Multi-View Vertebra Localization and Identification from CT Images

Han Wu[1], Jiadong Zhang[1], Yu Fang[1], Zhentao Liu[1], Nizhuan Wang[1],
Zhiming Cui [1,✉], and Dinggang Shen [1,2,3,✉]

[1] School of Biomedical Engineering, ShanghaiTech Univerisity, Shanghai, China
[2] Shanghai United Imaging Intelligence Co. Ltd., Shanghai, China
[3] Shanghai Clinical Research and Trial Center, Shanghai, China
cuizhm@shanghaitech.edu.cn, dgshen@shanghaitech.edu.cn

BACKGROUND

Problem: Automatic Localization and identification of vertebra from CT images are crucial in clinical practice, particularly for surgical planning, pathological diagnosis, and post-operative evaluation.

Challenges:

- Current SOTA methods solved vertebrae labeling task in patch-wise methods where cropped patches are always used as inputs for 3D model instead a whole 3D volume, which limits the amount of global information to be learned.
- The intrinsic sequential structure embedded along the spine vertebrae can hardly be well-captured with existing 3D patch-wise methods.
- Complex training and inference settings, with cropping and fusion patches strategy, requiring careful attention for its hyper-parameters(patch size, overlap etc.).

EVALUATION METRIC

Localization Error(L-Error): The average Euclidean distance between the ground truth centroids and predictions.

Identification Rate(Id-Rate): The ratio of correctly identified vertebrae to the total number of vertebrae.

METHOD

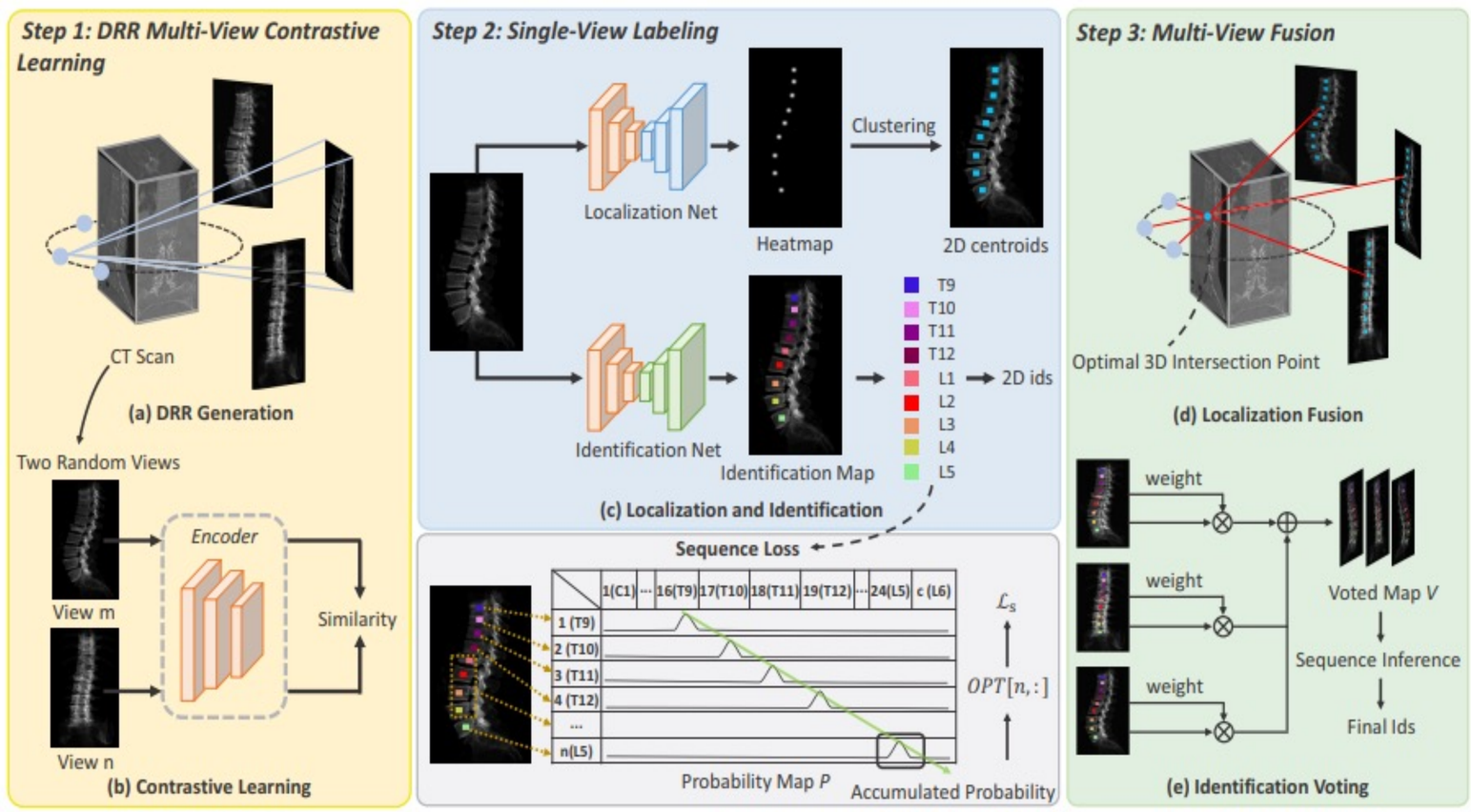


Fig.1. Overview of our proposed method

Step 1: Introduce DRR to transform a 3D labeling tasks into a set of 2D tasks and pretrain the backbone with proposed contrastive learning strategy.

Step 2: Finish the 2D labeling on single-view with two 2D networks where the whole projection images are used as input for global information and proposed Sequence Loss is used as extra supervision for sequential information.

