Out of Sample Performance of a Deep Learning Based Registration Quality Assurance Method

Xiaoxiao Zhou, Shaikat M Galib, Hyoung K. Lee, Geoffrey Hugo
Department of Radiation Oncology
Washington University in St. Louis School of Medicine

Computational Radiotherapy Lab (CORAL)
Department of Radiation Oncology
Washington University School of Medicine
Barnes Jewish Hospital
St. Louis, Missouri

INTRODUCTION

- Deformable image registration(DIR) is widely used in radiation oncology, despite lacking a ground-truth for validating individual results.
- Quality assurance(QA) of DIR in the clinic relies on either visual inspection, or performance in phantom, which are time-consuming and impractical.

AIM

- To develop a QA Deep Neural Network(DNN) for DIR of thoracic CT images. (previous work)
- To evaluate how this algorithm generalizes to diverse datasets and registration algorithms.

METHODS

a) Register two thoracic CT datasets(with annotated landmarks) with two DIR algorithms

Datasets		DIR algorithms	
Dirlab	Long4DCT(in-house)	B-spline	Dramms

b) Training data preparation

- Input:
Randomly resample 32*32*32 patches from fixed image / moving image/ jacobian map, and fuse them into one image as three channels.
- Gound truth:
Calculate mean landmark distance error of each patch, thresholded as 0 – acceptable / 1 – need to be review

c) Training and testing QA DNN model (Figure 2)

d) Performance evaluation (Robustness & Accuracy) (Table 1, Figure 1)

RESULTS

Training Data	Training Registration	Held-out Data	Held-out Registration	Held-out Accuracy	Held-out AUC	Out-of-sample Data	Out-of-sample Registration	Out-of-sample Accuracy	Out-of-sample AUC
Dirlab	B-spline	Dirlab	B-spline	0.95	0.99	Long4DCT	B-spline	0.88	0.94
						Long4DCT	DRAMMS	0.87	0.94
						Dirlab	DRAMMS	0.89	0.96
Dirlab	B-spline	Dirlab	B-spline	0.95	0.99	Long4DCT	B-spline	0.88	0.94
Dirlab	DRAMMS	Dirlab	DRAMMS	0.95	0.99	Long4DCT	DRAMMS	0.87	0.93
Dirlab	B-spline	Dirlab	B-spline	0.95	0.99	Dirlab	DRAMMS	0.92	0.97
Long4DCT	B-spline	Long4DCT	B-spline	0.91	0.97	Long4DCT	DRAMMS	0.90	0.96

