Temporal Convolutional Explorer Helps Understand 1D-CNN's Learning Behavior in Time Series Classification from Frequency Domain

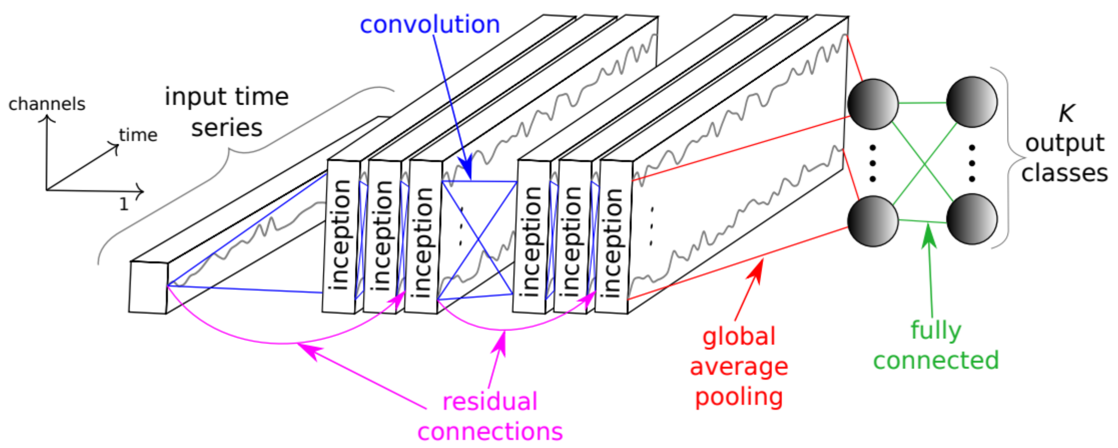
Junru Zhang[1], Lang Feng[1], Yang He[1], Yuhan Wu[1], Yabo Dong[1,*]

[1]College of Computer Science and Technology, Zhejiang University, Hangzhou, China

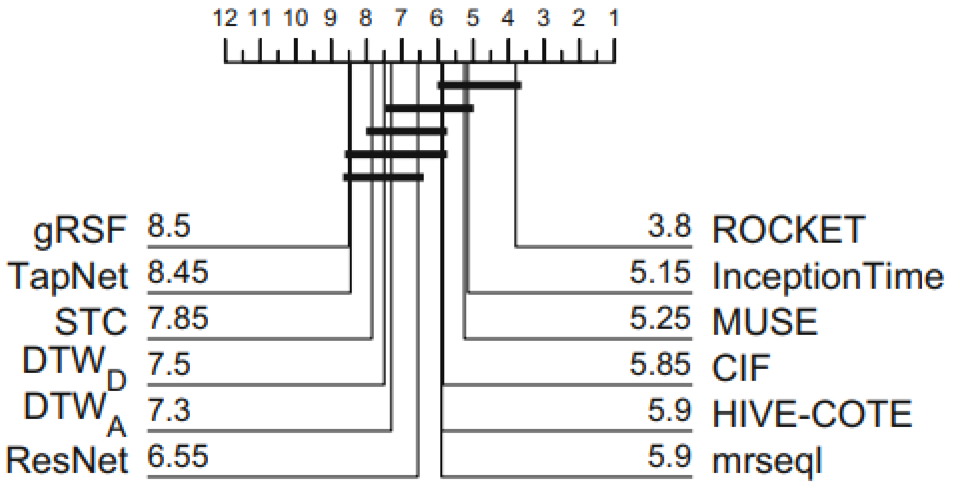
32nd ACM International Conference on Information and Knowledge Management

Background

1D-CNN-based methods like FCN, ResNet, and InceptionTime have proven effective for time series classification (TSC) tasks. However, unlike in the field of CV, the working mechanism of 1D-CNNs in time series has received relatively little attention.



InceptionTime Architecture (Image source: Hassan Ismail Fawaz et al. 2020)



Accuracy omparison on UEA Benchmark. InceptionTime outperforms more TSC problems. (Image source: A. P. Ruiz et al. 2020)

Motivation

We observed an intriguing phenomenon that deeper networks do not necessarily bring about better performance in TSC tasks. From Fig. 1, the results revealed that deeper networks do not always perform better. This unexpected finding has motivated us to investigate the complex learning mechanisms of deep 1D-CNNs.

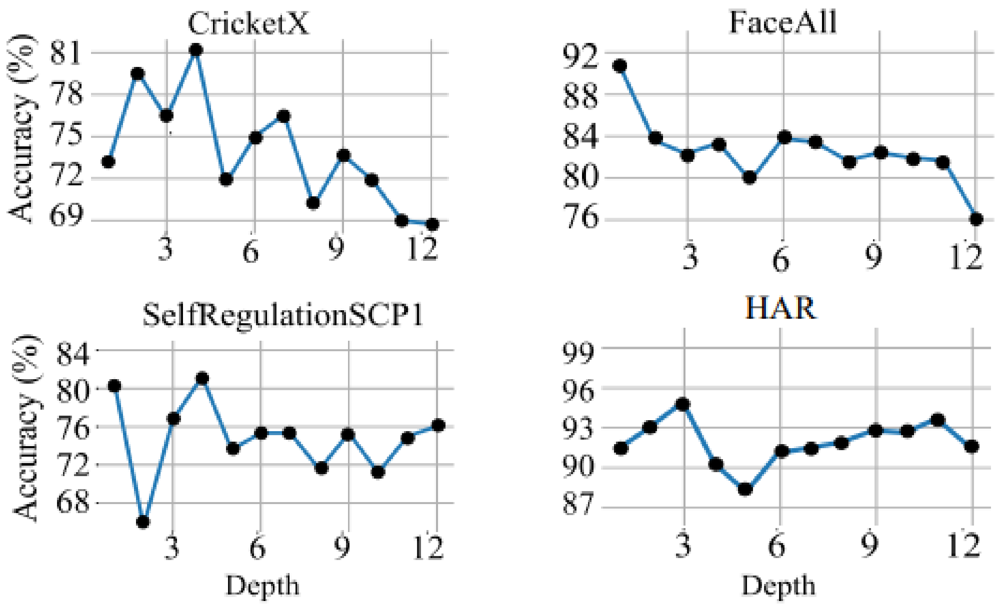


Fig 1: Accuracy Degradation: Deeper ≠ Better

CV Field

Recent studies (Wang et al. 2020, Xu et al. 2020, Xu & Zhou, 2021) have investigated the learning behavior of CNNs, particularly focusing on the correlation between the **frequency spectrum** of images and the behavior of deep CNNs.

