Smart Health 24 (2022) 100261

Contents lists available at ScienceDirect

Smart Health

journal homepage: www.elsevier.com/locate/smhl

EEG classification with spiking neural network: Smaller, better, more energy efficient☆

Zhanglu Yan *, Jun Zhou, Weng-Fai Wong

Department of Computer Science, National University of Singapore, Singapore

ARTICLE INFO

Keywords:
Spiking neural network
EEG
Emotion classification

ABSTRACT

Real-time emotion recognition with electroencephalograph (EEG) has been an active field of research in recent years. In particular, deep learning has been shown to be effective in emotion classification tasks. However, the monitoring of EEG signals is a continuous process, there is a need for energy-efficient emotion classification methods. Compared with artificial neural networks (ANNs), spiking neural networks (SNNs), in which weight multiplications are replaced by additions, are more energy efficient. In this paper, we propose a near-lossless transfer learning method for SNNs, specially designed for EEG signals. Data is preprocessed, and its power spectral density (PSD) is extracted to represent the frequency domain of the raw EEG signal. Using a 3-layer pretrained SNN, running on the DEAP dataset, we achieved an accuracy of 78.87% and 76.5% for valence and arousal dimensions, respectively. By training a model based on one dimension and fine-tuning on another, we achieve an even higher accuracy, 82.75% for the valence and 84.22% for the arousal. As far as we know, our results yield the smallest SNN with the highest accuracy for this task to date. The power of our SNNs for valence and arousal dimensions is 13.8% that of our CNN-based solutions. The framework was developed by PyTorch and is available under an open-source license.[1]

1. Introduction

Electroencephalograph (EEG) signal has been widely used to evaluate the electrical activities of the brain. Such technology can be useful as a brain–machine interface for a wide range of uses including entertainment, security, and medicine, for example, emotion-detecting VR games and software that help police test criminal suspects for signs of deception. The market for emotion-detection has been estimated to be roughly $21.6 billion in 2019 and is predicted to be $56 billion in 2024 (Gannouni, Aledaily, Belwafi, & Aboalsamh, 2021).

Patterns of brain waves are tracked and recorded by an EEG machine. Usually, electrodes made of metal are attached to the scalp. Electrical impulses detected through those electrodes are measured and analyzed. An EEG consists of five different kinds of waves: gamma (over 30 Hz), beta (12–30 Hz), alpha (7–12 Hz), theta (4–7 Hz), and delta (1–4 Hz) waves. According to Li and Lu (2009), the gamma and alpha waves contribute the most to emotion detection. Changes in the gamma band can be related to the emotions of happiness and sadness. A decrease in the alpha wave in different sides of the temporal lobe can also be related to the emotions of happiness and sadness (left for sadness and right for happiness). Beta and theta waves correspond to the active

☆ This research is supported in part by Programmatic grant no. A1687b0033 from the Singapore government's Research, Innovation, and Enterprise 2020 plan (Advanced Manufacturing and Engineering domain).

* Corresponding author.
E-mail addresses: zhangluyan@comp.nus.edu.sg (Z. Yan), zhoujun@comp.nus.edu.sg (J. Zhou), wongwf@comp.nus.edu.sg (W.-F. Wong).

[1] https://github.com/zhoujuncc1/shenjingcat/tree/master/examples/eeg_binary.

https://doi.org/10.1016/j.smhl.2021.100261
Received 29 July 2021; Received in revised form 11 November 2021; Accepted 26 December 2021
Available online 11 January 2022

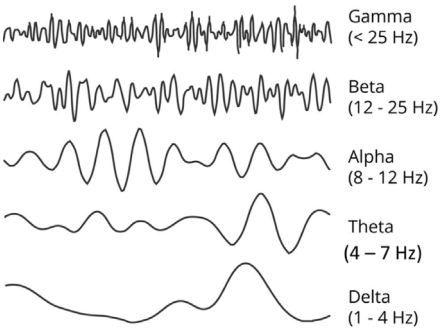


Fig. 1. EEG Waves (Sakhavi, 2017).

states of the mind and subconscious mind, respectively. In our paper, all five waves are used to detect the emotion for greater accuracy (Sakhavi, 2017). The shape of those waves is showed in Fig. 1.

However, most existing works use some deep and complex artificial neural networks for EEG detection that are hard to implement on resource-constrained platforms (Alhagry, Fahmy, & El-Khoribi, 2017; Li, et al., 2016; Zhang, Roy, & Jensen, 2001). Deep neural networks always suffer from high energy consumption, making them unsuitable for wearable or embedded devices such as the Oculus Quest 2 (VR helmet) and EMOTIVE (EEG detection device), where the promise of real-time emotion detection is now made possible. To enable such "always-on" features, we propose an energy-saving *spiking neural network* testing method that is custom designed for EEG signals.

A *spike-timing-dependent plasticity* (STDP) (Vigneron & Martinet, 2020) based SNN training algorithm NeuCube (Luo, Fu, et al., 2020) has been used for EEG emotion detection. However, to date, such approaches only work for shallow network structures (Kaiser, Mostafa, & Neftci, 2020; Shrestha & Orchard, 2018). More importantly, the accuracy of such training methods seems to be lower than traditional neural networks. For example, Luo, Fu, et al. (2020) used 1000 neurons for emotion classification and achieved an accuracy of 74% for arousal dimension. Our work uses *transfer learning* instead. First, a CNN is trained using the standard backpropagation method. Then, the weights of CNN are transferred to the equivalent SNN. The CQ training ('clamped and quantized') algorithm (Yan, Zhou, & Wong, 2021) is then used to reduce the accuracy difference between the CNN and SNN.

