Self-supervised Learning for ECG-based Emotion Recognition

Pritam Sarkar, Ali Etemad

Department of Electrical and Computer Engineering
Queen's University, Kingston, Canada

ICASSP 2020

Ambient Intelligence and
Interactive Machines (Aiim) Lab

Outline

- ❑ Problem and Motivation
- ❑ Related work
- ❑ Proposed Framework
- ❑ Datasets
- ❑ Results
- ❑ Analysis
- ❑ Summary

Problem and Motivation

Limitations of fully-supervised learning:

- Human annotated labels are required to learn data representations; the learned representations are often very task specific.
- Larger labelled data are required in order to train deep networks; smaller datasets often result in poor performance.

Advantages of self-supervised learning:

- Models are trained using automatically generated labels.
- Learned representations are high-level and generalized; therefore less sensitive to inter or intra instance variations (local transformations).
- Larger datasets can be acquired to train deeper and sophisticated networks.

Queen's UNIVERSITY Aiim Lab

Problem and Motivation

Limitations of fully-supervised learning:

- ❑ Human annotated labels are required to learn data representations; the learned representations are often very task specific.
- ❑ Larger labelled data are required in order to train deep networks; smaller datasets often result in poor performance.

Advantages of self-supervised learning:

- ❑ Models are trained using automatically generated labels.
- ❑ Learned representations are high-level and generalized; therefore less sensitive to inter or intra instance variations (local transformations).
- ❑ Larger datasets can be acquired to train deeper and sophisticated networks.

Queen's UNIVERSITY Aiim Lab

Literature Review

- *Healey et al., 2005*:
  - Stress detection during driving task
  - Time-frequency domain features
  - LDA classifier
- *Liu et al., 2009*:
  - Affect based gaming experience
  - Time-frequency domain features
  - RF, KNN, BN, SVM classifiers
- *Santamaria et al., 2018:*
  - Movie clips were used to elicit emotional state
  - Time/frequency domain features
  - Deep CNN classifier
- *Siddharth et al., 2019:*
  - Affect recognition
  - HRV and spectrogram features
  - Extreme learning machine classifier

