APPENDIX:

T-WAVENET: A TREE-STRUCTURED WAVELET NEURAL NETWORK FOR TIME SERIES SIGNAL ANALYSIS

Anonymous authors
Paper under double-blind review

In this supplementary material, we first describe the datasets, the *experimental settings*, and the *evaluation metrics*. Next, the *hyper-parameter analysis*, more details about the *ablation study* in the paper on **OPPOR** and more experimental results about the *frequency spectrum energy analysis* (FSEA) technique for various tree-structured network configurations on **UCI-HAR**, **BCICIV2a** and **NinaPro DB1** are presented. Finally, since real-world applications usually require deploying the DNN on resource-constrained IoT devices, we also present a *design space exploration* technique for *T-WaveNet* to evaluate the computational cost under different tree structures.

1 DATASETS

OPPOR consists of annotated recordings from on-body sensors when carrying out common gestures of kitchen activities. We follow the settings of the sporadic gestures task in the OPPORTUNITY challenge. We train the models on the data of all ADL and drill sessions for the Subject 1 and on ADL1, ADL2 and drill sessions for Subjects 2 and 3. The testing set consist of ADL4 and ADL5 for Subjects 2 and 3. **UCI-HAR** collects sensor data of 6 activities (walking, walking upstairs, walking downstairs, sitting, standing, laying) with a smartphone (Samsung Galaxy S II) on the waist. We follow the official dataset configuration [1], where 70% of the volunteers was selected for generating the training data and 30% the test data. **BCICIV2a** contains EEG signals from 9 healthy subjects performing four movement intention tasks (left hand, right hand, feet, and tongue), which are bandpass-filtered between 0.5Hz and 100Hz. We use the same settings as Zhang et al. (2020) to perform the leave-one-subject-out manner. **NinaPro DB1** contains sparse multi-channel sEMG recordings for hand prostheses, and we configure this dataset following Rahimian et al. (2020). We use subjects 1,3,4,6,7,8,9 for training and subjects 2,5,10 for testing.

2 EXPERIMENTAL SETUP

All the experiments are run on a single Nvidia GTX 1080 Ti. The batch size is 64 for all datasets. We use Adam optimizer Kingma & Ba (2015) with an initial learning rate of $3 * 10^{-4}$ and the decay rate is 0.95 for each epoch. The maximum number of training epoch is 100. θ in FSEA and λ in Eq. (11) of the paper is set to 2 and 0.1, respectively, for all the experiments.

3 EVALUATION METRICS

Following previous works, we use $Accuracy$, $weighted\ F1\ score$ (F_w) and $macro(mean)\ F1$ $score$ (F_m) as our evaluation metrics. The definitions are as follows:

$$Accuracy = \frac{Number\,of\,correct\,classification}{Total\,number\,of\,test\,samples} \tag{1}$$

$$F_w = 2\sum_{i=0}^{C-1} w_i \frac{precision_i \times recall_i}{precision_i + recall_i} \tag{2}$$

[1] https://archive.ics.uci.edu

$$F_m = \frac{2}{C}\sum_{i=0}^{C-1}\frac{precision_i \times recall_i}{precision_i + recall_i} \tag{3}$$

where i is the class index, $w_i = N_i / \sum_{i=0}^{C-1} N_i$ is the proportion of samples of the class, and N_i is the number of samples in i-th class. C is the total number of the class. Note that a few related work (e.g., Qian et al. (2019)) lists the $micro(mean)$ $F1$ as the evaluation metric; therefore, we also test this metric and show the superior result (**0.9324** *vs* 0.8366 Qian et al. (2019)) among others in OPPOR dataset.

4 HYPER-PARAMETER ANALYSIS

In the proposed model, there are two main hyperparameters: (i) energy splitting ratio ζ in *FSEA*, (ii) regularization parameter λ in Equation (11). Tables 1 and 2 show sensitivity analysis on them.

As we can see in Table 1, the default energy splitting ratio ζ is set to 2.0, which results in a more even energy distribution (The standard deviation is minimum). Note that the finer splitting ratio such as 1.0 would result in too many subbands, thereby incurring an over-fitting problem. On the contrary, a coarser division (e.g., 2.5) would lead to uneven information distribution, reducing feature representation capacity.

