EE798R
Enhancing Self-Supervised ECG Representation Learning

Dhruv
210338

November 7, 2024

Abstract

This document presents an improvement to a self-supervised learning model for ECG signal representation by incorporating residual connections into the convolutional neural network architecture.

1 Proposed Improvement: Residual Connections

I propose incorporating **Residual Connections** [1] into the convolutional neural network. Residual connections allow the network to learn residual functions with reference to the layer inputs, improving representation learning.

2 Mathematical Formulation

A standard convolutional block without residual connections is defined as:

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \mathcal{W}), \tag{1}$$

where $\mathbf{x}$ is the input, $\mathcal{F}(\mathbf{x}, \mathcal{W})$ represents the transformation function, and $\mathcal{W}$ denotes the layer weights.

In a residual block, a shortcut connection allows:

$$\mathbf{y} = \mathbf{x} + \mathcal{F}(\mathbf{x}, \mathcal{W}), \tag{2}$$

enabling the block to learn the residual mapping $\mathcal{F}(\mathbf{x}, \mathcal{W}) = \mathbf{y} - \mathbf{x}$.

When dimensions differ, a linear projection $\mathcal{W}_s$ aligns them:

$$\mathbf{y} = \mathcal{W}_s\mathbf{x} + \mathcal{F}(\mathbf{x}, \mathcal{W}), \tag{3}$$

where $\mathcal{W}_s$ (often a 1×1 convolution) adjusts $\mathbf{x}$ to the required size.

3 Benefits of Residual Connections

- **Mitigating Vanishing Gradients:** Residual connections allow gradients to flow directly through the network, reducing the vanishing gradient problem and enabling the training of deeper models.
- **Learning Identity Mappings:** They provide a path for inputs to bypass convolutional layers, helping the network learn identity mappings and preserve important features if needed.
- **Improved Training Dynamics:** Residual connections simplify optimization, allowing faster convergence and better performance for networks.

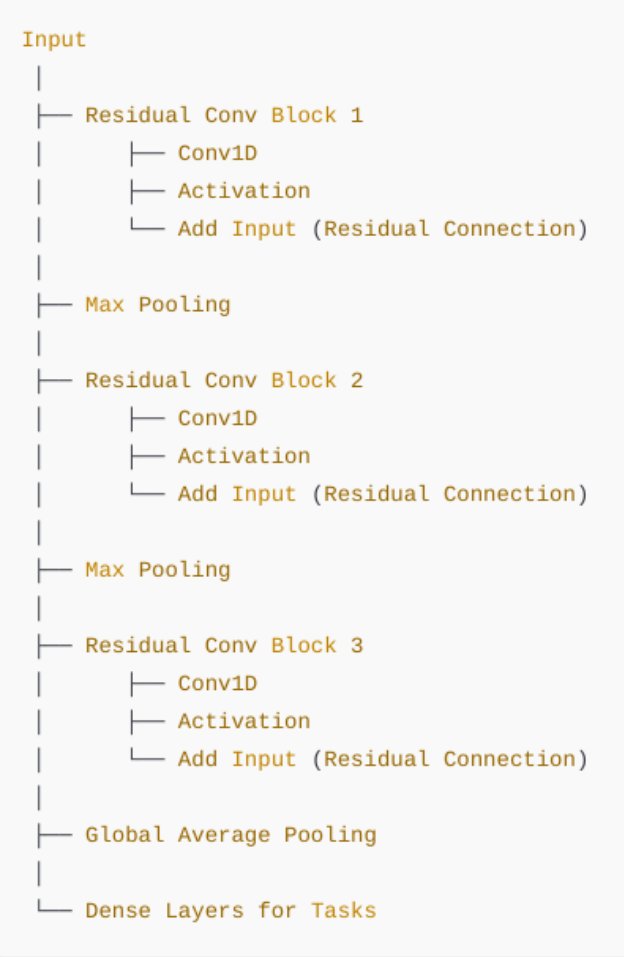

Figure 1: Modified Model with Residual Connections

4 Experiments and Results

Experiments to evaluate the impact of adding residual (skip) connections to different convolutional layers of our self-supervised ECG representation learning model. The configurations tested were:

1. **No Skip Connections**: The baseline model without any residual connections.
2. **Skip Connection in First Convolutional Block**: Adding a residual connection only in the first convolutional block.
3. **Skip Connections in First and Second Convolutional Blocks**: Adding residual connections in both the first and second convolutional blocks.
4. **Skip Connections in All Convolutional Blocks**: Adding residual connections in the first, second, and third convolutional blocks.

4.1 Training Time and Accuracy

For each configuration, the model was trained till it converged and measured the total training time and the accuracy achieved on the validation set. The results are summarized in Table 1.

Table 1: Comparison of Training Time and Accuracy for Different Configurations of Skip Connections

Configuration	**Training Time (hours)**	**Arousal**		**Valence**	
		Acc.	F1	Acc.	F1
No Skip Connections	~ 1.3	80.1	80.05	79.9	80.2
Skip Connection in First Block	~ 1	79.6	80.1	79.8	79.9
Skip Connections in First and Second Blocks	~ 1	81.5	81.5	81.3	81.1
Skip Connections in All Blocks	~ 0.7	82.8	82.9	82.4	82.8

4.2 Performance Improvements

Incorporating residual connections led to:

- Faster convergence during training.
- Improved accuracy on self-supervised transformation prediction tasks.
- Enhanced performance on downstream tasks such as emotion recognition.