Time/Frequency Domain Feature Extraction

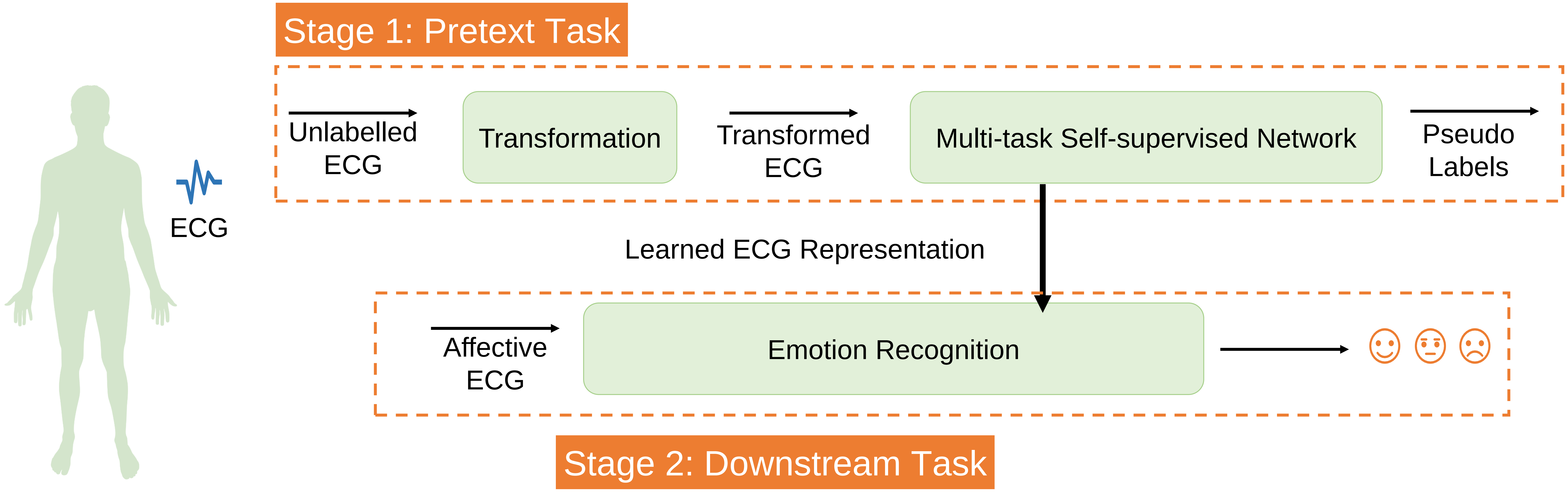
Fully-supervised Classifier

Emotion Recognition

