JDAT: Joint-Dimension-Aware Transformer with Strong Flexibility for EEG Emotion Recognition

Zeyu Wang, *Student Member, IEEE*, Ziqun Zhou, *Student Member, IEEE*,
Haibin Shen, Qi Xu, and Kejie Huang, *Senior Member, IEEE*

Abstract—Electroencephalography (EEG) emotion recognition, an important task in Human-Computer Interaction (HCI), has made a great breakthrough with the help of deep learning algorithms. Although the application of attention mechanism on conventional models has improved its performance, most previous research rarely focused on multiplex EEG features jointly, lacking a compact model with unified attention modules. This study proposes Joint-Dimension-Aware Transformer (JDAT), a robust model based on squeezed Multi-head Self-Attention (MSA) mechanism for EEG emotion recognition. The adaptive squeezed MSA applied on multidimensional features enables JDAT to focus on diverse EEG information, including space, frequency, and time. Under the joint attention, JDAT is sensitive to the complicated brain activities, such as signal activation, phase-intensity couplings, and resonance. Moreover, its gradually compressed structure contains no recurrent or parallel modules, greatly reducing the memory and complexity, and accelerating the inference process. The proposed JDAT is evaluated on DEAP, DREAMER, and SEED datasets, and experimental results show that it outperforms state-of-the-art methods along with stronger flexibility.

Index Terms—Electroencephalography (EEG), emotion recognition, Joint-Dimension-Aware Transformer (JDAT), squeezed Multi-head Self-Attention (MSA), state-of-the-art.

I. INTRODUCTION

AS one of the most significant human senses, emotion largely affects people's quality of life. At present, emotion recognition is mainly based on analysis of physical signals and physiological signals. Physical signals, such as facial expression [1] and speech [2], often generate inaccurate results due to the obscure boundaries between people's external actions. In contrast, physiological signals, including Electroencephalography (EEG) [3], Electromyography (EMG) [4], and Electrocardiography (ECG) [5], reflect human emotion more objectively. EEG is a non-invasive method of Human-Computer Interaction (HCI) [6], which has become efficient and indispensable in human emotion recognition, in light of its valuable multidimensional information. In the spatial dimension, earlier works stated that emotional changes

mostly affect the EEG signals on the frontal and temporal lobes [7], [8], and the experiments demonstrated that channel selection can improve the recognition efficiency [9]. In the spectral dimension, positive emotion generates high-frequency EEG signals, while negative emotion is exactly the reverse. Therefore, Differential Entropy (DE) [10] and Power Spectral Density (PSD) [11] features are commonly used to better reflect the spectral characteristics. In the temporal dimension, the signal activation strength of positive emotion is generally higher than that of negative emotion. Furthermore, from the perspective of neuroscience [12], brain activities will generate complicated cross-frequency coupling and phase-amplitude coupling across different brain lobes [13]. The above phenomena have proved the advantages of utilizing joint dimensional features of EEG to recognize human emotions.

Generally, existing machine learning methods in EEG emotion recognition have two main streams: traditional methods and deep learning methods. In traditional methods, K-Nearest Neighbor (KNN) [14], Bayesian Network (BN) [15], and Support Vector Machine (SVM) [16], [17] appear to be the most popular algorithms. However, these algorithms can not detect or extract complex EEG features, and they used to obtain general classification results. In recent years, deep learning methods have demonstrated its efficiency and flexibility for its excellent performance, especially Convolutional Neural Network (CNN) [18], Long Short-Term Memory (LSTM) [19], and Graph Neural Network (GNN) [20]. After intensive exploration, attention mechanism has been successfully adapted to EEG emotion recognition. Attention-based Convolutional Recurrent Neural Network (ACRNN) [21], 4D attention-based Neural Network (4D-aNN) [22], and Spatial-Spectral-Temporal based attention 3D dense Network (SST-EmotionNet) [23] are the most successful designs. However, they suffer from not only complex network structures with tremendous parameters, but also separate attention modules which can not utilize joint dimensional information of EEG. In addition, they used to achieve good results on specific datasets but are not adaptable to other datasets, due to the different experimental settings such as the channel selection and sampling frequency.

In this paper, we propose the Joint-Dimension-Aware Transformer (JDAT) with strong flexibility for EEG emotion recognition. It adopts the adaptive Multi-head Self-Attention (MSA) mechanism to focus on different dimensional EEG features mutually and jointly, without any hybrid structures or other attention mechanism. Its structure mainly consists of the Spatial-Spectral-Projection (SSP) Block and Multidimensional Trans-

This work was supported by the Major Scientific Research Project of Zhejiang Lab (No. 2019KC0AD02), the National Natural Science Foundation of China (U19B2043), and the National Key Research and Development Program of China (2020AAA0109002).

Z. Wang, Z. Zhou, H. Shen, and K. Huang are with the College of Information Science & Electronic Engineering, Zhejiang University, Hangzhou 310027, China (email: wangzeyu2020@zju.edu.cn, zhouziqun@zju.edu.cn, shen_hb@zju.edu.cn, huangkejie@zju.edu.cn).

Qi Xu is with the School of Artificial Intelligence, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, 16024, China (email: xuqi@dlut.edu.cn).

Fig. 1. Illustration of the models utilizing attention mechanism on different dimensions. Attn denotes Attention, and MSA denotes Multi-head Self-Attention. The models include (a): Attention-based models for EEG emotion recognition. (b): Transformer models for EEG classification, in tasks of seizure detection, motor imagination, and emotion recognition. (c): Our proposed JDAT which focuses on spatial, spectral, and temporal features jointly for emotion recognition.

former Encoder (MTE). The SSP Block serves to facilitate the training of Transformer and automatically assign the weights of different channels and frequency bands. The MTE simultaneously processes intersecting EEG information of space, frequency, time, and intensity, where the global features are well noticed. In MTE layers, the Spatial-Spectral MSA attends to mixed features of space and frequency, which is capable of discovering the complicated phase couplings and resonance across different brain lobes; the Temporal MSA handles the sequential space-frequency-fused information, which makes the model sensitive to the signal activation and waveform changes globally. Meanwhile, we revise the conventional MSA into a squeezed structure, applying the attention between a specific feature point and merged points to reduce the risk of overfitting. Performance of JDAT is evaluated on three datasets, including DEAP [24], DREAMER [25], and SEED [26], [27]. Compared with conventional models, our model achieves the State-Of-The-Art (SOTA) results in all datasets.

