BOTM: Echocardiography Segmentation via Bi-directional Optimal Token Matching

Zhihua Liu[1], Lei Tong[1] , Xilin He[2] , Huiyu Zhou[1]

[1]Biomedical Image Processing Lab, School of Computing and Mathematical Science, University of Leicester, UK
[2]Shenzhen University, China

Find Zhihua 👉

We are Under Review ✌ MICCAI 2024 Marrakesh MOROCCO

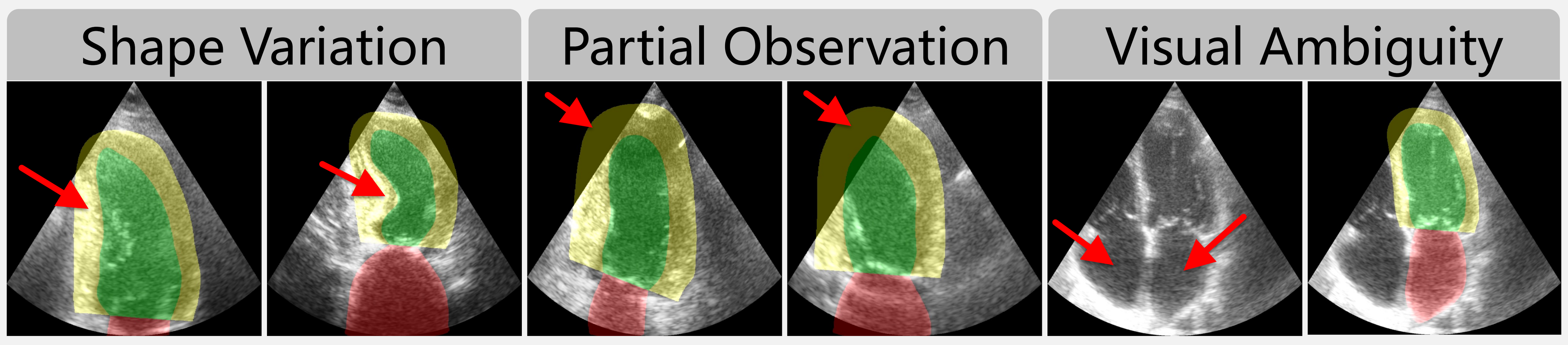
Echocardiography Segmentation

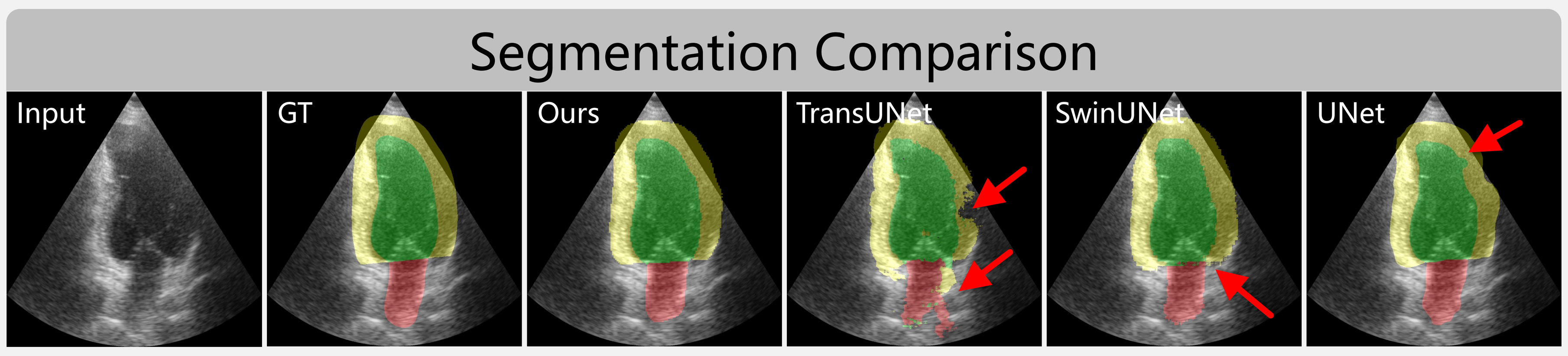
Background: Cardiac dysfunction, Echocardiography