Findings: CNNs often fit objective functions from low to high frequencies during the training process and that high-frequency components are difficult to encode.

Suggestion: CNNs were designed to fast learn a function with high frequencies to improve their performances.

Challenges: These studies conducted on CNNs for image learning may *not directly* apply to 1D-CNNs in time series.

Key Difference

- **Time series vs. Image**

- TSC tasks involve numerous channels with diverse attributes.
- Images typically have only three color channels.

- **1D-CNNs vs. CNNs**

- 1D-CNNs capture temporal features by recognizing cross-channel information.
- CNNs primarily focus on spatial features in images.

➡ Highlight the unique learning behavior of 1D-CNNs in TSC tasks compared to CNNs in image learning

Goal: Develop an explorer to understand the learning behavior of 1D-CNNs!

✓ Identify the underlying factors influencing the accuracy degradation phenomenon in TSC tasks

Design effective strategies to improve the performance of deep 1D-CNNs in TSC

Benefit the community to develop more suitable and powerful 1D-CNN-based classifiers for TSC tasks

Our work

Method: Temporal Convolutional Explorer (TCE)

Identify the frequency components of time series that are ***emphasized/overlooked*** by deeper convolutional layers

- ✓ Discovers that deeper 1D-CNNs tend to *distract* the focus from *low frequencies*, leading to accuracy degradation.
- ✓ Identify that the "*disturbing convolution*" is the primary cause of this problem.

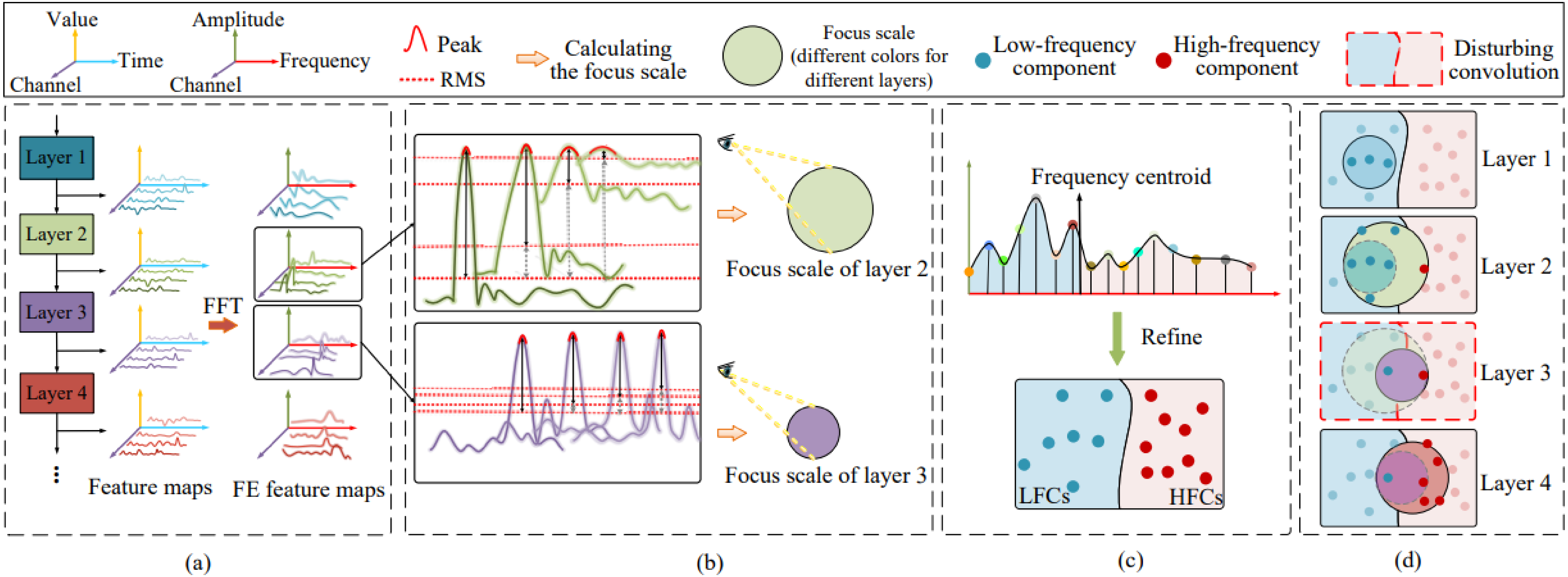


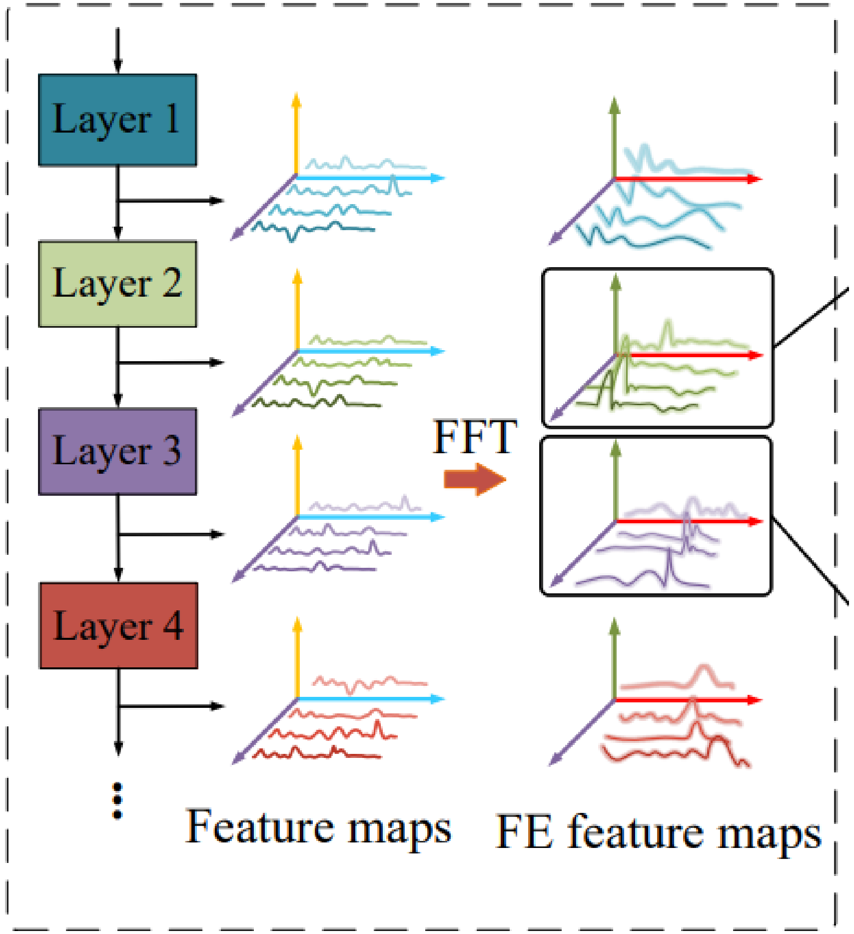
Fig 2: Illustration of TCE in four convolutional layers