The rest of the paper is organized as follows: Section II introduces related work, including some attention-based models for EEG emotion recognition and Transformer models for EEG classification. Section III describes the proposed methods in detail, including pre-processing methods and our proposed JDAT model. Section IV provides our experiments along with the results. Finally, the conclusion and discussion are made in Section V.

II. RELATED WORK

A. Attention-Based EEG Emotion Recognition

The early neural network models in EEG emotion recognition rarely adopted attention mechanism. For example, Hwang *et al.* [18] designed a simple CNN similar to LeNet [28], for automatic learning from DE features of EEG signals. Yang *et al.* [29] integrated LSTM layers into their parallel convolutional recurrent neural network for emotion recognition and achieved satisfactory results. Recently, attention mechanism has shown its strength in EEG emotion recognition. For instance, Tao *et al.* [21] proposed ACRNN, a fused model based on channel-wise attention and self-attention, to assign the weights of different channels and seek intrinsic

similarity in EEG signals. Xiao *et al.* [22] proposed 4D-aNN, which adopted spectral and spatial attention mechanism to adaptively focus on the corresponding features. However, as illustrated in Fig. 1(a), these independent attention modules in ACRNN and 4D-aNN only pay attention to the specific dimensional features, while not treating them in a correlation way. Prominently, SST-EmotionNet proposed by Jia *et al.* [23] is able to concentrate on spatial-spectral and spatial-temporal features. However, Fig. 1(a) shows that it is still unable to process the global features in a unified attention module. In addition, SST-EmotionNet's attention streams are constructed in parallel branches, which make the model complex and inflexible. Therefore, unified attention modules processing multidimensional features should be developed to avoid the above-mentioned issues.

B. Transformer for EEG classification

On the basis of MSA mechanism, Transformer has shown its strength in both Natural Language Processing (NLP) and Computer Vision (CV). Since the birth of BERT [30], Transformer has become the gold standard methods in most NLP tasks. Afterwards, ViT [31] has demonstrated the ability of Transformer in computer vision, in tasks of classification [32], segmentation [33], and object detection [34]. Nowadays, some researchers have made an effort to integrate Transformer into their models for EEG classification tasks, since Transformer can focus on sequential EEG data globally. Pedoeem *et al.* [35] proposed the Transformer Based Seizure detector (TABS), a hybrid model comprising convolutional layers, fully connected layers, and Transformer, which achieved a good result in seizure detection. Sun *et al.* [36] constructed several Transformer-based models for motor imaginary EEG classification (MI-Transformer), whose results showed that Transformer combined with CNN might serve as a powerful model. However, as illustrated in Fig. 1(b), these works are not aware of the value of joint dimensional EEG information. TABS only takes Transformer as the tool for processing time series. MI-Transformer are constructed with attention only in spatial or temporal dimension, instead of the fused dimensions. Recently, Liu *et al.* [37] built the EEG emotion Transformer (EeT),

Fig. 2. Illustration of pre-processing workflow of baseline removal and Continuous Wavelet Transform (CWT). The schematic diagram of recording EEG signals from a person's brain is taken from [39].

applying joint attention on spatial and temporal dimensions. However, as illustrated in Fig. 1(b), they ignored the valuable spectral information of EEG signals for emotion recognition. Some other works fused their hybrid structures with MSA layers, the core part of Transformer, in order to strengthen their networks with the help of self-attention [21], [38]. In this work, we take Transformer as the main frame of our proposed model, utilizing the multidimensional EEG features jointly and comprehensively for emotion recognition.

III. PROPOSED METHODS

A. Overview

Our proposed methods consist of two main parts: pre-processing and proposed model. The pre-processing aims to facilitates the training and operation of our model. Our proposed JDAT takes Transformer as the main frame, built with the unified squeezed MSA instead of fused attention mechanism. As shown in Fig. 1(c), it focuses on multiple dimensional EEG features and processes them jointly.

B. Pre-Processing

EEG signals should be pre-processed before being sent into the model. The pre-processing steps include baseline removal, Continuous Wavelet Transform (CWT), and data augmentation.

1) *Baseline Removal*: Recorded EEG signals usually contain pre-trial (baseline) and trial signals. Yang *et al.* [29] proposed that baseline removal can improve the accuracy of EEG emotion recognition. Tao *et al.* [21] also used this pre-processing method and achieved better results. As illustrated in the first step of Fig. 2, baseline removal can be described in detail as follows:

Firstly, pre-trial signals X_p and trial signals X_t are sliced into N and M segments with the same length, respectively.

$$X_p \rightarrow \{X_p^1, X_p^2, \dots, X_p^N\}, X_p^i \in \mathbb{R}^{C \times L} \quad (1)$$

$$X_t \rightarrow \{X_t^1, X_t^2, \dots, X_t^M\}, X_t^j \in \mathbb{R}^{C \times L} \quad (2)$$

where C and L denote the number of channels and the length of the segments, respectively.

Secondly, the mean value $\bar{X}_p$ of the N pre-trial signal segments is calculated, which denotes the mean baseline.

$$\bar{X}_p = \frac{\sum_{i=1}^N X_p^i}{N} \quad (3)$$

Finally, $\bar{X}_p$ is subtracted from each trial segment.

$$X_t^{j'} = X_t^j - \bar{X}_p \quad (4)$$

In this paper, L is set to the sampling points of 1 second, while C varies with the EEG channels provided by datasets.