- A primary cause for hospital admissions, growing global health concern
- Measuring of left ventricle changes to identify eligible patients
- Low-cost, rapid-acquisition, radiation-free, and non-invasiveness
- Supporting diagnostic decisions, risk stratification, surgical preparation

Challenge: Manual vs. Automated Segmentation

- Manual cardiac segmentation is time-consuming
- Highly depend on professional experiences, suffering observer varieties
- Speckle noise, shape variation, partial observation and visual ambiguity
- Disconnected boundaries, ambiguous localization and topological defect





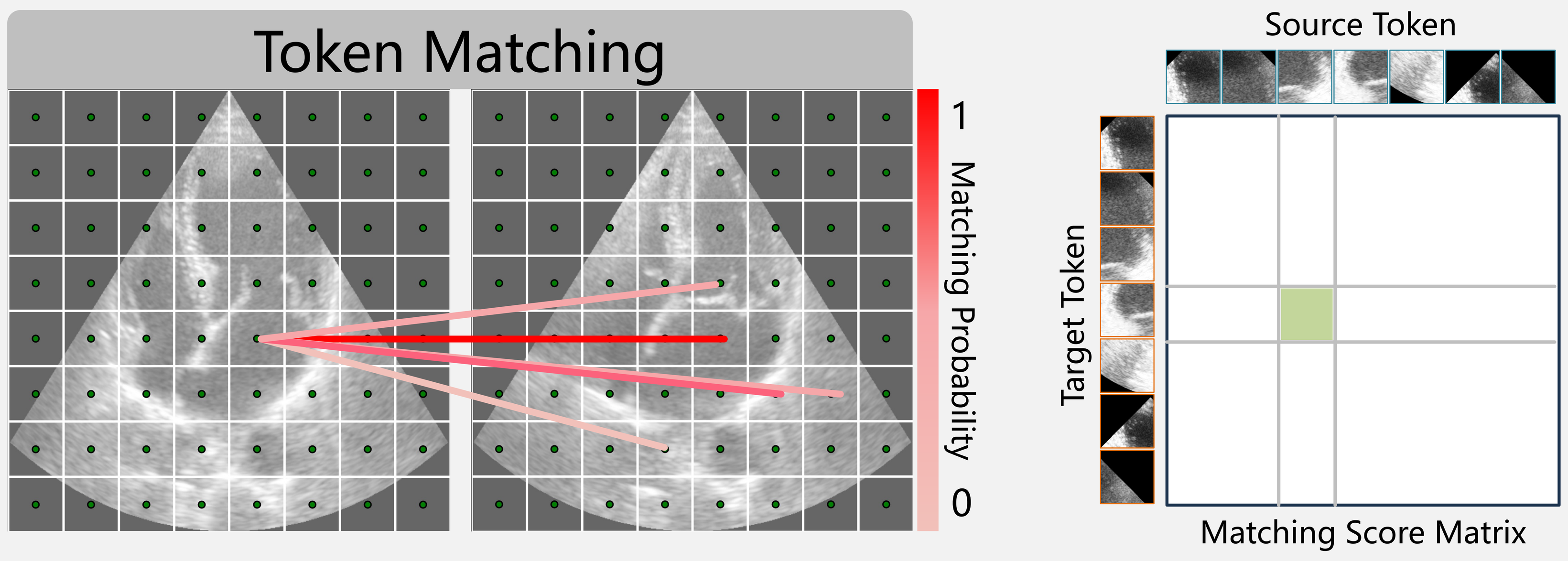
Myocardium | Left Ventricle | Left Atrium

Anatomical Consistency

- **Our motivation stems from the clinical need to ensure anatomical consistency**
- **Preserving intricate anatomical details, corresponding objects can retain identity across frame**