Frequency-Extracted (FE) Feature Maps

Serve as the basic feature maps for TCE

- Obtained by transforming the feature maps using Fast Fourier Transform (FFT).

- Offer the unique representations of frequency features extracted by convolutional kernels.



(a)

Focus Scale

A measure of the diversity of frequency patterns captured by each feature map within a convolutional layer.

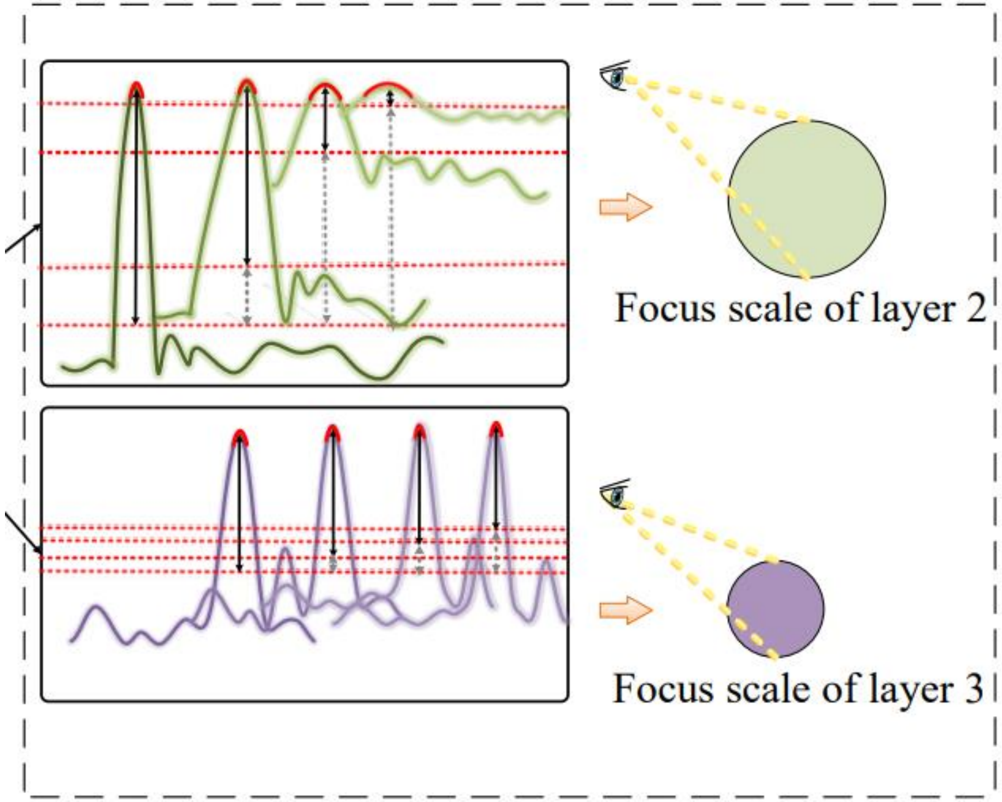
$$p_l^{c_l} = \frac{\max_{\omega}\{\mathcal{A}^{\omega}(\mathbf{o}_l^{c_l})\}}{\sqrt{\sum_{\omega \in \mathbf{I}} (\mathcal{A}^{\omega}(\mathbf{o}_l^{c_l}))^2}} \times \sqrt{\left\lfloor \frac{H_l - 1}{2} \right\rfloor + 1}, \tag{2}$$

$$v_l = \frac{\sum_{c_l=1}^{C_l} (p_l^{c_l} - \sum_{c_l'=1}^{C_l} p_l^{c_l'}/C_l)^2}{C_l}. \tag{3}$$

The variance of the ratio relationships between the peak and root-mean-square (RMS) of amplitude in the FE feature maps across all feature maps in a given convolutional layer.

Focus Scale

A *larger* focus scale implies that the layer responds to a *broader* range of frequency features, providing an intuitive understanding of the **frequency range** captured by the layer.



(b)

