IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 5 June 2024, accepted 14 June 2024, date of publication 17 June 2024, date of current version 24 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3416034

APPLIED RESEARCH

A Lightweight Network for Seismic Phase Picking on Embedded Systems

YADONGYANG ZHU[1], SHUGUANG ZHAO[1], WEI WEI[1], FUDONG ZHANG[2], AND FA ZHAO[2]
[1]School of Information Engineering, Beijing Institute of Petrochemical Technology, Beijing 102617, China
[2]College of Instrumentation and Electrical Engineering, Jilin University, Changchun 130026, China

Corresponding author: Fa Zhao (zhaofacy@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 42104175, in part by the Young Elite Scientists Sponsorship Program by Beijing Association for Science and Technology (BAST) under Grant BYESS2022093, and in part by the Research and Development Program of Beijing Municipal Education Commission under Grant KM202210017006.

ABSTRACT Phase picking is a critical task in seismic data processing, where deep learning methods have been applied to enhance its accuracy. While lightweight deep learning networks have been optimized for edge computing devices, there is a lack of networks developed explicitly for embedded systems. This paper presents a seismic phase picking model, a hybrid network integrating convolutional neural networks and Transformer, designed for embedded systems. Optimizing network parameters and computational resources, the model significantly reduces resource consumption while guaranteeing accuracy. It employs a multi-branch architecture. Specifically, the global branch employs a modified self-attention mechanism, effectively extracting global features through shared contextual information. The local branch retains local information from the input features. Such a multi-branch architecture facilitates effective interaction between global features and local details, thereby more efficiently capturing the relationships among features. The model can be configured into variants with different sizes to match various embedded systems. This research evaluated the model using the Stanford Earthquake Dataset, achieving a precision of 99.9% for the P-phase and 99.3% for the S-phase. On Raspberry Pi, the model reduced inference time by 58.1% compared to the earthquake transformer while maintaining comparable detection performance.

INDEX TERMS Phase picking, lightweight network, modified self-attention, embedded systems.

I. INTRODUCTION

In recent years, with the rapid advancement of science and technology, research in seismology has deepened. These studies not only aid researchers in better understanding the causes and characteristics of earthquakes but also provide scientific foundations for earthquake data analysis and inversion, the development of early warning systems, and disaster risk assessment [1], [2], [3]. Utilizing advanced monitoring equipment and data analysis techniques, researchers can more accurately predict seismic activity and formulate corresponding warning and emergency response plans, thereby effectively reducing the harm that earthquakes pose to society.

The associate editor coordinating the review of this manuscript and approving it for publication was Prakasam Periasamy.

In seismic data processing, phase picking is a crucial and challenging task. Traditional machine learning methods primarily rely on manually extracted features and classical algorithms, such as PAI-S/K, [4] FilterPicker, [5] and AIC [6]. These methods typically require manually set thresholds, are susceptible to noise, and usually cannot match the accuracy of manual picking. In contrast, deep neural networks can automatically learn complex features from large datasets without manual feature extraction. They outperform traditional methods in accuracy and generalization and have shown promising results in seismic data denoising, phase picking, and data inversion [7], [8]. However, performance enhancement is accompanied by increased model size and longer inference times. As distributed computing and IoT technologies are increasingly applied to geophysical instruments, the demand for on-site seismic phase

picking is rising, driving networks with fewer parameters and lower FLOPS.

Existing research on lightweight networks primarily targets mobile and edge devices. Although processors on edge devices are less potent than GPUs/TPUs, their computational capacity is still considerable. However, controllers in existing geophysical instruments mainly employ embedded CPUs/MPUs designed for low power consumption and high efficiency. Despite limitations in processing speed and parallel computation, embedded CPUs/MPUs have been applied intelligently in diverse domains, such as smart agriculture [9], [10], fisheries [11], and livestock management [12], [13]. Given their extensive use in geophysical exploration instruments, this research intends to design a seismic phase picking network, optimized for embedded systems.

When it comes to designing efficient computational networks, Convolutional Neural Networks (CNN) are favored concerning their processing speed and strong inductive capabilities, making them a cost-effective choice for lightweight networks. Nonetheless, CNNs are constrained by their limited local receptive fields, capturing local spatial correlations in images at each layer but failing to account for global interactions, which hinders further performance enhancement of the model [14], [15], [16]. Compared to CNNs, Transformers offers a novel paradigm by implementing global interactions through self-attention mechanisms rather than focusing solely on local spatial correlations. In computer vision, Transformers have demonstrated superior performance to CNNs in many existing studies. It effectively captures information across the entire image by decomposing images into a sequence of tokens and leveraging self-attention mechanisms. To enhance Transformer performance, the trend is toward increasing model parameters, which requires substantial computational resources [17], [18], [19].

On this basis, hybrid networks that integrate the strengths of Transformers in learning global context with the powerful inductive biases of CNNs have manifested significant performance improvements in image processing research [20], [21], [22], [23]. These networks facilitate global feature interaction while preserving local feature details.

This study revolves around applying a seismic phase picking Transformer (SPPFormer) on embedded systems, proposing a lightweight deep learning network that integrates CNN and Transformer. This hybrid network integrates the inductive learning of CNN with the global perception ability of the Transformer. SPPFormer uses a multi-branch architecture, where the global branch modifies the self-attention mechanism in the Transformer, optimizing network parameters and computational resources to achieve a balance of high accuracy and fast inference speed on embedded systems. The local branch retains local input features. This multi-branch design allows the network to learn global and local information, effectively capturing relationships between features and aiding in accurately identifying P-waves and S-waves in seismic signals. The contributions of this study are as follows:

1. A lightweight deep learning network named SPPFormer, which integrates CNN and Transformer, is presented for seismic phase picking on embedded systems.
2. The SPPFormer block employs a multi-branch structure that facilitates effective interaction between global attributes and local details, thereby more effectively capturing inter-feature relationships.
3. The SPPFormer can be configured into variants with different model sizes to match the computational capacities and resource constraints of embedded systems, enabling deployment on various devices.

II. RELATED WORKS

A. LIGHTWEIGHT NEURAL NETWORKS

With the flourishing applications of machine learning, there is a surging demand across diverse domains to harness the potential of deep neural networks effectively. This demand has spurred research into efficient network architectures. For instance, the MobileNet [24] drastically alleviates the computational burden of traditional convolution operations by introducing depthwise separable convolutions, making them more applicable to resource-constrained edge computing environments. MobileNetV2 [14] further refines the design by incorporating inverted residual modules. Building on this, research efforts have been targeted at proposing more efficient CNN architectures, such as Inception [25] and MnasNet [26], designed to optimize network performance and reduce computational cost. Another line of research focuses on the development of lightweight networks, such as ShuffleNetv1 [27], ESPNetv2 [28], GhostNet [29], MobileNeXt [30], EfficientNet [31], and TinyNet [32]. These modeling strategies aim to strengthen computational efficiency with as few parameters as possible. Despite this, these architectures typically emphasize capturing local spatial correlations within convolutional layers instead of global information exchange, which could be a limitation in their applications. Large matrix computations, particularly on CPUs, still pose a considerable computational overhead.