The DEAP dataset (Koelstra, et al., 2011) is a well-known EEG emotion detection dataset that is also used in our paper. We tested both the *arousal* and *valence* dimensions of EEG. As far as we know, our results outperform the state-of-the-art on this dataset.

In summary, the contributions of this paper are:

(1) A transfer learning method for SNN with a state-of-art accuracy (82.75% for valence dimension and 84.22% for arousal dimension), tested on the DEAP dataset.
(2) The power of our SNNs for the valence and arousal dimensions are 13.8% that of equivalent CNN-based solutions we implemented. This is indicative of the energy-saving that can be achieved by the use of our model.
(3) A fast and easy-to-program PyTorch-based spiking neural network framework with a method of transfer learning designed for EEG emotion detection that has been tested on the DEAP dataset.

The paper will next deal with the related works and the background of emotion classification on EEG. We will then present our approach in Section 4 including the preprocessing involved and the convolutional spiking neural network model used. In Section 5, we present our results and discussions on some of the key parameters of the model. This is followed by the conclusion.

2. Related works

It has been shown that raw EEG data cannot be used to achieve good accuracy on both the arousal and valence dimensions (Koelstra, et al., 2011). Preprocessing is needed. Naser and Saha (Daimi & Saha, 2014) use a *dual-tree complex wavelet packet transform* (DT-CWPT) to extract features from the alpha, gamma, and beta waves of EEG signals. Salama et al. employed *singular value decomposition* (SVD), QR factorization with column pivoting (QRcp), and F-ratio (Salama, El-Khoribi, Shoman, & Shalaby, 2018) for feature elimination. Tripathi, Acharya, Sharma, Mittal, and Bhattacharya (2017) proposed a novel feature-based emotion recognition model based on statistical features such as median, kurtosis coefficient, and standard deviation to improve the accuracy of detecting emotion. However, those combinations of complex preprocessing methods consume significant energy even before the classification task starts and is hard to implement on wearable devices or energy-constrained platforms. In this paper, we used only *power spectral density* (PSD) to represent the raw EEG signal in the frequency domain (Chen, et al., 2019).

While not energy efficient, traditional machine learning methods have achieved high accuracy on EEG classification. For example, *support vector machines* (SVM) (Daimi & Saha, 2014), deep *convolutional neural networks* (CNNs) (Li, et al., 2016), *recurrent neural networks* (RNN) (Alhagry et al., 2017) and *spatiotemporal recurrent neural network* (STRNN) (Zhang et al., 2001) demonstrated success for this task. There are also proposals that use *spiking neural networks* (SNNs). NeuCube uses a *spike-timing-dependent plasticity* (STDP)-based SNN training method for this task (Al-Nafjan, Alharthi, & Kurdi, 2020; Luo, Fu, et al., 2020).

Table 1
Structure of DEAP dataset.

Array name	Array shape	Array contents
Data	40* 40* 8064	Videos* Channels* Data
Labels	40*4	Valence, arousal, dominance, liking

40 channels: 32 different electrodes (record the electrical activity generated by the brain via electrodes placed on the scalp surface) and 8 peripheral channels.

In this paper, we use the method of transfer learning. Unlike direct SNN training, the issue of non-differentiability goes away because the model training is done using the CNN twin. We first train a CNN twin using the standard backpropagation method with both a batch norm and a dropout layer. The weights are then transferred to an equivalent SNN. Using this method, high accuracy and low energy consumption are achieved in both the valence and arousal detection tasks on the DEAP dataset.

3. Background

3.1. CNN vs SNN

Convolution neural networks (CNNs) have been successfully applied in medical image analysis. Examples include heartbeat detection with electrocardiogram (ECG) (Yan, Zhou, & Wong), breast cancer diagnosis (Gao, et al., 2018) and emotion detection with electroencephalograph (EEG) (Acharya, Oh, Hagiwara, Tan, & Adeli, 2018). CNNs always include convolution layers, pooling layers and fully connected layers. For example, in the convolutional layer l, inputs $\boldsymbol{x}^{l-1}$ are multiplied with weights $\boldsymbol{w}^l$ of layer l. After the weight sum, a *bias* $\boldsymbol{b}^l$ and non-linear activation function f are applied. *Batch normalization* (BN) and *dropout* (DP) layers are used to prevent overfitting. Mathematically:

$$x_i^l = DP\{BN\{f(\sum_{j=0}^{n}(w_{i,j}^l \cdot x_j^{l-1}) + b_i^l)\}\} \tag{1}$$