Frequency Centroid

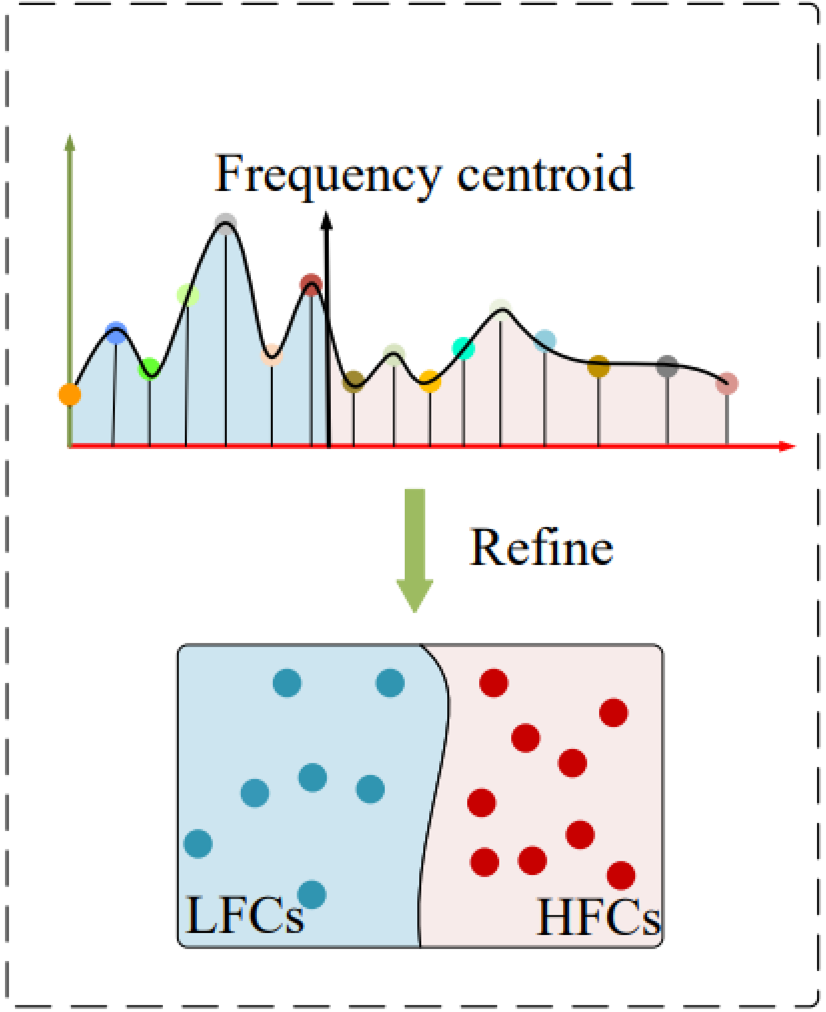
Represents the distribution centroid of the frequency components within a sequence:

$$\mathcal{F}^{fc}(\mathbf{s}) = \frac{\int_0^{\lfloor (T-1)/2 \rfloor} \omega \mathcal{A}^{\omega}(\mathbf{s}) \mathbf{d}\omega}{\int_0^{\lfloor (T-1)/2 \rfloor} \mathcal{A}^{\omega}(\mathbf{s}) \mathbf{d}\omega}, \tag{4}$$

It acts as a measure of central tendency for the signal's frequency components and helps to effectively identify and analyze spectral properties in diverse time series signals. A frequency centroid located at the center indicates a uniform frequency distribution, with no preference for lower or higher frequencies.

Frequency Centroid

A lower frequency centroid signifies a greater concentration of energy in the low-frequency range of the signal, while a higher frequency centroid emphasizes the presence of high-frequency information, often associated with rapid changes or noise disturbances.



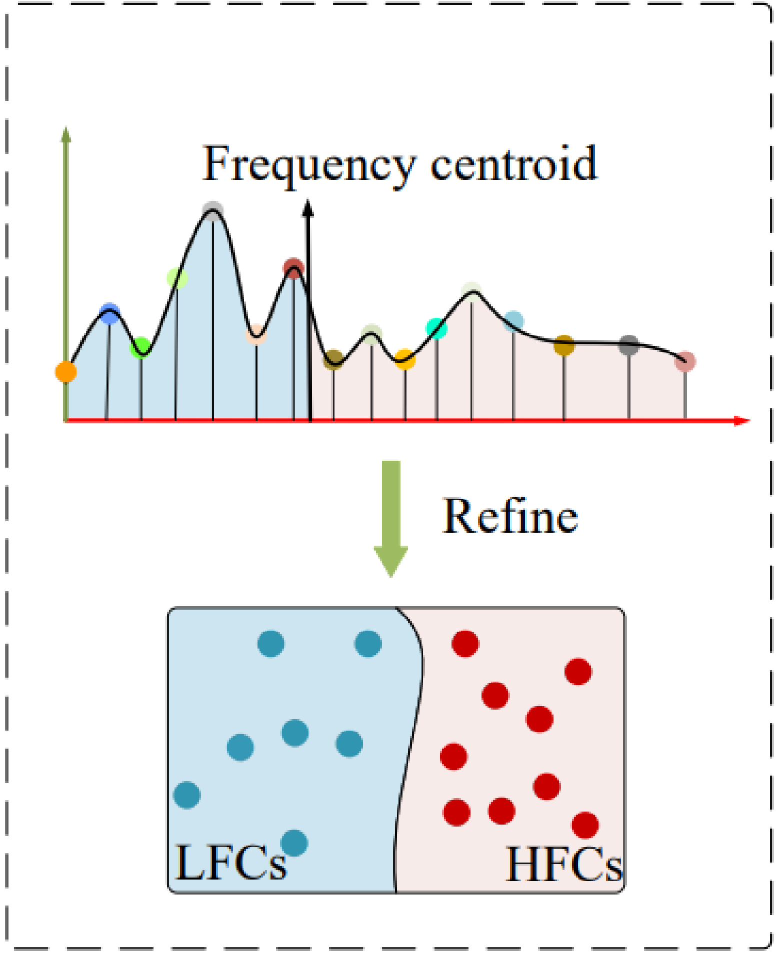
(c)

Frequency Centroid

Distinguish between low-frequency components (LFCs) and high-frequency components (HFCs) based on the relative positions of frequencies in relation to the centroid.

- Allow for the filtering of specific frequency components using the inverse FFT, enabling the observation of the learning behavior of the 1D-CNNs with respect to the targeted frequency components.