2) *Continuous Wavelet Transform*: CWT is able to capture signal frequency along the time axis, which can fully decompose a continuous time function and generate the “Time-Frequency window” [40]. Different from Fourier Transform [41], it constructs a Time-Frequency relationship with subtle resolution in both domains. CWT of an integrable function $f(t)$ can be expressed as:

$$F_w(a, b) = \frac{1}{\sqrt{|(b)|}} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-a}{b}\right) dt \quad (5)$$

where $\psi(t)$ is the wavelet function, and a, b are translational value and scale factor, respectively.

In this work, CWT is conducted separately on each channel's signal, where the frequency of wavelet ranges from 1 to F Hz. Therefore, each segment is converted into 3D data $I^j \in \mathbb{R}^{C \times F \times L}$, as illustrated in the second step of Fig. 2. These data embody richer and more obvious spatial, spectral, and temporal information with stronger connection, which are taken as the input of our proposed model. In this work, F is set to 50. Different wavelets are compared during our experiments and Complex Gaussian Derivative Wavelet with the best performance is chosen.

$$I^j = \text{CWT}(X_t^{j'}) \quad (6)$$

Previous works showed that slicing EEG signals into samples between 1 second and 10 seconds helped achieve good results in emotion recognition [9], [23]. In this work, the 1-second segments $I^j (j = 1, 2, \dots, M)$ after baseline removal and CWT are directly taken as the dataset samples. Thus,

Fig. 3. Overall structure diagram of JDAT. The tuples above the model represent the data dimensions at the current stages. The result is illustrated for three emotional states, while the neutral class is eliminated when there are only two emotional states.

Fig. 4. Illustration of the Spatial-Spectral convolutional layer in SSP Block. There are D kernels in total shown in different colors.

Fig. 5. Visualization of kernels in spatial-spectral convolutional layer. The four kernels are taken as the examples from our trained model.

our dataset is made up of 1-second non-overlapping samples, which contains more sliced data and avoids repeated training and testing.

3) *Data Augmentation*: Considering the limited amount of public EEG data and the difficulty of training Transformer, we apply data augmentation on the training set. Translation, grayscale, or generative models are not adopted because they will destroy distinct EEG features and affect the objective experimental results. Therefore, Gaussian noise with zero mean and a standard deviation of 0.01 is added to enrich the training set to avoid overfitting. The data augmentation technique is similar to the method in [9].

C. Proposed Model

As shown in Fig. 3, our proposed JDAT mainly consists of SSP Block and MTE. The SSP Block compresses and projects the input data to facilitate the operation of MTE. The MTE serves as its main frame, which is composed of several stages with the unified attention modules, which are based on squeezed MSA. Every stage except the last consists of one MTE layer and a process of temporal compression. These MTE layers have different scales to adapt the model to gradually compressed feature maps. The MTE is followed by fully connected layers to classify different emotional states. The specific structures of modules in JDAT are described in detail as follows.

1) *SSP Block*: In SSP Block, the input $I \in \mathbb{R}^{C \times F \times L}$ is projected and compressed into $I' \in \mathbb{R}^{L \times D}$, with L sequential tokens of dimension D .

In detail, a Spatial-Spectral convolutional layer with D kernels of the size $(F, 1)$ is applied on I , which is illustrated in Fig. 4. These kernels extract global information of space and frequency along the time axis, projecting the spatial and spectral dimensions into a mixed dimension. Thereafter, an activation of Rectified Linear Unit (ReLU) and a Fully Connected (FC) layer $\in \mathbb{R}^{D \times D}$ are followed to increase the non-linearity.

$$I' = \text{FC}(\phi(\text{SSconv}(I))) \quad (7)$$

where SSconv denotes the Spatial-Spectral convolutional layer, and $\phi(\cdot)$ represents the ReLU activation function. The parameter D is set to 128 optimized by the experiments.

The reason for fusing the spatial and spectral information lies in that, in the human brain, there exist complicated couplings and resonance between signals of different frequency bands and different lobes [42]. Joint attention on spatial and spectral features will help discover these complex relationships. In addition, the convolutional kernels automatically assign the weights of different channels and frequency bands. As shown in Fig. 5, we visualize four convolutional kernels from

Fig. 6. Illustration of the structure of one MTE layer, which consists of Spatial-Spectral MSA, Temporal MSA, and MLP. The internal structures of Spatial-Spectral MSA and Temporal MSA are illustrated below the MTE layer.

the trained model as the examples. From the visualization, we can see that different kernels focus on different spatial and spectral information. Attention on low-frequency bands is irregular and changeable for various channels. For example, Kernel 1 and 2 exhibit totally different weights between 1 and 20 Hz in most channels. In contrast, high-frequency bands are critical for emotion in some channels, but insignificant in other channels. For instance, Kernel 3 only concentrates on high-frequency bands in Channel 3, but discards information of other channels.

Above all, SSP Block not only promotes the deep attention and facilitates the training of the succeeding MTE, but also automatically extracts critical spatial and spectral information.

2) *Positional Embedding*: The Positional Embedding (PE) is added on I' like most Transformer-based models, which is a set of 2D learnable parameters.

$$Z_0 = I' + \text{PE} \quad (8)$$

where $Z_0 \in \mathbb{R}^{L \times D}$. Therefore, the matrix Z_0 containing all compressed tokens with positional information is obtained and served as the input of MTE, whose dimensions correspond to time and fused channel-frequency.

3) *MTE Layer*: Typically, the Transformer layer contains one MSA layer and one feedforward Multi-Layer Perceptron (MLP) [43], where the MSA is usually applied on the sequential dimension. The MSA mechanism is sensitive to the internal features of long-term series data for its strong global attention. The multiple attention heads allow it to learn relevant information in different representation subspaces and reduce the computational cost. The scaled dot-product attention is computed from queries, keys, and values (Q, K, V), all of which are linearly projected from the sequential data. Every

Fig. 7. Illustration of the comparison of self-attention mechanism between conventional MSA and our squeezed MSA. Here n denotes the number of feature points. (a) In the convention MSA, the correlations between all specific feature points are computed. (b) In our squeezed MSA, the correlations between feature points and pairs of merged points are computed.

