FINAL YEAR PROJECT REPORT

University of Birmingham

Project title: EEG-based Emotion Recognition using Multi-Channel Images with ResNet50

Author: Minchang Chung

Student ID: 1906881

Programme name: Artificial Intelligence and Computer Science

Project Supervisor: Hamid Dehghani

Word count: 4220

Abstract

Emotion recognition has emerged as a crucial topic in the field of human-computer interaction (HCI). Recently, with the development of deep neural networks, attempts to learn to predict human emotions are increasing. Among them, electroencephalography (EEG) got attention because there is no fake expression compared to other data such as facial images and gestures. Deep neural networks have proved promising results in EEG signal classification for brain-computer interfaces (BCIs). However, current studies often use one-dimensional EEG features, which fail to capture local information. In this paper, we aim to investigate EEG-based emotion recognition using ResNet50 by converting 1-dimensional EEG features into 2-dimensional images and combining them into multi-channel images. We proposed a methodology to generate EEG images based on EEG electrode placement. And EEG images were combined in four different approaches, two approaches include spatial features, and the other two include spatial and temporal features. Using these generated images, we trained ResNet50 to perform emotion recognition for emotion recognition. However, the experimental results were not successful. As a result of learning two images with only spatial features and two images with both spatial and temporal features, images with only spatial features showed 38% and 54% accuracy, but images with both features showed an accuracy of 29% and 39%. The reasons for this failure were that the model did not properly learn the temporal characteristics of EEG data and problems that occurred in the pre-processing process. These failures represent limitations and problems in emotion recognition research using EEG data. In the future, it is considered that improvements in the pre-processing method of EEG data and the design of the model are needed.

Contents:

1. Introduction

Emotion recognition has been frequently studied in recent decades to make a machine to classify human emotions. Human emotion data are commonly collected from facial expressions, gestures and voices. However, some of these data could be fake expressions because of their nonbiological signals [1]. To avoid this problem, researchers started to use physiological measurements such as functional resonance imaging(fMRI), electroencephalography (EEG), positron emission tomography(PET), or magnetoencephalography(MEG) which shows objective results compared to nonbiological signals.

There are two categories to study Emotion recognition: Multimodal and single-modal. Multimodal emotion recognition approaches combine two or more physiological and non-physiological aspects [37]. Liu et al. [21] proposed Multimodal emotion recognition on various datasets and achieved 85.3% on the SEED-V data set. Single-modal approaches generally use only one type of physiological signal. EEG is the most widely used in single-modal emotion recognition studies as it carries information about neural activities that underlies almost all other physiological and non-physiological reactions. In this study, we will conduct a single-modal emotion analysis using EEG data.

Electroencephalography (EEG) measures the potential difference between electrodes attached to a specific location on the scalp where the electrical signal is generated by neural activity in the brain, including the ability to respond quickly to external stimuli [2]. EEG has the advantage of being non-invasive and less expensive than other physiological measurements. After measurement, EEG signals are processed using signal processing techniques such as a bandpass filter to remove noise and artefacts. Then feature extraction is performed through statistical methods such as power spectral density (PSD) and differential entropy(DE). It is often used because it is the simplest method of measuring brain activity, but it is sensitive to the measurement environment, subjective to individuals, and has limitations in spatial resolution compared to other techniques in measuring neural activities such as MEG, fMRI, and PET.

Electroencephalogram (EEG) signals have been tested for emotion recognition through the application of machine learning methods such as SVM [12], ANN [13] etc., to learn the linearity or nonlinearity of the EEG dataset. In recent years, researchers have also investigated the use of deep neural networks (DNNs) [4], including recurrent neural networks (RNNs) [14] and convolutional neural networks (CNNs) [9], for EEG-based emotion recognition. One major challenge with applying CNNs to traditional EEG features is their lack of direct compatibility with the CNN structure, requiring further modifications and pre-processing steps for effective utilization.

In contrast to typical image classification tasks, EEG signals are time-frequency data that require consideration of the temporal information in the input to effectively capture the dynamics of the data [15]. Neglecting this critical feature can limit the performance of emotion recognition models, as the emotional state of a person is not solely determined by the current signal but rather by the past and future signals as well. Therefore, it is imperative to develop methodologies that can effectively capture the temporal dynamics of EEG signals, such as Long Short-Term Memory (LSTM) networks [15].

The objective of this paper is to explore different Multi-Channel EEG image inputs trained in ResNet50. Here, we introduce the methodology to visualize EEG signal to trainable image and 4 different approaches to generate input images to train the model: (1) a single brainwave channel (1-channel), (2) a combination of five brainwave channels (5-channel), and (3) stacking of single-channel or 5-channel images in chronological order (13-channel, 65channel). By adding time sequence features as channels in EEG images, we expect to handle sequential data without adding recurrent neural networks (RNNs) and be able to capture both spatial and sequential characteristics of EEG.

2. Background and Related Work

Electroencephalography (EEG) signals are widely used in various applications for analysing brain activity [3]. To ensure the accuracy of EEG signal analysis, proper pre-processing is critical. Cheah et al. [37] investigated the possibility to train plain emotion-related EEG signals on a convolutional neural network and conclude that plain EEG signals did not show superior performance compared to feature-based algorithms with the current size of publicly available EEG datasets.

The pre-processing of raw EEG signals typically involves several stages of signal processing, including filtering, artefact removal, and feature extraction [3]. Filtering is used to eliminate unnecessary frequency components in the raw EEG signals. The artefact removal stage removes unrelated signals such as eye blinks or muscle artefacts to ensure that only EEG signals are analysed. Finally, feature extraction is done to extract relevant features from the pre-processed EEG signal, and extracted features are used for further analysis.