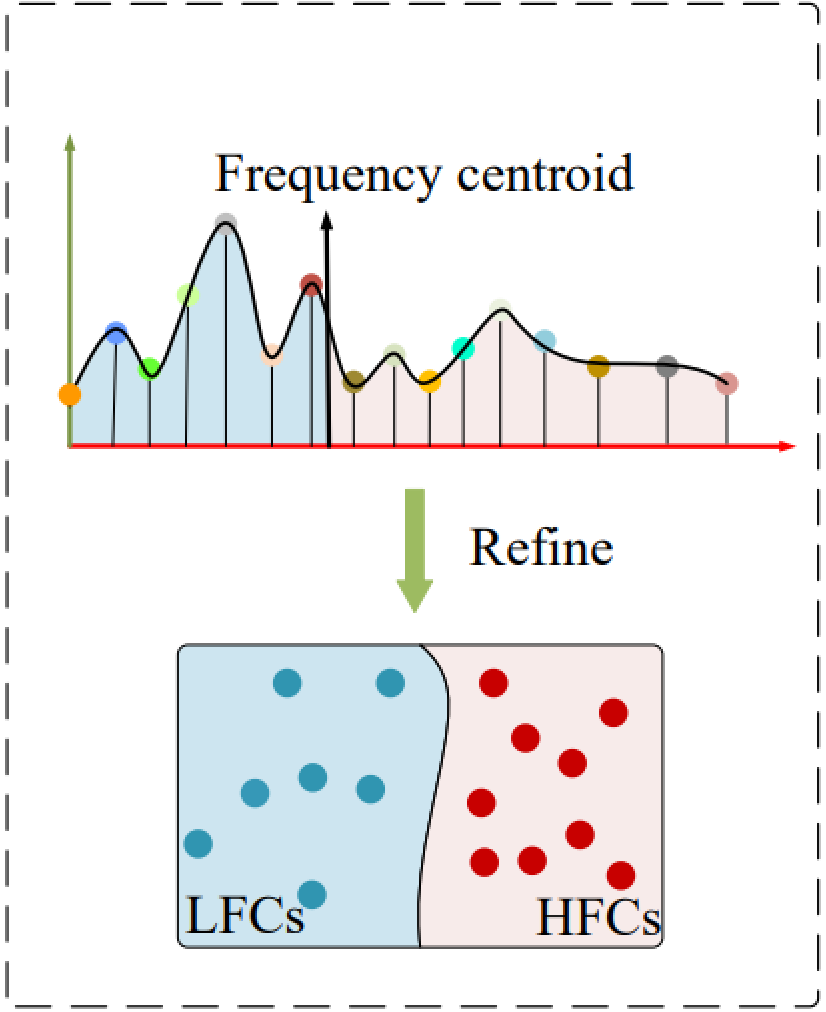
- Monitor the evolution of frequency centroids within the feature maps of each convolutional layer aids in understanding the learning bias of individual layers concerning frequency components.



(c)

Frequency Centroid

When a deeper convolutional layer focuses its attention on modeling HFCs within the signal, it triggers a concentrated activation of high-frequency energy, resulting in an elevated frequency centroid. Hence, the frequency centroid helps to understand the **learning direction** of deeper 1D-CNNs in terms of frequency.



(c)

Learning Behaviour

Focus scale & Frequency centroid

Learning range & Learning direction

The frequency components that deeper convolutional layers emphasize or overlook

The change in focus scales is defined as:

$$M_l = v_l - v_{l-1},$$

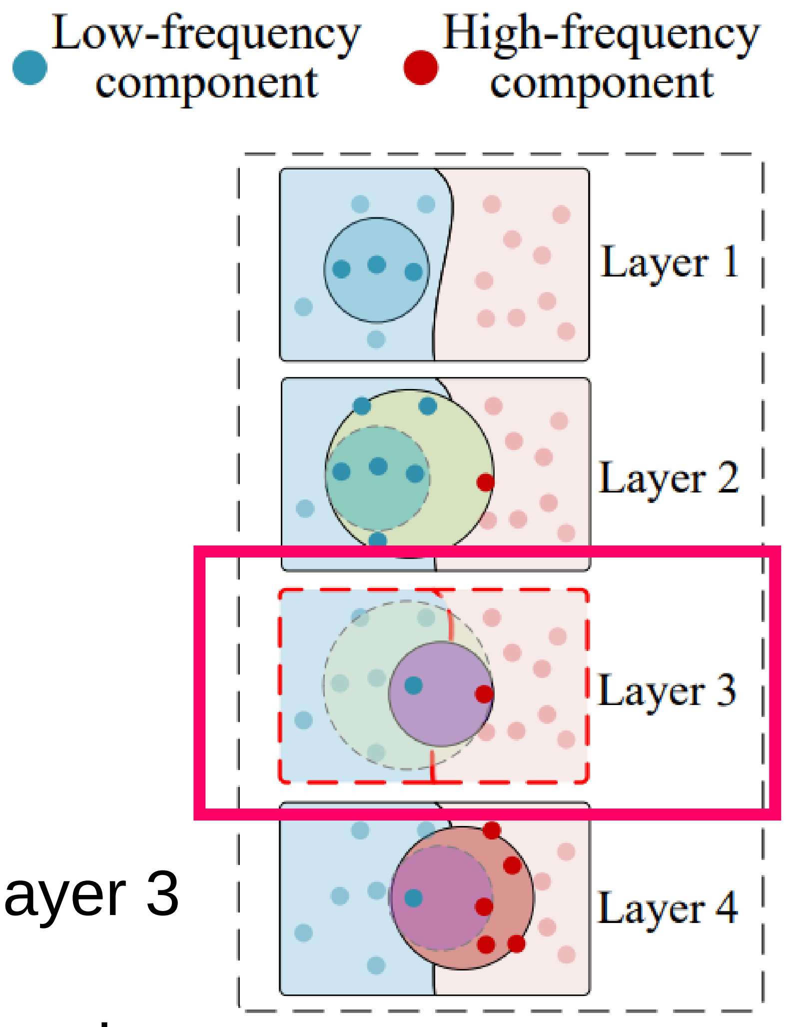
to determine whether a layer is losing focus on certain frequency components.

Learning Behaviour & Disturbing Convolution

REMARK 1. *1D-CNNs with the increase of depth tend to distract the focus from the LFCs. TCE can utilize the change in focus scales, which is defined by*

$$M_l = v_l - v_{l-1},$$

to describe the internal factor of this tendency: l-th convolutional layer with negative M_l causes the 1D-CNN to lose focus on certain frequency components, mainly the LFCs.



(d)

Fig 2(d): Focus scales of layer 2 and layer 4 increase, while layer 3 (referred to as the '**disturbing convolution**') decreases.
With the shift of deeper focus to HFCs, layer 3 drives the network to distract from LFCs.