5 Proposed Improvement: Attention-Based Dynamic Feature Selection

To improve focus on critical ECG features, we integrate an **attention-based feature selection** into the CNN. This involves a *Channel Attention Layer* added after each convolutional block, which refines feature maps by channel relevance. Given feature map $\mathbf{F} \in \mathbb{R}^{C \times L}$:

- **Squeeze**: Apply global average pooling to create a channel descriptor:

$$s_c = \frac{1}{L} \sum_{l=1}^{L} F_{c,l}.$$

- **Excitation**: Pass the descriptor through fully connected layers to compute channel weights:

$$a_c = \sigma\left(W_2 \delta\left(W_1 s_c\right)\right).$$

- **Recalibration**: Scale the original feature map to obtain the refined output:

$$\tilde{F}_{c,l} = a_c \cdot F_{c,l}.$$

5.1 Modification to the CNN Architecture

The modified CNN architecture includes the attention mechanism after each convolutional block:

- **Convolutional Block 1**:
  - Conv1D → ReLU → Dropout
  - Conv1D → ReLU → Dropout
  - **Channel Attention Layer**
  - Max Pooling
- **Convolutional Block 2**:
  - Conv1D → ReLU → Dropout
  - Conv1D → ReLU → Dropout
  - **Channel Attention Layer**
  - Max Pooling
- **Convolutional Block 3**:
  - Conv1D → ReLU → Dropout
  - Conv1D → ReLU → Dropout
  - **Channel Attention Layer**
  - Max Pooling

5.2 Expected Benefits

Integrating the attention mechanism is expected to offer the following benefits:

- **Improved Feature Representation**: By focusing on the most informative channels, the model can learn more discriminative features, potentially leading to higher accuracy.
- **Efficient Learning**: Emphasizing relevant features may accelerate convergence during training, possibly reducing the total training time.
- **Minimal Computational Overhead**: The attention mechanism introduces a small number of additional parameters and computations, maintaining the overall efficiency of the model.

6 Proposed Improvement: Contrastive Learning Approach

To enhance the representation learning capability of our ECG signal model, we propose integrating a **contrastive learning** framework into the existing self-supervised architecture. Contrastive learning aims to learn embeddings by bringing similar data points closer while pushing dissimilar ones apart in the feature space. This method leverages unlabeled data effectively, making it suitable for tasks where labeled ECG data is limited.

6.1 Contrastive Learning Framework

The core idea is to train the model to distinguish between positive pairs (different augmented versions of the same ECG signal) and negative pairs (augmented versions of different signals). By minimizing the distance between positive pairs and maximizing it between negative pairs, the model learns robust and discriminative representations.

6.1.1 Embedding Function

Let f_θ denote the neural network encoder parameterized by θ, mapping an input ECG signal x to an embedding $z = f_\theta(x)$. We utilize our existing CNN architecture as the encoder.

6.1.2 Contrastive Loss Function

We employ the *Normalized Temperature-Scaled Cross Entropy Loss* (NT-Xent Loss) for training:

$$\mathcal{L}_{\text{contrast}} = -\log \frac{\exp\left(\text{sim}(z_i, z_j)/\tau\right)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp\left(\text{sim}(z_i, z_k)/\tau\right)},$$

where:

- $\text{sim}(z_i, z_j) = \frac{z_i^\top z_j}{\|z_i\| \|z_j\|}$ is the cosine similarity between embeddings.
- τ is a temperature parameter controlling the concentration level of the distribution.
- N is the batch size.
- $\mathbb{1}_{[k \neq i]}$ is an indicator function that equals 1 if $k \neq i$.

6.2 Implementation Details

6.2.1 Data Augmentation

To create positive pairs, we apply two different random augmentations to each ECG signal, such as:

- **Scaling**: Changing the amplitude of the signal.
- **Shifting**: Shifting the signal in time.
- **Adding Noise**: Introducing random noise to the signal.
- **Time Warping**: Distorting the time intervals between signal points.

These augmentations help the model learn invariant features.

6.2.2 Modified Network Architecture

We add a *projection head* after the encoder, consisting of fully connected layers with non-linear activation functions, to map the embeddings to a space where the contrastive loss is applied. This projection head enhances the quality of the learned representations.

6.3 Benefits of Contrastive Learning

- **Improved Representation Quality**: The model learns more meaningful and discriminative features, which can enhance performance on downstream tasks like emotion recognition.
- **Data Efficiency**: Utilizes unlabeled ECG data effectively, reducing the reliance on large labeled datasets.
- **Flexibility**: The approach can be adapted to various types of ECG signal transformations and tasks.

6.4 Conclusion

Incorporating a contrastive learning framework into our self-supervised model provides a powerful means to improve ECG representation learning. This approach leverages the inherent structure of the data and has the potential to significantly enhance model performance with minimal additional complexity.

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. *Deep Residual Learning for Image Recognition.* In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.