MLP contains two Fully Connected (FC) layers with a Gaussian Error Linear Unit (GELU) activation function between them, where the middle dimension is usually expanded to better express rich spatial characteristics.

To better adapt the model to our task, we build our JDAT which focuses on the multiplex EEG information jointly, utilizing the unified attention modules. As shown in Fig. 6, three sublayers are built in our adjusted MTE layer, namely Spatial-Spectral MSA layer, Temporal MSA layer, and MLP. The residual connection and Layer Normalization (LN) follow every sublayers, which alleviate the problems of gradient dissipation, explosion, etc. The workflow of MTE layers can be formulated as follows:

$$Z_l^\alpha = \text{LN}(\text{SS-MSA}(Z_{l-1}) + Z_{l-1}) \quad l = 1, 2, 3, 4 \quad (9)$$

$$Z_l^\beta = \text{LN}(\text{T-MSA}(Z_l^\alpha) + Z_l^\alpha) \quad l = 1, 2, 3, 4 \quad (10)$$

$$Z_l^\gamma = \text{LN}(\text{MLP}(Z_l^\beta) + Z_l^\beta) \quad l = 1, 2, 3, 4 \quad (11)$$

where the SS-MSA and T-MSA denote Spatial-Spectral MSA layer and Temporal MSA layer, respectively. Besides, the number of attention heads in every Spatial-Spectral MSA layer is optimized to 8, while that in Temporal MSA layers are 8, 4, 2, and 1 from Stage 1 to Stage 4 to fit the gradually compressed structure, respectively. The expansion ratio of the middle layer in feedforward MLP is set to 4.

As illustrated in Fig. 6, the internal structure of our MSA layer is revised into a squeezed structure, to improve the generalization ability of the model. The conventional MSA layer applies the global attention between all specific tokens, while our squeezed MSA is applied between a specific feature point and a bunch of adjacent points, which is illustrated in Fig. 7. Specifically, we take the squeezed Temporal MSA as the example to illustrate the mechanism. The input sequence $Z \in \mathbb{R}^{L \times D}$ is squeezed by an 1D Average Pooling (1D-AP) layer with a reduction ratio R , generating $\tilde{Z}$ with a shorter temporal length. It is equivalent to merge the adjacent R tokens into a new token of the same dimension.

$$\tilde{Z} = \text{1D-AP}(Z) \quad (12)$$

where $\tilde{Z} \in \mathbb{R}^{(Lz/R) \times D}$. Then, Q is computed from the input Z , while K and V are computed from the squeezed data $\tilde{Z}$. The linear projection W_Q , W_K , and $W_V \in \mathbb{R}^{D \times d}$ transform the input sequence into the corresponding heads of the same dimension d .

$$Q = ZW_Q, K = \tilde{Z}W_K, V = \tilde{Z}W_V \quad (13)$$

where $Q \in \mathbb{R}^{Lz \times d}$, $K \in \mathbb{R}^{(Lz/R) \times d}$, and $V \in \mathbb{R}^{(Lz/R) \times d}$. Therefore, the number of tokens in K and V are also squeezed by the same ratio. The squeeze ratio R is optimized to 2 in our model, so the squeeze process is equivalent to merging pairs of adjacent feature points. The dimension of attention heads d equals to D divided by the number of heads. The computation process of squeezed MSA from Q , K , and V can be formulated as follows:

$$\begin{aligned} \text{Attention}(Q, K, V) &= \text{softmax} \left(\frac{QK^T}{\sqrt{d}} \right) V \\ &= \text{softmax} \left(\frac{ZW_Q W_K^T \tilde{Z}^T}{\sqrt{d}} \right) \tilde{Z} W_V \end{aligned} \quad (14)$$

The multiple attention heads will be concatenated together and projected into the original dimension D again. Therefore, the squeezed MSA computed from Q , K , and V holds the same shape as the conventional MSA. On one hand, our novel squeezed MSA layer can reduce the calculation time generated by the long-term sequence. On the other hand, self-attention between a specific feature point and merged points can reduce the risk of overfitting.

In our MTE layer, the Spatial-Spectral MSA layer adaptively seeks the essential information across different electrode channels and frequency bands, focusing on the complex features in integrated dimensions, such as cross-lobe coupling, frequency-intensity resonance, etc. Next, the Temporal MSA layer handles the long-time series of EEG data, which makes the model sensitive to the degree of signal activation and waveform changes. Meanwhile, every Temporal MSA layer utilizes the fused features discovered by the previous Spatial-Spectral MSA layer, whose output will be processed globally by the feedforward MLP and sent into the next stage again. It enables the model to repeatedly use intersecting features to reveal complicated brain activities. Our experiments have proved that the best performance is achieved with the combination of Spatial-Spectral MSA and Temporal MSA with the squeezed structures to avoid overfitting. Consequently, the MTE is able to focus its deep attention on multidimensional EEG information comprehensively.

4) *Temporal Compression*: Four stages with MTE layers are built in JDAT, where each stage except the last one is followed by a MAX Pooling (MAP) layer with the kernel size of 3 and the stride of 2, which acts as the temporal compression.

$$Z_l = \text{MAP}(Z_l^\gamma) \quad l = 1, 2, 3 \quad (15)$$

In stage 1, 2, and 3, the temporal compression gradually compresses the sequential dimensions of the output feature maps. We calculate the computational operations of stage n in JDAT and JDAT without the temporal compression, which

TABLE I
COMPUTATIONAL OPERATIONS OF JDAT AND
JDAT WITHOUT TEMPORAL COMPRESSION IN STAGE N.