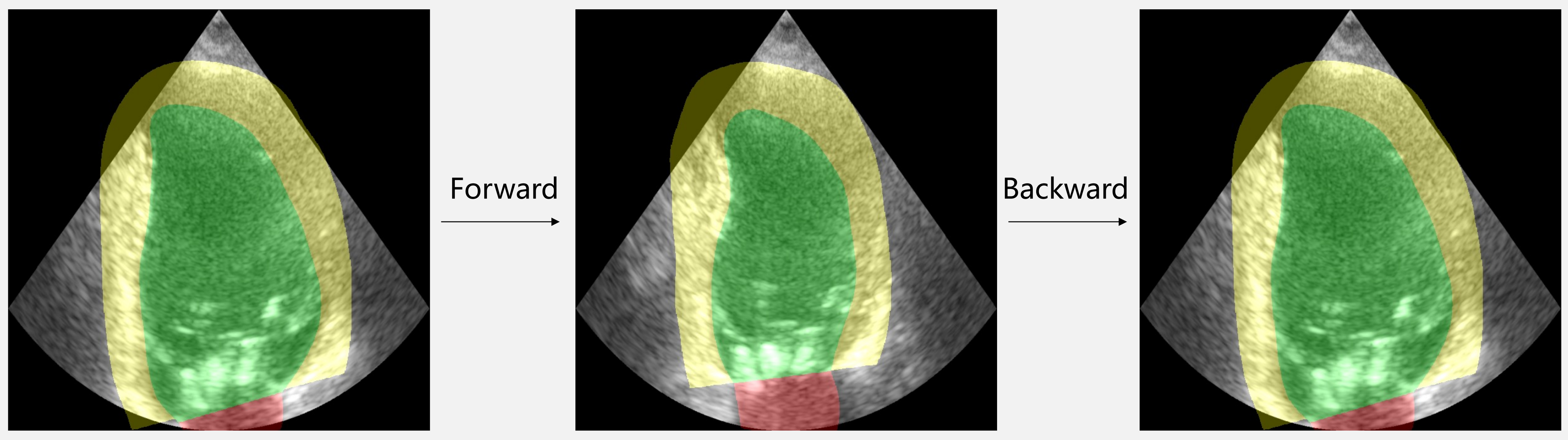
› Anatomical Consistency as Optimal Token Matching

Token-level anatomical consistency through a novel optimal transport (OT) perspective, without complicated pre/post-processing



› Bi-directional Transport Process

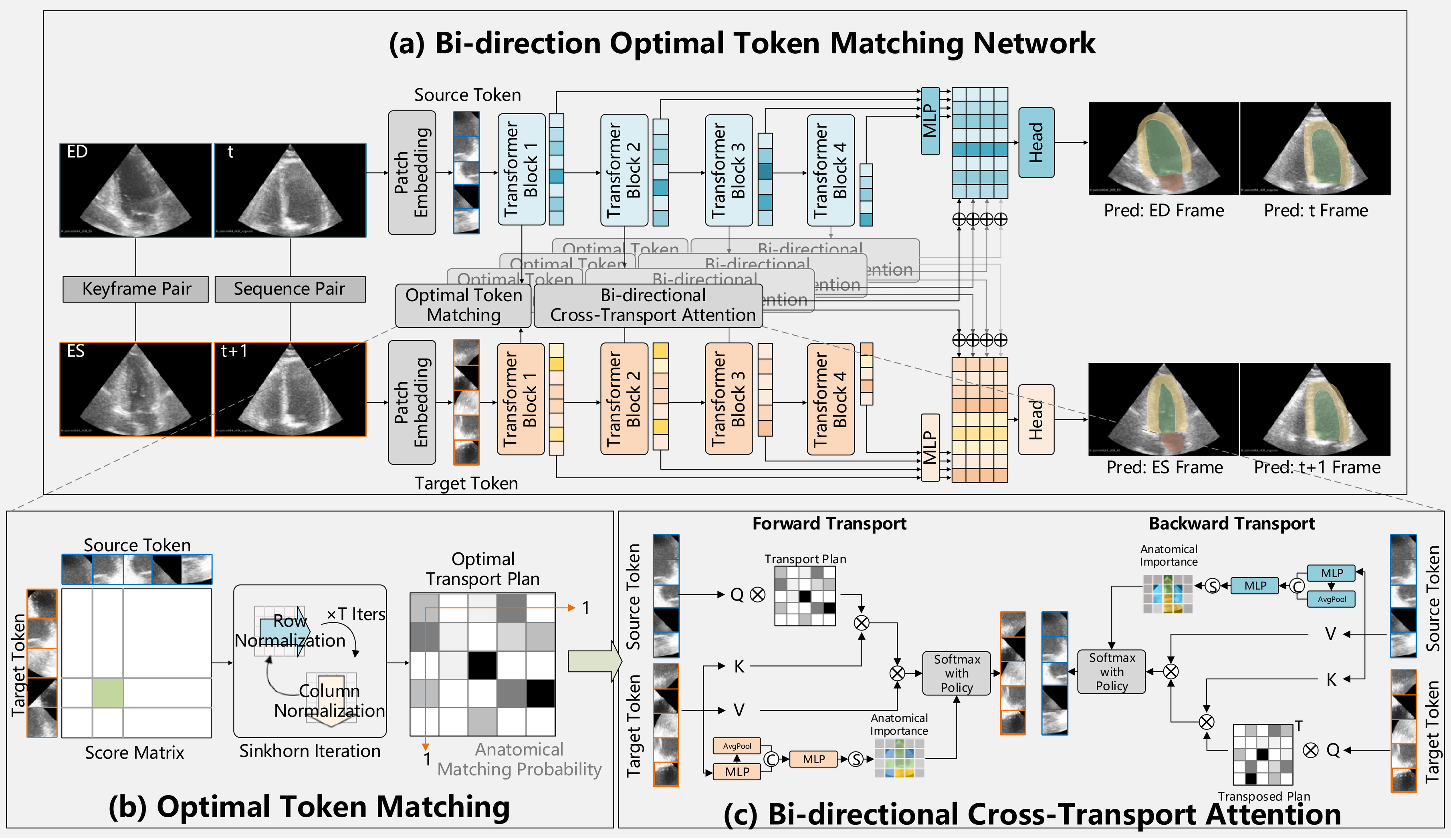
Temporal regulation by mimic cyclic cardiac motion