B. SELF-ATTENTION MECHANISMS

Recent advancements reveal that leveraging self-attention mechanisms and extensive datasets, Vision Transformer (ViT) [33] and its variants [34], [35], [36], [37], [38] have achieved unprecedented inference accuracy in numerous computer vision tasks [39], [40]. However, ViT-based networks often require significant computational power and memory resources. Researchers are exploring diversified approaches to optimize the operational efficiency of these networks. Notably, MobileViT [21] integrates the lightweight MobileNet [28] with the ViT structure, maintaining a lightweight profile while achieving impressive performance in visual recognition tasks. Another study by EdgeViT [41] introduces a local-global-local bottleneck design to effectively reduce model size, which promotes deployment in practical applications. Additionally, studies

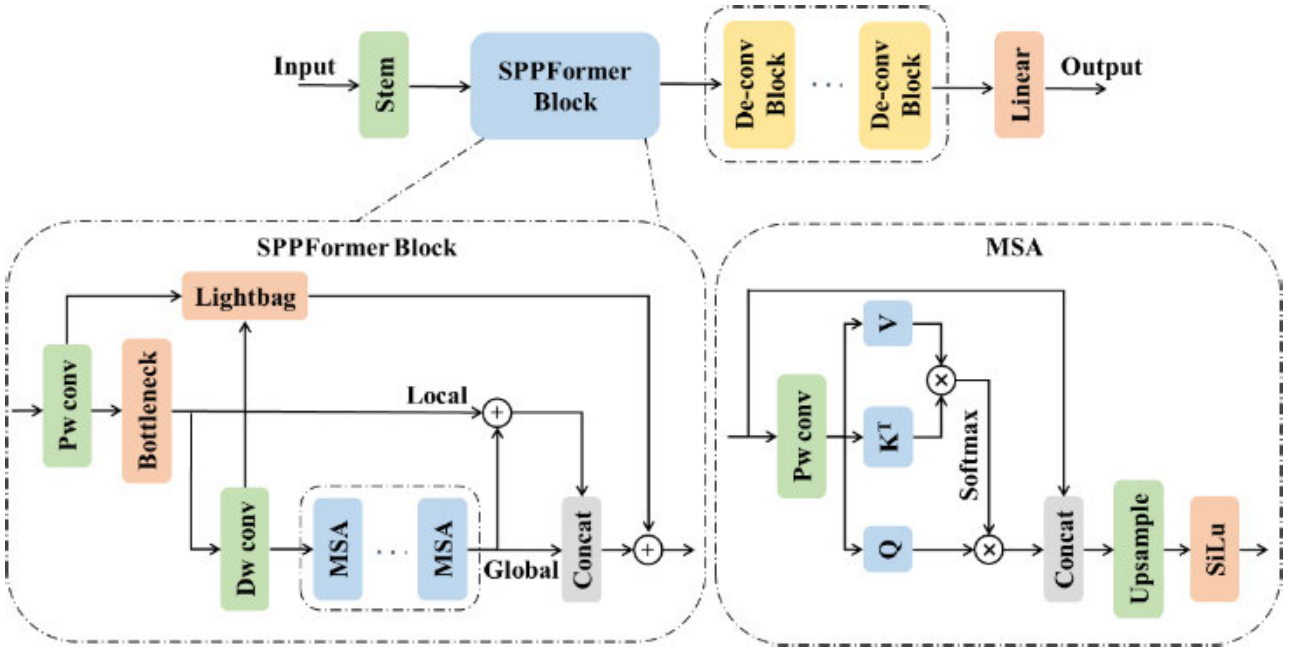

FIGURE 1. Overview of SPPFormer. The full details of each block were presented in the methods section.

integrating Transformers and CNNs, such as RepViT [42], manifest that such combinations can dramatically foster performance. Concurrently, the MobileFormer [43] model employs a parallel strategy, using MobileNet [24] to capture local features and a Transformer model for global features, integrated through a bidirectional bridging mechanism, thereby drastically lifting overall performance. Finally, EfficientFormer [23] focuses on a hybrid approach, combining the merits of convolutional and self-attention layers, ensuring both high recognition accuracy and network efficiency.

C. MODEL EVALUATION

It remains a challenge to find networks that reduce parameters and computational operations while maintaining the performance of hybrid CNN and Transformer architectures. Although some hybrid networks have made attempts at inference on edge devices, the chips used in relevant research are still designed for high computational performance. In lightweight model research, emphasis is mainly laid on model parameters and computational complexity. In contrast, research by Dehghani [44] and Vasu [45] reveals that inference time is equally critical when performing inference on single-board computers and does not always directly correlate with the model's parameter count and computational operations.

Moreover, in existing research, efforts to optimize model inference time are relatively scarce, as evidenced by networks such as MNASNet [26], MobileNetv3 [46], and ShuffleNetv2 [16]. This is particularly noticeable in the study of hybrid CNN and ViT architectures, where the mainstream focus lies on parameters and computational complexity rather than higher inference efficiency. A key reason for this trend is that assessing inference time must consider specific hardware characteristics. Different processor platforms have unique instruction sets and compilation tools.

III. MTEHODS

This research aims to develop a lightweight hybrid CNN and Transformer network to better balance inference time and accuracy when deploying the network on embedded systems.

A. PRINCIPLE OF MODEL

This section provides a comprehensive explanation of the proposed network. For phase picking on embedded systems, this research has designed a resource-efficient deep learning network that fuses CNNs with Transformers, which can account for the transformation and transition of different semantic features between CNNs and Transformers. The model precisely predicts each time point by mapping input time series samples to corresponding probability outputs. This approach efficiently processes time series data and reliably detects crucial information in seismic signals.

In the initial stage of the network architecture, unlike the Vision Transformer (ViT) [33], which linearly maps input data into tokens, this research employs a series of convolutional layers to process the input [23], [47], [48]. Given the computational challenges of the Transformer's attention mechanism, which scales quadratically with the rise of tokens, extensive feature data drastically raise computational complexity. To solve this, some studies have applied attention mechanisms only to local sub-regions or tokens of the feature data [49], [50], [51], [52]. While this reduces computational

complexity, it is at the price of ViT's key advantage of global information aggregation, potentially reducing inference accuracy, as global contextual information is crucial in the model.

The SPPFormer block developed in this study involves multiple parallel branches to address this. For one thing, the MSA is applied within the global branch, leveraging contextual information for global feature extraction. In addition, the local branch maintains the integrity of original features for further data processing. More detailed information is presented in section *C*. Moreover, this research proposes MSA, detailed in section D, to facilitate global feature extraction while ensuring compatibility with embedded systems.

B. OVERALL DESIGN ARCHITECTURE

Figure 1. illustrates the overall architecture of the model. Initially, a series of standard convolutional layers is adopted to transform the input data into features instead of applying the linear mapping method in Transformer. This approach captures the features of the seismic data with lower computational expense, involving three convolutional layers with 3×3 kernels and a stride of 2. The SPPFormer block utilizes a multi-branch design to preserve local information during feature extraction and concurrently learn local and global information. Furthermore, MSA is introduced to restore representational capacity lost during feature extraction to capture the relationships between features more effectively. The deconvolution module maps the extracted features onto a probabilistic sequence correlated with the P and S phases of seismic signals at each point in time.

C. SPPFORMER BLOCK

Let $X \in R^{L \times C}$ represent the input feature for the SPPFormer. Through $X^{PW} = PW(X)$, the features undergo point-wise convolution (PW) to produce output X^{PW}; then, they are passed through a bottleneck layer ($BottleNeck$) for feature dimensionality reduction and enhancement of the model's non-linear expressive capabilities, yielding the output

$$X^{Bottle} = BottleNeck(X^{PW}). \tag{1}$$

The updated features X^{Bottle}, are split into two branches: local and global. The local branch passes X^{Bottle} directly in a residual manner, while the global branch subjects X^{Bottle} to depth-wise convolution (DW) convolution, producing output X^{DW},

$$X^{DW} = DW(X^{Bottle}). \tag{2}$$

This is followed by a modified self-attention (MSA) mechanism to obtain X^{MSA},

$$X^{MSA} = MSA(X^{DW}). \tag{3}$$

The features extracted from the local and global branches are summed to yield $X^{Feature}$,

$$X^{Feature} = X^{Bottle} + X^{MSA}. \tag{4}$$

To preserve more information, X^{MSA} and $X^{Feature}$ are concatenated along the channel dimension to obtain X^{Concat},

$$X^{Concat} = X^{Feature} + X^{MSA}. \tag{5}$$

In the hybrid architecture that integrates CNNs with Transformers, noise often disrupts the low-level features extracted by CNNs, exerting an impact on global information transfer. To mitigate this, a skip-connection branch named lightbag is applied. It adds the features extracted by PW and DW convolutions, fuses them through a 1×1 convolution, and then adds the fused information to X^{Concat}. In this way, an appropriate receptive field is ensured during skip connections.

