Recognizing End-diastole and End-systole Frames via Deep Temporal Regression Network

Bin Kong[1], Yiqiang Zhan[2], Min Shin[1], Thomas Denny[3], and Shaoting Zhang[1]

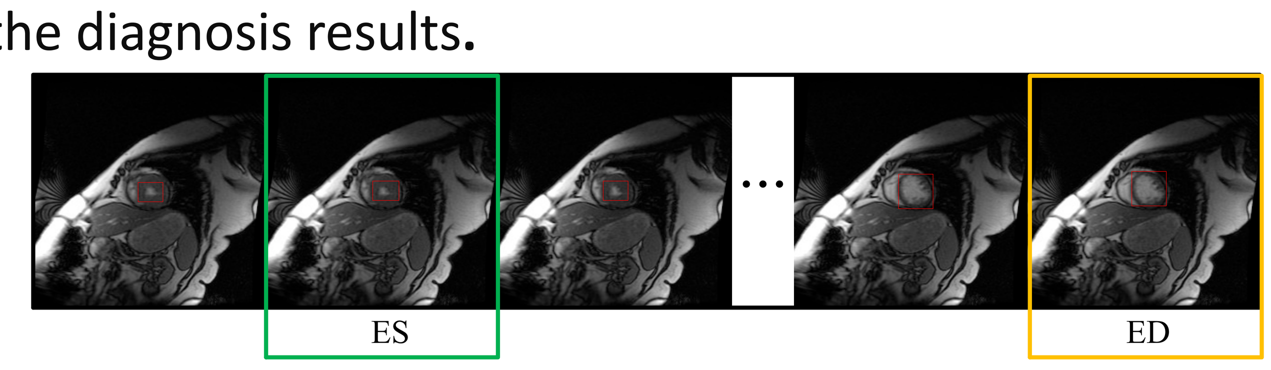
1. Department of Computer Science, UNC Charlotte, Charlotte, NC, USA
2. Siemens Healthcare, Malvern, PA, USA
3. MRI Research Center, Auburn University, Auburn, AL, USA

Background

- **Our goal:** Develop a system to automatically locate ED and ES frames in cardiac sequences with high precision.
- **Context behind this problem:** Accurately recognizing ED and ES frames is usually the first building block for most automated cardiac analysis systems. Theoretically, precisely and robustly locating these frames can greatly improve the final analysis results.
- **Motivation:** A single frame contains the spatial information such as the left ventricle volume size and a sequence contains the temporal information such as the left ventricle movement. Can we learn a single deep system to model both types of information?