End-Diastolic (ED) End-Systolic (ES) End-Diastolic (ED)

BOTM Pipeline

1. Paired Echocardiographic Image
2. Matching Score Estimation
3. Proxy Module with Attention Policy



Transport Plan Estimation with Sinkhorn

strongly convex by resorting to the original OT with entropy regularization

$$\underbrace{\mathrm{T}_l^*}_{\text{OT Plan}} = \underbrace{\underset{\mathrm{T}\in R^{h_s^l w_s^l \times h_t^l w_t^l}}{\operatorname{argmin}} \sum_{ij}}_{\text{Minimize matching difference}} \underbrace{\mathrm{T}_{ij}^l \mathrm{C}_{ij}^l}_{\text{Matching Cost}} + \underbrace{\epsilon H(\mathrm{T}^l)}_{\text{Entropy Regularization}} \quad \text{where } \mathrm{C} = 1 - \underbrace{\frac{\mathrm{X}_s^l \cdot \mathrm{X}_t^l}{\|\mathrm{X}_s^l\| \cdot \|\mathrm{X}_t^l\|}}_{\text{Anatomical Similarity}}$$

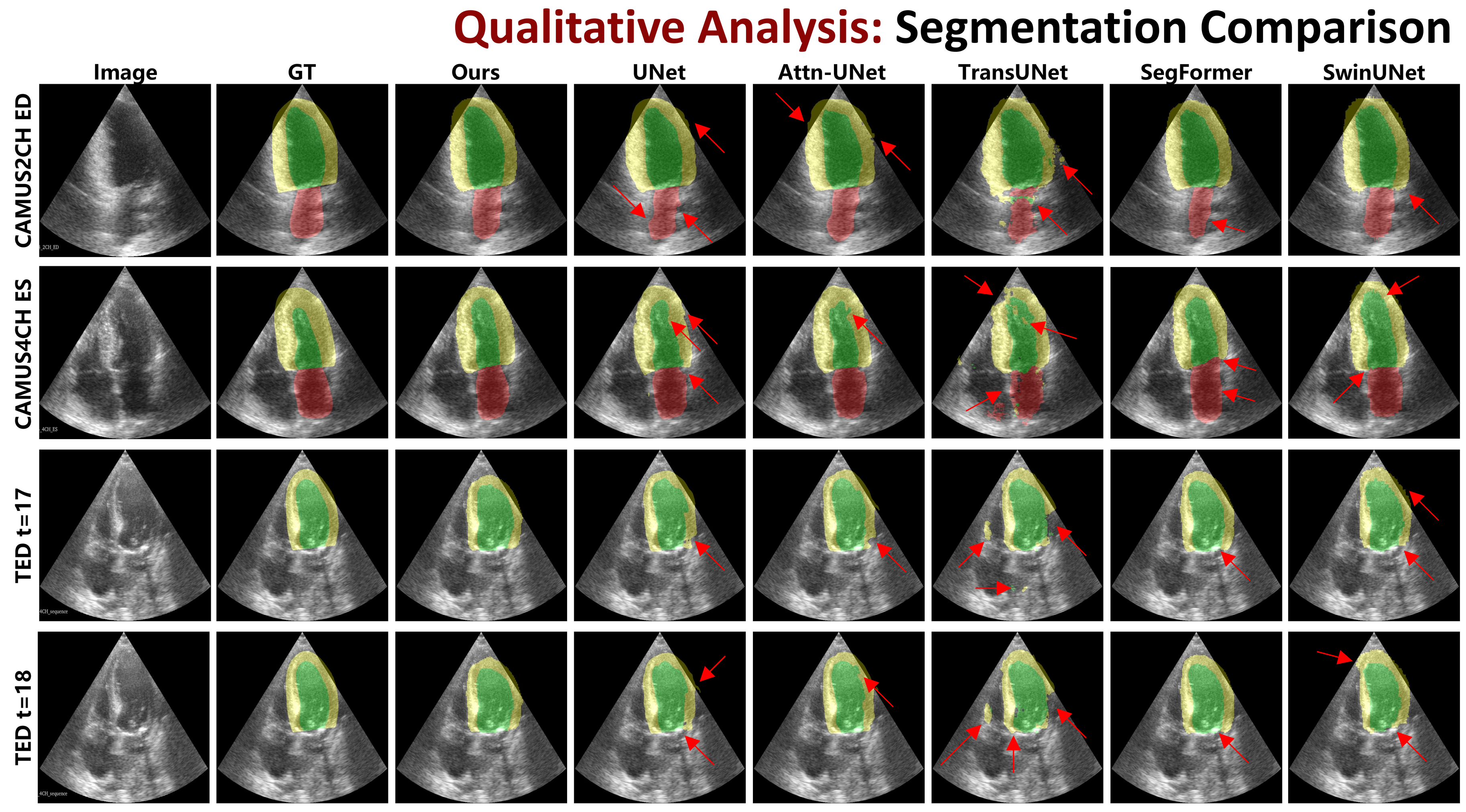
Training Settings

- A single NVIDIA A100 GPU, BatchSize of 8
- SGD (lr=0.001, momentum=0.9), 500 Epochs, Dice and Cross Entropy loss

Datasets*

- CAMUS: ES/ED Key Frame Segmentation
    - Apical 2 chamber (2CH) / 4 chamber view (4CH)
    - 450 patients (Training) / 50 patients (Test)
- TED: Video Segmentation
    - Apical 4 chamber view
    - 78 patients (Training) / 20 patients (Test)

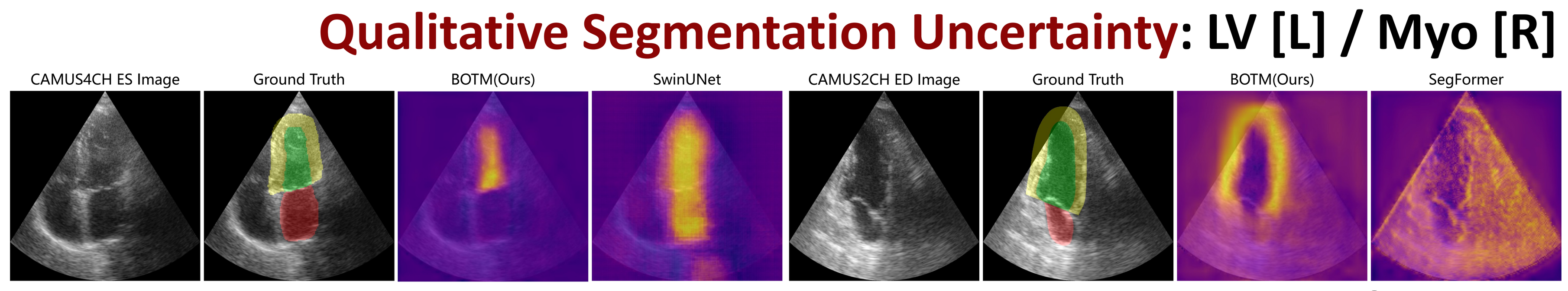
Qualitative Analysis: Segmentation Comparison