Data transmitted between CNN layers are in floating-point or some fixed-point number format. Unlike CNNs, the activation data transmitted between layers in spiking neural networks (SNNs) are the results of the *integrate-and-fire* model and are binary spikes, either 0 or 1. This allows us to replace the multiplications of weights with additions, with the potential of significantly reducing energy consumption. In each time step t, the *membrane potential* V_i^l at layer l is added with the weight sum (i.e., a masked summation) of the output spikes $s_j^{l-1}(t)$ of layer $l-1$ and the corresponding weights $w_{i,j}^l$. After adding the bias b_i^l at layer l, a 1 spike is then output if the membrane potential exceeds a predefined threshold, θ_i. Otherwise, it outputs '0'. $s_i^l(t)$ indicates the output spike trains of neuron i at layer l at time t. Mathematically:

If $V_i^l(t) >= \theta_i$:

$$V_i^l(t) = V_i^l(t-1) + \sum_j (w_{i,j}^l \cdot s_j^{l-1}(t) + b_i^l) - \theta_i \tag{2}$$

$$s_i^l(t) = 1 \tag{3}$$

Otherwise:

$$V_i^l(t) = V_i^l(t-1) + \sum_j (w_{i,j}^l \cdot s_j^{l-1}(t) + b_i^l) \tag{4}$$

$$s_i^l(t) = 0 \tag{5}$$

3.2. DEAP dataset

The DEAP Dataset (Koelstra, et al., 2011) is a publicly available dataset for emotion analysis. It is used extensively in research on EEG emotion classification. 32 participants' EEG and peripheral physiological signals are recorded in the DEAP dataset as they watched 40 music videos. Participants are required to perform a self-assessment questionnaire of their levels of arousal, valence, dominance, liking, and familiarity after watching different music videos. As described in the DEAP documentation, each participant's EEG was recorded with electrodes placed according to the international 10–20 system, with each electrode recording about 63 seconds of EEG signal. All data are downsampled to 128 Hz. *Electrooculogram* (EOG) artifacts are removed, and a bandpass frequency filter from 4.0–45.0 Hz was applied. The data was averaged to a common reference. The trials were reordered from presentation order to video (Experiment_id) order (Koelstra, et al., 2011).

There are two different perspectives to represent emotions. The first one (called *categorial*) indicates that basic emotions have evolved through natural selection. Plutchik proposed eight basic emotions: anger, fear, sadness, disgust, surprise, curiosity, acceptance, and joy (Plutchik, Kellerman, & Conte, 1979). All the other emotions can be formed by these basic ones. For example, disappointment is the composition of surprise and sadness.

An alternative perspective, called *dimensional*, is based on cognition. Emotions are mapped into valence and arousal dimensions, as shown in Fig. 2. Valence goes from very positive feelings to very negative (or unpleasant to pleasant), while arousal (also called activation) goes from states like sleepy to excited (Larsen & Diener, 1992).

For easy comparison with other works, we used the second perspective for EEG emotion classification of the DEAP dataset.

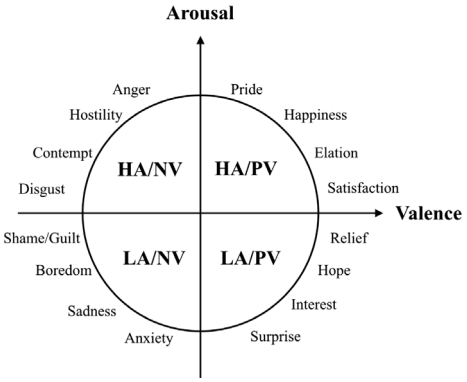


Fig. 2. Emotion Classification (Suhaimi, Yuan, Teo, & Mountstephens, 2018).

4. Approach

4.1. Power spectral density

Power Spectral Density(PSD), which served as a measure of signal's power content versus frequency, has been proven to useful in EEG feature extraction (Chen, et al., 2019). Welch's method is used here with Hamming window ($a_0 = 25/46$) to estimate the PSD (Welch, 1967). For L segment $x_i(n)$ of length M, we compute the periodogram $I_i(w)$, window function $w_i(n)$ and normalization parameter U_i as followed:

$$I_i(\omega) = \frac{1}{U_i}|\sum_{n=0}^{M-1} x_i(n)w(n)e^{-j\omega n}| \tag{6}$$

$$U_i = \frac{1}{M}\sum_{n=0}^{M-1} w^2(n) \tag{7}$$

$$w(n) = a_0 - (1-a_0)cos(\frac{2\pi}{M-1}) \tag{8}$$

$$P(e^{j\omega}) = \frac{1}{L}\sum_{i=1}^{L} I_i(\omega) \tag{9}$$

4.2. Convolutional neural network

In our experiment, we design our CNN according to the structure of LeNet, where pooling layers are added after convolutional layers. Instead of ReLU which is mostly used during activation function, we propose a "clamp ReLU" activation function $C(x)$ for higher CNN accuracy and lower CNN-SNN accuracy loss. After the clamp activation function, output CReLU(i) is restricted to $[0, 1]$. Mathematically,

$$\text{CReLU}(i) = \begin{cases} 0, & i < 0 \\ i, & 0 \le i \le 1 \\ 1, & i > 1. \end{cases} \tag{10}$$

