HUMAN ACTIVITY RECOGNITION WITH NEURAL NETWORKS

Universität Stuttgart

Fangwen Liao and Peizheng Li

Institute of Signal Processing and System Theory, University of Stuttgart, Germany

This project aims to identify human activities based on smartphone inertial sensors.
According to different output types, two approaches are implemented:
1. Label Prediction (S2L)
2. Sequense Prediction (S2S)

1. Input Pipelines

- Dataset: HAPT Dataset[1](motion sequence of 6 axes recorded by inertial sensors put on waist)
- Pipeline: TFRecord
- Preprocessing:
  - Remove of data without activity labels
  - Z-score normalization performed on each channel
  - Denoising and Feature Enhancement through filters, such as Gaussian, median, etc.(only S2S)
  - Window with different sizes and shifts
  - Balancing data by over-sampling of 6 postural transitions

2. Models

- **L**abel Prediction(S2L)
  - Deep-Conv-LSTM Model[2]:
    3 convolutional layers construct abstract feature map and then 1 LSTM layer extracts temporal features.

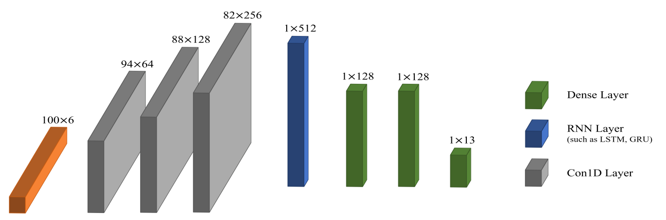

Figure 1: *Deep-Conv-LSTM*

- **S**equence Prediction(S2S)
  - Bidirectional RNN Model:
    This model is composed of several scale-reducing bidirectional LSTMs/GRUs plus some dense layers in between.

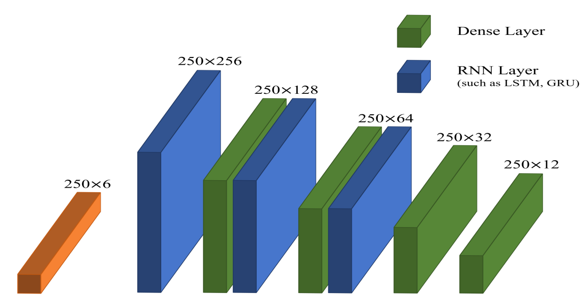

Figure 2: *Bidirectional RNN*

  - Seq2Seq Model with Attention:
    Sequence batch is first converted into a single feature tensor and then decoded into predictions.

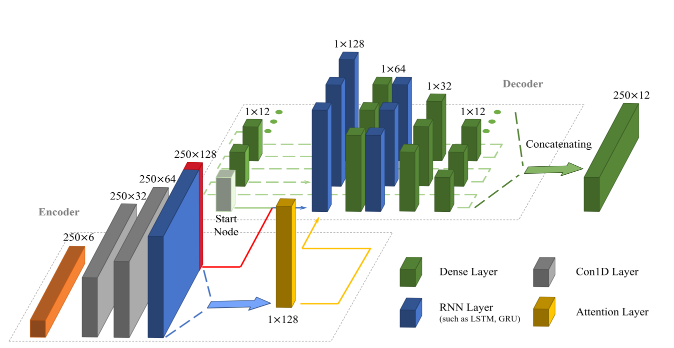

Figure 3: *Seq2Seq with Attention*

  - Conv1D Encoder-Decoder Model:
    Infomation in the time dimension is compressed into multi-channel features and then inversely amplified into predictions (similar to semantic segmentation).

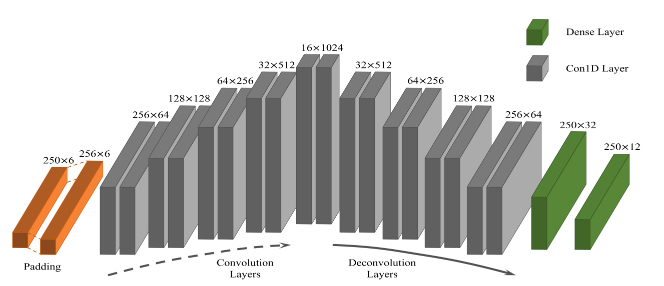

Figure 4: *Conv1D Encoder-Decoder*

- Ensemble Learning:
  There are two parallel branches:
  1. directly using RNN to process time-domain signals
  2. doing FFT and then using Conv1D to process corresponding frequency-domain signals
  Finally the outputs of the two branches are merged into a prediction.
- Postprocessing:
  This method is performing median filtering on the label sequence output by the model.

3. Evaluations

- **S**equence Prediction(S2S)
  - Preprocessing can not contribute to obvious improvement.