In Table 2, we can find that the hyperparameters λ is relatively stable. We choose λ = 0.1 and use it for all the datasets in our paper.

Table 1: The performance of different splitting ratios in the OPPOR dataset. $Std.$ denotes the standard deviation of the subbands' energy, and the lower values represent the more even energy distribution.

Splitting ratio	1.0	1.5	2.0	2.5
$Std.$	13.61	10.08	**9.43**	15.67
F_m	0.733	0.752	**0.763**	0.749

Table 2: The performance of different λ in the OPPOR dataset.

λ	0.05	0.1	0.5	1.0
F_m	0.757	**0.763**	0.761	0.756

5 MODEL SIZE AND TRAINING TIME

The model size and computing time in each dataset are shown in the table. As we can see, the training time depends on both the number of samples and the model size, which is not very much for a specific task. To deploy the proposed model to a resource-constrained device in real-world applications, we also present a design space exploration technique in Sec. 14.2.

Table 3: The model size and training time in different datasets.

Dataset	OPPOR	UCI-HAR	BCI2a	NinaPro
# Samples	46528	7296	129344	34048
Model size (M)	4.162	0.119	9.65	1.0
Training time s/epoch	67.66	7.46	287.75	14.82

6 EXPERIMENTS ON OTHER TYPES OF DATA

To demonstrate the effectiveness of the proposed methods on different types of data, we also present the results on other four datasets in Table 4. As we can see, the proposed method still achieves considerable results among other baselines. Since the data in Earthquakes and ElectricDevices also has these characteristics in their spectrum. We give their dominant energy range more dimensions in

the feature vector, which brings significant performance improvements compared to other baselines. As for the data in SmallKitchenAppliances, which shows a relatively homogeneous frequency distribution, therefore, the performance gain from our method is not much.

Table 4: Performance comparison on different types of data.

Datasets	T-WaveNet	MLP	FCN	ResNet	Encoder	MCNN	t-LeNet	MCDCNN	Time-CNN	TWIESN
Earthquakes	76.26	71.7	72.7	71.2	74.8	74.8	74.8	74.9	70.0	74.8
ElectricDevices	76.52	59.2	70.2	72.9	67.4	33.6	24.2	64.4	68.1	60.7
SmallKitchenAppliances	79.53	37.1	78.3	78.6	59.6	36.9	33.3	48.5	61.5	65.6

7 THE PERFORMANCE ON LIMITED NUMBER OF TRAINING DATA

To demonstrate the effectiveness of the proposed method on the limited number of training data, we present the results of different decreasing ratios. As we can see, compared with other baselines, the proposed method is more robust than others when the number of training data decreases. However, when the data is too little (< 0.5), all baselines will be affected significantly, but the T-WaveNet shows relatively less impact.

Table 5: The performance on limited number of training data.

Ratio	1.0	DEC	0.9	DEC	0.75	DEC	0.5	DEC	0.25	DEC
ConvLSTM	0.672	-	0.577	-14.1%	0.422	-37.2%	0.211	-68.6%	0.164	-75.6%
FilterNet	0.743	-	0.703	-5.4%	0.623	-16.2%	0.322	-56.6%	0.190	-74.4%
T-WaveNet	0.763	-	0.734	-3.8%	0.677	-11.2%	0.483	-36.7%	0.260	-65.9%

8 IMPACT OF FREQUENCY SPECTRUM ENERGY ANALYSIS

Table 6 shows that the two-phases subband splitting scheme in *FSEA* is necessary. Though formants-guidance splitting can get the most informative bands, each subband's energy (information) may vary significantly, which reduces the network's representation capability. On the other hand, the energy-guidance splitting process can obtain the relative even energy distribution, but the impact of formants is ignored. Therefore, the combination of them shows much better results.

Table 6: The effectiveness of FSEA for OPPOR dataset.