Step 3: Final 3D localization and identification results are obtained by the proposed multi-view fusion strategy with a least-square algorithm and a multi-view weighted voting.

RESULTS

Method	Test Dataset		Hidden Dataset	
	Id-Rate(%)	L-Error(mm)	Id-Rate(%)	L-Error(mm)
Payer C.[17]	95.65	4.27	94.25	4.80
Lessmann N.[17]	89.86	14.12	90.42	7.04
Chen M. [17]	96.94	4.43	86.73	7.13
Sekuboyina A.[18]	89.97	5.17	87.66	6.56
Ours	**98.12**	**1.79**	**96.45**	**2.17**

Table 1. Results on VerSe19 dataset

Baseline	Pre-train	Sequence Loss	Voting	Id-Rate(%)	
				Test Dataset	Hidden dataset
✓				84.00	83.45
✓	✓			85.58	86.52
✓	✓	✓		89.41	90.54
✓	✓	✓	✓	**98.12**	**96.45**

Table 2. Ablation on each components

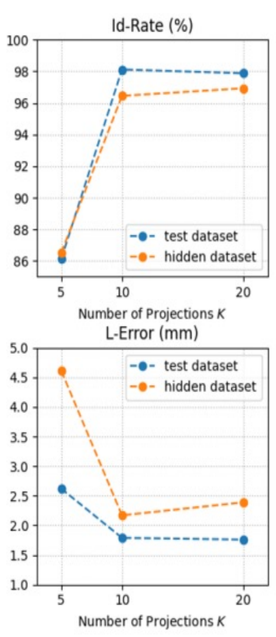
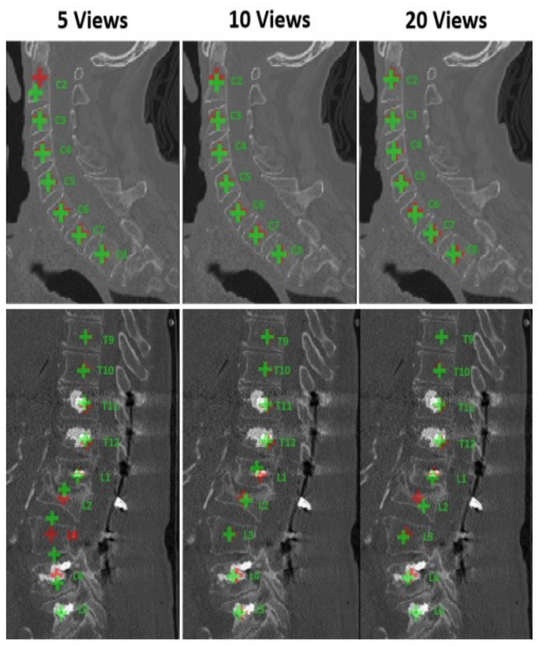


Fig.2. Results on different number of views



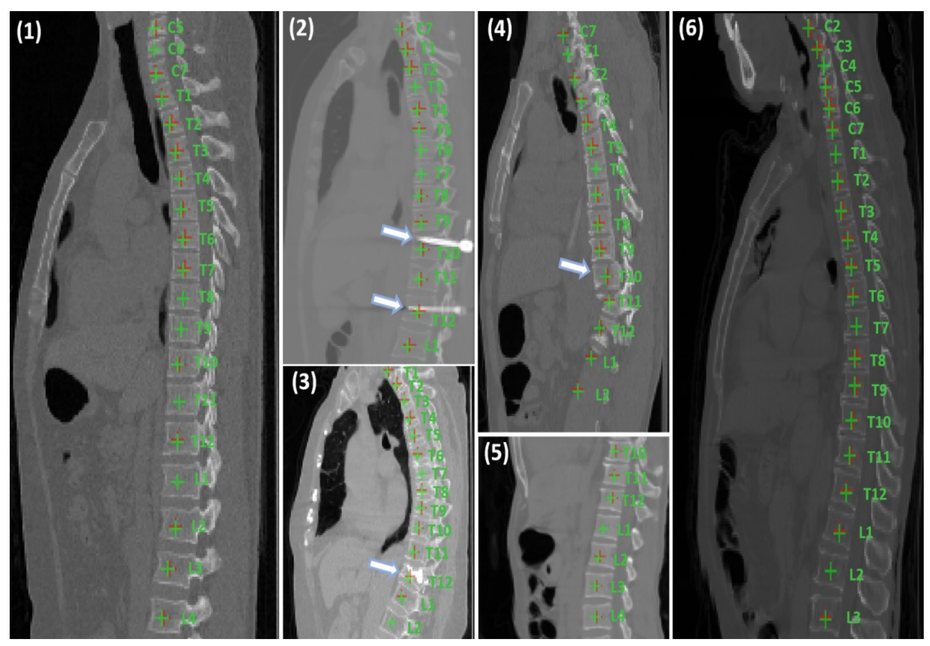


Fig.3. Qualitative results on typical challenging cases

CONCLUSIONS

A novel framework using multi-view DRR is proposed for 3D vertebrae labeling:

- We achieve superior results against other SOTA methods with just two 2D networks.
- Global information is naturally captured on 2D projecting images and sequential information is well-learnt under the supervision of our Sequence Loss.
- Multi-view fusion and our contrastive learning prove to be effective in the final labeling results

REFERENCE

[1]. Sekuboyina, A. et al.; Verse: A vertebrae labelling and segmentation benchmark for multi-detector ct images. Medical image analysis 73, 102166 (2021)

[2]. Cui, Z. et al.; Vertnet: Accurate vertebra localization and identification network from ct images. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24. pp. 281–290. Springer (2021)

CODE