Queen's UNIVERSITY

Aiim Lab

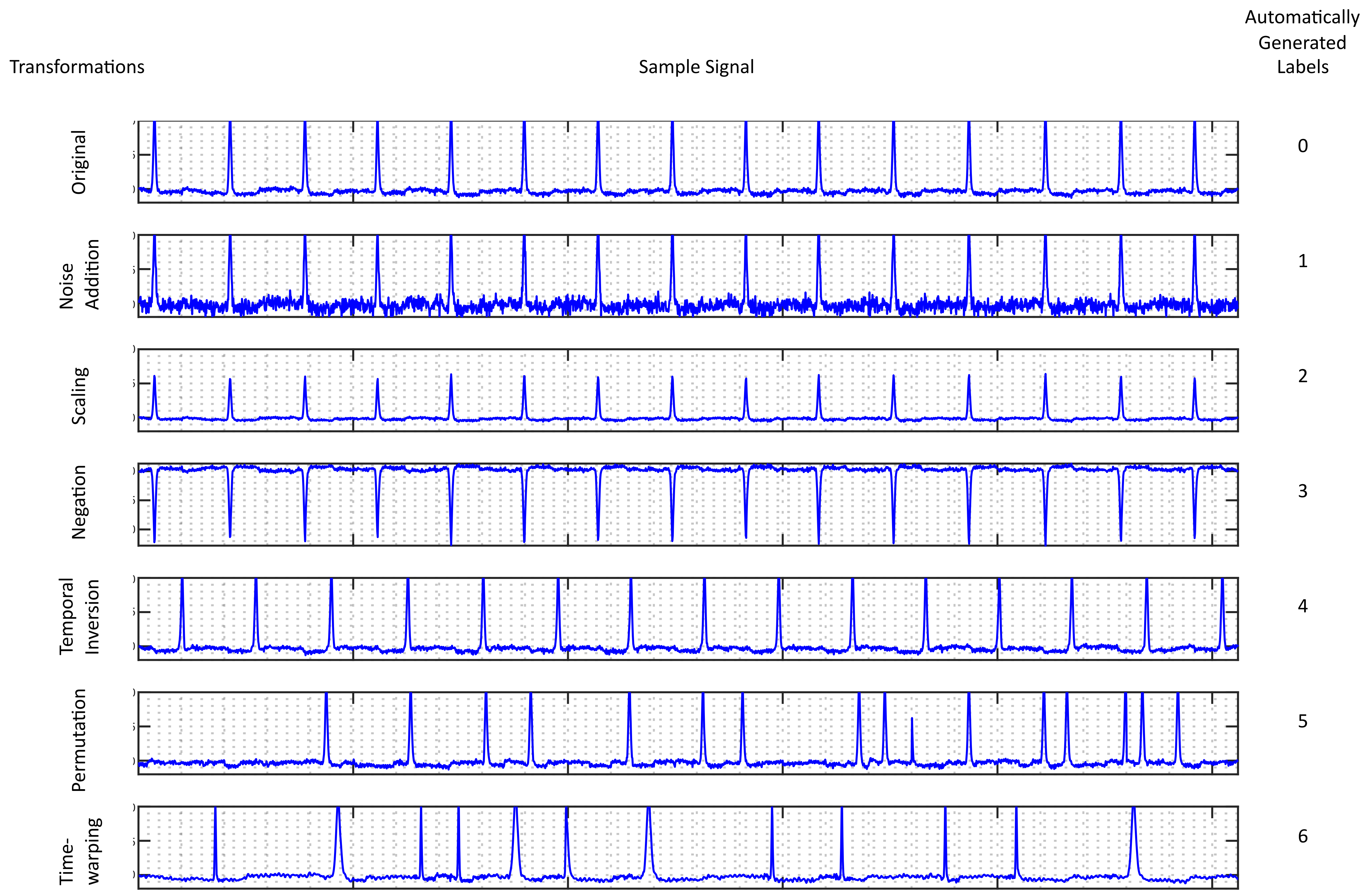
Proposed Framework



Our proposed framework.