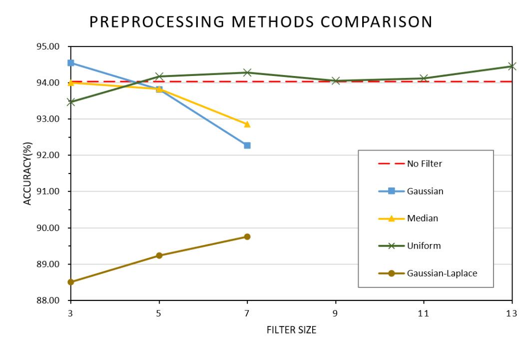

Figure 5: *Preprocessing Comparison*

  - Bidirectional RNN can consider both past and future together, having a better recognition capability.
  - Seq2Seq model has a relative lower accuracy.
  - As the depth increases, the accuracy of Conv1D Encoder-Decoder exceeds that of RNNs.
  - Ensemble Learning through Time- and Frequency-Domain Combination can reduce overfitting.
  - Postprocessing can remove the salt-and-pepper-noise in predictions.

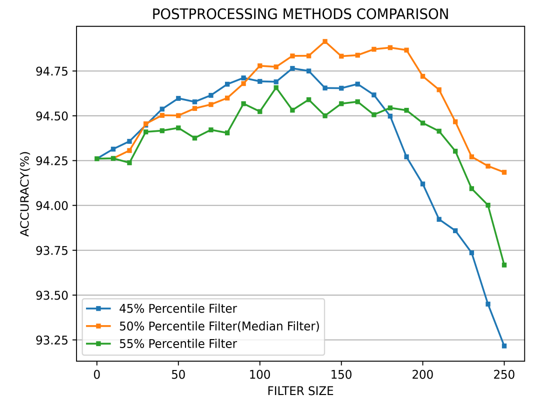

Figure 6: *Postprocessing Comparison*

  - Highest test accuracy: 96.50%
- **L**abel Prediction(S2L)
  - Average test accuracies for lengths 100, 50, 25:

Table 1: *Test Accuracies*

Size	100	50	25
Acc	96.5%	94.5%	89.5%

  - Average F1-scores of Size 100[3]:
    The F1 scores of the first 6 classes are significantly higher.

Table 2: *F1-scores of length 100*

1-3	98.6%	98.8%	98.5%
4-6	95.8%	96.0%	99.2%
7-9	81.0%	72.7%	62.1%
10-12	76.0%	57.7%	68.0%

  - Average Confusion Matrix[3]:
    Most elements are located diagonally with a few mistakes in postural transitions.

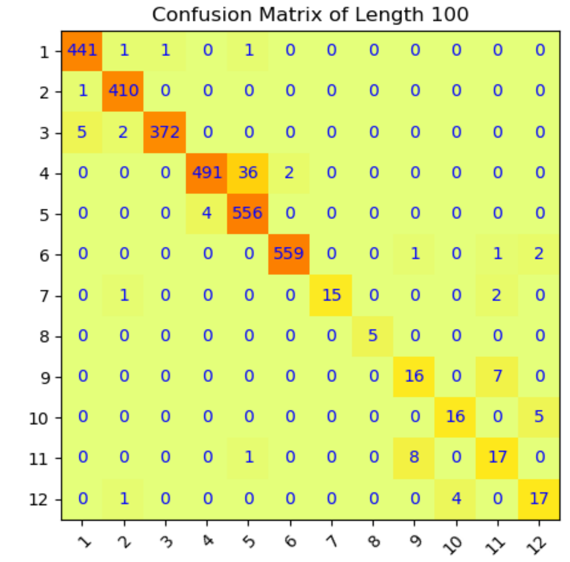

Figure 7: *Confusion Matrix*

- **V**isualization:
  Most activities can be accurately recognized.

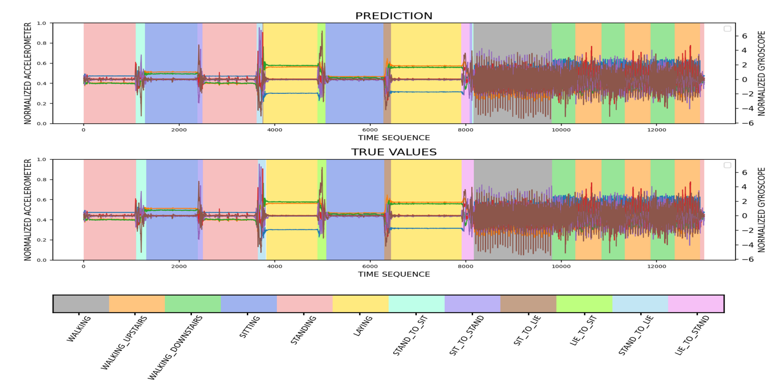

Figure 8: *Visualization*

4. Conclusions

- Compare to postural transitions, activities recognition achieves relatively higher accuracy.
- Activities containing Stand and Sit are more likely to be confused with each other, which also appears in postural transitions Sit to Lie and Stand to Lie.
- Imbalance and imperfection of the data set is the main obstacle to performance improvement.

[1]https://archive.ics.uci.edu/ml/datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions
[2]Ordonez, Francisco Javier, and Daniel Roggen. "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition." Sensors 16.1 (2016): 115
[3]Activity Labels: 1-WALKING, 2-WALKING UPSTAIRS, 3-WALKING DOWNSTAIRS, 4-SITTING, 5-STANDING, 6-LAYING, 7-STAND TO SIT, 8-SIT TO STAND, 9-SIT TO LIE, 10-LIE TO SIT, 11-STAND TO LIE, 12-LIE TO STAND