Ultimately, the output of the SPPFormer is designated as X^{Block},

$$X^{Block} = lightbag(X^{PW} + X^{DW}) + Fuse(X^{Concat}). \tag{6}$$

D. MODIFIED SELF-ATTENTION

The multi-branch design compensates for the local feature information loss resulting from attention computation. The representational power is somewhat diminished as attention operations are conducted on low-resolution features. This research modified the attention module to address this, introducing a modified self-attention mechanism (MSA). The module's input and output have the exact dimensions, permitting the stacking of multiple modules to consolidate the model's feature extraction capacity.

The structure of MSA is as follows: First, $X_i \in R^{L \times C_i}$ is defined as the input to the mechanism. In this research, the standard output of a CNN is adopted as the input, specifically a 5×5 standard convolution. The input for attention, comprising queries, keys, and values, is formed by distributing the output of the convolutional layer accordingly.

This research aims to enhance the expressive capacity by facilitating the aggregation of spatial information within the input feature while mitigating the complexity of training. To counterbalance the rise in computational costs, the study eliminated the separate linear transformations applied to queries and keys, instead representing them as X'_Q, X'_K, X'_V,

$$X'_Q = reshape(X[0, \frac{C_i}{3}]), \tag{7}$$

$$X'_K, X'_V = Split(reshape(X[\frac{C_i}{3}, C_i])). \tag{8}$$

where reshape alters the channel count of the input features, and split bisects the input tensor by channel numbers. The output of the MSA is X^{MSA},

$$X^{MSA} = X + SiLU(UP(X'')). \tag{9}$$

The $SiLU$ activation function is defined as: $f(x) = x * sigmoid(x)$, with $sigmoid(x)$ being the standard sigmoid function, mapping values between 0 and 1. The characteristics of $SiLU$ cover non-linearity and continuous differentiability, and it is defined across the entire range from negative to positive infinity. X'' is defined as follows:

$$X'' = Concat(X, Attention(X'_Q, X'_k, X'_v)). \tag{10}$$

Concat denotes the operation of joining two tensors along their channel dimension. *Attention* is defined as follows:

$$Attention(Q, K, V) = Softmax(\frac{QK^T}{\sqrt{d_k}})V. \quad (11)$$

where d represents the number of channels in query and key.

IV. EXPERIMENTS RESULTS

Experiments were conducted to evaluate the SPPFormer and compare with existing networks on the Stanford Earthquake Dataset (STEAD).

A. EXPERIMENTAL SETTING

The SPPFormer is primarily designed for embedded systems commonly applied in seismic exploration equipment. Additionally, its performance on CPUs and GPUs was examined for comparison. This paper uses the following three processors and platforms during experiments.

- An ARM Cortex-A53 single-board computer operating at 1.4GHz, specifically the Raspberry Pi 3 Model B+. Although categorized as low-end, the ARM Cortex-A53 is a quad-core 64-bit processor. In addition, the Raspbian OS and PyTorch 1.6.0 were adopted to execute the model.
- An Intel Xeon CPU E5-2686 V4 processor based on the Broadwell architecture, with 18 cores and 36 threads, and a maximum single-core turbo frequency of 3.0GHz. Ubuntu 18.04.6 LTS and PyTorch 1.10.1 were utilized to run the network.
- An NVIDIA GeForce RTX 3090 GPU features 24GB of GDDR6X memory on a 384-bit bus. It has a core clock speed of 1,395MHz and a boost clock speed of 1,695MHz, with 10,496 CUDA cores and 328 third-generation Tensor cores for machine learning. The system utilizes Ubuntu 18.04.6 LTS and PyTorch 1.10.1.

This research implemented and tested all networks using the PyTorch framework, with all networks operating based on the Seisbench library. Inference time for data processing was measured by inputting batches into the network. The batch size was set to 1024 on the GPU and CPU and 32 on Raspberry Pi, and the runtime on each platform was recorded. Multiple inferences were executed to record the average latency to ensure accuracy. During measurement, unrelated applications were shut down. All computations were performed using 32-bit floating-point operations. Model training was implemented on an RTX 3090 GPU platform, with inference times being subsequently evaluated on varying platforms.

B. DATASETS

This study trained and tested networks using the STEAD datasets [53], developed by Stanford University to advance the application of machine learning in seismology. The dataset consists of high-quality global seismic and non-seismic signals recorded at a sampling rate 100 across three orthogonal components. It is collected worldwide and features diverse seismic characteristics with detailed metadata, including event location, magnitude, and station information. Each waveform record lasts one minute, comprising approximately 1.05 million three-component seismic records and around 100,000 noise samples. The diverse seismic events facilitate the network's development, training, and evaluation. The large-scale, precisely labeled STEAD dataset supports the advancement of robust deep learning networks. Stored in efficient HDF5 files with comprehensive CSV metadata, STEAD allows for rapid data retrieval and access. During training, the dataset is randomly split into 70% training, 10% validation, and 20% testing sets, maintaining the original 100Hz sampling rate.

C. EVALUATION

During training, this research approached phase picking as a sequence-to-sequence learning problem. Regarding evaluation criteria, model output is considered a True Positive (*TP*) if the peak probability differs from the actual event time by no more than 0.35 seconds. Conversely, it is deemed a False Positive (*FP*) if the difference exceeds 0.35 seconds. Additionally, if the model fails to generate a probability peak matching the actual event, it is classified as a False Negative (*FN*).

Based on these definitions, performance metrics were calculated for the classification model, including Precision (*Pr*), Recall (*Re*), F1-Score (*F1*), and the Matthews correlation coefficient (*MCC*), with the following formulas:

$$Pr = \frac{TP}{TP + FP} \quad (12)$$

$$Re = \frac{TP}{TP + FN} \quad (13)$$

$$F1 = 2 \times \frac{Pr \times Re}{Pr + Re} \quad (14)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (15)$$

The metrics above jointly provide a comprehensive framework for performance evaluation, guiding further improvements and judgments of the seismic phase picking model efficacy.

D. TRAINING DETAILS

This research examined six size variants, namely SPPFormer-XXS, XS, S, M, B, and L, as illustrated in Table 1. For SPPFormer-XXS, -XS, -S, and -M, the feature extraction module maps inputs to 32 SPPFormer-XXS, -XS, -S, -M with a varying number of stacked self-attention modules set at 1, 2, 3, and 4, respectively. The SPPFormer-XXS, B, and L map input features 32, 64, and 96 dimensions, each utilizing a single enhanced attention mechanism. In the meantime, the number of deconvolution module layers is adjusted to match the varying dimensions. All networks were trained and tested on a normalized dataset.

TABLE 1. Design of sppformer variants with different model sizes.