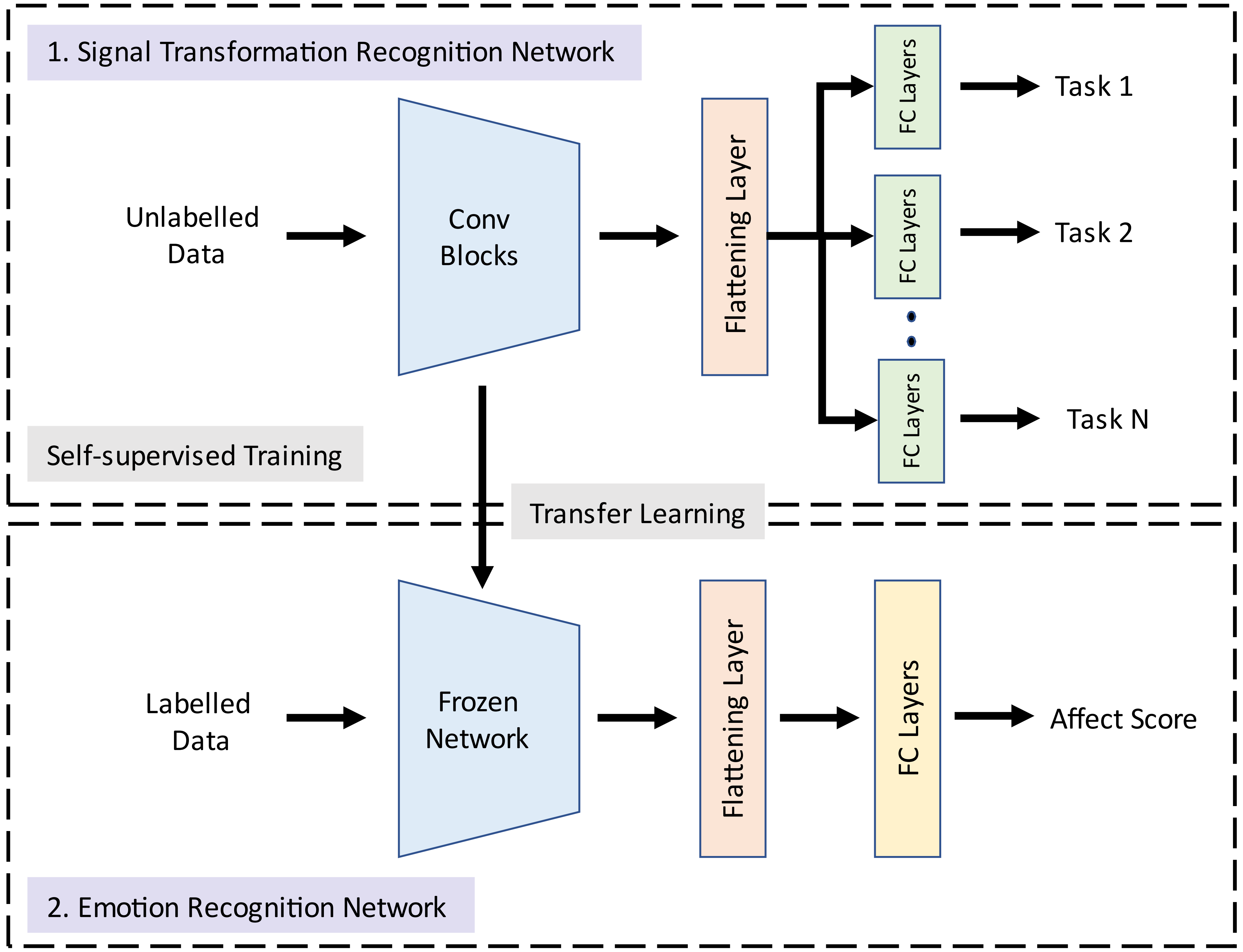
Transformations

- Noise Addition [SNR]
- Scaling [scaling factor]
- Negation
- Temporal Inversion
- Permutation [no. of segments]
- Time-warping [no. of segments, stretching factor]



A sample of an original ECG signal with the six transformed signals along with automatically generated labels are presented.

Proposed Architecture



The proposed self-supervised architecture is presented.

Table 1. The architecture of the signal transformation recognition network is presented.

Module	Layer Details	Feature Shape
Input	—	2560×1
Shared Layers	$[conv, 1 \times 32, 32] \times 2$	2560×32
	$[maxpool, 1 \times 8, stride = 2]$	1277×32
	$[conv, 1 \times 16, 64] \times 2$	1277×64
	$[maxpool, 1 \times 8, stride = 2]$	635×64
	$[conv, 1 \times 8, 128] \times 2$	635×128
	global max pooling	1×128
Task-Specific Layers	$[dense] \times 2$ $\times$ *7 parallel tasks*	128
Output	—	2

Datasets

We use 2 public datasets: AMIGOS and SWELL

- AMIGOS:
  - Affect attributes: Arousal, Valence
  - Total Participants: 40
  - Movie clips were shown to participants.
  - Shimmer sensors were used to capture ECG signal at 256 Hz.

- SWELL:
  - Affect attributes: Arousal, Valence, Stress
  - Total Participants: 25
  - Participants performed office tasks.
  - TMSI devices were used to capture ECG signal at 2048 Hz.