Several methods and techniques of pre-processing to remove noise and artefacts have been proposed in the literature to improve the quality of measured EEG signals such as ICA [30], LDS [18], Wavelet Transform [31] etc. These EEG pre-processing techniques can significantly affect the accuracy and authenticity of the derived results. Jiang et al. [34] investigated various techniques of noise and artefact removal and demonstrated the advantages and disadvantages of each technique. They concluded that there is no optimal technique for noise and artefact removal. Therefore, it is crucial to employ appropriate pre-processing techniques to analyse EEG signals and obtain valuable insights into brain function.

2.1 EEG Feature Extraction

Electroencephalogram (EEG) feature extraction is a critical step for analysing EEG. There are various methods to extract features from EEG and two commonly used EEG feature extraction methods are power spectral density (PSD) and differential entropy (DE).

PSD is a measure of the power of a signal at different frequencies, computed by taking the Fourier transform of a signal and squaring the magnitude of the resulting complex numbers [19]. It has been widely used for EEG-based emotion recognition as work done by Zheng et al. [4], which utilized PSD to extract features from EEG signals for emotion recognition.

Differential entropy, on the other hand, is a measure of the randomness or uncertainty of a signal, computed by integrating the negative logarithm of the probability density function of a signal [38]. DE has also been extensively used for EEG-based emotion recognition, as seen in the work by Zheng et al. [4], which employed DE to extract features from EEG signals for emotion recognition and showed improvements compared to PSD features.

However, both PSD and DE features are represented as one-dimensional vectors which have limitations in capturing the topology information of the neural activity, which is essential for understanding brain function and emotion regulation.

2.2 EEG emotion recognition using Deep learning

Various studies on EEG-based emotion recognition are done on traditional machine learning techniques such as PCA [16], QDA [17], and LDA [18] because of the lack of large and available datasets to train deep learning. One of the major challenges in EEG-based emotion recognition using deep learning is the limited availability of large datasets. Additionally, assigning emotional labels to large datasets is a critical problem, and there are two standard approaches for doing so; valence-arousal dimension, and fine emotional labels such as neutral, joy, anger, sadness, etc.[4]

Despite the limitation of the lack of a large dataset, in recent studies, deep learning techniques have demonstrated promising results in EEG-based recognition tasks, particularly in emotion recognition. Wang et al. [5] addressed the issue of data shortage in EEG-based emotion recognition by proposing a data augmentation method. They compared the performance of three different models, namely SVM [20], ResNet [6], and LeNet [7]. However, their method did not address the issue of low resolution in EEG data with topology information according to the recorded location of each channel. Moreover, despite using a deep learning approach, their method showed poor performance in emotion recognition.

One approach to handling temporal aspects of EEG is to use recurrent neural networks (RNNs) or variants of RNNs such as long short-term memory (LSTM) networks [27]. Another approach to handle sequential aspects is to use CNNs with a 1D convolutional layer, which can be applied to time-series data directly [28]. Cheah et al.[32] proposed ResNet18 model with 1d-kernel to train spatial dimension and temporal dimension separately. However, these approaches don't consider the spatial characteristics and temporal characteristics of EEG signals simultaneously. Bashivan et al. [10] proposed a method for classifying EEG signals using a deep neural network inspired by computer vision techniques. They generated EEG images that preserved topology by utilizing PSD features and trained a convolutional neural network (CNN) to learn spatial features. To learn temporal information, they used a long-short-term memory (LSTM) model. They demonstrated that their deep neural network achieved higher accuracy than traditional classifiers in a working memory task. However, it should be noted that the EEG dataset must contain enough temporal information to successfully learn sequential information.

3. Dataset

In this part, we explain detailed information about the EEG dataset used in this experiment. There are several EEG datasets for emotion recognition, and we have chosen the SEED-V dataset which has 5 emotion labels (happy, sad, fear, disgust, and neutral). SEED-V dataset is originally designed to study Muti-modal emotion recognition and contains Eye-Movement data and EEG signal data. As our study aimed to focus on EEG-based emotion recognition, we did not consider the Eye-Movement data.

3.1 SEED-V dataset

SEED datasets are well designed EEG datasets for emotion classification. Among various SEED datasets, the SEED-V dataset [8] classified EEG data into 5 labels (happy, sad, disgust, fear, and neutral) which have more labels compared to other datasets. A 62-channel ESI NeuroScan System was used to collect EEG data from 16 subjects (10 female and 6 males) and collected raw EEG data were downsampled to 200 Hz. There are 15 emotional movie clips (three clips X five emotion labels) that intensify the emotion of each participant. Each participant watched fifteen movie clips three times with at least three days between experiments, a total of 45 experiments were recorded. Fig.1 shows the specific process of how raw EEG signals were collected.

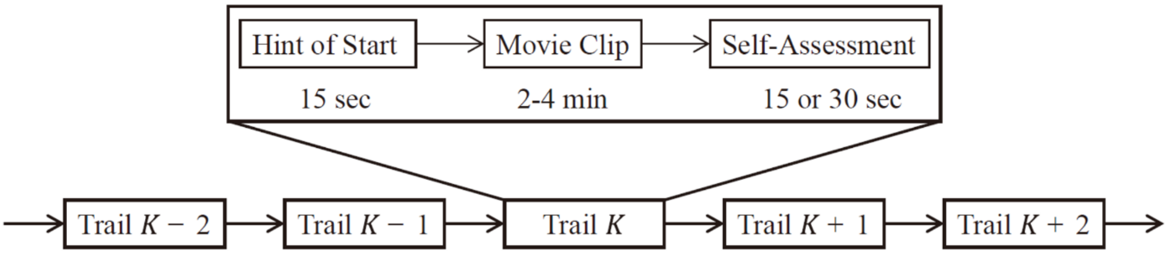


Fig.1. Experiment protocol

4. Preprocessing

The main objective of this project is to classify emotions using Multi-channel EEG images and ResNet50. The EEG dataset used in this study had already provided extracted features that had been pre-processed by applying a bandpass filter between 1 Hz and 75 Hz to remove noise and artefacts. The Differential Entropy (DE) feature was then extracted across five frequency bands: delta (1-3 Hz), theta (4-7 Hz), alpha (8-13 Hz), beta (14-30 Hz), and gamma (31-50 Hz) using the Short-Time Fourier Transform (STFT) with a non-overlapping 4-second Hanning window. DE has been presented in numerous studies as a consistent feature that has demonstrated exceptional performance in EEG signal analysis [4][9][21]. Given that the EEG signal where the time series ***X*** follows a Gaussian distribution N(μ, σ2), DE can be straightforwardly defined in Fig.2 [26].