Sublayer	JDAT	JDAT w/o TC
SS-MSA	$\frac{LD^3}{2^{n+1}} + \frac{L^2 D^2}{2^{n+3}} + \frac{3L^2 D}{2^{2n-2}}$	$\frac{LD^3}{4} + \frac{L^2 D^2}{16} + 3L^2 D$
T-MSA	$\frac{3LD^2}{2^{n-1}} + \frac{L^2 D^2}{2^{2n+2}} + \frac{L^3 D}{2^{3n-1}}$	$3LD^2 + \frac{L^2 D^2}{16} + \frac{L^3 D}{4}$
MLP	$2^{4-n} LD^2$	$8LD^2$

is listed in Table I. The shape of the input of MTE is (L, D) , and we have also considered the specific numbers of attention heads in MSA layers. In stage 1, the two models hold the same module structures, while JDAT efficiently reduces the computational operations in higher stages. Compared with the model without temporal compression, it saves nearly 40% of the total operations with the same depth, which will dramatically boost the inference speed. The experiment also shows that the performance is maintained with the temporal compression.

5) *Classification*: Different from popular Transformer models like BERT [30] and ViT [31], our JDAT discards the class token, a learnable embedding prepended to the sequential tokens. Instead, the last stage of MTE is extended with a Global Average Pooling (GAP) layer, transforming the matrix Z_4^γ into a vector of dimension D . Finally, two dense FC layers are implemented on the head of the model to make the classification.

$$\text{Result} = \text{Dense}(\text{GAP}(Z_4^\gamma)) \quad (16)$$

This modification is more suitable for our JDAT with a gradually compressed structure, and the experiments also show that using GAP performs better than prepending the class token.

IV. EXPERIMENTS

A. Datasets

Our experiments were conducted on three benchmarking datasets, namely DEAP [24], DREAMER [25], and SEED [26], [27]. These datasets contain EEG signals stimulated by outside multimedia, whose details are listed in Table II.

In DEAP and DREAMER, we regard valence and arousal as our classification tasks and set the thresholds to divide the labels into two emotional classes. The videos with rating values no larger than the threshold are labeled as negative, while that larger than the threshold are labeled as positive. In SEED, there are no baseline removal and label division. Also, only one label is provided for these videos with three emotional classes (positive, neutral, and negative), and we treat it as the valence task. The EEG signals in SEED are down-sampled to 128Hz to fit our model. Some other details of the public datasets, such as ages, male/female, and signal filters are clear in their public documentation and will not be described here.

TABLE II
DETAILS OF DEAP, DREAMER, AND SEED DATASETS.

Dataset	DEAP	DREAMER	SEED
# Subjects	32	23	15
# Videos	40	18	45
# Channels	32	14	62
Frequency	128Hz	128Hz	200Hz
Pre-Trial	3s	4s	—
Trial	60s	65-393s	240s
Valence	✓	✓	✓
Arousal	✓	✓	—
Values	1 – 9	1 – 5	—
Threshold	5	3	—
# Classes	2	2	3

B. Settings

We trained and tested JDAT on an NVIDIA RTX 1080Ti GPU, implemented with the PyTorch framework [44]. During training, the Adam optimizer [45] was used for back-propagation. The learning rate was set to 0.001. For experiment on each subject's samples, we used 10-fold cross-validation to evaluate our model. To prevent overfitting, dropout was implemented on SSP Block, MTE, and classification layers, with dropout rates of 0.1, 0.1, and 0.5, respectively.

C. Results

We compare the performance of our JDAT with SOTA baseline models listed below on DEAP, DREAMER, and SEED datasets. These works all conduct subject-dependent experiments. They use the same methods to divide labels in DEAP and DREAMER datasets, and most of them use the same cross-validation method.

- **4D-CRNN**: Combined model of CNN and RNN with 4D input structure [46].
- **4D-aNN**: 4D attention-based neural network [22].
- **ACRNN**: Attention-based convolutional recurrent neural network [21].
- **SST-EmotionNet**: Spatial-Spectral-Temporal based attention 3D dense network [23].
- **gcForest**: Deep forest model named multi-Grained Cascade Forest [47].
- **MLF-CapsNet**: Multi-level features guided capsule network [48].
- **EeT**: EEG emotion Transformer which applies attention on temporal and spatial dimensions [37].

The 4D-CRNN and 4D-aNN both combine CNN and LSTM layers with attention modules together. They adopt 4D data as the input, which are transformed from the EEG signals

or extracted features. SST-EmotionNet focuses on spatial-spectral and spatial-temporal features using parallel blocks, whose attention is separately applied in different combined dimensions. ACRNN adopts the preprocessed EEG signals as the input instead of extracted features, containing channel-wise attention, CNN layers, LSTM layers, and self-attention. gcForest is based on the deep forest algorithm, whose critical scanning module mines the spatial and temporal information of EEG signals to classify emotions. MLF-CapsNet is an end-to-end capsule network, combining multi-level features extracted from different convolutional layers to form primary capsules. EeT is a Transformer-based model, which focuses on spatial-temporal EEG information jointly, but ignores the spectral information. Most of the above baseline models are conducted only on one or two datasets, because their structures are usually not applicable to other datasets.

The accuracies and standard deviations of these baseline models and our proposed JDAT on three datasets are presented in Table III. Overall, deep learning methods combined with attention mechanism exhibit better performance and flexibility. The traditional method, gcForest, achieves good result on DEAP dataset but performs inferior on DREAMER dataset. The standard deviation varies with different models, but on the whole it decreases as the accuracy increases.

Obviously, our JDAT achieves the highest accuracies in all datasets. In SEED dataset, JDAT improves the accuracy to 97.30%, outperforming other attention-based models including EeT by more than 1%. In DEAP dataset, it obtains accuracies above 98.5% in both valence and arousal tasks, which performs much better than the networks with combined attention modules, namely ACRNN and 4D-CRNN. In DREAMER dataset, it also performs slightly better than ACRNN. According to our analysis, SEED dataset stores the largest amount of data for each subject, which facilitates the training of our Transformer-based model. Meanwhile, SEED and DEAP datasets contain EEG signals of 62 and 32 channels respectively, while only 14 channels are included in DREAMER dataset. Therefore, the performance improvement in SEED and DEAP is relatively significant, which indicates that JDAT is more suitable for learning EEG data with richer information. But in short, JDAT has demonstrated the strong and flexible learning ability, since it operates well in various EEG datasets with different experimental settings and different information.