The gradient (first derivative) of CReLU$'(i)$ is computed as

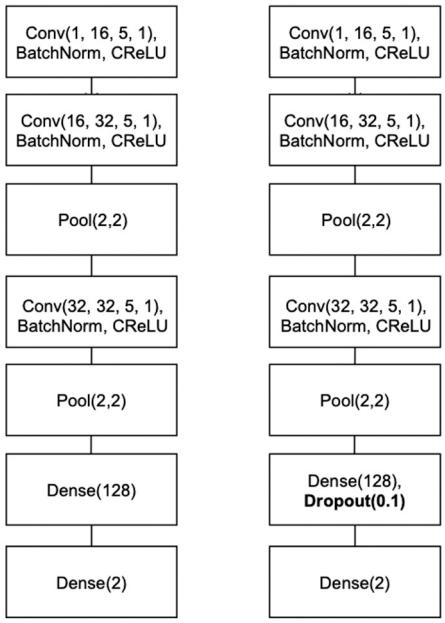
$$\text{CReLU}'(i) = \begin{cases} 0, & i < 0 \\ 1, & 0 \le i \le 1 \\ 0, & i > 1, \end{cases} \tag{11}$$

where i is the input.

After Clamp function, quantization function is added for the output of each convolutional layers as following to reduce accuracy drop during CNN-SNN conversion:

$$Q_T(x) = \frac{\lfloor x \cdot k \rfloor}{k}, \tag{12}$$

where k is the scaling factor.



(a) CNN for Arousal. (b) CNN for Valence.
Conv (c_{in}, c_{out}, k, s) means the size of stride is s, the input channel is c_{in}, the output channel is c_{out} and the kernel size is k; Pool(x,y) means the size of a stride is x * y; Dense(x) means the size of output is x.

Fig. 3. The CNN structure for Arousal and Valence classification.

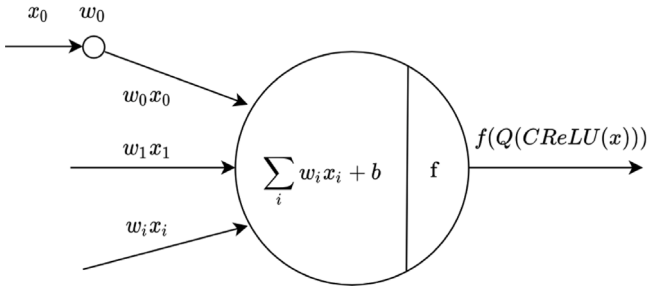


Fig. 4. Neuron of CNN.

We apply quantization to both the input data and layers activations. Hence, we need to consider the gradients for back propagation. Since $Q_T(x)$ is non-differentiable at turning points and $Q'_T(x) = 0$ elsewhere, we refer to *straight-through estimator* (Bengio, Léonard, & Courville, 2013) and approximate its gradient simply as

$$Q'_T(x) \approx 1. \tag{13}$$

BatchNorm layers have been added after each convolutional layer. To prevent overfitting, a dropout of 0.1 has been added after the layer convolutional layer.

The details of CNNs for both valence and arousal dimensions have been shown in Fig. 3.

4.3. Convolutional spiking neural network

In a CNN, as showed in Fig. 4, we need to multiply x_i with w_i and then do a sum. However, as we know, compared with additions, multiplications use much more energy (Wang, Zhou, Wong, & Peh, 2020).

But in spiking neural networks, shown in Fig. 5, all the data transmitted among layers are binary spikes. For example, for inputs from x_1 to x_n: only x_1 and x_3 are 1, the other inputs are all 0. Instead of multiplications, we can simply add w_1 to w_3. More generally,

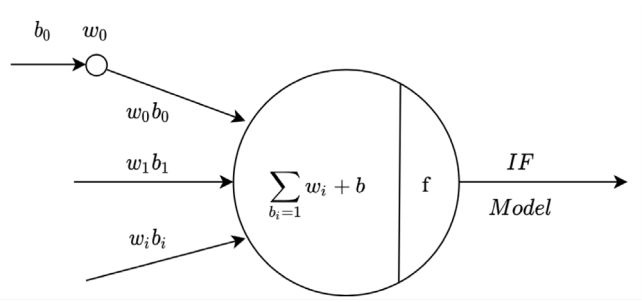


Fig. 5. Neuron of SNN.

we only need to sum up the weights (floating-point numbers) whose corresponding inputs are 1. Weights whose corresponding inputs are 0 can just be ignored. This simplification makes SNNs ideal for energy-constrained platforms.

The numerical EEG raw signals need to be encoding into binary spikes train for the input of the SNN neuron. Instead of a randomly spike rate encoding method, we propose a new input encoding algorithm to further reduce the accuracy drop, showed in Algorithm 1. First, we initiate a membrane voltage V and set it as 0. Then for each time step, we add the membrane voltage with input x. If V is larger than 1, we set the s at that time step as 1. Otherwise, we set it as zero. For example, if the input data is 0.5 and T is 10. Then the spike train becomes five zeros and five ones. Generally, each normalized numeric EEG signal is encoded as a spike train of length which includes $x_i T$ ones and (1 - x_i)T zero orderly. Then, the encoded spike trains are used as the first input layer of spiking neural networks.

Algorithm 1: Spike Input Encoding Algorithm.

Require: Input data $x \in [0, 1]$. Threshold 1
Ensure: Spike train **s** of length T.
  Membrane voltage $V = 0$
  for $i = 0$ to $T - 1$ **do**
    $V = V + x$
    if $V >= 1$ **then**
      $\mathbf{s}_i = 1$
      $V = V - 1$
    else
      $\mathbf{s}_i = 0$
    end if
  end for

For the length of time window T of both valence and arousal dimension, T is set as 80 and 40 respectively in our experiments.