$$\begin{aligned} h(x) &= -\int_{\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) dx \\ &= \frac{1}{2}\log(2\pi e\sigma^2) \end{aligned}$$

Fig.2. DE equation

4.1 Feature map

As mentioned in the introduction, EEG data is a continuous signal that measures voltage changes over time [2]. The extracted EEG features are represented as one-dimensional and a two-dimensional coordinate system is required to image this, but if proper imaging methods are not applied, high-dimensional information such as the neural activity of the entire brain may be lost.

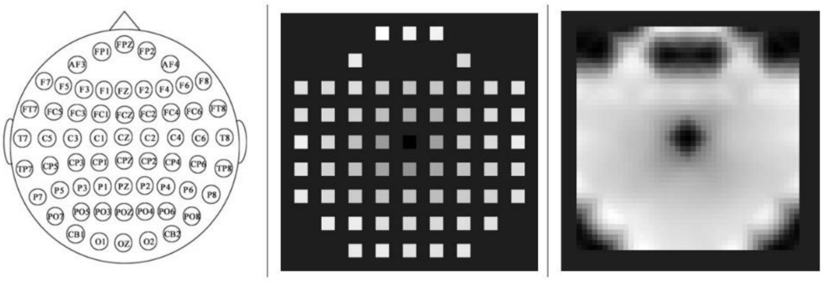


Fig.3. 10-20 electrode placement of SEED dataset and the feature map to convert EEG to image (left to right)

In the SEED-V dataset, with reference to the 10-20 electrode placement system, points of the electrodes in the feature map were set corresponding to each channel in order to visualize one-dimensional vectors as a 2D image, as depicted in Fig.3. Subsequently, the values for each channel were allocated at their respective points. Prior to the interpolation process, the entire dataset underwent normalization using the global mean and standard deviation (6.404 and 2.272, respectively) to adjust the scales of each participant. This step facilitated better comparability and consistency among the features. Following the normalisation of the dataset, the Clough-Tocher cubic interpolation between the scattered points was performed to fill in the empty spaces. As proposed by Hwang et al. [9], the Clough-Tocher cubic interpolation method demonstrated superior results when compared to alternative interpolation techniques such as the nearest neighbour, and linear interpolation.

4.2 Approaches

From the proposed methodology, we generated Multi-channel EEG images. Specifically, we created 1-channel images using a single waveband, 5-channel images by combining 5 different wavebands, 13-channel images by stacking 1-channel images over time, and 65-channel images by stacking 5-channel images over time. The rationale behind combining 5 wave bands into 5 channels was to reflect the brain's overall activity by capturing the activity in different brain regions where each wave band originates. Moreover, to illustrate the time dynamics of the EEG signal, we connected one-channel and five-channel images along a time dimension in which the minimum length of the video clip is 13, respectively, to generate 13 and 65-channel images. This allowed us to capture the temporal information of the EEG signal in the input. The proposed method offers an effective approach to capturing both spatial and temporal features of EEG signals, which are essential for accurate emotion recognition.

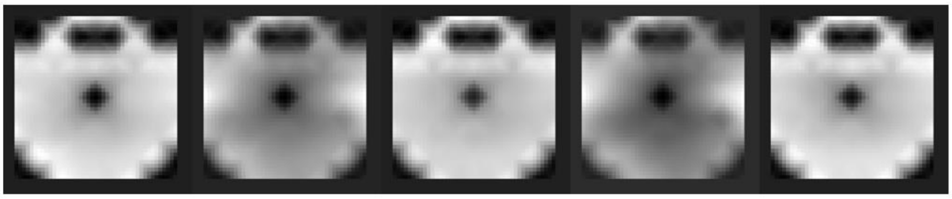


Fig.4. Generated EEG images using DE features of one participant. The five images represent five frequency bands; alpha, beta, delta, gamma, and theta (left to right)

The generated EEG images shown in Fig.4 have a resolution of 50(w) X 50 (h) X 1(N), 50(w) X 50 (h) X 5(N), 50(w) X 50 (h) X 13(M), and 50(w) X 50 (h) X 65(N x M); w and h are the width and height of generated images, N is the number of frequency bands, M is the minimum length of 15 movie clips. We named images 1ch, 5ch, 13ch, and 65ch respectively.

5. ResNet50

Kaiming et al. [6] introduced the ResNet architecture, which has exhibited outstanding performance in a range of image recognition tasks. The fundamental concept of residual learning involves incorporating residual connections (skip connections) to tackle the vanishing gradient issue that occurs in traditional convolutional neural networks (CNNs). ResNet-50 is composed of 50 layers including bottleneck layers which have the advantage of reducing computational complexity and the number of parameters in the network.