Table 1: **Overall Accuracy and Area under Curve(AUC) of our model**

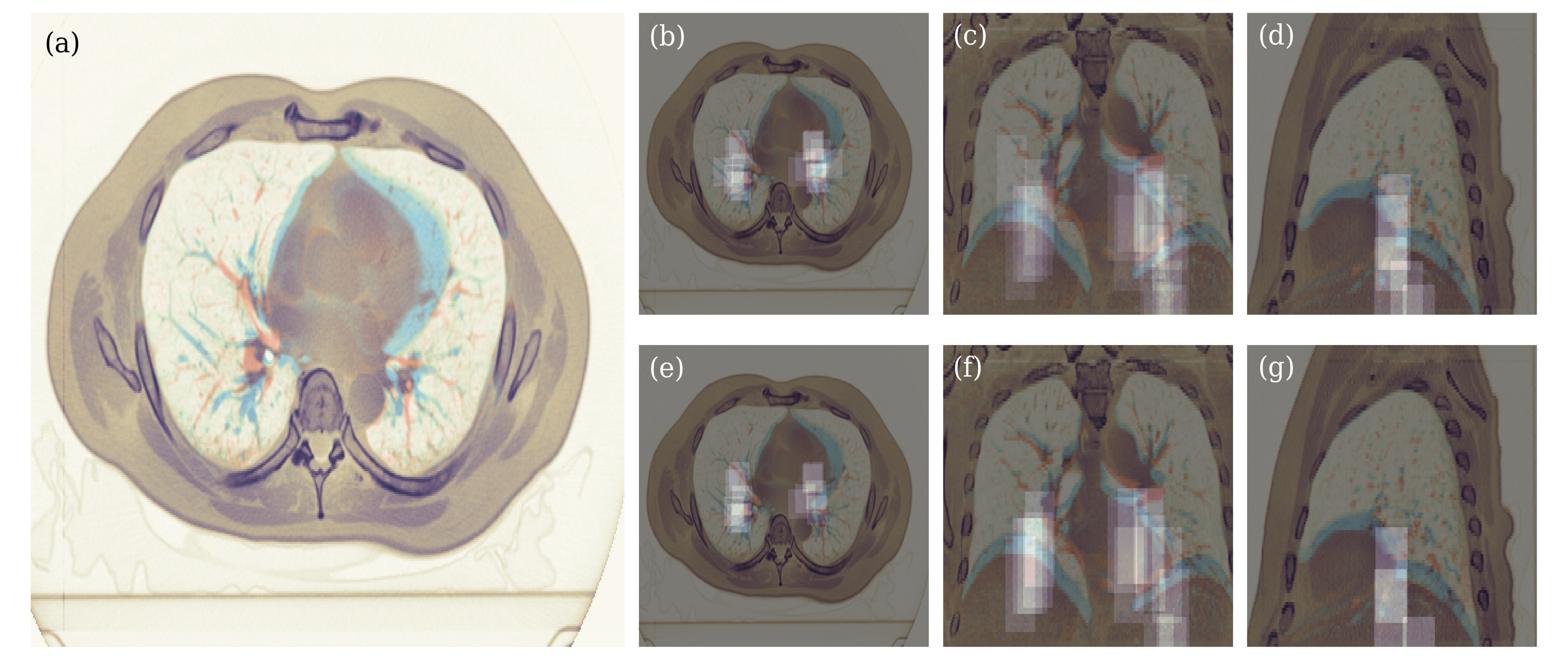


Figure 1:
(a), overlapping of original fixed and registered moving image
(b)-(d), overlay of predicted label map in 3 axes
(e)-(g), overlay of true label map in 3 axes