Stage	Type	Block	SPPFormer						Resolution		
			XXS	XS	S	M	B	L	XXS/XS/S/M	B	L
1	Embedding	Patch Embed	×3 (k=3×3, s=2)						(B,3,6000)	(B,3,6000)	(B,3,6000)
			dim.32	dim.32	dim.32	dim.32	dim.64	dim.96	(B,32,750)	(B,64,325)	(B,96,188)
2	SPPFormer Block	PW-Conv	×1	×1	×1	×1	×1	×1	(B,32,750)	(B,64,325)	(B,96,188)
		BottleNeck	×1	×1	×1	×1	×1	×1	(B,32,750)	(B,64,325)	(B,96,188)
		DW-Conv	×1	×1	×1	×1	×1	×1	(B,32,750)	(B,64,325)	(B,96,188)
		Stochastic	dim.32	dim.32	dim.32	dim.32	dim.64	dim.96	(B,32,750)	(B,64,325)	(B,96,188)
		MSA	×1	×2	×3	×4	×1	×1	(B,32,750)	(B,64,325)	(B,96,188)
		Fusion	×1	×1	×1	×1	×1	×1	(B,32,750)	(B,64,325)	(B,96,188)
3	Decoder	ConvT	×3	×3	×3	×3	×4	×5	(B,16,1500)	(B,32,750)	(B,64,325)
									(B,8,3000)	(B,16,1500)	(B,32,750)
									(B,3,6000)	(B,8,3000)	(B,16,1500)
										(B,3,6000)	(B,8,3000)
											(B,3,6000)

TABLE 2. Inference performance of different networks on different platforms.

Model	Model Size	PARAMS	Raspberry Pi(s)	CPU(s)	GPU(ms)
SPPFormer-XXS	140.1KB	35.0K	614	304	31461
SPPFormer-XS	202.4KB	51.0K	723	335	37524
SPPFormer-S	264.7KB	66.9K	838	505	47902
SPPFormer-M	327.0KB	82.9K	917	551	68679
SPPFormer-B	558.7KB	141.0K	474	248	26097
SPPFormer-L	1339.7KB	340.0K	424	219	22974
PhaseNet[54]	161.8KB	35.7K	197	57	16100
EQTransformer[55]	1544.2KB	376.2K	1012	216	33700
DPPPicker[56]	425.0KB	85.2K	941	194	21200
LEQNet[57]	862.2KB	210.3K	897	228	30700
SEANet[58]	213.0KB	73.9K	487	174	31400

For training SPPFormer, a batch size of 1024 and an initial learning rate of 0.01 were set. In addition, the research adopted the Adam optimizer with an exponential decay learning rate strategy and trained the model for 100 epochs on the STEAD dataset. On the premise of guaranteeing randomness and consistency in data processing, this research randomized the shuffling of the training set and normalized the input data. Data augmentation techniques applied during training included random offsets, addition of Gaussian noise, insertion of random blank gaps, and random zeroing of one or two channels.

E. MODEL COMPLEXITY

Table 2 compares SPPFormer and other seismic phase picking networks on model complexity. Inference efficiency was tested on a dataset of 12,656 samples using NVIDIA GeForce RTX 3090 GPU, Intel CPU, and Raspberry Pi. The results demonstrate that GPUs significantly expedite inference in SPPFormer, whereas CPU and Raspberry Pi require more time. Take the

SPPFormer-M as an example on the Raspberry Pi. It can process 14 samples, each 60 seconds long, in one second, adequately serving seismic exploration and relevant applications.

Increasing the number of MSA in SPPFormer, without changing the input feature dimension, brings about more model parameters and longer inference time. On the contrary, enlarging the input feature dimension while not changing the MSA number leads to a further increase in parameters but a decline in inference time. This is partly because the increase in feature dimensions dramatically boosts the parameters, but large matrix multiplications are executed faster on hardware compared to several smaller ones, thereby shortening inference time. Similarly, larger matrices reduce the number of memory access operations, decreasing inference time. Apart from model size and parameters, the number of floating-point operations (FLOPs) required by the model is also worth considering. Figure 2. compares several learning-based models from the three dimensions of model size, parameters, and FLOPs. When deploying models in practice, selecting a variant that aligns with the device's memory capacity, power consumption constraints, and inference time requirements is essential.

The identification and localization of the P-phase and S-phase in seismic data is illustrated in Table 3. To quantify the performance of the model, a set of statistical metrics is employed, including mean and standard deviation (std), measured in seconds, as well as the mean absolute error (MAE). These metrics shed light on the accuracy and reliability of the model. The mean reflects the average deviation between

TABLE 3. Performance of different networks on phase picking.

Model	P-phase						S-phase						MCC
	Precision	Recall	F1	MEAN	STD	MAE	Precision	Recall	F1	MEAN	STD	MAE	
SPPFormer-XXS	0.992	0.991	0.982	0.985	0.383	0.141	0.979	0.956	0.969	0.048	0.739	0.207	0.985
SPPFormer-XS	0.995	0.989	0.983	0.985	0.312	0.140	0.983	0.965	0.977	0.07	0.571	0.173	0.985
SPPFormer-S	0.997	0.993	0.985	0.990	0.329	0.114	0.972	0.965	0.978	0.084	0.311	0.178	0.990
SPPFormer-M	0.998	0.998	0.991	0.996	0.132	0.092	0.989	0.971	0.984	0.03	0.134	0.132	0.996
SPPFormer-B	0.998	0.997	0.987	0.995	0.155	0.082	0.989	0.968	0.982	0.006	0.206	0.121	0.995
SPPFormer-L	0.990	0.999	0.994	0.997	0.051	0.072	0.993	0.973	0.985	0.003	0.107	0.089	0.997
PhaseNet	0.966	0.967	0.968	0.984	0.117	0.097	0.962	0.949	0.948	0.025	0.137	0.158	0.984
EQTransformer	0.991	0.994	0.992	0.995	0.039	0.019	0.990	0.963	0.983	-0.009	0.118	0.022	0.995
DPPPicker	0.983	0.982	0.983	0.983	0.070	0.080	0.980	0.967	0.979	0.014	0.357	0.12	0.983
LEQNet	0.980	0.963	0.962	0.983	0.055	0.166	0.971	0.955	0.958	-0.043	0.274	0.146	0.983
SEANet	0.986	0.993	0.990	0.993	0.038	0.074	0.988	0.973	0.984	-0.06	0.128	0.017	0.993

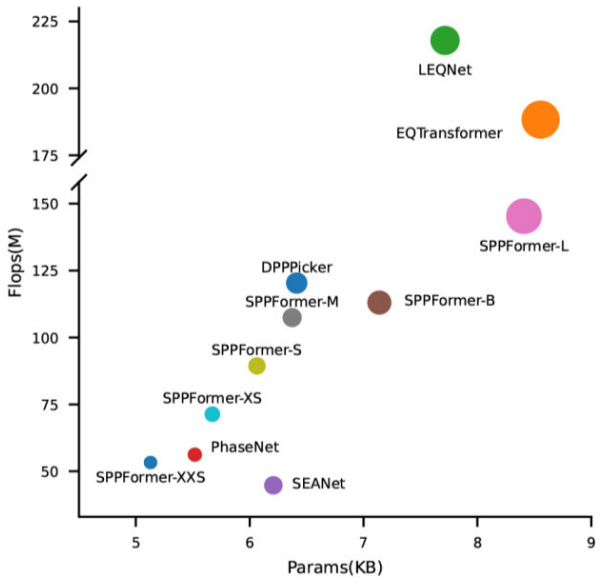

FIGURE 2. Parameters and FLOPs. The scale of the abscissa axes is logarithmic. The radius of the graph in the figure is positively correlated with the corresponding model size.

the model's picking time and the actual phase arrival time of seismic events, while the standard deviation can reveal the variability of these picking times. Additionally, the MAE represents the average of all individual prediction errors and is an intuitive measure of prediction accuracy.

The comparative networks utilized during tests are derived from official pre-trained models, which have been optimized and evaluated through retraining on the seisbench. After a detailed comparative analysis, the results demonstrate that the SPPFormer-L exhibits performance comparable to Earthquake Transformer (EQTransformer) [55]. Moreover, SPPFormer-L shows superior performance in most metrics compared to other competing networks.

