UNIVERSITY OF LEICESTER | THE UNIVERSITY of EDINBURGH | 香港中文大學 The Chinese University of Hong Kong | IMPERIAL | AstraZeneca

BOTM: Echocardiography Segmentation via Bi-directional Optimal Token Matching

AI and Machine Learning Group

Paper | Code | Dataset | @Zhihua

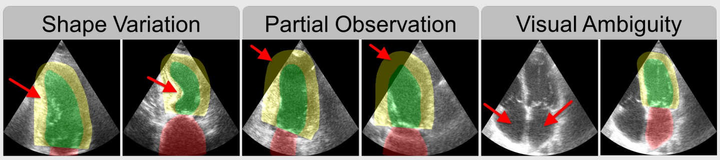
Echocardiography Segmentation

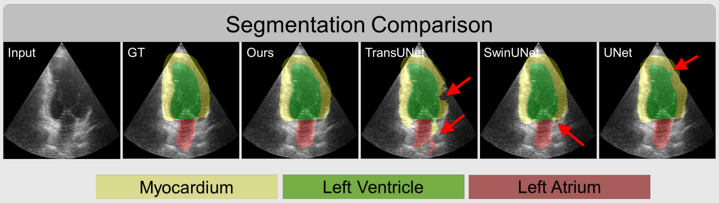
Background: Cardiac dysfunction, Echocardiography

- 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



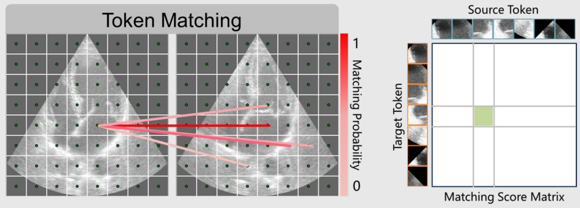


Anatomical Consistency

- Our motivation comes from the clinical need to ensure anatomical consistency
- Preserving intricate anatomical details, so that corresponding objects can retain identity across frames

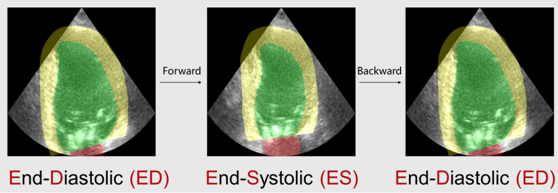
1. Optimal Token Matching

Token-level anatomical consistency through a novel optimal transport (OT) perspective.



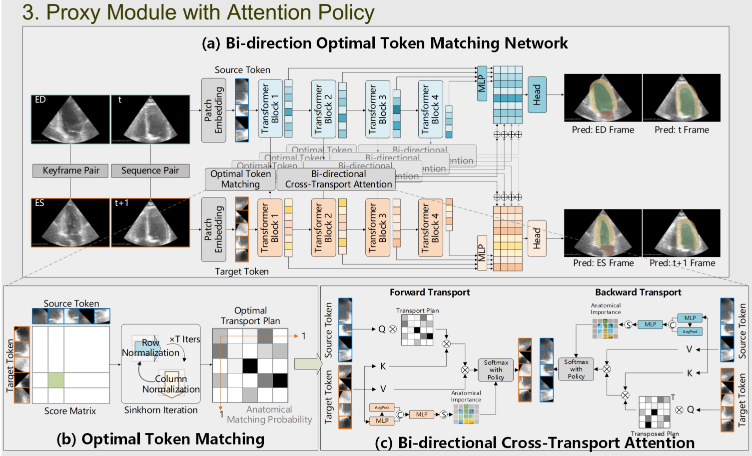
2. Bi-directional Transport Process

Temporal regulation by mimic cyclic cardiac motion



Pipeline: Bi-directional Optimal Token Matching (BOTM)

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

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

*Cuturi M. Sinkhorn distances:Lightspeed computation of optimal transport. Advances in neural information processing systems.2013;26

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

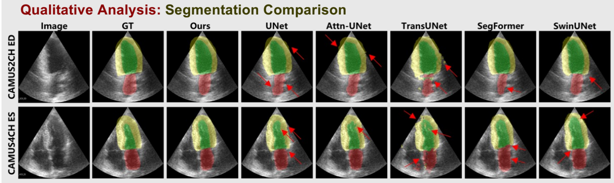
- Apical 2 chamber (2CH) / 4 chamber view (4CH)
- 450 patients (Training) / 50 patients (Test)

TED: Video Segmentation

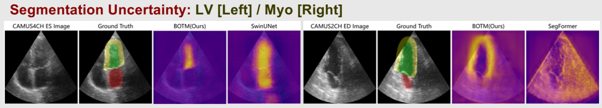
- Apical 4 chamber view
- 78 patients (Training) / 20 patients (Test)

Results#

Qualitative Analysis: Segmentation Comparison



Segmentation Uncertainty: LV [Left] / Myo [Right]



Generalization Study: CAMUS4CH [Left] / TED [Right]

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%	RandomFrameDropout 30%	RandomFrameDropout 50%	RandomFrameDropout 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

Please refer our paper for more results and technical details

Zhihua Liu (UoE/UoL), Lei Tong (AZ/UoL), Xilin He (CUHK), Che Liu (IC), Rossella Arcucci (IC), Chen Jin (AZ), Huiyu Zhou (UoL)
Contact: Zhihua Liu (zliu7@ed.ac.uk), Huiyu Zhou (hz143@leicester.ac.uk)

BMVC 2025