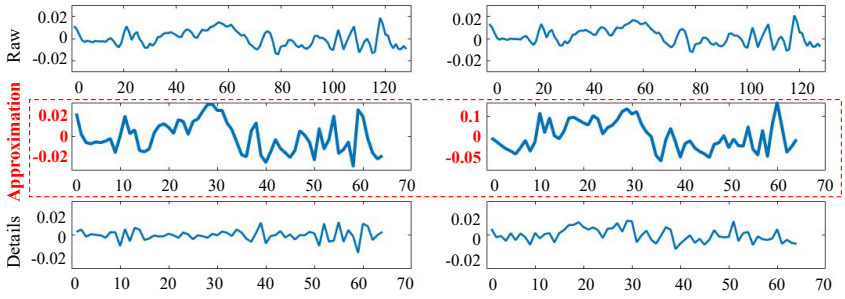
FSEA process	F_w	F_m
w/o Energy-guidance	0.921±0.011	0.745±0.008
w/o Formants-guidance	0.918 ±0.005	0.742±0.004
full model	**0.931±0.013**	**0.763±0.011**

9 IMPACT OF FREQUENCY BISECTION OPERATOR

Our frequency bisection operator is built with an INN-based wavelet transform, and Fig. 1 (b) shows that the proposed module inherits the wavelet's frequency decomposition function. In contrast with the traditional Haar wavelet basis, our learned wavelet coefficients also facilitate the classification task by increasing the value of the approximation part (low-frequency component) to enhance its impact. This is also in line with our domain knowledge that the low-frequency components of the sensor signal usually reflect the intrinsic property of the activities.

10 ROBUSTNESS EVALUATION

To validate the robustness of the networks to time series signals obtained from an unseen subject, we perform Leave-One-Subject-Out (LOSO) cross-validation on the OPPOR dataset. The results are listed in Table 7. As can be observed, the proposed *T-WaveNet* achieves better inference performances when testing on data recorded on the unseen subject compared with *FilterNet*, showing much higher



(a) Haar (b) INN-based

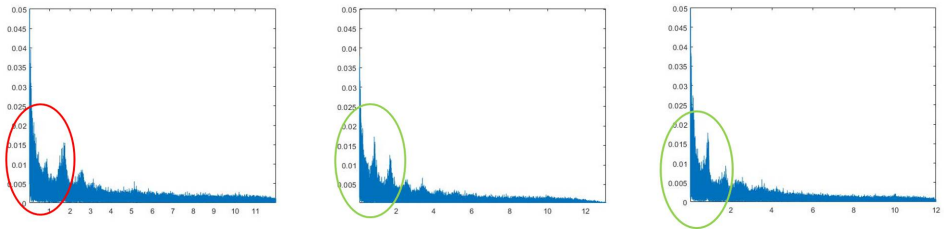
Figure 1: **Comparison of frequency bisection operator design in signal decomposition**. The raw signal (*walking* in UCI HAR) is decomposed into two sub-series corresponding to the low- and high-frequency subbands (approximation and details) by **(a) traditional *Haar*** and **(b) INN-based wavelet transform** (our solution). As can be observed, the approximation value in our learning-based solution is greatly enhanced.

Table 7: Leave-one-subject-out cross-validation in OPPOR dataset.

Subject#	FilterNet		*T-WaveNet*	
	F_w	F_m	F_w	F_m
1	0.7969	0.4869	0.8598	**0.6675**
2	0.8196	0.5260	0.8699	0.7147
3	0.8055	0.4806	0.8620	0.7016
Average	0.8073	0.4978	0.8639	0.6926

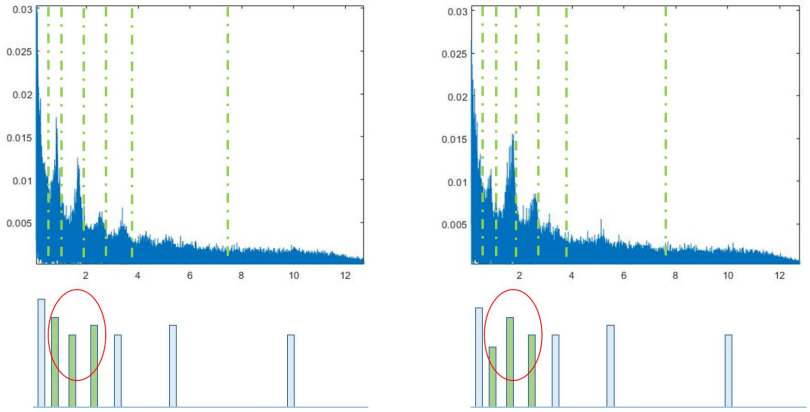
robustness. This is because the feature fusion module takes into consideration the personalized heterogeneity in the data and learns to weigh the frequency subbands adaptively for different subjects.