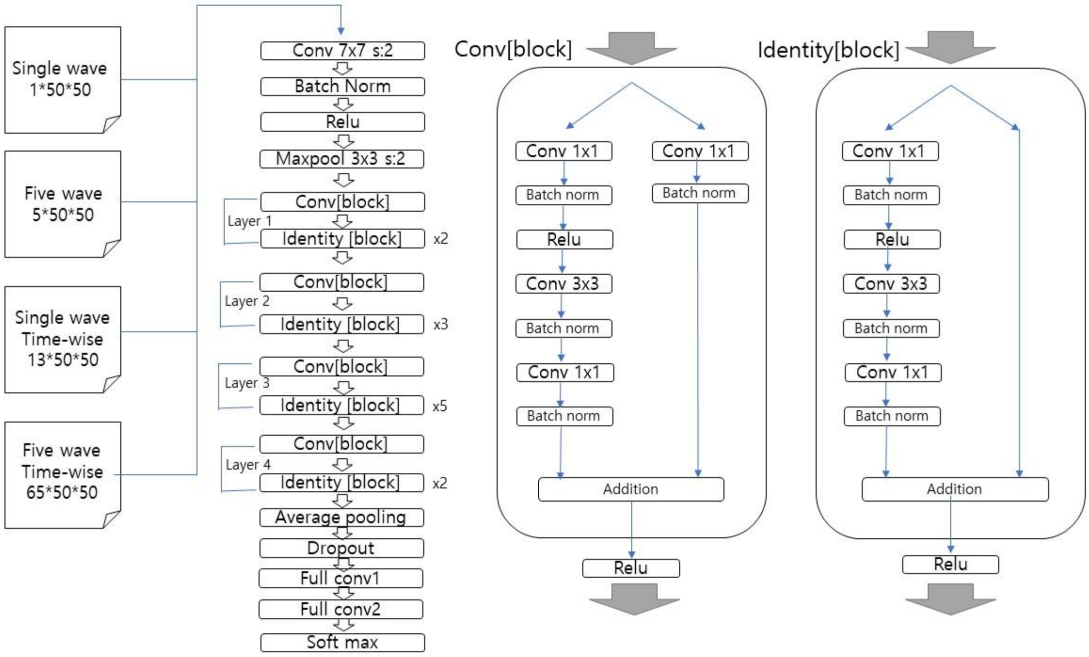


Fig.5. Our ResNet50 architecture

After generating SEED-V EEG images, we learn ResNet50 with the SoftMax classification layer which is shown in Fig.5**.** The convolutional kernels in ResNet50 are all 2-dimensional kernels which have 3 by 3 kernels and 1 by 1 kernels throughout the whole process except for the first convolutional layer which has 7 by 7 kernels. Layer 1 to 4 illustrated in Fig.5 represents the Bottleneck layers. We set the Dropout layer to avoid overfitting. Two Fully Connected layers are implemented after the Dropout layer with five output nodes in the last layer which correspond to the five emotion classes. At last, the network ends with the SoftMax classification layer that demonstrates strength in multi-class classification.

In our implementation of ResNet50, we modified the first convolutional layer by reducing the number of input channels from 64 to 32. Specifically, we set the 'in_channels' parameter to 32 in the constructor of the ResNet50 class. The remaining layers of the network were kept unchanged from the original ResNet50 architecture except for input dimensions. This modification was made to reduce the computational complexity of the model while maintaining its overall performance.

The experiments are implemented using PyTorch (1.13.1) and learned the ResNet50 by using Adam optimizer (learning rate: 0.01, Batch size: 16). The input and output dimensions are presented in Table.1. We used matplotlib to make confusion matrixes and visualize features.

Layer	Type	Input dimension	Output dimension
1	Convolution	1, 5, 13, 65	32
2	Batch normalize	32	32
3	Relu	32	32
4	Maxpooling	32	32
5	Layer1	32	64
6	Layer2	64	128
7	Layer3	128	256
8	Layer4	256	512
9	AvgPool	512	256
10	Dropout	256	256
11	Fully Connected 1	256	50
12	Fully Connected 2	50	5
13	SoftMax	5	5

Table.1. The dimensions of our ResNet50 model

5.1 Experimental Setup

As previously mentioned, the SEED-V dataset comprises a total of 45 experimental records per subject, stemming from the combination of 3 videos per label, 5 labels, and 3 repetitions. However, the specific order of these records remains unknown [8]. Despite this, the records were organized into three groups: the first 5, middle 5, and last 5 records of each session. For the purposes of this study, two of these groups were designated as the training set, while the remaining group was utilized as the test set. It is important to note that label distribution was unbiased across all groups. In our experimental design, the first two sessions were assigned as the training set, and the final session was designated as the test set.

6. Results and Discussion

6.1 Result

Classifier	Dataset	Labels	Feature	Channels	ACC(%)
SVM [4]	SEED	3	DE	62	83.99
DBN [4]	SEED	3	DE	62	86.08
CNN [9]	SEED	3	TP-DE	62	90.41
SVM [8]	SEED-V	5	DE	62	69.50
ResNet(ours, 1ch)	SEED-V	5	TP-DE	62	38
ResNet(ours, 5ch)	SEED-V	5	TP-DE	62	54
ResNet(ours, 13ch)	SEED-V	5	TP-DE	62	29
ResNet(ours, 65ch)	SEED-V	5	TP-DE	62	39

Table.2. The comparison with different classifiers

We evaluated our emotion recognition methods by comparing other classifier results. Table.2 summarizes the best accuracy of our emotion recognition results compared with other classifiers. The results showed that 5ch images showed better accuracy (54%) compared to our other experiments. However, these results show that our methods didn't show improvement compared to other classifiers.

