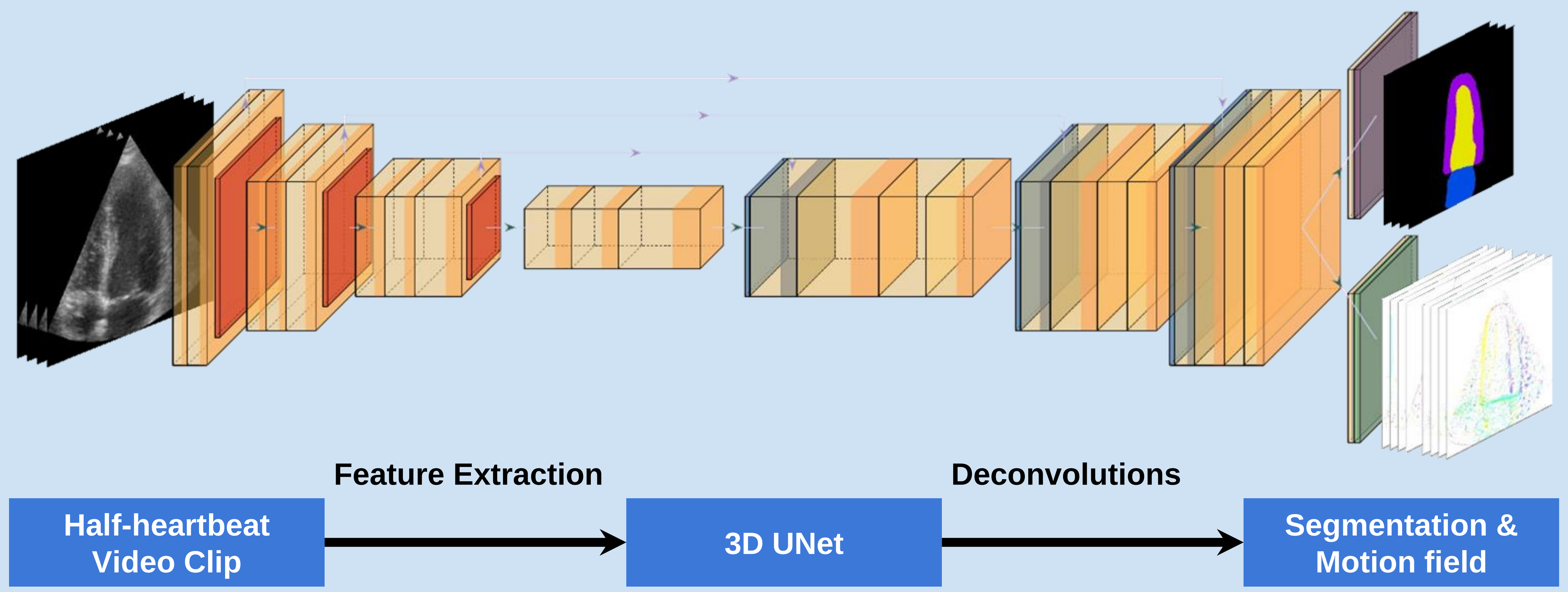
Temporally-Coherent Video Segmentation of Echocardiography

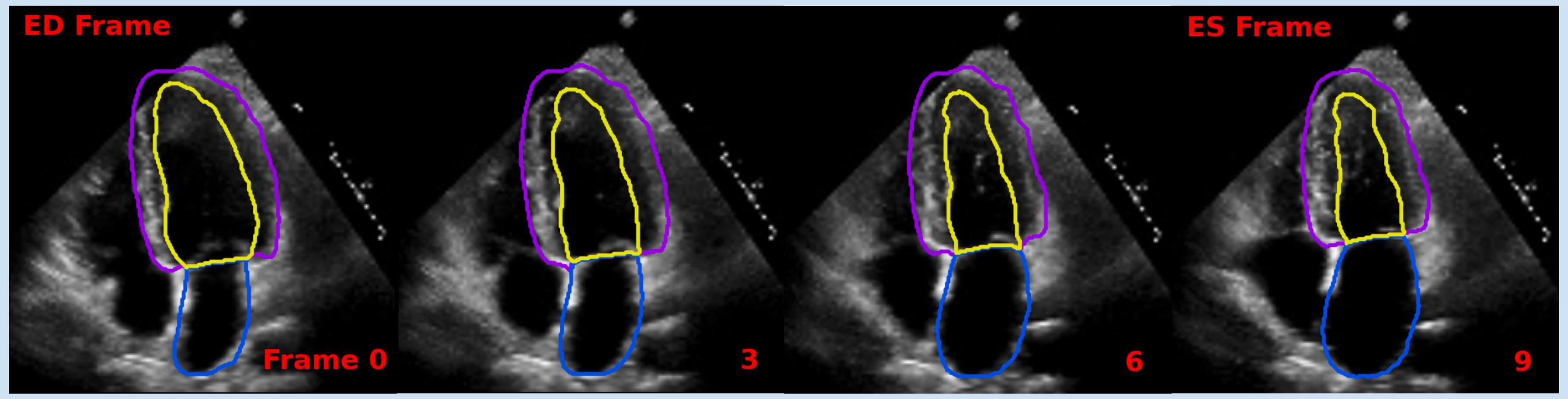
Yida Chen
Advisors: Dr. Joshua Stough[1], Dr. Xiaoyan Zhang[2], Dr. Christopher Haggerty[2]
[1]*Department of Computer Science Bucknell University, Lewisburg, PA*
[2]*Translational Data Science and Informatics, Geisinger Health Group*