Queen's UNIVERSITY Aiim Lab

Results

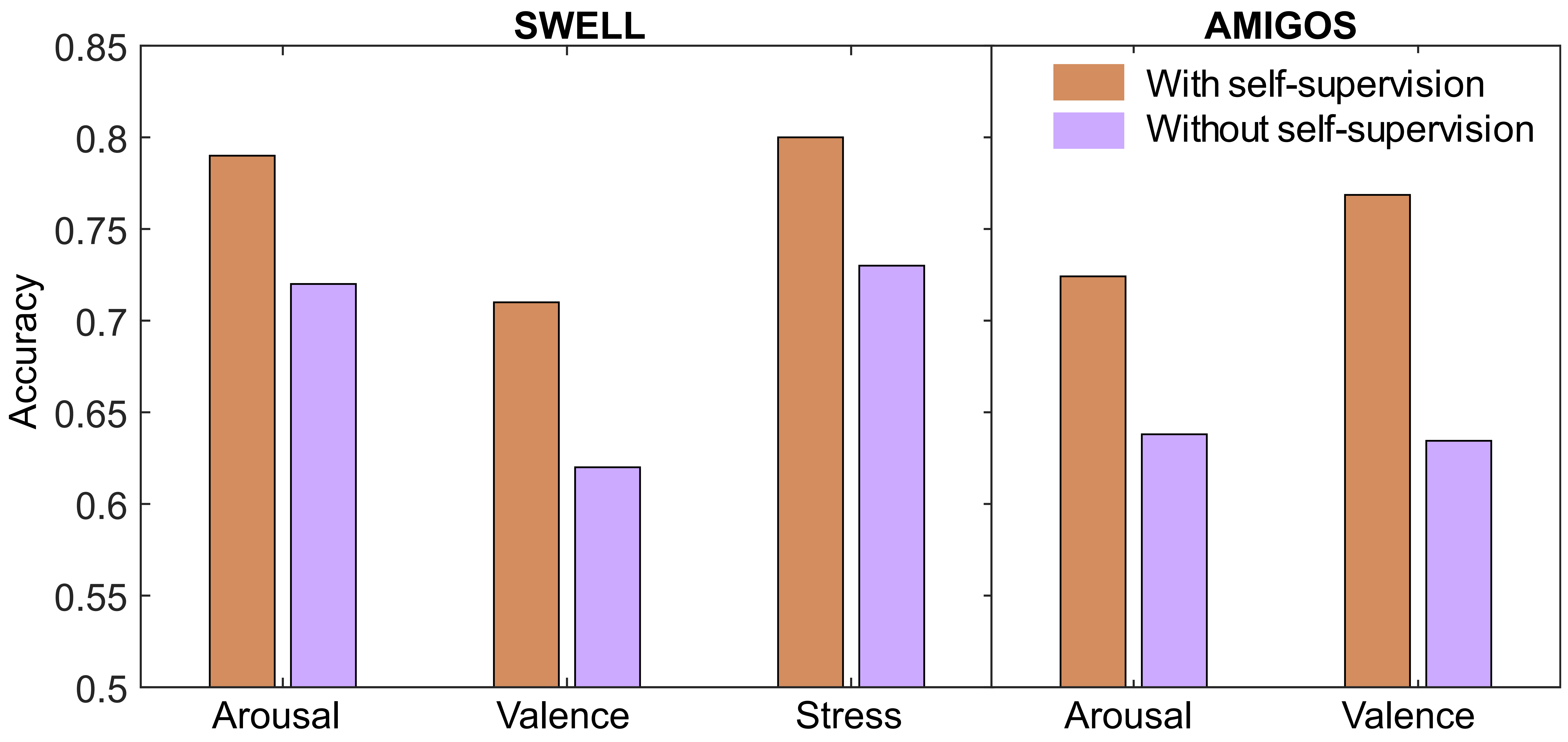
Table 2. The results of our self-supervised method on the SWELL dataset are presented and compared to prior work as well as the emotion recognition network without the self-supervised step.

Ref.	Method	Stress	Arousal		Valence	
			Acc.	F1	Acc.	F1
[24]	SVM	0.641				
[23]	SVM	0.864				
[22]	BBN	0.926				
Our	CNN w/o self-sup.	**0.984**	0.958	**0.957**	0.961	0.956
	CNN with self-sup.	0.983	**0.960**	0.956	**0.963**	**0.958**

Table 3. The results of our self-supervised method on the AMIGOS dataset are presented and compared to prior work as well as the emotion recognition network without the self-supervised step.

Ref.	Method	Arousal		Valence	
		Acc.	F1	Acc.	F1
[11]	GNB		0.545		0.551
[21]	CNN	0.81	0.76	0.71	0.68
Ours	CNN w/o self-sup.	0.837	0.828	0.809	0.808
	CNN with self-sup.	**0.858**	**0.851**	**0.840**	**0.837**

Analysis



Performance of our method with and without the self-supervised learning step using 1% of the labels in the datasets are presented.

Queen's UNIVERSITY Aiim Lab

Summary

- We proposed a novel ECG-based self-supervised learning framework for affective computing for the first time.

- We achieved state-of-the-art results on 2 public datasets (AMIGOS and SWELL).

- We showed that for a very limited amount of labelled data our self-supervised model perform considerably better compared to the fully-supervised model.

Queen's UNIVERSITY Aiim Lab

Thank you!

If you have any questions please reach me at:
pritam.sarkar@queensu.ca
www.pritamsarkar.com

Queen's UNIVERSITY Aiim Lab