As shown in Algorithm 2, we first train a CNN with both the clamp and quantization function. The weight and bias of each ANN layer are stored and transferred to its twin SNN layer. (l_s.weights = l_a.weights and l_s.bias = l_a.bias) After weight and bias sharing, each SNN layer l_s^1 receives an spike input $s^{l_s^{i-1}}$ from its previous SNN layer. (The spike input encoding method is shown in Algorithm 1.) Next, we obtain the spike output of l_{th} layer by sending $s^{l_s^{i-1}}$ to IF model (according to Eqs. (2)–(5)). Finally, we return the sum of all the spikes in the last SNN layer for classification.

According to Eq. (2), using the integrate-and-fire model in SNN, the spike rate of one neuron $|s_i|$ with respect to its input is

$$|s_i^l| = \lfloor \frac{\sum_j w_{i,j}^l |s_j^{l-1}| + b_i T}{\theta_i} \rfloor = \frac{\sum_j w_{i,j}^l |s_j^{l-1}| + b_i T}{\theta_i} - r_i^l \tag{14}$$

which will lead to a residual:

$$r_i^l = \frac{\sum_j w_{i,j}^l |s_j^{l-1}| + b_i T}{\theta_i} - \lfloor \frac{\sum_j w_{i,j}^l |s_j^{l-1}| + b_i T}{\theta_i} \rfloor$$

Such a residual will accumulate over the layers and cause a loss of accuracy from ANN to SNN. By using the clamp and quantization function in Section 4.2, a similar residual term $r' = x - Q(x)$ is introduced to the forward pass so that the activation becomes

$$x_i^l = Q_T(\sum_{j=0}^{n} w_{i,j}^l \cdot x_j^{l-1} + b_i^l) = \sum_{j=0}^{n} w_{i,j}^l \cdot x_j^{l-1} + b_i^l - r_i'^l \tag{15}$$

Omitting the expansion of equations, and taking $\theta_i^l = 1$, it is easy to derive that $r_i'^l \approx r_i^l$. Hence, in theory, without the accumulated error, the accuracy drop from ANN to SNN is near zero.

Algorithm 2: SNN Inference Algorithm .

Require: Spike encoded input **s**; ANN layers $\{l_a^1, \ldots, l_a^N\}$; SNN layers $\{l_s^1, \ldots, l_s^N\}$;Sum of Spike $s_{sum} = 0$
Ensure: Spike encoded input **s** of length T.
Transfer weight and bias of ANN to SNN:
l_s.weights = l_a.weights
l_s.bias = l_a.bias

Inference of SNN:
for $i = 1$ to N **do**
$\mathbf{s}^{l_s^i} = \mathbf{IF}(\mathbf{s}^{l_s^{i-1}})$ // Equation 2-5
end for
for $i = 1$ to T **do**
$s_{sum} = s_{sum} + \mathbf{s}^{l_s^N}[i]$
end for
Return s_{sum}/T

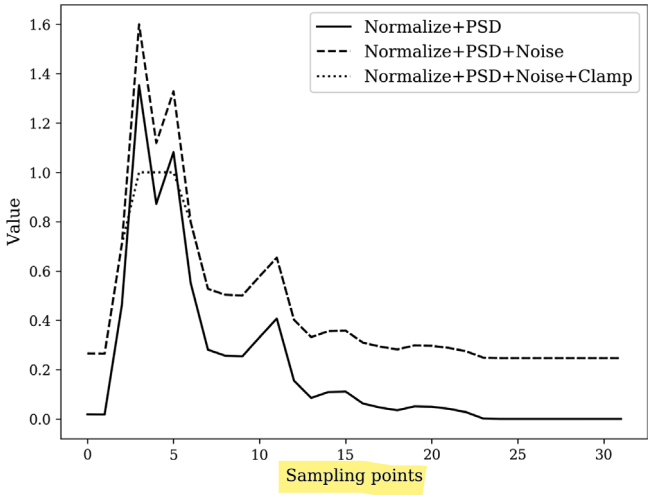


Fig. 6. Accuracy Change of Two-classes/Four-classes SNN.

5. Result and discussion

5.1. Dataset and evaluation platform

The DEAP dataset records 40 different channels comprising 32 electrodes and 8 peripheral channels. In our experiment, we only use data from electrodes placed on the scalp surface. As shown by Candra, et al. (2015), the best sampling window for valence and arousal is 3–12 s. As shown in Table 1, the 8064 sampling points represent one minute worth of EEG monitoring. So, for each channel, we divide the whole 8064 samples into 10 segments, each having about 807 sampling points. For each segment $\boldsymbol{x}$, PSD extraction is done after normalization. A scale factor of k and Gaussian noise G is used for better accuracy. As mentioned in Section 4, the input of an SNN is binary. Therefore, we added a clamp function to lower the accuracy loss from transfer learning (Yan et al., 2021). Mathematically,

$$x_i = \min(\max((kP(\frac{x_i - \boldsymbol{x}_{\min}}{\boldsymbol{x}_{\max} - \boldsymbol{x}_{\min}}) + G), 0), 1) \tag{16}$$

The preprocessed signal is showed in Fig. 6. We adapt the same method with Chen, et al. (2019), who uses 80% data for training and 20% for testing. We show the data pre-processing method of one segment in the volunteer one in Fig. 6, where Normalize indicates a normalization of raw data; PSD means Power Spectral Density explained in Section 4.1; Noise means Gaussian noise; Clamp function clamps the data to [0, 1].