Although our proposed *T-WaveNet* is more robust than existing techniques, its performance also varies with different individuals, especially when using Subject #2 and Subject #3 as the training set and leaving Subject #1 as the testing set. Therefore, we visualize the frequency distributions of different subjects in Fig. 2. The figure shows that the frequency components of Subject #1 in the range of [0,2] are quite different from the other two subjects. Therefore, if we use such a setting, the distribution of the training set is different from the testing set, which leads to relatively poor performance. At the same time, as our *T-WaveNet* disentangles the diverse information from different signal frequency components, such heterogeneity can be confined to a local range with our feature representation, as shown in Fig. 3.



(a) Subject 1 (b) Subject 2 (c) Subject 3

Figure 2: The spectrum of different subjects in the OPPOR dataset. The frequency components in the red circles are quite different from those in the green circle.



(a) Training set (S2,S3) (b) Testing set (S1)

Figure 3: The main difference between the training set and the testing set is highlighted with the red circle, and the heterogeneity problem is confined locally.

11 Evaluation of INN-based Wavelet Transform

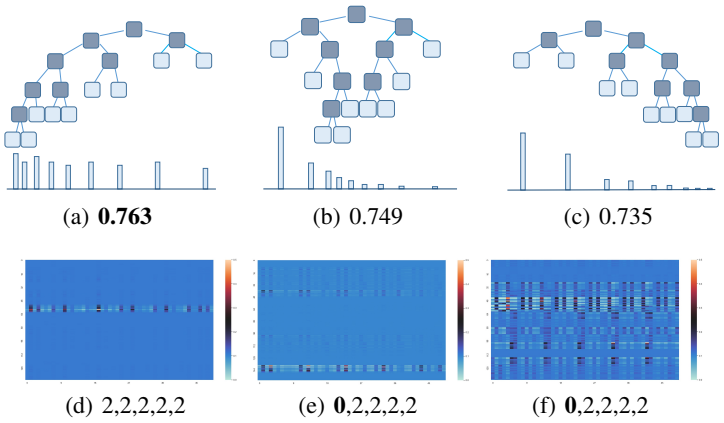
From the Eq. (3-6) in the paper, we can observe that the standard Lifting Scheme equations are particular forms of our INN-based wavelet equations. Therefore, we experiment on the deep version of the Lifting Scheme, in which we realize the Predictor P and Updater U in Eq.$(3)(4)$ of the paper using the same deep modules as ϕ, ψ,ρ and η. Table 8 shows that our method consistently outperforms the deep Lifting Scheme version across all the network structure settings, showing the INN-based operator's superiority.

Table 8: L-k means the module has k convolution layers; C-k represents the output channel size of the first layer is k times of the input; D-k is the dilation size of the first convolution layer to enlarge the receptive fields.

Configuration	P, U		ϕ,ψ,ρ,η	
	$F_m\%$	$F_w\%$	$F_m\%$	$F_w\%$
default:(L-2, C-3, D-1)	92.6	73.5	**93.1**	**76.3**
(L-1, C-3, D-1)	91.2	69.4	**91.9**	**71.4**
(L-3, C-3, D-1)	91.5	71.6	**92.4**	**73.5**
(L-2, C-1, D-1)	92.7	74.5	**92.8**	**75.5**
(L-2, C-5, D-1)	91.6	72.1	**92.7**	**75.7**
(L-2, C-3, D-3)	92.0	73.6	**92.7**	**75.1**
(L-2, C-3, D-5)	92.0	72.0	**92.5**	**73.3**

12 Evaluation of the Feature Fusion Module

The feature fusion module is realized by the multi-head self-attention mechanism. In Fig. 4(d-f), we concatenate the attention map in each head of the self-attention module of one sample in the vertical direction and five samples in the horizontal direction. As we can see the attention scores in Fig. 4(d-f), the proposed feature fusion module can fine-tune the results by assigning different weights to each frequency subband feature. However, compared with (e) and (f) showing diverse attention heads, the attention scores in (d) are almost homogeneous except for one head. It illustrates that the uneven energy distribution((b),(c)) will make the fusion module take lots of effort to learn the attention weights; thus, it should increase the burden of the learning process but with limited improvements, verifying our claim in the paper.