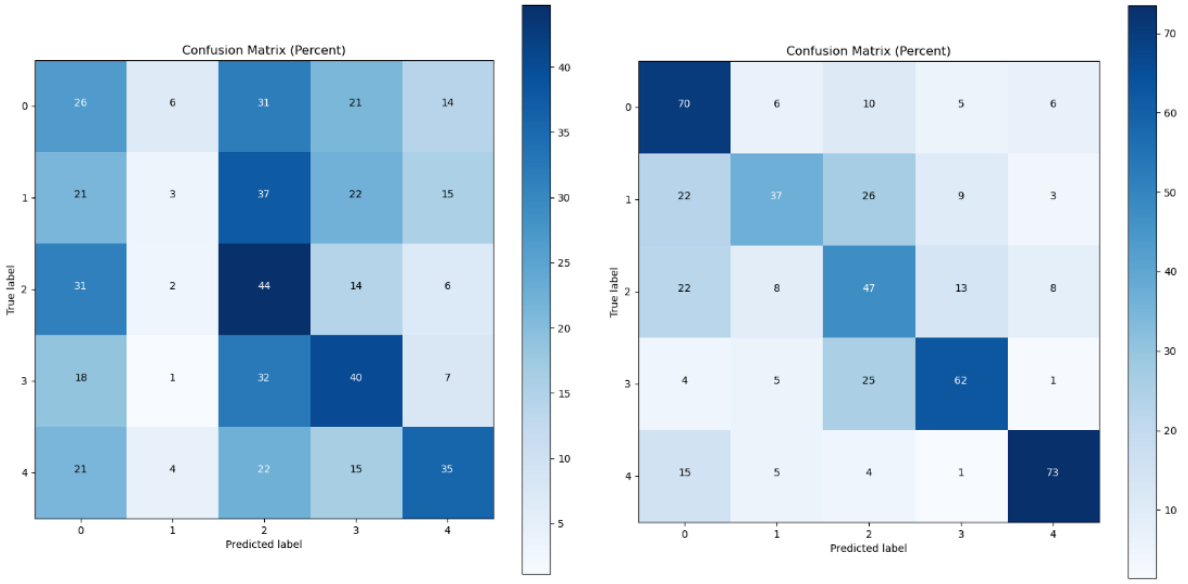


Fig.6. Confusion Matrix of our result (1ch, 5ch image, left to right). Each label corresponds to the following emotions: 0: disgust, 1: fear, 2: sad, 3: neutral, and 4: happy.

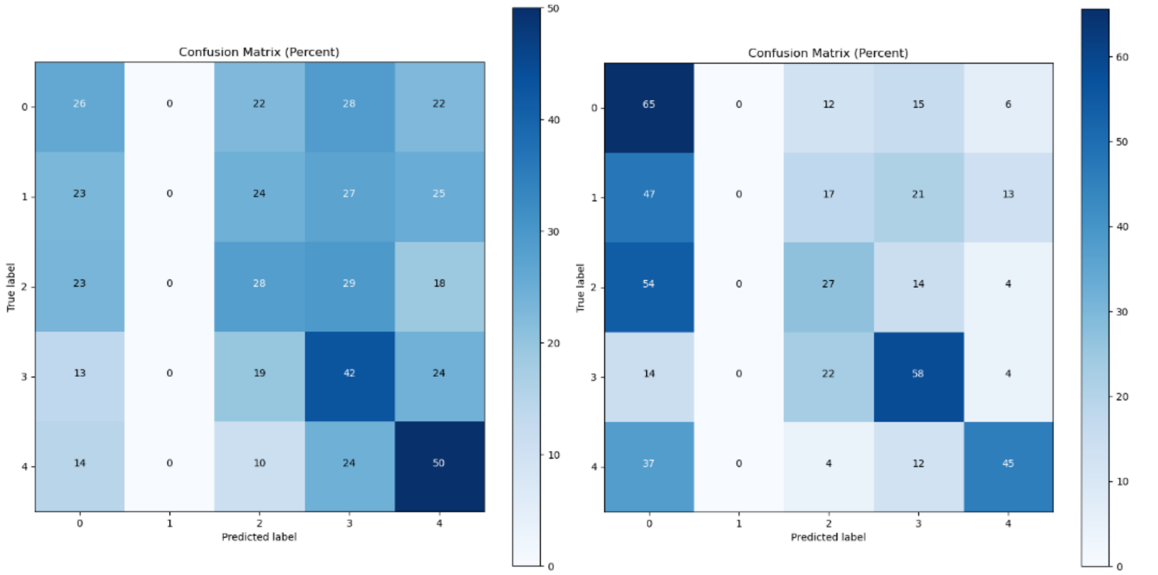


Fig.7. Confusion matrix of our result (13ch, 65ch image, left to right). Each label corresponds to the following emotions: 0: disgust, 1: fear, 2: sad, 3: neutral, and 4: happy.

The confusion matrix for the single-wave band shows an overall poor predictive performance. On the other hand, the five-wave band exhibits better predictive performance than the single-wave band, particularly for the emotions of disgust, neutral, and happy showed 70%, 62%, and 73% respectively. The confusion matrix of the 13-channel showed 42% and 50% accuracy, and the 65-channel showed 54% and 45% accuracy in predicting neural and happy emotions. Compared to the 13-channel results, the 65-channel results showed 65% accuracy in the emotion of disgust and made better predictions. However, when considering the inclusion of temporal features into the spatial dimension, both the single and five-wave band models demonstrate an inability to predict the emotion of fear.

6.2 Discussion

In our experiment, we employed a proposed methodology to transform EEG features into multi-channel 2D images and subsequently trained a ResNet50 model on these images. Despite the experimental results not meeting our expectations, we believe this could be attributed to limitations in the dataset, our model, and the pre-processing procedure.

The reason of 1-channel and 13-channel results for emotion recognition showed relatively unsuccessful prediction compared to the other results could be the fact that certain neural patterns associated with emotions vary depending on the specific frequency band. Wang et al.[4] investigated neural patterns associated with positive, negative, and neutral emotions using DE features extracted from five frequency waves. They found that the energy of beta and gamma frequency bands increased on positive emotions and decreased on the other emotions. They also concluded that these two frequency bands showed better results for emotion recognition. In our experiments, we performed emotion recognition in all five frequency bands without such differentiation, which may have contributed to the poor prediction of both single-channel and 13-channel images.