Different configurations of SPPFormer, when compared with other seismic phase picking networks, consistently achieve similar or even better picking accuracy for both P and S phases. SPPFormer-L, in particular, has a higher accuracy than the top-performing EQTransformer in picking both phases and a remarkable inference time advantage. Enhancing the input feature dimensions in SPPFormer can yield better phase picking results than merely stacking MSA. Although this enhancement significantly enlarges the model's parameters, it can reduce inference time. Conversely, maintaining a small input dimension while stacking MSA brings about a minor parameter increase with improved inference performance, albeit at the cost of longer inference time.

F. PREDICTION ON SEISMIC DATA

Figure 3. illustrates the phase picking performance of SPPFormer-L on test sets using Raspberry Pi across various noise levels. The network performs well in phase picking with both high and low signal-to-noise ratios (SNR) in Figures 3a and 3b while exhibiting strong robustness to highly noisy data, as presented in Figure 3c. This can be attributed to its structural design, which mitigates its sensitivity to noise in seismic data compared to traditional phase picking networks. Additionally, in Figure 3d, when the E-component is missing, and only two valid components are available, SPPFormer-L also achieves excellent prediction results.

Figure 4 illustrates the picking results of SPPFormer-L and comparison models on the STEAD test set. SPPFormer-L and EQTransformer demonstrate similar picking performance, significantly outperforming other models. Phasenet shows good picking capability at high SNR but performs poorly on P-wave picking in low SNR conditions. LEQNet excels at P-wave picking but fails to pick S-waves in both scenarios. SEANet is capable of picking both P-waves and S-waves, but the P-wave picking time deviates from the actual time by approximately 0.5 seconds. DPPPicker performs poorly

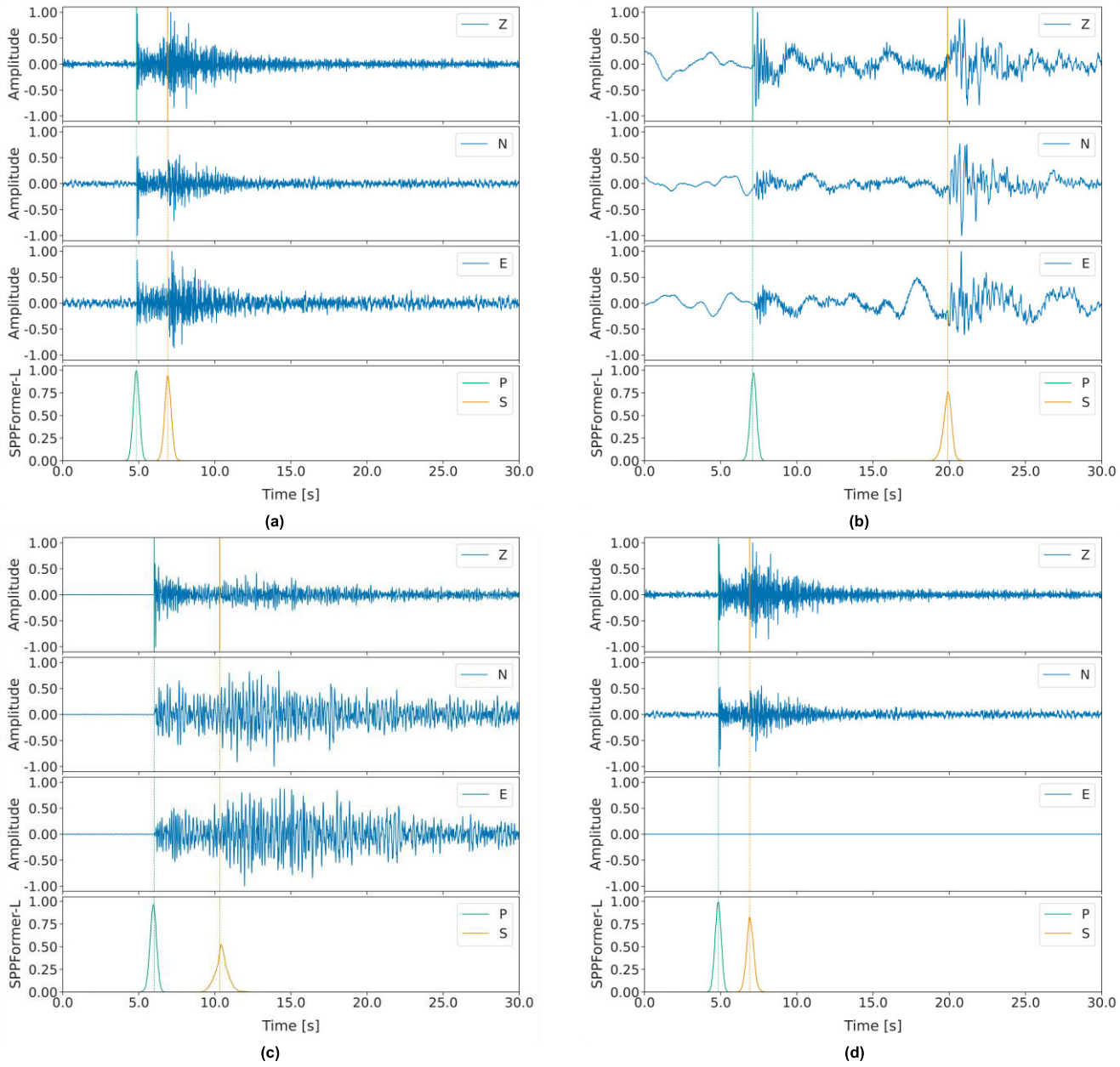

FIGURE 3. Panels a-d show four representative waveforms of SPPFormer applied to the test set. Each waveform has a length of 60 seconds with 100 samples per second. Each panel displays only a 30-second-long waveform containing P and S phases. The top of each panel shows three channel waveforms, and the bottom shows the output prediction of phase picking. The vertical color-coded lines in panels are the manually selected arrival times from the catalog.

in both scenarios, with ineffective picking if the threshold is set too high.

Phase picking is highly dependent on the geological structure, meaning signals observed in different geographical backgrounds may exhibit distinct characteristics. To enhance the network's generality and validate its adaptability and reliability under varying geological conditions, we conducted thorough testing and analysis using continuous waveform data provided by the EarthScope Consortium (Figures 5a and b) and National Institute of Geophysics and Volcanology (INGV) (Figures 5c and d). EarthScope Consortium, which operates the U.S. National Science Foundation's geodetic and seismic facilities, is dedicated to supporting transformative global geophysical research and education. INGV is the Italian National Institute, provides extensive seismic data resources crucial for seismological research and disaster risk assessment.

Data preprocessing is a crucial step to ensure the accuracy of the network. Before testing, necessary preprocessing steps were performed on the raw data, including denoising, filtering, and resampling, to ensure data quality meets the analytical requirements. Additionally, to enhance the training