D. Ablation Studies

To illustrate the distinct advantages of our JDAT, we implement the following variants for comparison:

- **JDAT w/o TC** which removes the Temporal Compression in Stage 1, 2, and 3.
- **JDAT w/o LS** which removes the Length Squeeze in every MSA layer, which adopts conventional MSA layers.
- **JDAT w/o SSP** which removes the SSP Block.
- **JDAT w/o T-MSA** which removes the Temporal MSA layer in every MTE layer.
- **JDAT w/o SS-MSA** which removes the Spatial-Spectral MSA layer in every MTE layer.

TABLE III
PERFORMANCE COMPARISON BETWEEN BASELINE MODELS AND JDAT ON DEAP, DREAMER, AND SEED DATASETS.

Model	Dataset	# Classes	Cross-validation	Valence		Arousal	
				ACC (%)	STD (%)	ACC (%)	STD (%)
ACRNN	DEAP	2	10-fold	93.72	3.21	93.38	3.73
4D-CRNN	DEAP	2	5-fold	94.22	2.61	94.58	3.69
gcForest	DEAP	2	10-fold	97.69	1.22	97.53	1.52
MLF-CapsNet	DEAP	2	10-fold	97.97	1.67	98.32	1.24
JDAT	DEAP	2	10-fold	98.51	1.24	98.55	1.13
gcForest	DREAMER	2	10-fold	89.03	5.56	90.41	5.33
MLF-CapsNet	DREAMER	2	10-fold	94.59	3.77	95.26	3.63
ACRNN	DREAMER	2	10-fold	97.93	1.73	97.98	1.92
JDAT	DREAMER	2	10-fold	98.01	2.13	98.09	1.89
4D-CRNN	SEED	3	5-fold	94.74	2.32	N/A	N/A
SST-EmotionNet	SEED	3	None	96.02	2.17	N/A	N/A
4D-aNN	SEED	3	5-fold	96.10	2.61	N/A	N/A
EeT	SEED	3	5-fold	96.28	4.39	N/A	N/A
JDAT	SEED	3	10-fold	97.30	1.74	N/A	N/A

Among these variants, JDAT without SSP Block also removes the Continuous Wavelet Transform (CWT) in the pre-processing. Instead, the EEG signals after baseline removal are directly taken as the training data, and an additional FC layer is implemented to adapt the data dimensions to the MTE. The other variants maintain the SSP Block and the same pre-processing method. All variants hold the same classification layers as the complete JDAT.

Firstly, we conduct experiments on the DEAP, DREAMER, and SEED datasets to compare the performance of these variants, where the accuracies and F1-scores are tested. F1-score embodies quality of the model as it considers the generalization ability when the data samples are unevenly distributed. The experimental results of JDAT and its variants are shown in Fig. 8. In DEAP and DREAMER, the results are shown as the mean values of results in valence and arousal tasks.

The ablation study shows that our proposed JDAT achieves better performance than the variant without temporal compression in SEED dataset, and comparable results in the other two datasets. The temporal compression greatly reduces the computational operations, and it still enables the model to maintain the overall performance. In addition, it is obvious that without the length squeeze in MSA layers, the accuracies and F1-scores drop further by about 2.5%. It demonstrates that self-attention between all specific feature points are unnecessary, and our squeezed MSA can decrease the risk of

overfitting. Furthermore, the lack of SSP Block will have a significant negative impact on the model, especially in SEED dataset, where it reduces the accuracy and F1-score by around 3% and 4%, respectively. Therefore, the SSP Block exhibits its strong ability to extract critical spatial and spectral information globally, and facilitate the operation of MTE. It is worth noting that JDAT without Spatial-Spectral MSA layers or Temporal MSA layers reduces the accuracy by around 3% - 4.5%. The lack of MSA layers shows the largest negative influence on the performance, which demonstrates the advantages of joint attention on spatial, spectral, and temporal EEG features. The Spatial-Spectral MSA seems to be more significant, since the attention on space and frequency plays an important role. In summary, all of these modules are essential in our JDAT, and the best performance is achieved by the combination of them.

Meanwhile, to demonstrate the flexibility and practicality of our designed model, we also test the parameters, Giga Floating Point operations (GFLOPs), and inference time during the ablation study. We choose the variant without temporal compression and the variant without length squeeze for comparison, as they hold the same module structures and model depth as the complete JDAT. The results are listed in Table IV, where the GFLOPs and inference time are tested per data sample in SEED dataset. Obviously, the temporal compression and length squeeze both alleviate the parameters and GFLOPs, and speed up the inference. The temporal compression in stage 1, 2, and 3 greatly reduces the total

Fig. 8. Ablation study on JDAT, (a): DEAP dataset; (b): DREAMER dataset; (c): SEED dataset. In the experiments, the accuracies and F1-Scores are tested. In DEAP and DREAMER datasets, the results are shown as the mean values of results in valence and arousal tasks.

GFLOPs by around 40%, and saves 0.12ms for the inference of one data sample. It helps compress the sequential data size step by step, instead of keeping the same shape during forward propagation. The above experiments have demonstrated that the performance is maintained with the temporal compression. Meanwhile, the length squeeze in the MSA layer further alleviates the parameters and GFLOPs, while still greatly improving the overall performance. With the combination of squeezed MSA and temporal compression, the lightweight and fast JDAT are more suitable for mobile hardware devices.

V. CONCLUSION AND DISCUSSION

In this paper, we have proposed JDAT, utilizing Transformer for EEG emotion recognition. It adopts the unified architecture based on adaptive self-attention mechanism, solving the problem that attention was always applied on scattered dimensional

TABLE IV
PARAMETERS, GFLOPs AND INFERENCE TIME OF JDAT AND ITS TWO VARIANTS ON SEED DATASET.