The CNNs in this paper were implemented and trained using Pytorch on a server equipped with an Nvidia Tesla P100 card and tested in CPU mode on a laptop with a 2.6 GHz 6-Core Intel Core i7 CPU with integrated graphics and 8 GB of memory. Power was measured using the Intel Power Gadget. We took the difference in the system power when running the CNN and when the system was idle as the CNN power. This difference is 7.74 watts, which is still considerable. We then used Shenjing (Wang et al.,

Table 2
Confusion matrix of the CNN for arousal dimension.

	High arousal (HA)	Low arousal (LA)
HA	**725**	191
LA	227	**689**

Table 3
Confusion matrix of the SNN for arousal dimension.

	High arousal (HA)	Low arousal (LA)
HA	**756**	160
LA	270	**646**

Table 4
Accuracy from CNN to SNN for arousal dimension.

	HA		LA		Average	FLOPs
	Se	P+	Se	P+		
CNN	79.1%	76.2%	75.2%	78.3%	77.2%	7.74 W
SNN	82.5%	73.7%	70.5%	80.1%	76.5%	1.07 W

Table 5
Confusion matrix of the CNN for valence dimension.

	Positive valence (PV)	Negative valence (NV)
PV	**715**	201
NV	188	**728**

2020), a customized SNN accelerator architecture, to estimate the power of the converted SNN. The power for valence dimension and arousal dimension was estimated to be 1.07 watt. This is much smaller than the CNN-based solutions.

5.2. Classification for arousal dimension

For the arousal dimension, we first developed a CNN classification method according to Section 4.2, and transferred its weight to SNN as explained in Section 4.3. We present the performances of both convolutional neural networks and spiking neural networks here using *sensitivity* (Se), *positive predictivity* (P+) (Altman & Bland, 1994) and average accuracy. In the following equations, 'TP', 'FN', and 'FP' stand for 'true positives', 'false negatives' and 'false positives', respectively:

$$\mathrm{Se} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{17}$$

$$\mathrm{P+} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{18}$$

With the data preprocessed as described in Section 4.1, we trained the CNN until it converges in about 50 epochs. The scaleing factor was set as 3000 here. The confusion matrix of both CNN is showed in Table 2. Table 4 gives the Se and P+ of SNN. Correct predictions are highlighted in bold.

From Table 3, Se and P+ for high arousal detection is 82.5% and 73.7%, respectively. For low arousal detection, they are 70.5% and 80.1%, respectively. The average accuracy for arousal detection is 76.52%.

F1 scores of 77.62% and 77.86% were achieved for CNN and SNN of arousal dimension.

5.3. Classification for valence dimension

For classifying the valence dimension, we used a CNN shown in Fig. 3. We first trained the CNN for 50 epochs with a learning rate of 1, then decrease the learning rate to 0.1 and trained for another 10 epochs. Unlike the arousal dimension, the SNN designed for the valence dimension could achieve higher accuracy with a time window size of 40.

The weight of the CNN is transferred into a spiking neural network with a threshold of 1 in all layers. The scaling factor for data preprocessing is set as 3000. We present the confusion matrix along with the Se and P+ of CNN and SNN in Table 5 and Table 6.

F1 scores of 78.61% and 79.25% were achieved for CNN and SNN in the arousal dimension.

Table 6
Confusion matrix of the SNN for valence dimension.

	Positive valence (PV)	Negative valence (NV)
PV	**739**	177
NV	210	**706**

Table 7
Accuracy from CNN to SNN for Valence Dimension.

	PV		NV		Average	Power
	Se	P+	Se	P+		
CNN	78.1%	79.2%	79.5%	78.4%	78.77%	7.74 W
SNN	80.7%	77.9%	77.0%	80.0%	78.87%	1.07 W

Table 8
The classification accuracies under DEAP dataset using different methods.

	Methods	Valence	Arousal
Our 3-layer SNN	Pre-trained SNN	78.87%	76.5%
Our 3-layer SNN	Transfer learning + Pre-trained SNN	82.75%	84.22%
Our 5-layer SNN	Pre-trained SNN	79.70%	77.24%
Yuling (Luo, Fu, et al., 2020)	SNN	78%	74%
Tan (Tan, Šarlija, & Kasabov, 2021)	SNN	67.76%	78.07%
Abeer (Al-Nafjan et al., 2020)	SNN	84.62%	61.54%
Tripathi (Tripathi et al., 2017)	CNN	75.78%	73.12%
Mert (Mert & Akan, 2018)	KNN/ANN	72.87%	75.00%
Kumar (Kumar, Khaund, & Hazarika, 2016)	SVM	61.17%	64.84%

Table 9
Accuracy change of SNN for valence and arousal dimension.