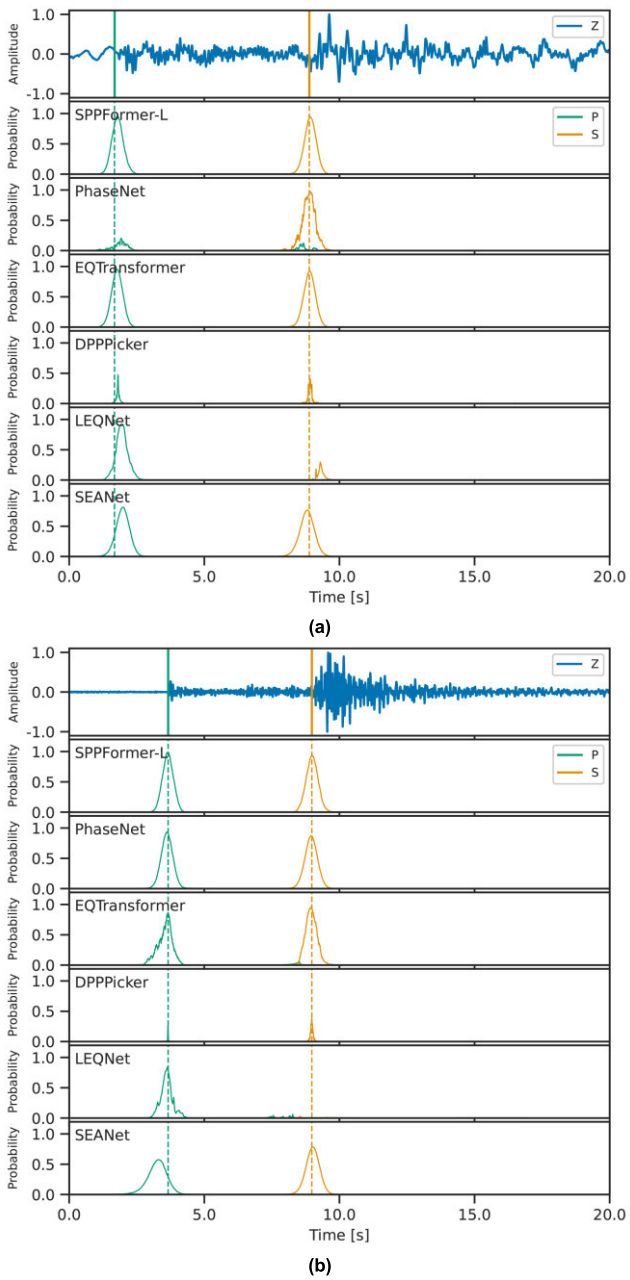

FIGURE 4. Picking results of SPPFormer-L and comparison models on the STEAD test set. Panels a and b have SNR of 13.2 and 30.5, respectively. Each waveform is 60 seconds long with 100 samples per second. Each panel displays only a 20-second segment containing P and S phases. Only the vertical component is shown above, although all three components are used as model inputs.

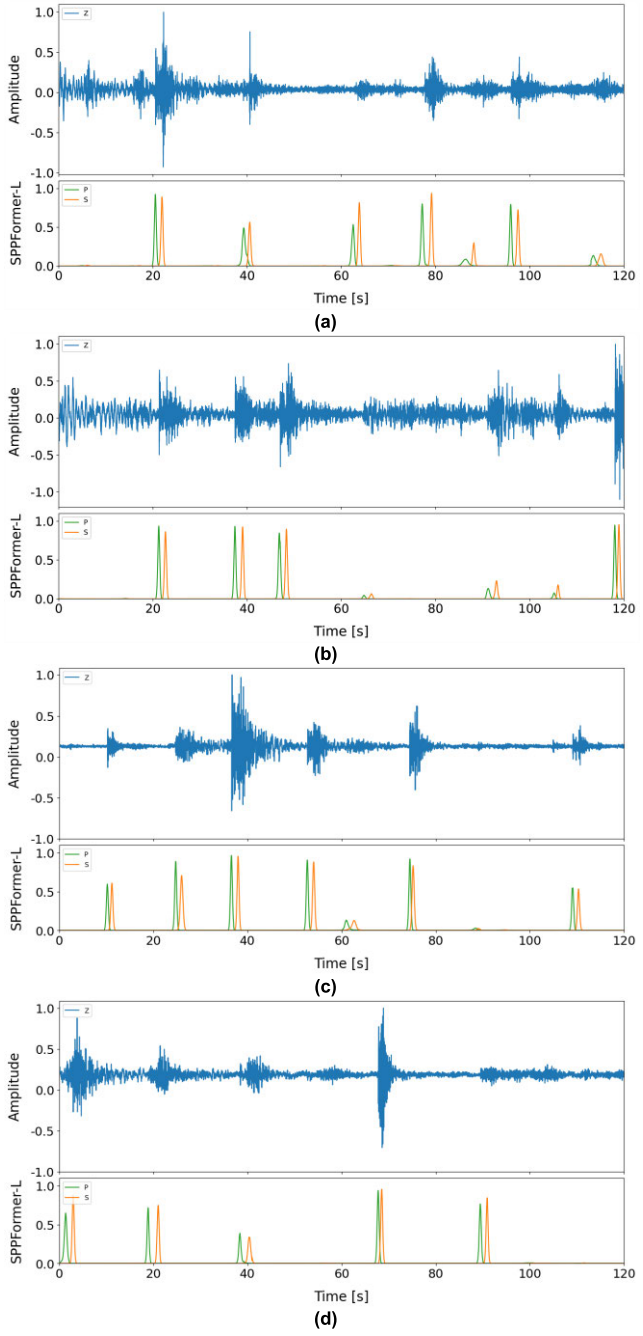

FIGURE 5. Test results on continuous waveform data. The model's performance on continuous waveform data is demonstrated by applying the model to a representative segment of continuous waveforms from the EarthScope Consortium (Panels a, b) and INGV (Panels c, d). Panel a and c have multiple events within a two-minute window. Panel b and d have events occurring near the edges. Only the vertical component is depicted above, while all three components are taken as model inputs.

and prediction efficiency of the network, the preprocessed continuous data was segmented into 1-minute windows. In practical implementation, appropriate overlap between these windows is introduced to increase data continuity and reduce boundary effects. These 1-minute data windows can be individually or batch-wise input into the SPPFormer. Before feeding the data into the network for prediction, normalization was applied to eliminate the influence of varying magnitudes and dimensions of the data, thus improving the accuracy of event picking.

The training data for SPPFormer consists solely of samples from individual phase picking events. However, the model has demonstrated strong adaptability and robustness, as shown in Figures 5a and c. Even in scenarios where multiple seismic events are present within a single window and distributed across different time points, this model can still accurately identify and pick these events. In addition,

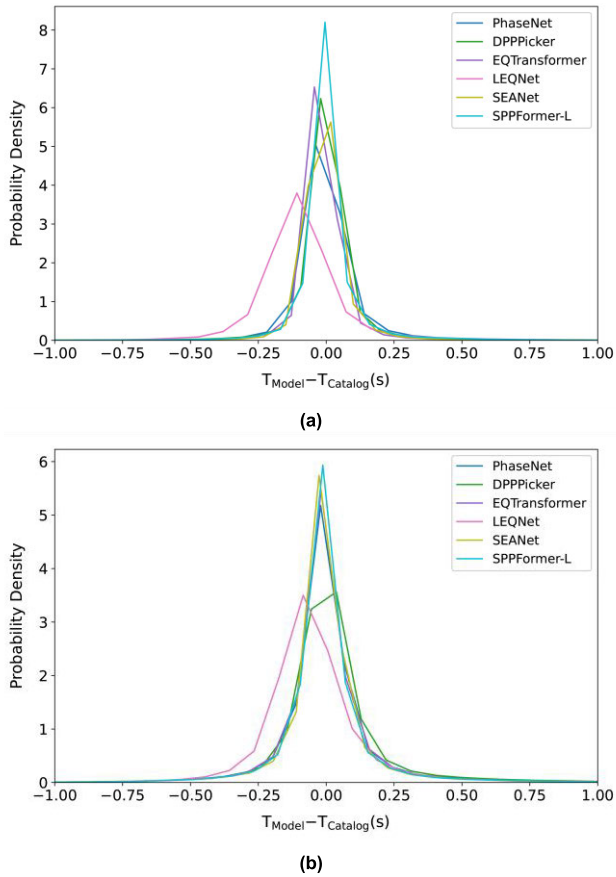

FIGURE 6. The KDE plots of the time residuals of the predicted and true phase arrivals. Panel a is the distribution of P-picking residuals. Panel b is the distribution of S-picking residuals.

as shown in Figures 5b and d, the proposed network in this study can accurately pick complex events at the edges of the signals. This indicates that the SPPFormer presented in this research can achieve precise spatiotemporal localization of events within seismic signals, even when these events are close to the data window's starting or ending boundaries.