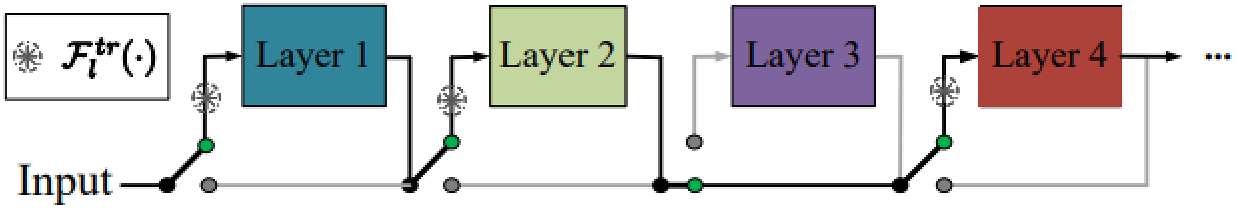
Findings through experimental exploration using TCE:

- LFCs significantly enhance the overall generalizability of 1D-CNNs. However, despite deeper 1D-CNNs demonstrating a stronger capacity to learn LFCs, they often underutilize this potential due to **distraction from lower frequencies**.

- This distraction narrows the range of captured low frequency features, **leading to a decrease in accuracy**.

- This phenomenon is attributed to the '**disturbing convolution**', identified as the convolutional layer with a negative M_l.

Regulatory Framework



To mitigate 1D-CNN's suboptimal learning behavior and accuracy degradation, we propose a regulatory framework.

- Consist of TCE and a gating mechanism (a switch structure).

- Identify 'disturbing convolutions,' enabling selective bypassing or retention of convolutional layers.

- Alleviate the negative effects of disturbing convolutions as well as can reduce the consumption of memory and computational overhead.

Experiments

➢ Verification of Learning Behavior using TCE:

- **Datasets**: two UTSC tasks (CricketX and FaceAll) and two MTSC tasks (SelfRegulationSCP1 and HAR)
- **Architecture**: ResNet with different depths

➢ Regulatory Framework Evaluation:

- **Datasets**: all datasets from three archives (UCR, UEA, UCI), encompassing 128 UTSC and 31 MTSC datasets.
- **Architectures**: three competitive 1D-CNN backbones, namely FCN, ResNet, and InceptionTime.

Learning Ability for LFCs: A quantitative analysis using the frequency centroid of TCE on frequency components.

- As depth increases, 1D-CNNs struggle to generalize well on HFCs, leading to noticeable discrepancies between training and test accuracy, while differences are less pronounced and decreasing for LFCs.

.

LFCs significantly contribute to the generalization performance of 1D-CNNs in TSC tasks, and deeper networks are better equipped to leverage and generalize LFCs.

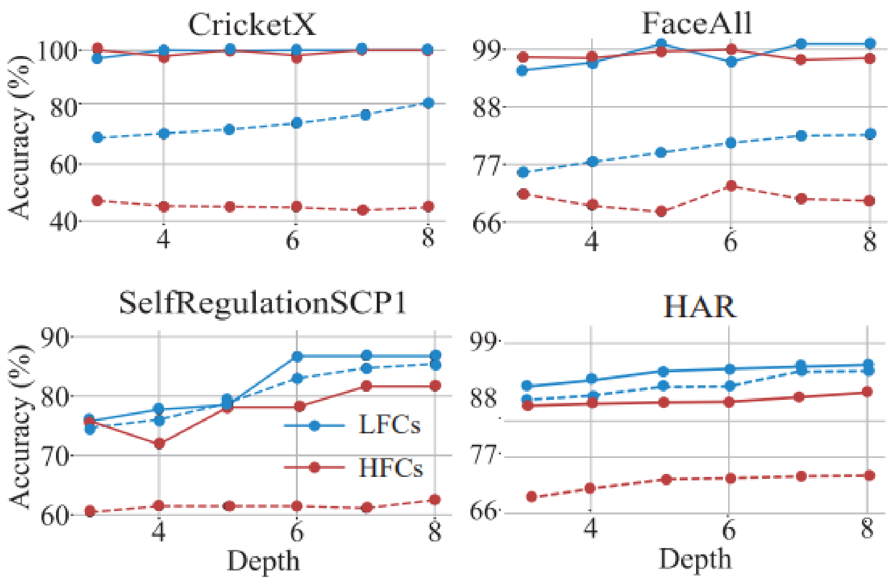


Fig 5: Training and Test Accuracies of ResNet at Different Depths. LFCs (in blue) and HFCs (in red). Solid lines represent training accuracy, while dotted lines represent test accuracy.

Learning Ability for LFCs

- Deeper networks have faster learning speeds for LFCs and achieve the target training loss with fewer epochs.

Combined with the good generalization on LFCs as shown in Fig. 5, ***deeper 1D-CNNs exhibit stronger learning ability for LFCs in time series data.***

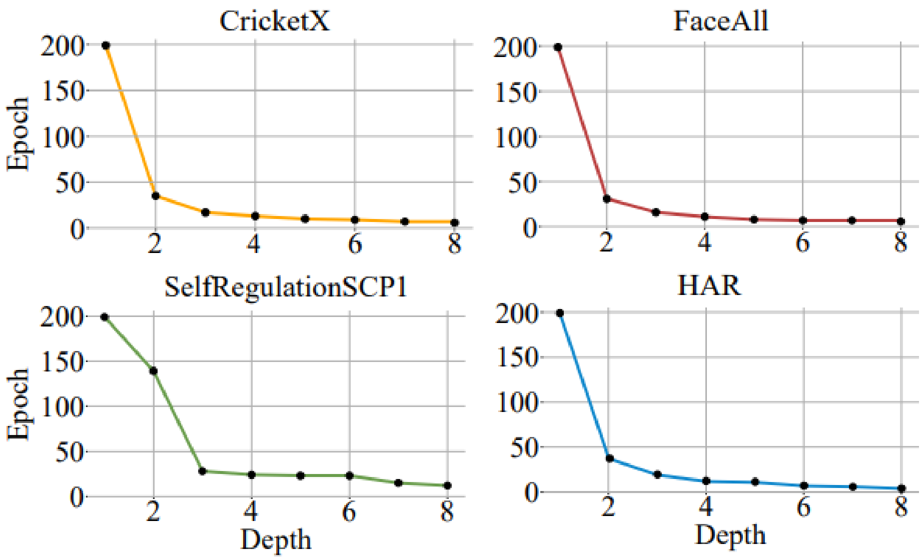


Fig 6: The required training epochs for ResNets with different depths to achieve the target loss on LFC

Why the stronger learning ability for highly generalizable LFCs has not improved the performances of the deeper ResNet ?