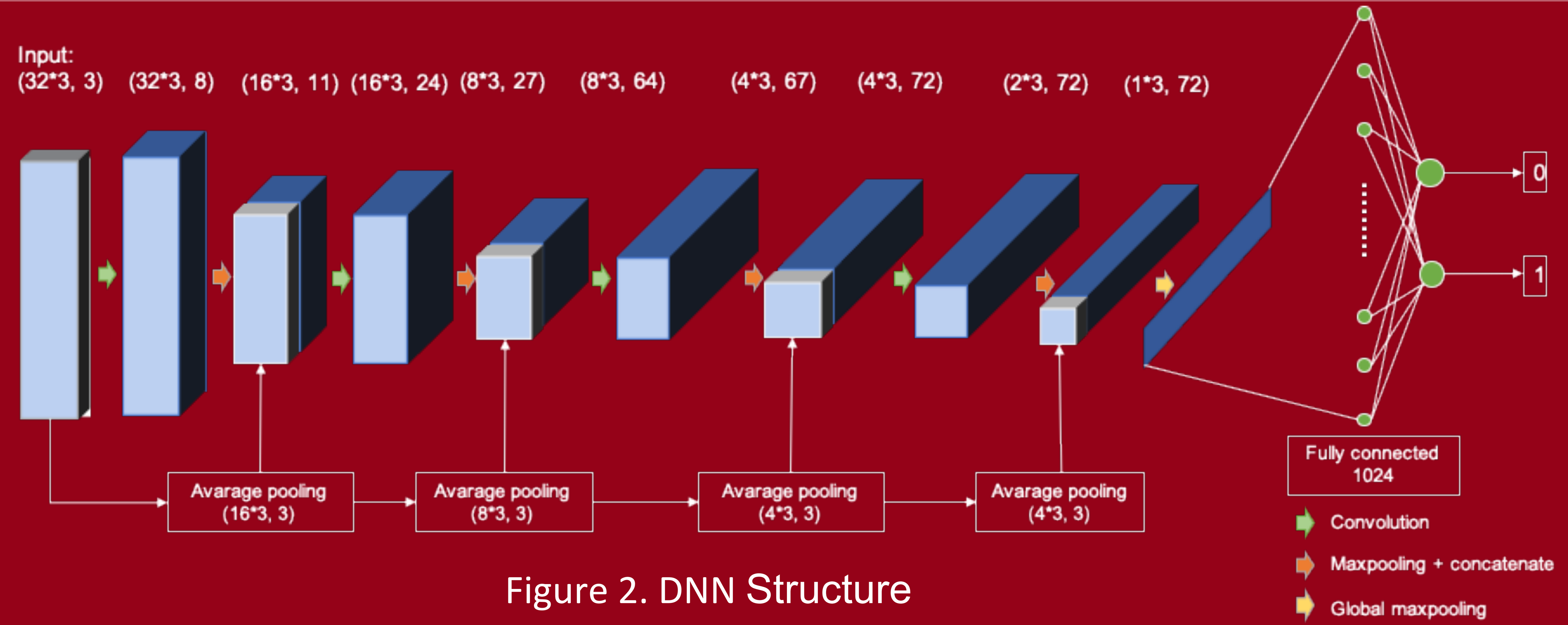
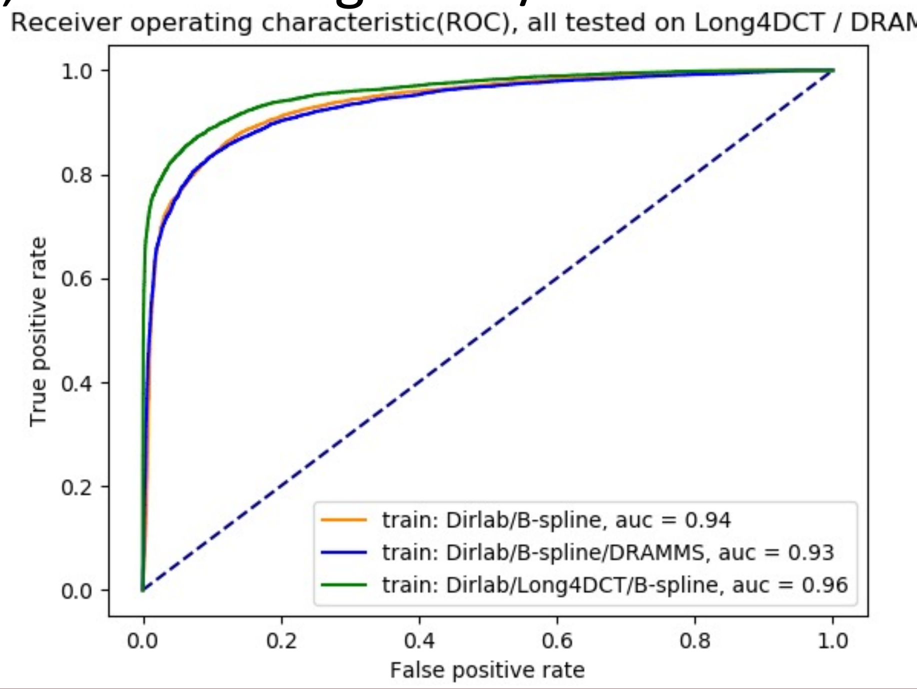


Figure 2. DNN Structure

RESULTS

- The Area Under Curve (AUC) for training Dirlab / B-spline tested on held out Dirlab / B-spline was 0.99, on Long4DCT / B-spline was 0.94, on Dirlab / DRAMMS was 0.96, and on Long4DCT / DRAMMS was 0.94.
- Training on multiple registration algorithms Dirlab / B-spline / DRAMMS gave AUC 0.99 for Dirlab / B-spline, 0.94 for Long4DCT / B-spline, 0.99 for Dirlab / DRAMMS, and 0.93 for Long4DCT / DRAMMS.
- By training on Dirlab / Long4DCT / B-spline, the AUC for testing on Dirlab / B-spline was 0.99, for Long4DCT / B-spline was 0.97, for Dirlab / DRAMMS was 0.97, and for Long4DCT / DRAMMS was 0.96.



CONCLUSIONS

- Our registration QA algorithm showed reliable inference ability across various datasets and DIR, however, training with diverse datasets provided slightly better performance than by adding DIR algorithms.
- Enriching training data by different dataset combinations obtained some improvement, demonstrating that our model's robustness has the potential for improvement.

REFERENCES

- Galib, Shaikat M., et al. "A fast and scalable method for quality assurance of deformable image registration on lung CT scans using convolutional neural networks." Medical physics 47.1 (2020): 99-109.

Contacts: xiaoxiao.zhou@wustl.edu, gdhugo@wustl.edu