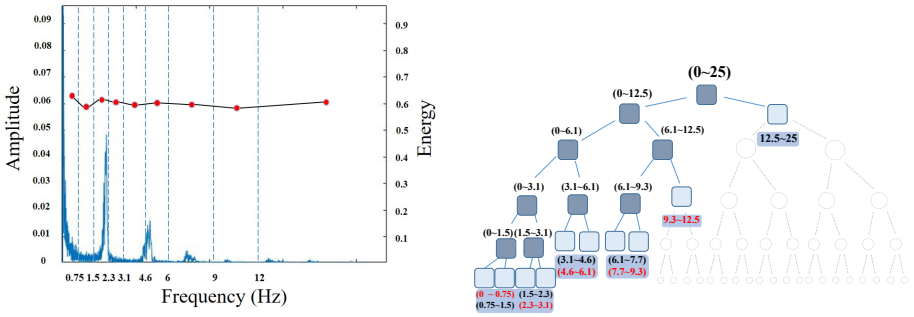
(a) **0.763** (b) 0.749 (c) 0.735

(d) 2,2,2,2,2 (e) **0**,2,2,2,2 (f) **0**,2,2,2,2

Figure 4: Different tree-structures in the OPPOR dataset. (d-f) are the attention maps of the corresponding network structures (a-c), respectively. The number below the attention maps are the predicated results, and the wrong prediction is highlighted in bold.

13 Frequency Spectrum Energy Analysis

To further verify the effectiveness of our *FSEA* method, we train the variants of the *T-WaveNet* in other three datasets.



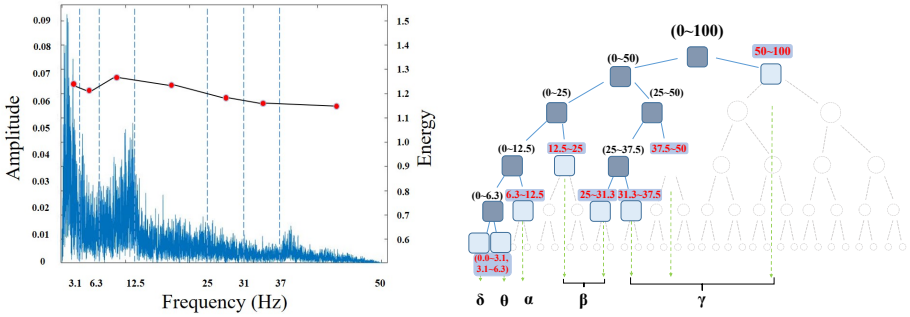
(a) FSEA (b) UCI-HAR

Figure 5: (a) The frequency spectrum energy analysis of the UCI-HAR. (b) The network structure. The dominant frequency subbands based on formants are highlighted with red color. (unit:Hz)

UCI-HAR: The spectrum of UCI-HAR (Fig.5(a)) shows that the dominant signal energy distributes in the frequency range below 5 Hz. We construct the tree-structured network based on the set of frequency subbands, as shown in Fig.5(b). In contrast, other variants such as Fig.10(a-h,j) with uneven energy distribution limit the network representation capability.