Model	Params ($\times 10^6$)	GFLOPs ($\times 10^{-1}$)	Time (ms)
w/o TC	1.50	1.70	0.44
w/o LS	1.33	1.11	0.36
JDAT	1.32	1.05	0.32

features in previous works. The main structure of JDAT only consists of SSP Block and MTE. SSP Block captures and fuses spatial and spectral information, which facilitates the operation of MTE. The MTE is built with multidimensional squeezed MSA layers, which is sensitive to the complicated couplings and resonance in EEG signals. The squeezed MSA helps improve the generalization ability of the model, reducing the risk of overfitting. Temporal compression in each stage greatly reduces the parameters and computational operations, and the model maintains the overall performance. JDAT exhibits its distinct advantage of automatic and comprehensive joint-attention on features in multiple dimensions, without selecting channels or extracting spectral features. The SOTA results on three datasets verify its strength and flexibility, and the ablation studies demonstrate the necessity of essential modules of JDAT. In future, we will utilize JDAT in other EEG classification tasks, such as motor imagination and seizure detection.

REFERENCES

- [1] K. Anderson and P. W. McOwan, "A real-time automated system for the recognition of human facial expressions," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 36, no. 1, pp. 96–105, 2006.
- [2] F. Dellaert, T. Polzin, and A. Waibel, "Recognizing emotion in speech," in *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96*, vol. 3. IEEE, 1996, pp. 1970–1973.
- [3] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from eeg data using machine learning approach," *Neurocomputing*, vol. 129, pp. 94–106, 2014.
- [4] B. Cheng and G. Liu, "Emotion recognition from surface emg signal using wavelet transform and neural network," in *Proceedings of the 2nd international conference on bioinformatics and biomedical engineering (ICBBE)*, 2008, pp. 1363–1366.
- [5] F. Agraftioti, D. Hatzinakos, and A. K. Anderson, "Ecg pattern analysis for emotion detection," *IEEE Transactions on affective computing*, vol. 3, no. 1, pp. 102–115, 2011.
- [6] A. Al-Nafjan, M. Hosny, A. Al-Wabil, and Y. Al-Ouali, "Classification of human emotions from electroencephalogram (eeg) signal using deep neural network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 9, pp. 419–425, 2017.
- [7] C. D. Salzman and S. Fusi, "Emotion, cognition, and mental state representation in amygdala and prefrontal cortex," *Annual review of neuroscience*, vol. 33, pp. 173–202, 2010.
- [8] X. Deng, M. Yang, and S. An, "Differences in frontal eeg asymmetry during emotion regulation between high and low mindfulness adolescents," *Biological Psychology*, vol. 158, p. 107990, 2021.
- [9] Y. Cimtay and E. Ekmekcioglu, "Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset eeg emotion recognition," *Sensors*, vol. 20, no. 7, p. 2034, 2020.