The dataset we utilized included repeated stimuli, which may have led to a reduction in neural activity and hindered the model's ability to learn consistent patterns for each emotional label [11]. Furthermore, the dataset was limited to 16 subjects, resulting in a relatively small sample size, even though we increased its size by converting the EEG signal into images. To overcome the lack of public EEG datasets, Bo et al. [29] demonstrated generative adversarial networks (GANs) for data augmentation and improved the performance of emotion recognition based on the deep learning model.

One potential limitation of the dataset relates to the emotion labels employed. Previous emotion recognition studies have typically used either valence-arousal labelling or 3-label classification (positive, negative, neutral) for emotion recognition tasks. Hwang et al. [9] reported an accuracy of 90.41% on the SEED dataset with three labels. However, the SEED-V dataset incorporated five labels: happy, sad, fear, disgust, and neutral. The labels sad, fear, and disgust can be grouped as negative emotions, diverging from prior labelling schemes. We observed relatively poor classification results for sad, fear, and disgust emotion labels from Fig.6,7, which suggests that the labelling of emotional states in the dataset may have impacted the performance of our model.

Another possible contributing factor to our model's suboptimal performance is the image generation process. We employed the SEED-V dataset's electrode placement and evenly scattered points to create a feature map when converting EEG signals into images. However, failure to accurately plot the distances between each electrode may result in a loss of spatial information during EEG image generation. Bhavsar et al. [22] demonstrated the importance of preserving topological information by presenting the EEG features on the actual location of electrodes. This approach ensures the correlation of the actual location between the EEG signal and maintains the topology information of the signal.

In addition, the normalisation process may also influence the suboptimal classification results. In our study, we normalised the scale of each participant's EEG data using global mean and standard deviation, which was intended to ensure consistency and comparability. However, this normalisation process may have resulted in a loss of unique neural patterns for each participant. Therefore, it is necessary to investigate whether normalisation of EEG data is required in EEG-based emotion recognition tasks, and if necessary, further investigation is required to determine the optimal normalisation method, which can improve the classification performance.

Despite incorporating temporal features into the 3D image input using our proposed method for generating spatial features, our ResNet50[6] model may not have effectively learned the temporal features of the EEG data. This could be due to the fact that the ResNet50 architecture, which is primarily designed for spatial feature learning, may not be equipped to handle the complexity of temporal dynamics present in EEG-generated images, leading to suboptimal performance in our study. Previous studies have addressed this issue by selecting models that incorporate temporal features, such as CNN-LSTM or 3D CNN, to learn the spatial and temporal features of EEG data. For example, Bashivan et al. [10] proposed CNN-LSTM to investigate the spatial and temporal features of EEG data and achieved 90% accuracy. Salama et al. [23] proposed a 3D Convolution Neural Network (3D-CNN) to investigate multi-channel EEG data for emotion recognition and achieved 88.49% accuracy on the DEAP dataset.

Lastly, our study's suboptimal results could be due to the use of a bandpass filter for noise and artefact removal in the EEG data [24]. Bandpass filters, which allow specific frequency ranges to pass through while blocking remaining signals, are commonly used to extract frequency ranges of interest in EEG data processing. Nonetheless, this approach may not adequately address all noise and artefact issues present in the data. Consequently, alternative noise and artefact removal techniques, such as Linear Dynamic Systems (LDS) [18] and Independent Component Analysis (ICA) [25], could improve results and enhance our model's overall performance. Future research should explore these alternative methods to more effectively address the challenges posed by noise and artefacts in EEG data.

7. Conclusion

Emotion recognition has emerged as a crucial topic in the field of human-computer interaction (HCI). In particular, EEG-based emotion analysis has acquired attention due to its objective nature compared to facial images, body gestures, and vocal cues. Additionally, EEG measurement offers relatively easier access compared to other brainwave measurement techniques, such as fMRI.

In this study, we proposed Multi-channel EEG image generation techniques for EEG-based emotion recognition using ResNet50. Among several EEG feature extraction methods, DE features yielded the most promising results [4,8,9]. To train the model, we utilized the ResNet50 architecture, which incorporates residual connections to reduce gradient vanishing/exploding issues commonly encountered in traditional CNN approaches. To adapt the one-dimensional EEG DE features for training, we generated images based on electrode positions and employed the Clough-Tocher cubic interpolation method [10], which demonstrated the best performance in filling the gaps between electrodes. We combined the generated images into 1-channel, 5-channel, 13-channel, and 65-channel images which represent single waveband, five wavebands, time-wise stacked single wavebands, and time-wise stacked five wavebands. Finally, we evaluated and compared our proposed methodology with other classifiers.

Our study has revealed that the ResNet50 architecture employed in our experiment was unable to effectively capture the temporal characteristics of the EEG data. Moreover, our proposed methodology did not consider the inter-electrode distance, resulting in a certain degree of spatial feature loss [22]. Furthermore, the pre-processing method in this experiment which used a bandpass filter to remove noise and artefacts was not optimal to obtain relevant frequency from the original EEG signal. In addition, the SEED-V dataset utilized in our study consisted of 15 videos repeated three times, resulting in a total of 45 experimental results that may have led to suboptimal performance because repeated exposure to the same video contributes to a reduction in neural activity [11]. These limitations, including the experimental design, the chosen model, and the constraints of the dataset, have resulted in an unsuccessful outcome.

Based on the results of our study, our model ResNet50 didn't show good results in terms of the spatial aspect of EEG and was unable to learn temporal aspects of EEG data. This is because the ResNet50 model does not learn the temporal aspects, but also it is speculated that the EEG pre-processing has contributed to the decrease in overall accuracy.