Challenges

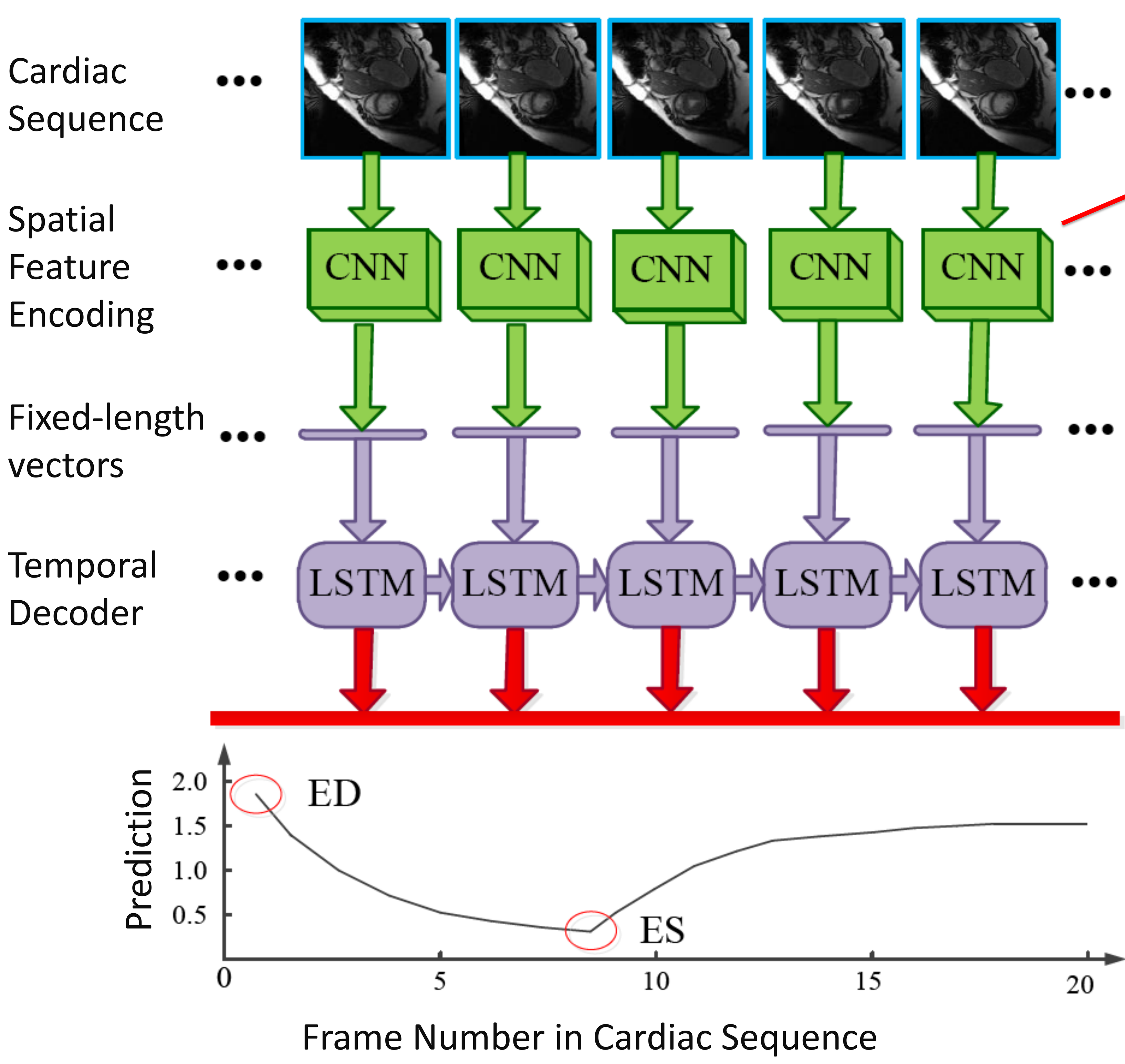
- The temporal relationships in cardiac cycles are very complex and differences among consecutive cardiac frames are subtle. And mislabeling even one frame may affect the diagnosis results.

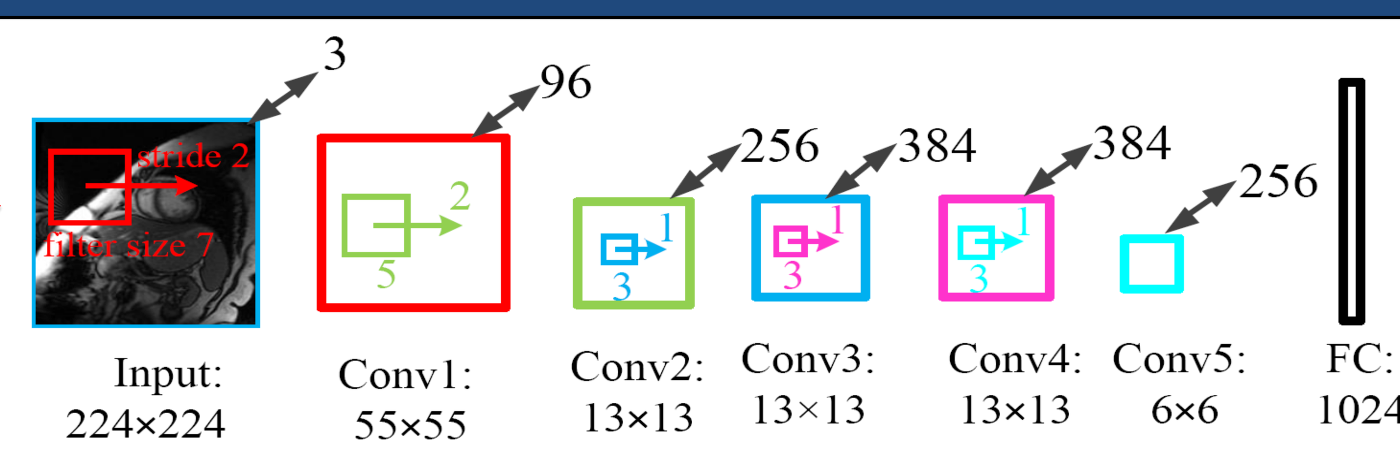


Neighboring frames have very subtle difference!

- Semantic gap between low-lever pixel information and high-level cardiac dynamic information.