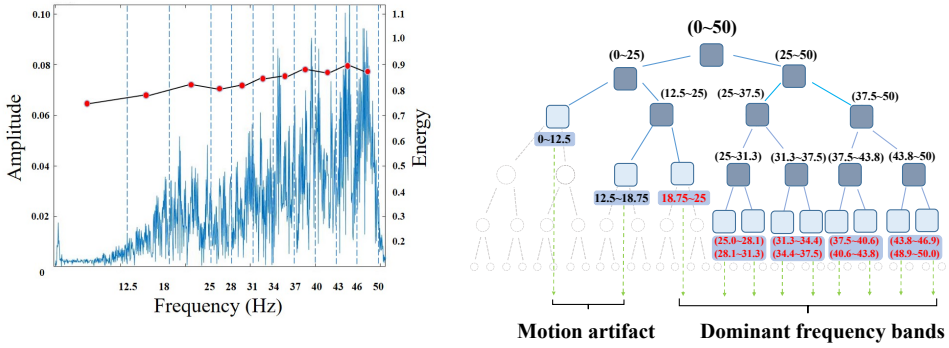
BCICIV2a: In addition to the result of the frequency spectrum energy analysis shown in Fig. 6(a), we also consider the domain knowledge of the EEG signal, *i.e.*, the multi-wave ($\alpha, \beta, \gamma, \delta$ and θ) analysis Abhang et al. (2016), to configure the tree-structured network as Fig.6(b). To demonstrate the effectiveness of the proposed subbands splitting strategies, we also experiment on other network variants. From the results in Fig. 11, we can observe that our proposed tree structure (Fig. 11(g)) performs the best. We attribute this to the even energy distribution among the decomposed frequency subbands obtained using our frequency spectrum energy analysis scheme.

NinaPro DB1: Unlike the other three datasets where their dominant energy is distributed in low-frequency subbands, the sEMG's is between 20Hz and 2000Hz during muscle contraction, and the motion artifact noise falls mainly in 0-20Hz (Fig. 7(a)). Considering that the sampling rate of NinaPro



(a) FSEA (b) BCICIV2a

Figure 6: (a) The frequency spectrum energy analysis of the BCICIV2a. (b) The network structure. The dominant frequency subbands based on formants are highlighted with red color. (unit:Hz)



(a) FSEA (b) NinaPro DB1

Figure 7: (a) The frequency spectrum energy analysis of the NinaPro DB1. (b) The network structure. The dominant frequency subbands based on formants are highlighted with red color. (unit:Hz)

DB1 is only 100Hz, based on the Nyquist theorem, the frequency range of the collected data is in 0-50Hz. Therefore, the energy of the sEMG signals mainly lies in a narrow frequency range(20-50Hz). Accordingly, the resulted structure is shown in Fig.7(b). Note that although the motion artifact contains less information, it can still enhance the model robustness. Moreover, the various tree structures in Fig. 12 show the performance of different energy distributions. The tree structure in Fig. 12(h) is obtained by the proposed frequency spectrum energy analysis, which achieves the highest accuracy.

14 Design Space Exploration

Real-world applications usually require deploying the DNN on resource-constrained IoT devices. However, it is challenging because one needs to seek a trade-off between the performance and available resources. To maximally preserve the accuracy, we present a design space exploration technique for *T-WaveNet* to achieve Pareto optimal solution under the resource constraints of the IoT devices.

14.1 Method

The computation and memory cost of *T-WaveNet* is mainly determined by the number of frequency bisection operators *FBO* nodes with gate "1" in the tree structure, because nodes with gate "0" will

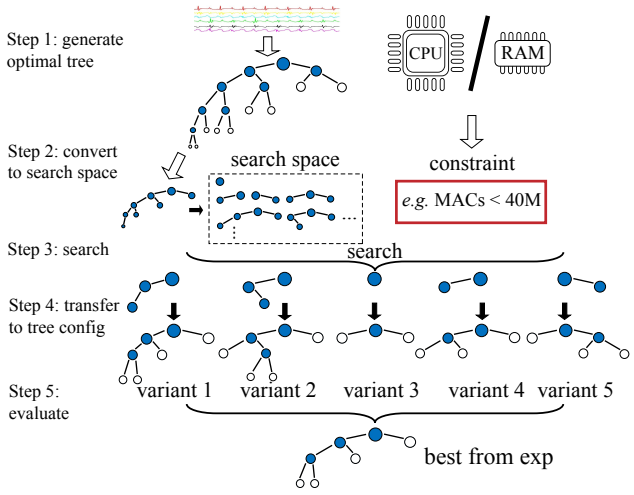


Figure 8: Design space exploration. Blue nodes are with gate "1", and white nodes are with gate "0".

directly bypass the input without extra computation. In view of this, we can reduce the resource requirement of *T-WaveNet* by re-configuring the number of nodes of the optimal tree structure.