In the future study, we will study more accurate EEG-based emotion recognition considering the problems raised. First, we plan to pre-process raw EEG signals directly instead of using pre-processed data. There are many pre-processing techniques, but we will perform noise and artefact removal using Linear Dynamics Systems (LDS) [18], which is often used recently in the field of emotion recognition and extracting DE features. Based on this pre-processed data, we will generate an image considering the electrode distances in topology. ResNet50 has advantages in learning spatial features, but attempts to incorporate temporal features into spatial features have failed, so we plan to build a 3D ResNet model by adding time dimension without concatenating temporal features into channels to potentially improve the performance of emotion recognition tasks.

By addressing the limitations and challenges identified in this study based on these plans, we plan to improve the overall performance of EEG-based emotion recognition. If these plans do not improve performance, we will explore other pre-processing methods such as Wavelet Transform [31] and alternative models such as CNN-LSTM [35], and Vision Transformer (ViT) [36] to further investigate and implement ways to capture both temporal and spatial features. To address the shortage of datasets, we will look for additional EEG emotion recognition datasets to train the models. Additionally, another methodology called computer-generated Holography can be investigated for visualising 1D EEG data to the 2D image. Topic et al. [33] compared topographic feature map and holographic feature map representation of EEG signals and proposed that holographic feature map showed better results.

Reference

[1] Saxen, F., Werner, P. and Al-Hamadi, A. (2017). Real vs. Fake Emotion Challenge: Learning to Rank Authenticity from Facial Activity Descriptors. *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*. doi:https://doi.org/10.1109/iccvw.2017.363.

[2] Ernst Niedermeyer, Schomer, D.L. and Silva (2011). *Niedermeyer's electroencephalography : basic principles, clinical applications, and related fields*. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health.

[3] Saeid Sanei and Chambers, J.A. (2013). *EEG Signal Processing*. John Wiley & Sons.

[4] Wei-Long Zheng and Bao-Liang Lu (2015). Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. *IEEE Transactions on Autonomous Mental Development*, 7(3), pp.162–175. doi:https://doi.org/10.1109/tamd.2015.2431497.

[5] Wang, F., Zhong, S., Peng, J., Jiang, J. and Liu, Y. (2018). Data Augmentation for EEG-Based Emotion Recognition with Deep Convolutional Neural Networks. *MultiMedia Modeling*, pp.82–93. doi:https://doi.org/10.1007/978-3-319-73600-6_8.

[6] He, K., Zhang, X., Ren, S. and Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.770–778. doi:https://doi.org/10.1109/cvpr.2016.90.

[7] Lecun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), pp.2278–2324. doi:https://doi.org/10.1109/5.726791.

[8] Li, T.-H., Liu, W., Zheng, W.-L. and Lu, B.-L. (2019). *Classification of Five Emotions from EEG and Eye Movement Signals: Discrimination Ability and Stability over Time*. [online] IEEE Xplore. doi:https://doi.org/10.1109/NER.2019.8716943.

[9] Hwang, S., Hong, K., Son, G. and Byun, H. (2019). Learning CNN features from DE features for EEG-based emotion recognition. *Pattern Analysis and Applications*, 23(3), pp.1323–1335. doi:https://doi.org/10.1007/s10044-019-00860-w.

[10] Bashivan, P., Rish, I., Yeasin, M. and Codella, N. (2015). *Learning Representations from EEG with Deep Recurrent-Convolutional Neural Networks*. [online] arXiv.org. Available at: https://arxiv.org/abs/1511.06448 [Accessed 7 Apr. 2019].

[11] Grill-Spector, K., Henson, R. and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, [online] 10(1), pp.14–23. doi:https://doi.org/10.1016/j.tics.2005.11.006.

[12] Suykens, J., Vandewalle, J. Least Squares Support Vector Machine Classifiers. Neural Processing Letters 9, 293–300 (1999). https://doi.org/10.1023/A:1018628609742

[13] Mert, A., Akan, A. Emotion recognition from EEG signals by using multivariate empirical mode decomposition. Pattern Anal Applic 21, 81–89 (2018). https://doi.org/10.1007/s10044-016-0567-6

[14] Zhang, T., Zheng, W., Cui, Z., Zong, Y. and Li, Y. (2019). Spatial–Temporal Recurrent Neural Network for Emotion Recognition. *IEEE Transactions on Cybernetics*, 49(3), pp.839–847. doi:https://doi.org/10.1109/tcyb.2017.2788081.

[15] Ozdemir, M., Degirmenci, M., Izci, E. and Akan, A. (2021) EEG-based emotion recognition with deep convolutional neural networks. Biomedical Engineering / Biomedizinische Technik, Vol. 66 (Issue 1), pp. 43-57. https://doi.org/10.1515/bmt-2019-0306

[16] Wold, S., Esbensen, K. and Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, [online] 2(1-3), pp.37–52. doi:https://doi.org/10.1016/0169-7439(87)80084-9.

[17] McLachlan, G.J. (2005). *Discriminant Analysis and Statistical Pattern Recognition*. John Wiley & Sons.

[18] Balakrishnama S, Ganapathiraju A (1998) Linear discriminant analysis-a brief tutorial. Inst Signal Inf Process 18:1–8

[19] Stoica, P. and Moses, R.L. (2005). *Spectral Analysis of Signals*. Prentice Hall.

[20] Suykens, J., Vandewalle, J. Least Squares Support Vector Machine Classifiers. Neural Processing Letters 9, 293–300 (1999). https://doi.org/10.1023/A:1018628609742

[21] Liu, W., Qiu, J.-L., Zheng, W.-L. and Lu, B.-L. (2022). Comparing Recognition Performance and Robustness of Multimodal Deep Learning Models for Multimodal Emotion Recognition. *IEEE Transactions on Cognitive and Developmental Systems*, 14(2), pp.715–729. doi:https://doi.org/10.1109/tcds.2021.3071170.