Motivation

- Diagnosis of cardiovascular disease requires accurate quantification of Ejection Fraction (EF)
- A recently published video segmentation convolutional neural network (CNN) [1] improves EF estimation on a curated dataset
- We generalize this method on clinical echocardiography and improve results over previous frame segmentation network

CLAS Architecture

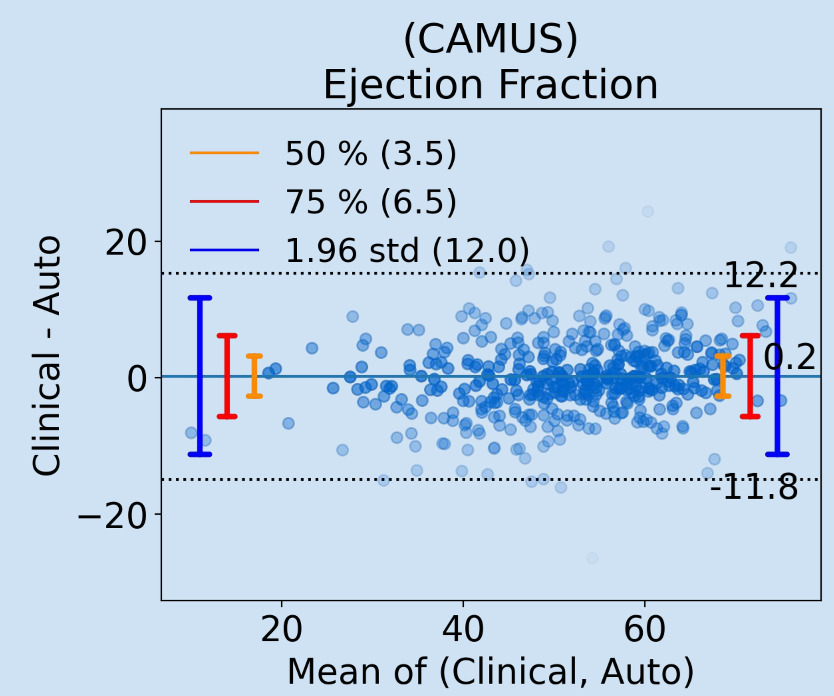




- CNN, CLAS [1], infers on sampled 10-frame ED to ES video clips
- Segmentation and motion tracking tasks use a shared feature extractor, 3D UNet
- Output frame-level multi-structural segmentation and bi-directional pixel-wise motion estimation.

10-Fold Cross-Validation

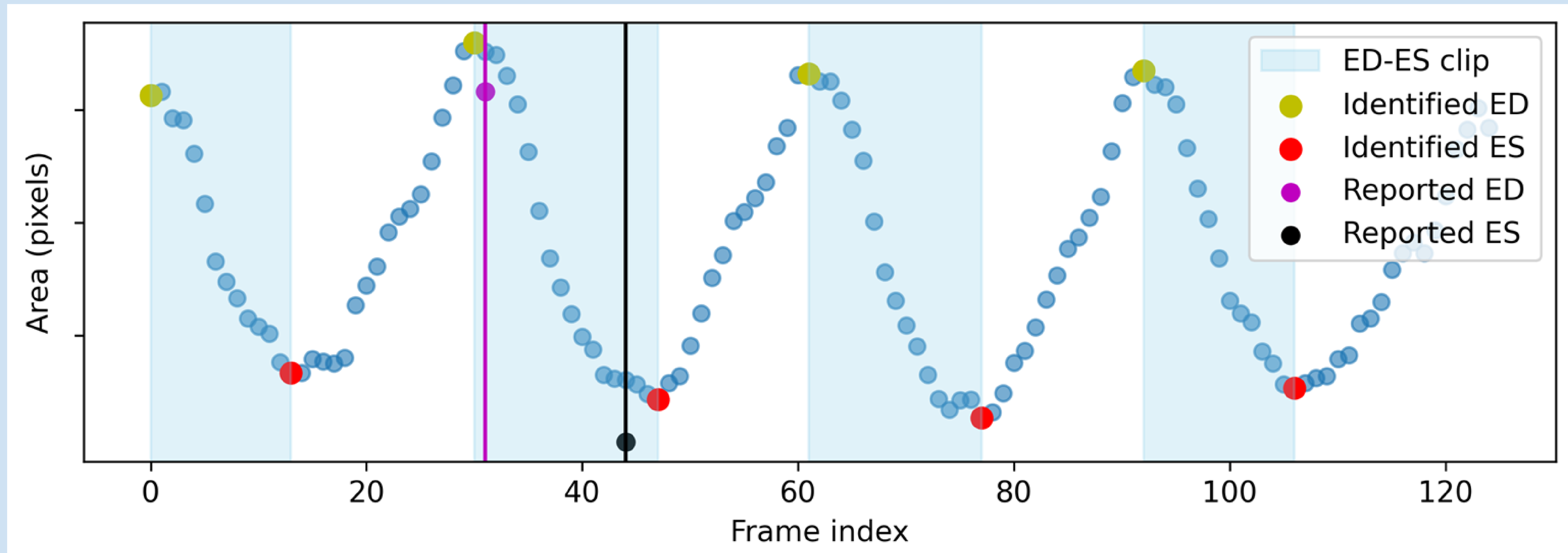
- We compare the CLAS with a frame segmentation CNN [2] on 450 patients in CAMUS dataset [3]
- Apply data-augmentation in training
- Augmented-CLAS (A-CLAS) achieves consistent dice score on segmenting LV_{endo} (0.93), LV_{epi} (0.95) and LA (0.88).
- A-CLAS significantly improves estimation on EDV, ESV, EF