Analysis on Learning Behavior

- Deeper layers exhibit significantly elevated frequency centroids, indicating a more pronounced activation of high-frequency energy.

*As the time series data progresses through successive layers, the network **progressively** focuses its attention on capturing **HFCs**.*

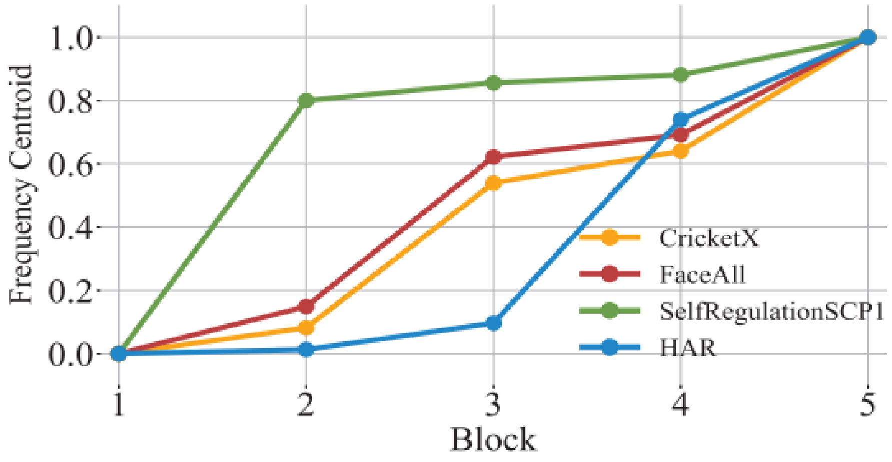


Fig 7: The average frequency centroid of feature maps for each convolutional block in deeper ResNet.

- Accuracy does *not gradually improve* as expected *when adding LFCs to HFCs* of four datasets.

Contrary to the previously demonstrated ability of deeper ResNet to generalize well on LFCs alone, it struggles to fully utilize its learning capacity when these low frequencies are restored from signals containing HFCs.

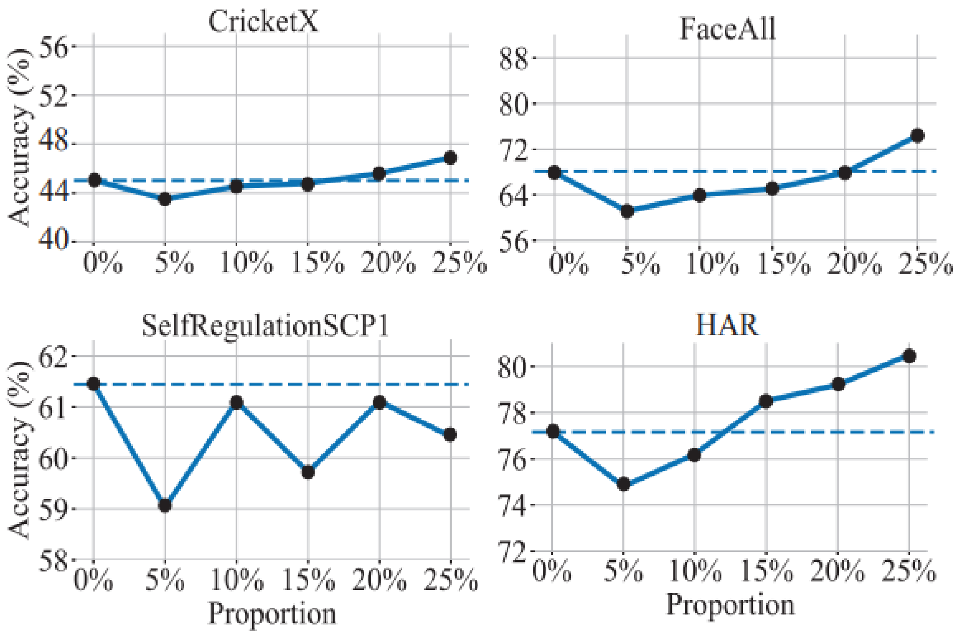


Fig 8: Accuracy of deeper ResNet on four datasets with different proportions of LFCs added to HFCs.

The presence of significant HFCs can lead to the inability of deeper ResNet to fully leverage low-frequency information, consequently impairing its overall performance.

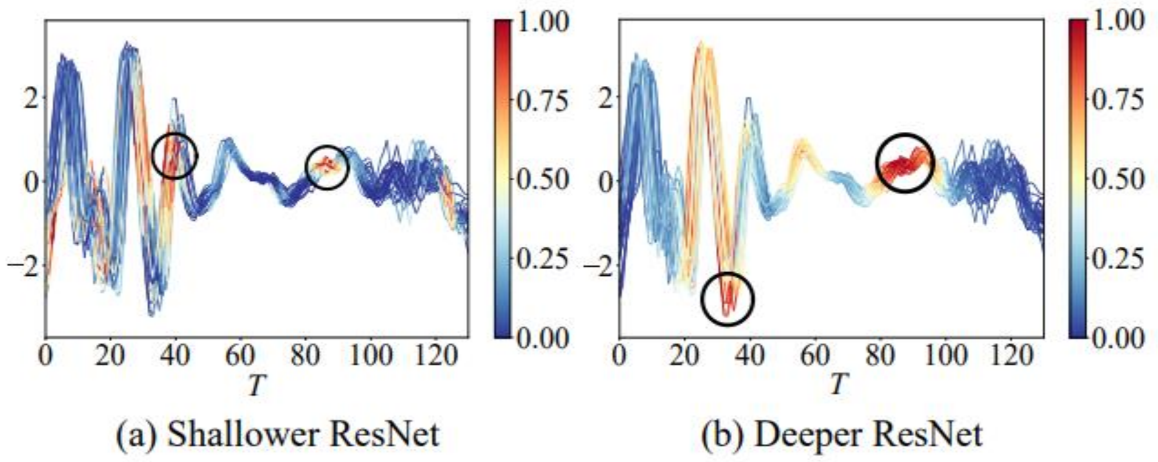


Fig 9: Grad-CAM of shallower and deeper ResNets on FaceAll instances in class 3. The shallower ResNet focuses on areas with subtle changes, while the deeper ResNet emphasizes regions with shorter wavelengths or rapid oscillations.