1) PERFORMANCE OF SPPFORME

The model's picking error is a crucial performance metric, which refers to the deviation between the predicted arrival time of seismic waves by the model and the manually picked. This paper conducts kernel density estimate (KDE) analysis to examine the distribution of picking errors. In the test results presented in Figure 6, density curves of picking errors for each model are visualized using KDE plots. The curve's steepness unveils the concentration or dispersion of picking errors, while the variation in height displays the distribution of the model's phase picking errors in the validation data. The data points are primarily concentrated in a specific region if the density curve exhibits a peak.

As shown in Figure 6, except for LEQNet, the time residuals of the remaining models for the P and S phases exhibit a Gaussian distribution centered around zero, with most residuals below 0.1s. Observing the density curves of picking errors for each model shows that PhaseNet, DPPPicker, EQTransformer, and SEANet have similar density curves, revealing that they have roughly the same distribution. Despite this, LEQNet's picking errors deviate significantly from other models, suggesting that it compromises the model's accuracy to a greater extent when lightweight EQTransformer. Additionally, among all the models, SPPFormer-L has the steepest KDE curve, and the most negligible picking errors correspond to the peak, indicating that its picking errors are more concentrated. Comparing Figure 6a and b, it can be inferred that SPPFormer-L performs better in picking P phases than picking S phases, suggesting that its predictions are closer to manually selected arrival times, leading to better fitting of the labeled data.

V. CONCLUSION

This research proposed a seismic phase picking network called SPPFormer, specifically designed to operate in resource-constrained embedded systems. This network optimizes the trade-off between inference accuracy and computational speed, ensuring practicality and efficiency on low-power devices. The architecture of SPPFormer effectively merges CNN's capability in feature extraction and the global perception advantage of Transformer in capturing long-range dependencies. Additionally, this study modified the standard attention mechanism of the Transformer to enhance its representation capability in complicated pattern recognition of seismic signals. Moreover, to validate the effectiveness of the SPPFormer, a series of seismic phase recognition experiments were carried out on the Raspberry Pi. The experimental results demonstrate that the SPPFormer network maintains a high level of picking accuracy and exhibits fast inference speed when executed on this hardware. The current research lays a theoretical foundation and provides experimental evidence for developing efficient seismic monitoring systems in embedded or edge computing environments.

While the SPPFormer achieves a commendable balance between inference accuracy and computational speed, it is not without its limitations, particularly regarding the computational complexity inherent in Transformer models. The quadratic computational requirement of the Transformer's self-attention mechanism can pose challenges, especially on resource-constrained embedded systems where processing power and energy efficiency are critical. Despite the modifications made to enhance the representation capability for seismic signal patterns, the model's scalability may be limited by the increased computational demand as sequence lengths grow. Future work could optimize the self-attention mechanism to mitigate this complexity, ensuring that the SPPFormer remains practical for real-time seismic monitoring on low-power devices.

REFERENCES

[1] H. Qiu, L. Su, B. Tang, D. Yang, M. Ullah, Y. Zhu, and U. Kamp, "The effect of location and geometric properties of landslides caused by rainstorms and earthquakes," *Earth Surf. Processes Landforms*, vol. 49, no. 7, pp. 2067–2079, Mar. 2024.

[2] H. Huang, M. Li, W. Zhang, and Y. Yuan, "Seismic behavior of a friction-type artificial plastic Hinge for the precast beam–column connection," *Arch. Civil Mech. Eng.*, vol. 22, no. 4, pp. 201:1–201:20, Sep. 2022.

[3] W. Du and G. Wang, "Fully probabilistic seismic displacement analysis of spatially distributed slopes using spatially correlated vector intensity measures," *Earthq. Eng. Struct. Dyn.*, vol. 43, no. 5, pp. 661–679, Apr. 2014.

[4] C. D. Saragiotis, L. J. Hadjileontiadis, and S. M. Panas, "PAI-S/K: A robust automatic seismic p phase arrival identification scheme," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 6, pp. 1395–1404, Jun. 2002.

[5] A. Lomax, C. Satriano, and M. Vassallo, "Automatic picker developments and optimization: FilterPicker—A robust, broadband picker for real-time seismic monitoring and earthquake early warning," *Seismolog. Res. Lett.*, vol. 83, no. 3, pp. 531–540, May 2012.

[6] N. Maeda, "A method for reading and checking phase time in auto-processing system of seismic wave data," *J. Seismolog. Soc. Jpn.*, vol. 38, no. 3, pp. 365–379, 1985.

[7] Y. Yang, Z. Wang, N. Liu, J. Wang, S. Pang, R. Liu, and J. Gao, "Physically driven self-supervised learning and its applications in geophysical inversion," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 4503211.

[8] N. Liu, Y. Lei, R. Liu, Y. Yang, T. Wei, and J. Gao, "Sparse time–frequency analysis of seismic data: Sparse representation to unrolled optimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5915010.

[9] V. Mazzia, A. Khaliq, F. Salvetti, and M. Chiaberge, "Real-time apple detection system using embedded systems with hardware accelerators: An edge AI application," *IEEE Access*, vol. 8, pp. 9102–9114, 2020.

[10] Y. Zhang, J. Yu, Y. Chen, W. Yang, W. Zhang, and Y. He, "Real-time strawberry detection using deep neural networks on embedded system (rtsd-net): An edge AI application," *Comput. Electron. Agricult.*, vol. 192, Jan. 2022, Art. no. 106586.

[11] J.-I. Watanabe, Y. Shao, and N. Miura, "Underwater and airborne monitoring of marine ecosystems and debris," *J. Appl. Remote Sens.*, vol. 13, no. 4, Oct. 2019, Art. no. 044509.

[12] N. Al-Thani, A. Albuainain, F. Alnaimi, and N. Zorba, "Drones for sheep livestock monitoring," in *Proc. IEEE 20th Medit. Electrotechnical Conf. (MELECON)*, Jun. 2020, pp. 672–676.

[13] C. Y. Kuan, Y.-C. Tsai, J.-T. Hsu, S. T. Ding, and T. T. Lin, "An imaging system based on deep learning for monitoring the feeding behavior of dairy cows," in *Proc. ASABE Annu. Int. Meeting*, Boston, MA, USA, Jul. 2019, p. 1.

[14] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[15] J. Zhang, X. Li, J. Li, L. Liu, Z. Xue, B. Zhang, Z. Jiang, T. Huang, Y. Wang, and C. Wang, "Rethinking mobile block for efficient attention-based models," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 1389–1400.

[16] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient CNN architecture design," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 116–131.

[17] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *North American Chapter of the Association for Computational Linguistics*. Minneapolis, MN, USA: Association for Computational Linguistics, 2019.

[18] H. Wu, B. Xiao, N. Codella, M. Liu, X. Dai, L. Yuan, and L. Zhang, "CvT: Introducing convolutions to vision transformers," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 22–31.

[19] A. Vaswani, "Attention is all you need," in *Proc. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010.

[20] M. Maaz, "EdgeNeXt: Efficiently amalgamated CNN-transformer architecture for mobile vision applications," in *Proc. Comput. Vis. Workshops*, 2023, pp. 3–20.

[21] S. Mehta and M. Rastegari, "MobileViT: Lightweight, general-purpose, and mobile-friendly vision transformer," 2021, *arXiv:2110.02178*.

[22] P. K. Anasosalu Vasu, J. Gabriel, J. Zhu, O. Tuzel, and A. Ranjan, "FastViT: A fast hybrid vision transformer using structural reparameterization," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 5762–5772.

[23] Y. Li, "EfficientFormer: Vision transformers at MobileNet speed," in *Proc. Adv. Neural Inf. Process.*, vol. 35, 2022, pp. 12934–12949.