Time steps	30	40	50	60	70	80
Arousal Acc(%)	74.23	75.05	75.54	75.81	76.1	76.52
Valence Acc(%)	77.29	78.87	78.98	79.14	79.03	78.98

5.4. Going deeper for SNN

In a bid to improve the accuracy of our models, we increased the number of layers of the SNNs from 3 layers to 5 layers. Two more convolutional layers with 512 and 1024 channels were added. The time window size was set to 800.

Furthermore, we explored the relationship between the arousal dimension and the valence dimension. By training a model based on one dimension and fine-tuning on another, we achieved an accuracy of 82.75% for the valence dimension and 84.22% for the arousal dimension using our 3-layer SNN with a time window of 80 (see Table 8) . We compared both our 3-layer SNN and 5-layer SNN with the start-of-art works, showed in Table 8. Our 5-layer SNN achieved an accuracy of 79.7% for the valence dimension and 77.24% for the arousal dimension. This shows that our framework supports a deeper structure as well.

As a reference, the best-known result for the same task on an ANN with an unlimited resource is 89.49% and 92.86% for valence and arousal dimension, respectively (Luo, Wu, et al., 2020). However, they used a complex and deep 10-layer network that combines convolutional layers, sparse autoencoder layers, and densely connected layers. It consumes significantly more power. As an indication, the average FLOP numbers of our SNNs for valence and arousal dimension are 1% and 2%, respectively, that of this 10-layer network. For other state-of-the-art studies related to the SNN-based EEG, Tan proposed a short-term emotion recognition framework based on spiking neural network (SNN) modeling of spatio-temporal EEG patterns. SNN accuracies of 67.76% and 78.07% were achieved in valence and arousal dimension respectively (Tan et al., 2021). Based on the NeuCube-based spiking neural network, Abeer achieved 84.62% SNN accuracy on valence dimension and 61.54% on arousal dimension using 60 EEG samples (Al-Nafjan et al., 2020). As far as we know, our SNNs on EEG outperform the current state-of-art SNNs.

5.5. Time window size analysis

The length of the binary spike train, i.e., the time window size, has a significant impact on the accuracy in SNNs. Generally, up to certain limits and subjected to the law of siminishing returns, larger time window sizes will lead to higher accuracies.

We profiled the accuracy of SNNs for both arousal and valence dimensions with different spike train lengths. The results are showed in Table 9. Using this result, we chose a spike train length of 80 for the SNN of arousal dimension and a spike train length of 40 for the SNN of valence dimension, as they give the best accuracy at a reasonable cost.

5.6. Validation on EEG brainwave feeling emotions dataset

We implemented and validated the proposed network architecture and SNN transfer learning method on one more publicly available EEG Brainwave Feeling Emotion dataset (Bird, Ekart, Buckingham, & Faria, 2019). We used the preprocessed EEG data provided by Feeling Emotions dataset (Bird, Faria, Manso, Ekárt, & Buckingham, 2019). 80% of the dataset was used for training and 20% for testing. Using the same network structure shown in Fig. 3, comparable accuracies were achieved for both CNN (96.87%) and SNN (95.67%).

6. Conclusion

In this paper, we have introduced a transfer learning method for implementing spiking neural networks for the task of emotion classification using EEG signals. We have tested it on the DEAP dataset and achieved an accuracy of 82.75% and 84.22% for valence and arousal dimensions, respectively. As far as we know, this outperforms the current state-of-art SNNs, and whilst using smaller time window sizes. The power of our SNNs for valence and arousal dimensions are 13.8% that of equivalent CNN-based solutions we implemented. This makes our SNN suitable for implementation on wearable devices and energy-constrained platforms.

CRediT authorship contribution statement

Zhanglu Yan: Conceptualization, methodology, software, data curation, writing – original draft, Writing – review & editing, visualization, investigation. **Jun Zhou:** Software, writing – review & editing, formal analysis. **Weng-Fai Wong:** Supervision,writing – review & editing, funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., & Adeli, H. (2018). Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Computers in Biology and Medicine*, *100*, 270–278.

Al-Nafjan, A., Alharthi, K., & Kurdi, H. (2020). Lightweight building of an electroencephalogram-based emotion detection system. *Brain Sciences*, *10*(11), 781.

Alhagry, S., Fahmy, A. A., & El-Khoribi, R. A. (2017). Emotion recognition based on EEG using LSTM recurrent neural network. *Emotion*, *8*(10), 355–358.

Altman, D. G., & Bland, J. M. (1994). Statistics notes: Diagnostic tests 2: predictive values. *The BMJ*, *309*(6947), 102.

Bengio, Y., Léonard, N., & Courville, A. (2013). Estimating or propagating gradients through stochastic neurons for conditional computation.

Bird, J. J., Ekart, A., Buckingham, C. D., & Faria, D. R. (2019). Mental emotional sentiment classification with an eeg-based brain-machine interface. In *Proceedings of theinternational conference on digital image and signal processing*.

Bird, J. J., Faria, D. R., Manso, L. J., Ekárt, A., & Buckingham, C. D. (2019). A deep evolutionary approach to bioinspired classifier optimisation for brain-machine interaction. *Complexity*, *2019*.

Candra, H., Yuwono, M., Chai, R., Handojoseno, A., Elamvazuthi, I., Nguyen, H. T., et al. (2015). Investigation of window size in classification of EEG-emotion signal with wavelet entropy and support vector machine. In *2015 37th Annual international conference of the IEEE engineering in medicine and biology society* (pp. 7250–7253). IEEE.