In conclusion, ***despite their strong learning ability for LFCs, deeper 1D-CNNs distract the focus from low frequencies leading to the accuracy degradation issue.***

Why the stronger learning ability for highly generalizable LFCs has not improved the performances of the deeper ResNet ?

Analysis on Disturbing Convolution: A repeated experiment for comparison

- When adding LFCs to HFCs, the accuracies gradually improved.

Deeper ResNet can refocus on the LFCs by skipping disturbing convolutions, and its learning ability for LFCs can be effectively leveraged for accuracy improvement.

Confirm TCE's insights: ***The disturbing convolution is the potential driving factor for the accuracy degradation phenomenon caused by distraction on LFCs, and skipping over it can alleviate this issue.***

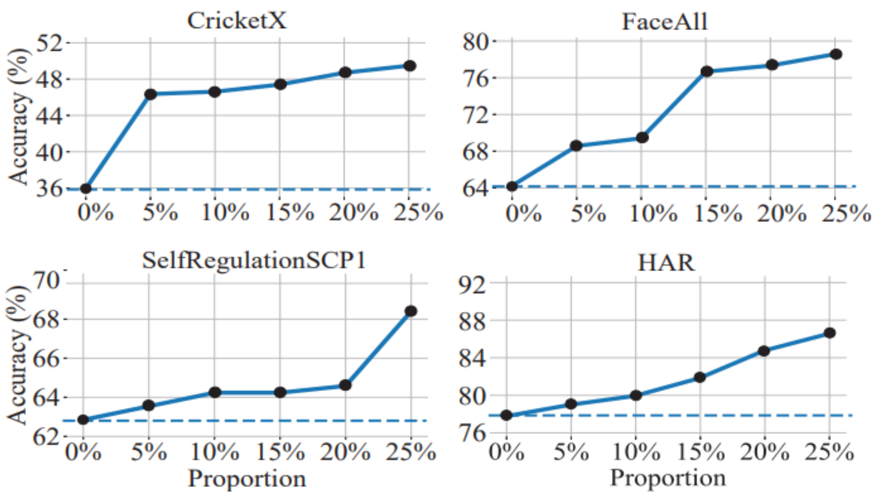
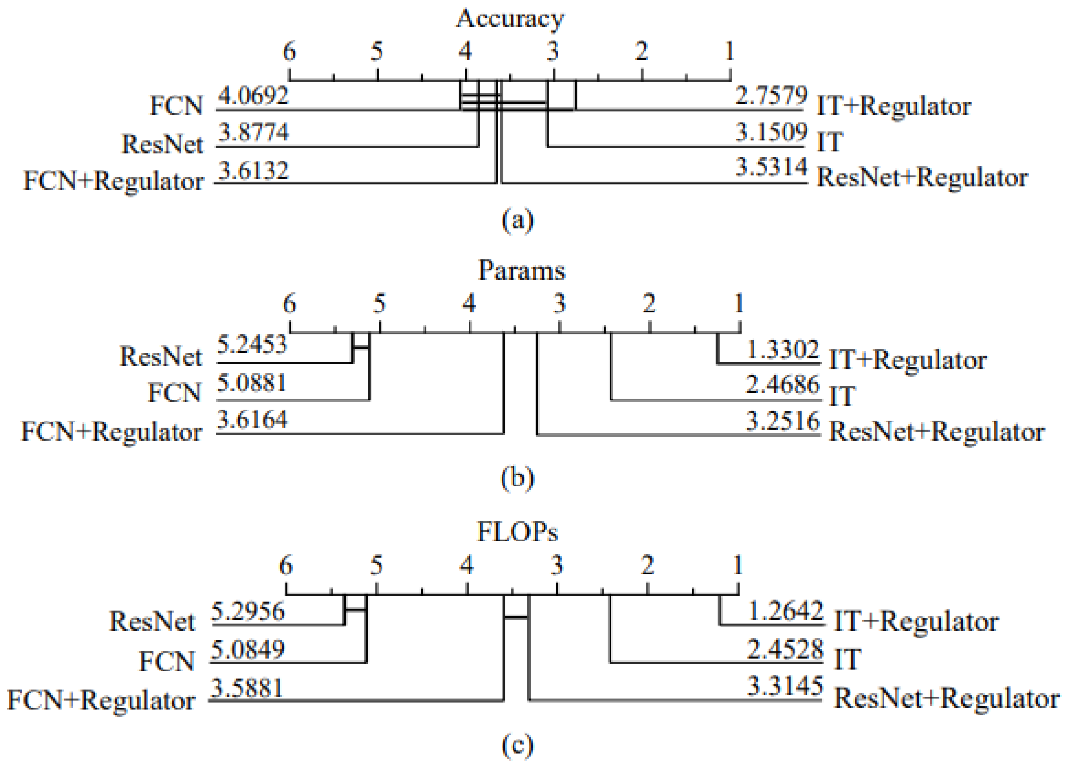


Fig 11: Accuracy of deeper ResNet on four datasets with different proportions of LFCs added to HFCs, after skipping the specified disturbing convolution.

Performances of Our Regulatory Framework



Our framework enables 1D-CNNs to significantly reduce memory and computational resource consumption while maintaining comparable or higher accuracy.

This success demonstrates that TCE can enhance the performance of 1D-CNNs in time series classification tasks.

Summary

- We empirically explore 1D-CNN learning behavior in TSC tasks, identifying the distraction of deeper CNNs from LFCs and the role of disturbing convolutions in accuracy degradation.

- We propose a regulatory framework to address this issue, supported by extensive experiments.

- Our goal is to uncover bottlenecks in deep 1D-CNNs for TSC tasks and share insights with the community. In future work, we aim to promote the design of robust TSC networks and enhance the theoretical understanding of 1D-CNNs.

Questions ?

Presenter: Junru Zhang
junruzhang@zju.edu.cn

Discussion and cooperation are welcome!