[24] A. G. Howard, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[25] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[26] M. Tan, B. Chen, R. Pang, V. Vasudevan, M. Sandler, A. Howard, and Q. V. Le, "MnasNet: Platform-aware neural architecture search for mobile," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2815–2823.

[27] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6848–6856.

[28] S. Mehta, M. Rastegari, L. Shapiro, and H. Hajishirzi, "ESPNetv2: A light-weight, power efficient, and general purpose convolutional neural network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9182–9192.

[29] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More features from cheap operations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1577–1586.

[30] D. Zhou, Q. Hou, Y. Chen, J. Feng, and S. Yan, "Rethinking bottleneck structure for efficient mobile network design," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 680–697.

[31] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[32] G. Chen, Y. Wang, H. Li, and W. Dong, "TinyNet: A lightweight, modular, and unified network architecture for the Internet of Things," in *Proc. ACM SIGCOMM Conf. Posters Demos*, Aug. 2019, pp. 248–260.

[33] A. Dosovitskiy, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[34] L. Beyer, P. Izmailov, A. Kolesnikov, M. Caron, S. Kornblith, X. Zhai, M. Minderer, M. Tschannen, I. Alabdulmohsin, and F. Pavetic, "FlexiViT: One model for all patch sizes," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 14496–14506.

[35] C. R. Chen, Q. Fan, and R. Panda, "CrossViT: Cross-attention multi-scale vision transformer for image classification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 347–356.

[36] J. Jiao, Y.-M. Tang, K.-Y. Lin, Y. Gao, J. Ma, Y. Wang, and W.-S. Zheng, "DilateFormer: Multi-scale dilated transformer for visual recognition," *IEEE Trans. Multimedia*, vol. 25, pp. 8906–8919, 2023.

[37] H. Touvron, M. Cord, A. Sablayrolles, G. Synnaeve, and H. Jégou, "Going deeper with image transformers," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 32–42.

[38] L. Zhu, X. Wang, Z. Ke, W. Zhang, and R. Lau, "BiFormer: Vision transformer with bi-level routing attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 10323–10333.

[39] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, Z. Yang, Y. Zhang, and D. Tao, "A survey on vision transformer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 87–110, Jan. 2023.

[40] W. Zhang, Z. Huang, G. Luo, T. Chen, X. Wang, W. Liu, G. Yu, and C. Shen, "TopFormer: Token pyramid transformer for mobile semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 12073–12083.

[41] Z. Chen, F. Zhong, Q. Luo, X. Zhang, and Y. Zheng, "EdgeViT: Efficient visual modeling for edge computing," in *Proc. Int. Conf. Wireless Algorithms, Syst., Appl.* Cham, Switzerland: Springer, 2022, pp. 393–405.

[42] A. Wang, H. Chen, Z. Lin, H. Pu, and G. Ding, "RepViT: Revisiting mobile CNN from ViT perspective," 2023, *arXiv:2307.09283*.

[43] Y. Chen, X. Dai, D. Chen, M. Liu, X. Dong, L. Yuan, and Z. Liu, "Mobile-former: Bridging MobileNet and transformer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5260–5269.

[44] M. Dehghani, A. Arnab, L. Beyer, A. Vaswani, and Y. Tay, "The efficiency misnomer," 2021, *arXiv:2110.12894*.

[45] P. K. A. Vasu, J. Gabriel, J. Zhu, O. Tuzel, and A. Ranjan, "MobileOne: An improved one millisecond mobile backbone," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7907–7917.

[46] A. Howard et al., "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Seoul, South Korea, 2019, pp. 1314–1324.

[47] B. Graham, A. El-Nouby, H. Touvron, P. Stock, A. Joulin, H. Jégou, and M. Douze, "LeViT: A vision transformer in ConvNet's clothing for faster inference," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 12239–12249.
[48] T. Xiao, M. Singh, E. Mintun, T. Darrell, P. Dollár, and R. Girshick, "Early convolutions help transformers see better," in *Proc. NIPS*, Dec. 2021, pp. 30392–30400.
[49] A. Hassani, S. Walton, J. Li, S. Li, and H. Shi, "Neighborhood attention transformer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 6185–6194.
[50] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.
[51] X. Pan, T. Ye, Z. Xia, S. Song, and G. Huang, "Slide-transformer: Hierarchical vision transformer with local self-attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 2082–2091.
[52] Z. Tu, H. Talebi, H. Zhang, F. Yang, P. Milanfar, A. Bovik, and Y. Li, "MaxViT: Multi-axis vision transformer," in *Proc. 17th Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2022, pp. 459–479.
[53] S. M. Mousavi, Y. Sheng, W. Zhu, and G. C. Beroza, "STanford EArthquake dataset (STEAD): A global data set of seismic signals for AI," *IEEE Access*, vol. 7, pp. 179464–179476, 2019.
[54] W. Zhu and G. C. Beroza, "PhaseNet: A deep-neural-network-based seismic arrival time picking method," *Geophys. J. Int.*, vol. 216, no. 1, pp. 261–273, Jan. 2019.
[55] S. M. Mousavi, W. L. Ellsworth, W. Zhu, L. Y. Chuang, and G. C. Beroza, "Earthquake transformer—An attentive deep-learning model for simultaneous earthquake detection and phase picking," *Nature Commun.*, vol. 11, no. 1, pp. 1–12, Aug. 2020.
[56] H. Soto and B. Schurr, "DeepPhasePick: A method for detecting and picking seismic phases from local earthquakes based on highly optimized convolutional and recurrent deep neural networks," *Geophys. J. Int.*, vol. 227, no. 2, pp. 1268–1294, Aug. 2021.
[57] J. Lim, S. Jung, C. JeGal, G. Jung, J. H. Yoo, J. K. Gahm, and G. Song, "LEQNet: Light earthquake deep neural network for earthquake detection and phase picking," *Frontiers Earth Sci.*, vol. 10, pp. 1–7, Jul. 2022.
[58] X. Hou, Y. Zheng, M. Jiang, and S. Zhang, "SEA-Net: Sequence attention network for seismic event detection and phase arrival picking," *Eng. Appl. Artif. Intell.*, vol. 122, Jun. 2023, Art. no. 106090.

YADONGYANG ZHU received the B.S. degree in measurement-control technology and instrumentation and the Ph.D. degree in test measurement technology and instrument from Jilin University, Changchun, China, in 2010 and 2018, respectively. He is currently a Lecturer with the School of Information Engineering, Beijing Institute of Petrochemical Technology. His research interests include geophysical instrument, artificial intelligence, the Internet of Things, and edge computing.

SHUGUANG ZHAO received the B.S. degree from Shanxi Agriculture University, Jinzhong, China, in 2022. He is currently pursuing the M.S. degree with the School of Information Engineering, Beijing Institute of Petrochemical Technology. His research interests include machine learning and time series analysis.

WEI WEI received the B.S. degree in information and computing science from China University of Mining and Technology, Beijing, in 2010, and the M.S. degree in software engineering and the Ph.D. degree in control science and engineering from Beijing Technology and Business University, in 2013 and 2019, respectively. After that, she is currently a Lecturer with the College of Information Engineering, Beijing Institute of Petrochemical Technology. Her research interests include emotion analysis and human–machine collaboration.

FUDONG ZHANG received the B.S. degree in measurement-control technology and instrumentation and the Ph.D. degree in the test measurement technology and instrument from Jilin University, Changchun, China, in 2011 and 2019, respectively. He is currently a Lecturer with the College of Instrumentation and Electrical Engineering, Jilin University. His research interests include distributed optical fiber sensing technology and instrument development, near infrared spectroscopy technology, and instrument development.

FA ZHAO received the M.S. degree from Jilin University, Changchun, China, in 2018, where he is currently pursuing the Ph.D. degree with the College of instrumentation and Electrical Engineering. His research interests include fracturing fracture imaging technology, reservoir geomechanics, microseismic monitoring technology, and equipment.