Our Method





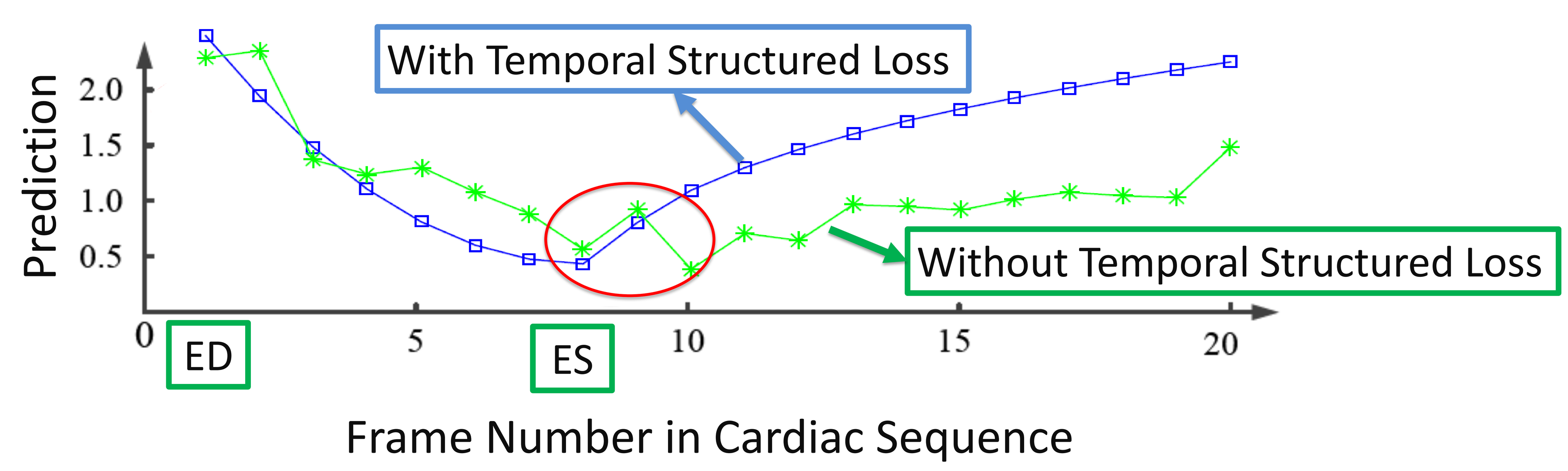
- We optimize the following objective function with two loss terms

$$\lambda^* = \arg\min_{\lambda}\{\sum_{S}\sum_{k}\|y_k - \eta_k(S,\lambda)\|^2 + \alpha L_{reg}(\lambda) + \beta L_{temp}\}$$

Difference between GT and prediction | Regularization term

$$L_{temp} = \frac{1}{2T}\{\delta(y_k > y_{k-1})\max(0,\eta_{k-1}-\eta_k) + \delta(y_k < y_{k-1})\max(0,\eta_k-\eta_{k-1})\}$$

Temporal Structured Loss | Indicator function



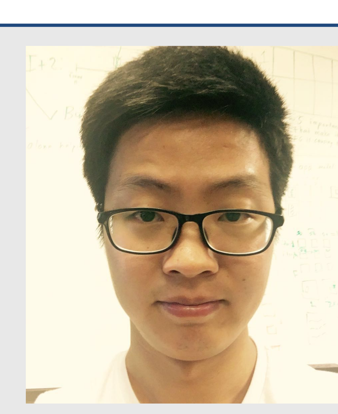
Experimental Results

- ~ 5,600 cardiac sequences have been acquired from 420 patients.
- Four-fold cross validation is performed on this datasets.
- Time and aFD (average frame difference) are computed for all the methods to compare the performance of these methods and traditional methods.

Methods		Seg-based: Level Set	Seg-based: Graph Cut	Reg-based: CNN + Reg	TempReg-Net (without TSC)	TempReg-Net
aFD	ED	1.54	2.27	1.30	0.47	0.38
	ES	1.24	1.65	1.97	0.52	0.44
STD	ED	1.93	2.89	1.77	0.49	0.39
	ES	1.64	1.96	2.42	0.53	0.46
Time (s)		2.9	3.5	1.5	1.4	1.4

Take Home Messages

- By integrating CNN and LSTM, our system can simultaneously extract both spatial and temporal information from cardiac sequences.
- By enforcing structured constraint upon the predicted results, in the deep learning architecture, the accuracy is greatly improved.
- The average difference between the ground truth the predictions is less than 0.5 frames, outperforming traditional methods by a big margin.



Contact Email:
bkong@uncc.edu
szhang16@uncc.edu