- MAE (A-CLAS vs. Frame):
  - EDV: 8.7 ml vs. 9.9 ml
  - ESV: 6.3 ml vs 6.6 ml
  - EF: 4.6% vs 5.3 %
- Bland-Altman plot of difference between automatic and clinical measurements vs mean.
- Limit of Agreement (LOA) of A-CLAS is 12.0 (vs 15.6 of human).

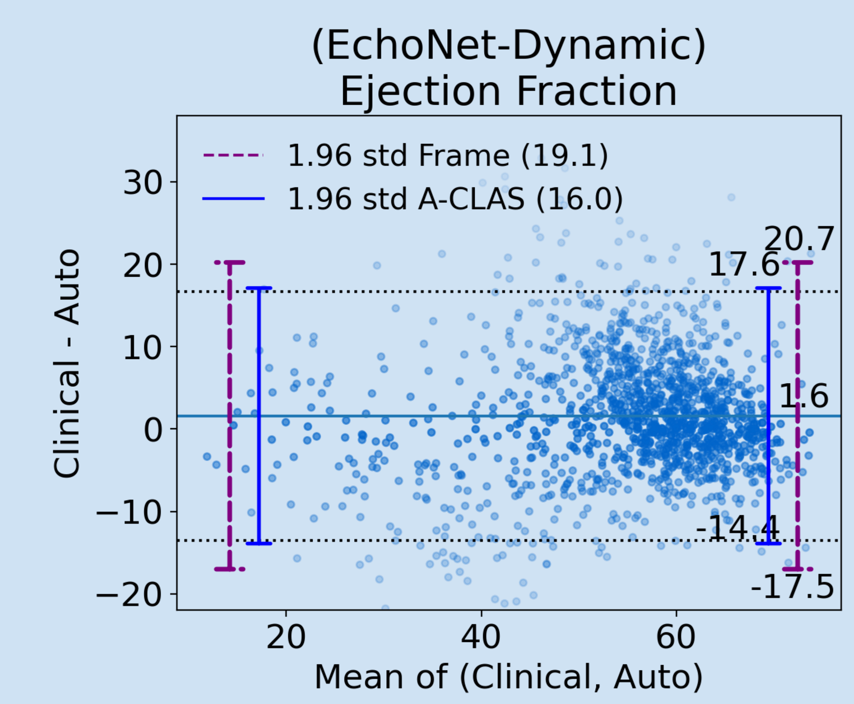
Generalize on Clinical Data

- Apply CAMUS-trained A-CLAS without tuning on the clinical EchoNet-Dynamic dataset [4]
- Clinical echocardiography contains multiple heartbeats
- A-CLAS segments all identified ED-ES half-heartbeat clips
- Use average of derived EFs from all clips



Results

- Over 1274 test patients of EchoNet-Dynamic
- Dice scores on LV_{endo} (ED/ES):
  - A-CLAS: 0.91/0.88
  - Frame: 0.91/0.88



- On EF estimation (A-CLAS vs Frame)
- Narrower LOA:
- 1.60% ± 16.0 (vs 19.1)
- Higher cross-correlation:
- 0.77 (vs 0.70)

Conclusions

- Video segmentation considers temporal coherence in addition to spatial features when segmenting Echocardiography
- Improved estimation on EDV, ESV, and EF
- Well generalized to a larger unseen clinical dataset

References

- [1] Wei, H., et al.: Temporal-consistent segmentation of echocardiography with co-learning from appearance and shape. MICCAI (2020).
- [2] Stough, J. V., et al.: Left ventricular and atrial segmentation of 2d echocardiography with convolutional neural networks. SPIE Medical Imaging (2020).
- [3] Leclerc, S., et al.: Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. IEEE Trans Med Imaging (2019).
- [4] Ouyang, D., et al.: Video-based ai for beat-to-beat assessment of cardiac function. Nature 580(7802), 252–256 (2020)

Acknowledgement

- Bucknell University Department of Computer Science
- Bucknell Geisinger Research Initiative.
- Ciffolillo Healthcare Technology Inventor Program