Fig. 8 shows the procedure of the design space exploration. First, given a particular signal, the optimal tree architecture is obtained through *FSEA*. The search space is constructed from all the sub-trees of the optimal tree structure after removing all the nodes with gate "0" from it, where each sub-tree corresponds to a configuration of *T-WaveNet* following the *FSEA* principle. Next, we select all the sub-trees that fulfill the resource constraints from the search space, and complement the children of each candidate with nodes of gate "0". Finally, the best variant is determined by evaluating the performance of all the candidates through experiments and choose the best one.

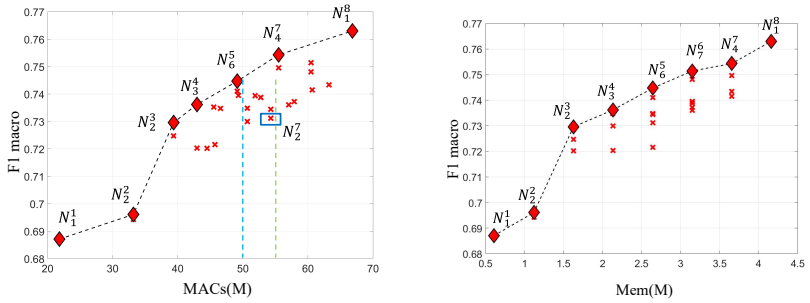
14.2 Performance-Cost Trade-Off Evaluation

With the design space exploration method mentioned above, we study the computational cost and the recognition performance of different candidates in the search space on the OPPOR dataset. The default *T-WaveNet* structure of this dataset given by *FSEA* is shown in Fig.1(b) in the paper. The performance of different candidates under various *MACs* and *Memory* constraints are depicted in Fig. 9. The best-performed models under different constraints are denoted using the red diamonds, which form the Pareto front that optimally trades off F_m with *MACs* or *Memory* constraints. The other models are marked using red crosses. One can pick the optimal model structure under certain constraint on the Pareto front. The reason is detailed as follows.

Generally speaking, the candidate models that utilize more computational resources would achieve better performance. For example, when MACs constraint is 50M (Fig.9 (a)), the structure N_6^5 whose MACs almost reaches 50M achieves the highest F1 score among structures that satisfy the constraint. However, this is not always the case. For instance, if the MACs constraint is 55M, the N_2^7 with the maximum MACs value is superior than the N_6^5. This is because the node configuration of N_6^5 gives a more balanced energy distribution compare with the N_2^7, which would benefit the feature learning. Therefore, in this case, we would still choose the structure N_6^5 on the Pareto front, which achieves higher F1 score with less MACs. We can draw similar conclusions for the *Memory* constraint based on Fig.9 (b).

15 Discussion

In this material, we provide more details about the experimental settings and discuss different modules proposed in *T-WaveNet* in an interpretability way. Moreover, we also evaluate the effectiveness of the two-phase frequency spectrum energy analysis on other three datasets (**UCI-HAR**, **BCICIV2a** and **NinaPro DB1**). The results show that *FSEA* can provide a set of frequency subbands with

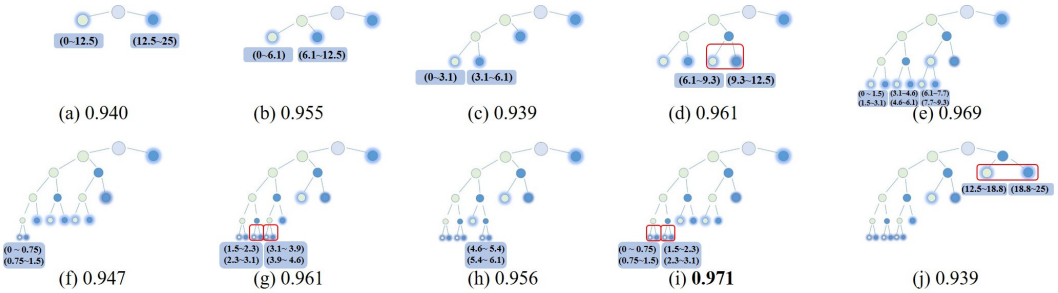


(a) MACs (b) Memory

Figure 9: **Performance-cost trade-off** of different model configurations. The points N_j^i represent different network architectures, where i denotes the number of nodes and j indexes one of the variants.

.