- [10] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for eeg-based emotion classification," in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2013, pp. 81–84.
- [11] C. A. Frantzidis, C. Bratsas, C. L. Papadelis, E. Konstantinidis, C. Pappas, and P. D. Bamidis, "Toward emotion aware computing: an integrated approach using multichannel neurophysiological recordings and affective visual stimuli," *IEEE transactions on Information Technology in Biomedicine*, vol. 14, no. 3, pp. 589–597, 2010.
- [12] L. Squire, D. Berg, F. E. Bloom, S. Du Lac, A. Ghosh, and N. C. Spitzer, *Fundamental neuroscience*. Academic press, 2012.
- [13] J. Daume, T. Gruber, A. K. Engel, and U. Frieze, "Phase-amplitude coupling and long-range phase synchronization reveal frontotemporal interactions during visual working memory," *Journal of Neuroscience*, vol. 37, no. 2, pp. 313–322, 2017.
- [14] G. Rigas, C. D. Katsis, G. Ganiatsas, and D. I. Fotiadis, "A user independent, biosignal based, emotion recognition method," in *International Conference on User Modeling*. Springer, 2007, pp. 314–318.
- [15] K.-E. Ko, H.-C. Yang, and K.-B. Sim, "Emotion recognition using eeg signals with relative power values and bayesian network," *International Journal of Control, Automation and Systems*, vol. 7, no. 5, p. 865, 2009.
- [16] P. Rani, C. Liu, N. Sarkar, and E. Vanman, "An empirical study of machine learning techniques for affect recognition in human-robot interaction," *Pattern Analysis and Applications*, vol. 9, no. 1, pp. 58–69, 2006.
- [17] C. Zong and M. Chetouani, "Hilbert-huang transform based physiological signals analysis for emotion recognition," in *2009 IEEE international symposium on signal processing and information technology (ISSPIT)*. IEEE, 2009, pp. 334–339.
- [18] S. Hwang, K. Hong, G. Son, and H. Byun, "Learning cnn features from de features for eeg-based emotion recognition," *Pattern Analysis and Applications*, vol. 23, no. 3, pp. 1323–1335, 2020.
- [19] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion recognition based on eeg using lstm recurrent neural network," *Emotion*, vol. 8, no. 10, pp. 355–358, 2017.
- [20] D. Zeng, K. Huang, C. Xu, H. Shen, and Z. Chen, "Hierarchy graph convolution network and tree classification for epileptic detection on electroencephalography signals," *IEEE Transactions on Cognitive and Developmental Systems*, 2020.
- [21] W. Tao, C. Li, R. Song, J. Cheng, Y. Liu, F. Wan, and X. Chen, "Eeg-based emotion recognition via channel-wise attention and self attention," *IEEE Transactions on Affective Computing*, 2020.
- [22] G. Xiao, M. Ye, B. Xu, Z. Chen, and Q. Ren, "4d attention-based neural network for eeg emotion recognition," *arXiv preprint arXiv:2101.05484*, 2021.
- [23] Z. Jia, Y. Lin, X. Cai, H. Chen, H. Gou, and J. Wang, "Sst-emotionnet: Spatial-spectral-temporal based attention 3d dense network for eeg emotion recognition," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 2909–2917.
- [24] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 18–31, 2011.
- [25] S. Katsigiannis and N. Ramzan, "Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices," *IEEE journal of biomedical and health informatics*, vol. 22, no. 1, pp. 98–107, 2017.
- [26] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for EEG-based emotion classification," in *6th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2013, pp. 81–84.
- [27] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Transactions on Autonomous Mental Development*, vol. 7, no. 3, pp. 162–175, 2015.
- [28] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [29] Y. Yang, Q. Wu, M. Qiu, Y. Wang, and X. Chen, "Emotion recognition from multi-channel eeg through parallel convolutional recurrent neural network," in *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2018, pp. 1–7.
- [30] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- [31] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
- [32] 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," *arXiv preprint arXiv:2103.14030*, 2021.
- [33] R. Strudel, R. Garcia, I. Laptev, and C. Schmid, "Segmenter: Transformer for semantic segmentation," *arXiv preprint arXiv:2105.05633*, 2021.
- [34] Z. Sun, S. Cao, Y. Yang, and K. M. Kitani, "Rethinking transformer-based set prediction for object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 3611–3620.
- [35] J. Pedoeem, S. Abittan, G. B. Yosef, and S. Keene, "Tabs: Transformer based seizure detection," in *2020 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. IEEE, 2020, pp. 1–6.
- [36] J. Sun, J. Xie, and H. Zhou, "Eeg classification with transformer-based models," in *2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech)*. IEEE, 2021, pp. 92–93.
- [37] J. Liu, L. Zhang, H. Wu, and H. Zhao, "Transformers for eeg emotion recognition," *arXiv preprint arXiv:2110.06553*, 2021.
- [38] W. Qu, Z. Wang, H. Hong, Z. Chi, D. D. Feng, R. Grunstein, and C. Gordon, "A residual based attention model for eeg based sleep staging," *IEEE journal of biomedical and health informatics*, vol. 24, no. 10, pp. 2833–2843, 2020.
- [39] S. Nagel, "Towards a home-use bci: fast asynchronous control and robust non-control state detection," Ph.D. dissertation, Universität Tübingen, 2019.
- [40] D. Jordan, R. Miksad, and E. Powers, "Implementation of the continuous wavelet transform for digital time series analysis," *Review of scientific instruments*, vol. 68, no. 3, pp. 1484–1494, 1997.
- [41] R. N. Bracewell and R. N. Bracewell, *The Fourier transform and its applications*. McGraw-Hill New York, 1986, vol. 31999.
- [42] R. M. Reinhart and J. A. Nguyen, "Working memory revived in older adults by synchronizing rhythmic brain circuits," *Nature neuroscience*, vol. 22, no. 5, pp. 820–827, 2019.
- [43] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," pp. 5998–6008, 2017.
- [44] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, pp. 8026–8037, 2019.
- [45] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [46] F. Shen, G. Dai, G. Lin, J. Zhang, W. Kong, and H. Zeng, "Eeg-based emotion recognition using 4d convolutional recurrent neural network," *Cognitive Neurodynamics*, vol. 14, no. 6, pp. 815–828, 2020.
- [47] J. Cheng, M. Chen, C. Li, Y. Liu, R. Song, A. Liu, and X. Chen, "Emotion recognition from multi-channel eeg via deep forest," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 2, pp. 453–464, 2020.
- [48] Y. Liu, Y. Ding, C. Li, J. Cheng, R. Song, F. Wan, and X. Chen, "Multi-channel eeg-based emotion recognition via a multi-level features guided capsule network," *Computers in Biology and Medicine*, vol. 123, p. 103927, 2020.

Zeyu Wang (S'21) received the B.S. degree from Soochow University, Suzhou, China, in 2020. He is currently pursuing the Ph.D. degree in College of Information Science & Electronic Engineering at Zhejiang University. Now, he is an IEEE student member. His research interests include deep learning, biological signal processing, brain-computer interface, and computer vision.

Qi Xu received his B.Eng. degree in the College of Computer Science and Technology from Zhejiang University of Technology in 2015, and Ph.D. degree in the College of Computer Science and Technology from Zhejiang University in 2021. Since 2021, he has been a tenure-track associate professor in School of Artificial Intelligence, Dalian University of Technology. He was ever granted as the honorary visiting fellow in the Centre for Systems Neuroscience, University of Leicester, U.K in 2019. His research interests include brain-inspired computing, neuromorphic computing, neural computation, computational neuroscience and cyborg intelligence.

Ziqun Zhou (S'19) received the B.S. degree in Information and Communication Engineering from Harbin Engineering University, Harbin, China, in 2018. Now, he is pursuing the Ph.D. degree in College of Information Science & Electronic Engineering at Zhejiang University. His research interests include deep learning and medical biological information processing.

Kejie Huang (Senior Member, IEEE) received the Ph.D degree from the Department of Electrical Engineering, National University of Singapore, Singapore, in 2014. He has been a principal investigator at College of Information Science & Electronic Engineering, Zhejiang University (ZJU) since 2016. Prior to joining ZJU, he spent five years in the industry including Samsung and Xilinx, two years in the Data Storage Institute, Agency for Science Technology and Research (A*STAR), and another three years in Singapore University of Technology and Design (SUTD), Singapore. He has authored or co-authored more than 50 scientific papers in international peer-reviewed journals and conference proceedings. He holds more than 10 granted patents, and another 20 pending ones. His research interests include hardware acceleration for deep learning, architecture development for post Moore's law era, and computer vision. He is an IEEE Senior Member, the Associate Editor of IEEE TCASII, and the reviewer of many international journals such as Nature Communications, IEEE TCAS, TVLSI, EDL.

Haibin Shen is currently a Professor with Zhejiang University, a member of the second level of 151 talents project of Zhejiang Province, and a member of the Key Team of Zhejiang Science and Technology Innovation. His research interests include learning algorithm, processor architecture, and modeling. His research achievement has been used by many authority organizations. He has published more than 100 papers on academic journals, and he has been granted more than 30 patents of invention. He was a recipient of the First Prize of Electronic Information Science and Technology Award from the Chinese Institute of Electronics, and has won a second prize at the provincial level.