Quantitative Analysis: Strong Learning Ability

	Methods	mDice ↑	mHD ↓	mIoU ↑	Spe ↑	MAE ↓
CAMUS2CH	UNet [23]	0.893/0.900	10.112/7.509	0.855/0.850	0.989/0.990	0.022/0.019
	Attn-UNet [19]	0.894/0.901	10.362/10.091	0.838/0.853	0.989/0.990	0.022/0.019
	TransUNet [5]	0.889/0.875	9.718/12.313	0.832/0.770	0.978/0.979	0.042/0.040
	SegFormer [27]	0.845/0.840	18.741/20.064	0.741/0.731	0.990/0.990	0.026/0.025
	SwinUNet [2]	0.890/0.893	8.296/9.514	0.843/0.821	0.986/0.988	0.028/0.025
	***BOTM* (Ours)**	**0.908/0.934**	**7.334/5.861**	**0.887/0.884**	**0.994/0.990**	**0.017/0.015**
CAMUS4CH	UNet [23]	0.891/0.894	10.526/11.027	0.873/0.864	0.993/0.992	0.014/0.015
	Attn-UNet [19]	0.886/0.902	13.506/9.786	0.865/0.871	0.992/0.992	0.014/0.014
	TransUNet [5]	0.850/0.859	13.627/17.522	0.774/0.789	0.986/0.987	0.031/0.030
	SegFormer [27]	0.875/0.885	15.786/10.759	0.795/0.802	0.991/0.992	0.018/0.017
	SwinUNet [2]	0.872/0.890	9.425/11.872	0.860/0.856	0.988/0.987	0.022/0.021
	***BOTM* (Ours)**	**0.920/0.912**	**6.656/6.616**	**0.897/0.889**	**0.995/0.992**	**0.011/0.015**
TED	UNet [23]	0.901	11.316	0.844	0.986	0.022
	Attn-UNet [19]	0.904	9.570	0.858	0.987	0.021
	TransUNet [5]	0.850	13.627	0.774	0.986	0.031
	SegFormer [27]	0.886	14.289	0.736	0.987	0.024
	SwinUNet [2]	0.872	13.149	0.792	0.984	0.030
	***BOTM* (Ours)**	**0.923**	**7.519**	**0.893**	**0.993**	**0.014**

Qualitative Segmentation Uncertainty: LV [L] / Myo [R]



Generalization Study: CAMUS4CH [L] / TED [R]

Methods	RandomBlur 10%	RandomBlur 30%	RandomBlur 50%	RandomGaussNoise 10%	RandomGaussNoise 30%	RandomGaussNoise 50%
UNet [23]	0.902/0.897	0.802/0.839	0.712/0.698	0.897/0.885	0.804/0.767	0.674/0.625
TransUNet [5]	0.869/0.874	0.778/0.813	0.693/0.742	0.881/0.879	0.769/0.803	0.712/0.794
***BOTM* (Ours)**	0.906/0.892	0.895/0.887	0.862/0.858	0.900/0.907	0.873/0.887	0.832/0.841

Methods	RandomFrameDropout 10%	30%	50%	70%
UNet [23]	0.901	0.877	0.849	0.810
TransUNet [5]	0.869	0.833	0.802	0.734
***BOTM* (Ours)**	0.912	0.893	0.875	0.851

Contact: ZL208@leicester.ac.uk

*Available at: https://humanheart-project.creatis.insa-lyon.fr/

UNIVERSITY OF LEICESTER