Appendix: One-Shot Medical Landmark Localization by Edge-Guided Transform and Noisy Landmark Refinement

Zihao Yin[1], Ping Gong[2], Chunyu Wang[3], Yizhou Yu[4], and Yizhou Wang[5,6,7]

[1] Center for Data Science, Peking University
[2] Deepwise AI Lab
[3] Microsoft Research Asia
[4] The University of Hong Kong
[5] Center on Frontiers of Computing Studies, Peking University
[6] School of Computer Science, Peking University
[7] Inst. for Artificial Intelligence, Peking University
{silvermouse, yizhou.wang}@pku.edu.cn, gongping@deepwise.com,
chnuwa@microsoft.com, yizhouy@acm.org

1 More Visualizations

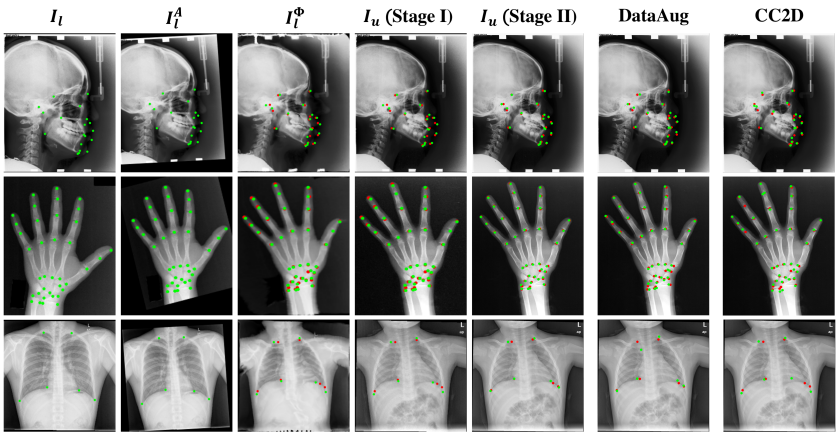

Fig. 1. Displayed images from left to right, are the exemplar (I_{src}), intermediate warped results (I_l^A, I_{src}^Φ) and the unlabeled target image I_u. **Green dots** denote locations of the transformed exemplar landmarks. **Red dots** denote the ground truth landmark locations of I_u.

As in Fig. 1, we show more challenging cases with differences not only in spatial structure, but also in appearance. Pseudo landmarks inferred by edge-guided transform generally conform to the global anatomical constraints and their local biases can be further corrected in stage II.

We further conduct experiments on the head dataset, using the same exemplar with CC2D [1] as in Tab. 1 and still achieve superior performance compared to their reported results.

2 Perspective Transformation

To adopt perspective transformation as the learned global alignment, we regress two more parameters (e_1, e_2) based on the last two elements of $o \in \mathbb{R}^8$ to make some perturbations based on the affine transformation A as follows:

$$o = \tanh(\text{MLP}(H)) \tag{1}$$

$$e_1 = o_7 * \epsilon, \ e_2 = o_8 * \epsilon \tag{2}$$

where ϵ controls the pertubation intensity and here we set ϵ as 0.1. The perspective transformation matrix $P \in \mathbb{R}^{3\times3}$ is computed as follows:

$$\begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ e_1 & e_2 & 1 \end{pmatrix}$$

where A is estimated in the way mentioned in the main manuscripts. In Tab. 2, we conduct experiments on the hand dataset, which shows more complex variations in global structures than the head dataset, and observe a further improvement brought by perspective transformation.

Table 1. Results on the head dataset using the same exemplar with CC2D.

backbone	**Head**				
	MRE↓	SDR↑(%)			
	(mm)	2mm	2.5mm	3mm	4mm
CC2D-SSL	4.67	40.42	47.68	55.54	68.38
CC2D-TPL	2.72	49.81	58.73	68.18	81.01
Ours-stage I	2.91	38.38	50.23	60.74	75.75
Ours-stage II	2.22	54.15	66.11	75.73	88.19

Table 2. Comparison between learned affine and perspective transform on the hand dataset.

Global Alignment Type	**Local Step** N_Φ	**Hand**			
		MRE↓	SDR↑(%)		
		(mm)	2mm	4mm	10mm
affine (L)	2	2.13	60.93	89.43	99.21
perspective (L)	2	2.10	62.27	89.53	99.32

3 Benefits of More Local Steps

Although the registration model is trained with one global step and two subsequent steps of local deformations, the optimal value of N_Φ in the test phase tends to be larger as in Tab. 3 and varies across different datasets.

Table 3. Ablation of N_Φ for local deformation in the test phase.

N_Φ	Head					Hand				Chest			
	MRE↓	SDR↑(%)				MRE↓	SDR↑(%)			MRE↓	SDR↑(%)		
	(mm)	2mm	2.5mm	3mm	4mm	(mm)	2mm	4mm	10mm	(px)	3px	6px	9px
1	2.85	38.55	50.30	61.68	77.79	3.18	39.43	74.84	97.33	12.52	8.00	22.67	44.67
2	**2.70**	**42.78**	**54.88**	**65.03**	**81.01**	2.24	58.68	88.39	98.89	10.48	9.67	31.33	54.67
3	2.75	41.71	54.59	64.23	79.94	**2.13**	**60.93**	**89.43**	99.21	10.25	12.00	34.33	57.33
4	2.82	39.68	52.51	63.22	78.72	2.16	59.86	88.92	**99.28**	**10.16**	**12.33**	**39.00**	**60.33**

4 Ablation of EMA for pseudo landmarks

To incorporate landmark locations into the registration learning, we maintain an exponential moving average of model predictions on the training set. For all experiments in the main manuscript, we set τ to 0.9. We further ablate its impact on the MRE of stage I. When $\tau = 0.0$, the pseudo landmarks are simply model predictions in each epoch, without moving average applied. The moving average of pseudo landmarks are only used to generate gaussian heatmaps for computing the masked similarities of edge structures around landmarks, with no gradients for backpropagation. As in Tab. 4, our method is robust to τ.

Table 4. Ablation of τ in EMA

τ	Head				
	MRE↓	SDR↑(%)			
	(mm)	2mm	2.5mm	3mm	4mm
0.0	2.76	40.27	53.12	63.39	80.04
0.9	2.70	42.78	54.88	65.03	81.01
0.99	2.74	41.01	53.54	65.03	80.63

References

1. Qingsong Yao, Quan Quan, Li Xiao, and S Kevin Zhou. One-shot medical landmark detection. *arXiv preprint arXiv:2103.04527*, 2021.