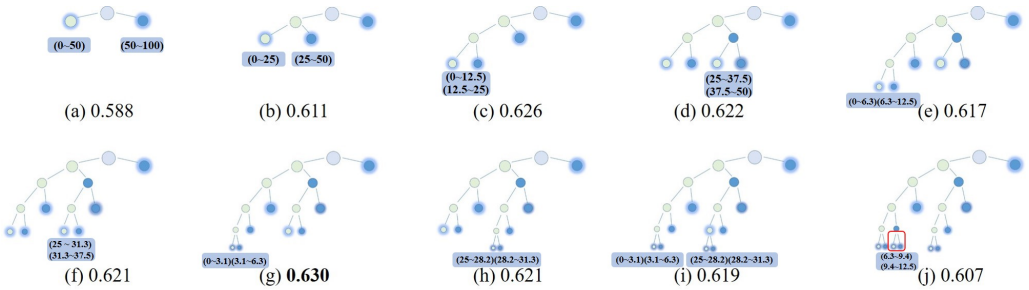
approximately even energy distribution, which can be utilized to construct an effective tree structure for feature extraction. Please note that the leaves' even energy distribution, rather than the tree structure's depth, is essential for feature learning of the tree-structured network. See examples in Fig. 10(j), Fig. 11(i)(j) and Fig. 12(i)(j). Finally, we present a design space exploration technique for *T-WaveNet* to achieve Pareto optimal solution under given computation or memory constraints.



(a) 0.940 (b) 0.955 (c) 0.939 (d) 0.961 (e) 0.969

(f) 0.947 (g) 0.961 (h) 0.956 (i) **0.971** (j) 0.939

Figure 10: The *Accuracy* results of different energy divisions in the UCI-HAR dataset.(unit:Hz)



(a) 0.588 (b) 0.611 (c) 0.626 (d) 0.622 (e) 0.617

(f) 0.621 (g) **0.630** (h) 0.621 (i) 0.619 (j) 0.607

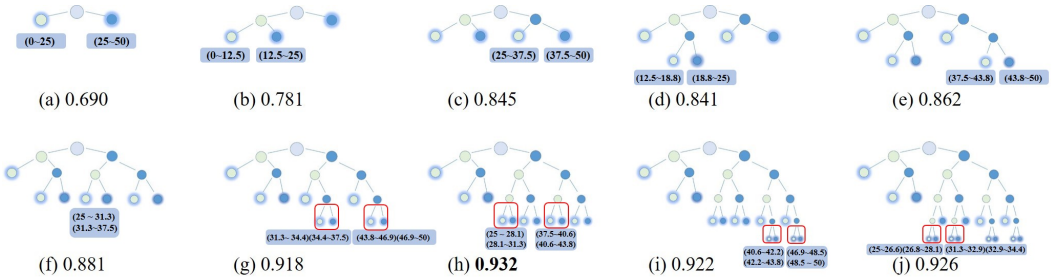
Figure 11: The *Accuracy* results of different energy divisions in the BCICIV2a dataset.(unit:Hz)

REFERENCES

Priyanka A Abhang, Bharti W Gawali, and Suresh C Mehrotra. *Introduction to EEG-and speech-based emotion recognition*. Academic Press, 2016.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Machine Learning*, 2015.

H. Qian, S. Pan, B. Da, and C. Miao. A novel distribution-embedded neural network for sensor-based activity recognition. *IJCAI 2019*, pp. 5614–5620, 2019.



(a) 0.690 (b) 0.781 (c) 0.845 (d) 0.841 (e) 0.862

(f) 0.881 (g) 0.918 (h) **0.932** (i) 0.922 (j) 0.926

Figure 12: The *Accuracy* results of different energy divisions in the NinaPro DB1 dataset.(unit:Hz)

S. Rahimian, E.and Zabihi, S. F. Atashzar, A. Asif, and A. Mohammadi. Xceptiontime: Independent time-window xceptiontime architecture for hand gesture classification. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1304–1308, 2020.

D. Zhang, K. Chen, D. Jian Zhang, K D. Chen, and L. Yao. Motor imagery classification via temporal attention cues of graph embedded eeg signals. *IEEE Journal of Biomedical and Health Informatics*, 2020.