Chen, J., Zhang, P., Mao, Z., Huang, Y., Jiang, D., & Zhang, Y. (2019). Accurate EEG-based emotion recognition on combined features using deep convolutional neural networks. *IEEE Access*, *7*, 44317–44328.

Daimi, S. N., & Saha, G. (2014). Classification of emotions induced by music videos and correlation with participants' rating. *Expert Systems with Applications*, *41*(13), 6057–6065.

Gannouni, S., Aledaily, A., Belwafi, K., & Aboalsamh, H. (2021). Emotion detection using electroencephalography signals and a zero-time windowing-based epoch estimation and relevant electrode identification. *Scientific Reports*, *11*(1), 1–17.

Gao, F., Wu, T., Li, J., Zheng, B., Ruan, L., Shang, D., et al. (2018). SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis. *Computerized Medical Imaging and Graphics*, *70*, 53–62.

Kaiser, J., Mostafa, H., & Neftci, E. (2020). Synaptic plasticity dynamics for deep continuous local learning (DECOLLE). *Frontiers in Neuroscience*, *14*, 424.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., et al. (2011). Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing*, *3*(1), 18–31.

Kumar, N., Khaund, K., & Hazarika, S. M. (2016). Bispectral analysis of EEG for emotion recognition. *Procedia Computer Science*, *84*, 31–35.

Larsen, R. J., & Diener, E. (1992). *Promises and problems with the circumplex model of emotion*. Sage Publications, Inc.

Li, M., & Lu, B.-L. (2009). Emotion classification based on gamma-band EEG. In *2009 Annual international conference of the IEEE engineering in medicine and biology society* (pp. 1223–1226). IEEE.

Li, X., Song, D., Zhang, P., Yu, G., Hou, Y., & Hu, B. (2016). Emotion recognition from multi-channel EEG data through convolutional recurrent neural network. In *2016 IEEE international conference on bioinformatics and biomedicine* (pp. 352–359). IEEE.

Luo, Y., Fu, Q., Xie, J., Qin, Y., Wu, G., Liu, J., et al. (2020). EEG-based emotion classification using spiking neural networks. *IEEE Access*, *8*, 46007–46016.

Luo, Y., Wu, G., Qiu, S., Yang, S., Li, W., & Bi, Y. (2020). EEG-based emotion classification using deep neural network and sparse autoencoder. *Frontiers in Systems Neuroscience*, *14*, 43.

Mert, A., & Akan, A. (2018). Emotion recognition from EEG signals by using multivariate empirical mode decomposition. *Pattern Analysis and Applications*, *21*(1), 81–89.

Plutchik, R., Kellerman, H., & Conte, H. R. (1979). A structural theory of ego defenses and emotions. In *Emotions in personality and psychopathology* (pp. 227–257). Springer.

Sakhavi, S. (2017). *Application of deep learning methods in brain-computer interface systems* (Ph.D. thesis), National University of Singapore (Singapore).

Salama, E. S., El-Khoribi, R. A., Shoman, M. E., & Shalaby, M. A. W. (2018). EEG-based emotion recognition using 3D convolutional neural networks. *International Journal of Advanced Computer Science and Applications*, *9*(8), 329–337.

Shrestha, S. B., & Orchard, G. (2018). Slayer: Spike layer error reassignment in time. In *Advances in neural information processing systems* (pp. 1412–1421).

Suhaimi, N. S., Yuan, C. T. B., Teo, J., & Mountstephens, J. (2018). Modeling the affective space of 360 virtual reality videos based on arousal and valence for wearable EEG-based VR emotion classification. In *2018 IEEE 14th international colloquium on signal processing & its applications* (pp. 167–172). IEEE.

Tan, C., Šarlija, M., & Kasabov, N. (2021). NeuroSense: Short-term emotion recognition and understanding based on spiking neural network modelling of spatio-temporal EEG patterns. *Neurocomputing*, *434*, 137–148.

Tripathi, S., Acharya, S., Sharma, R. D., Mittal, S., & Bhattacharya, S. (2017). Using deep and convolutional neural networks for accurate emotion classification on DEAP dataset. In *Proceedings of the thirty-first AAAI conference on artificial intelligence* (pp. 4746–4752).

Vigneron, A., & Martinet, J. (2020). A critical survey of STDP in spiking neural networks for pattern recognition (preprint).

Wang, B., Zhou, J., Wong, W.-F., & Peh, L.-S. (2020). Shenjing: A low power reconfigurable neuromorphic accelerator with partial-sum and spike networks-on-chip. In *2020 Design, automation & test in europe conference & exhibition*. IEEE, [Online]. Available: https://arxiv.org/abs/1911.10741.

Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, *15*(2), 70–73.

Yan, Z., Zhou, J., & Wong, W.-F. Energy efficient ECG classification with spiking neural network. *Biomedical Signal Processing and Control*, *63*, Article 102170.

Yan, Z., Zhou, J., & Wong, W.-F. (2021). Near lossless transfer learning for spiking neural networks.

Zhang, X.-S., Roy, R. J., & Jensen, E. W. (2001). EEG complexity as a measure of depth of anesthesia for patients. *IEEE Transactions on Biomedical Engineering*, *48*(12), 1424–1433.