0.8225
0.8864
0.9543
0.9361
0.8397
0.8901
0.9527
0.9327
0.8505
ACC
AUC
TPR
TNR

Discussion

- Some failed classification case are limit cases
- The performance of the model was comparable with intra rater variability (test-retest accuracy of 93%)
- All methods had better performance in TNR when using references