[22] Bhavsar, R., Sun, Y., Helian, N., Davey, N., Mayor, D. and Steffert, T. (2018). The Correlation between EEG Signals as Measured in Different Positions on Scalp Varying with Distance. *Procedia Computer Science*, 123, pp.92–97. doi:https://doi.org/10.1016/j.procs.2018.01.015.

[23] Salama, E.S., A.El-Khoribi, R., E.Shoman, M. and A.Wahby, M. (2018). EEG-Based Emotion Recognition using 3D Convolutional Neural Networks. *International Journal of Advanced Computer Science and Applications*, 9(8). doi:https://doi.org/10.14569/ijacsa.2018.090843.

[24] Cohen, M.X. (2014). *Analyzing neural time series data : theory and practice*. Cambridge, Massachusetts: The Mit Press.

[25] Aapo Hyvärinen, Juha Karhunen and Oja, E. (2004). *Independent Component Analysis*. John Wiley & Sons.

[26] Duan, R.-N., Zhu, J.-Y. and Lu, B.-L. (2013). *Differential entropy feature for EEG-based emotion classification*. [online] IEEE Xplore. doi:https://doi.org/10.1109/NER.2013.6695876.

[27] Chowdary, M.K., Anitha, J. and Hemanth, D.J. (2022). Emotion Recognition from EEG Signals Using Recurrent Neural Networks. *Electronics*, 11(15), p.2387. doi:https://doi.org/10.3390/electronics11152387.

[28] Mai, N.-D., Lee, B.-G. and Chung, W.-Y. (2021). Affective Computing on Machine Learning-Based Emotion Recognition Using a Self-Made EEG Device. *Sensors*, 21(15), p.5135. doi:https://doi.org/10.3390/s21155135.

[29] Bo Pan, Wei Zheng, "Emotion Recognition Based on EEG Using Generative Adversarial Nets and Convolutional Neural Network", Computational and Mathematical Methods in Medicine, vol. 2021, Article ID 2520394, 11 pages, 2021. https://doi.org/10.1155/2021/2520394

[30] Tharwat, A. (2018). Independent component analysis: An introduction. *Applied Computing and Informatics*. doi:https://doi.org/10.1016/j.aci.2018.08.006.

[31] Bajaj N (2021) Wavelets for EEG Analysis. Wavelet Theory. IntechOpen. DOI: 10.5772/intechopen.94398.

[32] Cheah, K.H., Nisar, H., Yap, V.V., Lee, C.-Y. and Sinha, G.R. (2021). Optimizing Residual Networks and VGG for Classification of EEG Signals: Identifying Ideal Channels for Emotion Recognition. *Journal of Healthcare Engineering*, 2021, pp.1–14. doi:https://doi.org/10.1155/2021/5599615.

[33] Topic, A. and Russo, M. (2021). Emotion recognition based on EEG feature maps through deep learning network. *Engineering Science and Technology, an International Journal*. doi:https://doi.org/10.1016/j.jestch.2021.03.012.

[34] Jiang, X., Bian, G.-B. and Tian, Z. (2019). Removal of Artefacts from EEG Signals: A Review. *Sensors (Basel, Switzerland)*, [online] 19(5), p.987. doi:https://doi.org/10.3390/s19050987.

[35] Wang, J., Yu, L.-C., Lai, K. and Zhang, X. (2016). *Dimensional Sentiment Analysis Using a Regional CNN-LSTM Model*. [online] pp.225–230. Available at: https://aclanthology.org/P16-2037.pdf.

[36] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J. and Houlsby, N. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv:2010.11929 [cs]*. [online] Available at: https://arxiv.org/abs/2010.11929.

[37] Cheah, K.H., Nisar, H., Yap, V.V., Lee, C.-Y. and Sinha, G.R. (2021). Optimizing Residual Networks and VGG for Classification of EEG Signals: Identifying Ideal Channels for Emotion Recognition. *Journal of Healthcare Engineering*, 2021, pp.1–14. doi:https://doi.org/10.1155/2021/5599615.

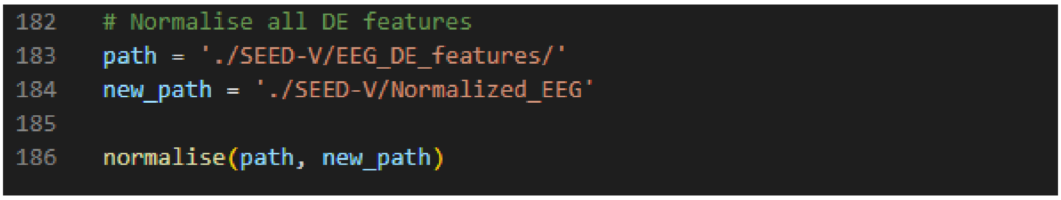
[38] Duan, R.-N., Zhu, J.-Y. and Lu, B.-L. (2013). *Differential entropy feature for EEG-based emotion classification*. [online] IEEE Xplore. doi:https://doi.org/10.1109/NER.2013.6695876.

Appendices

- How to run the codes

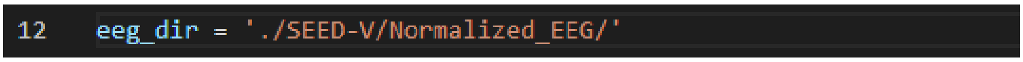
First of all, the dataset (SEED-V) is a public dataset but needs “The License Agreement” to be submitted on their website (https://bcmi.sjtu.edu.cn/home/seed/index.html). And the sequence to run the code is “utils.py”, “DataGenerator.py”, and “ResNet50.py”

"utils.py" contains the codes to create a confusion matrix, normalise DE feature, make a CSV file for DataLoader, and split train and test set. “utils.py” will automatically normalise the data. If the data path is different, change the path of DE features and set a new path to save normalised data.



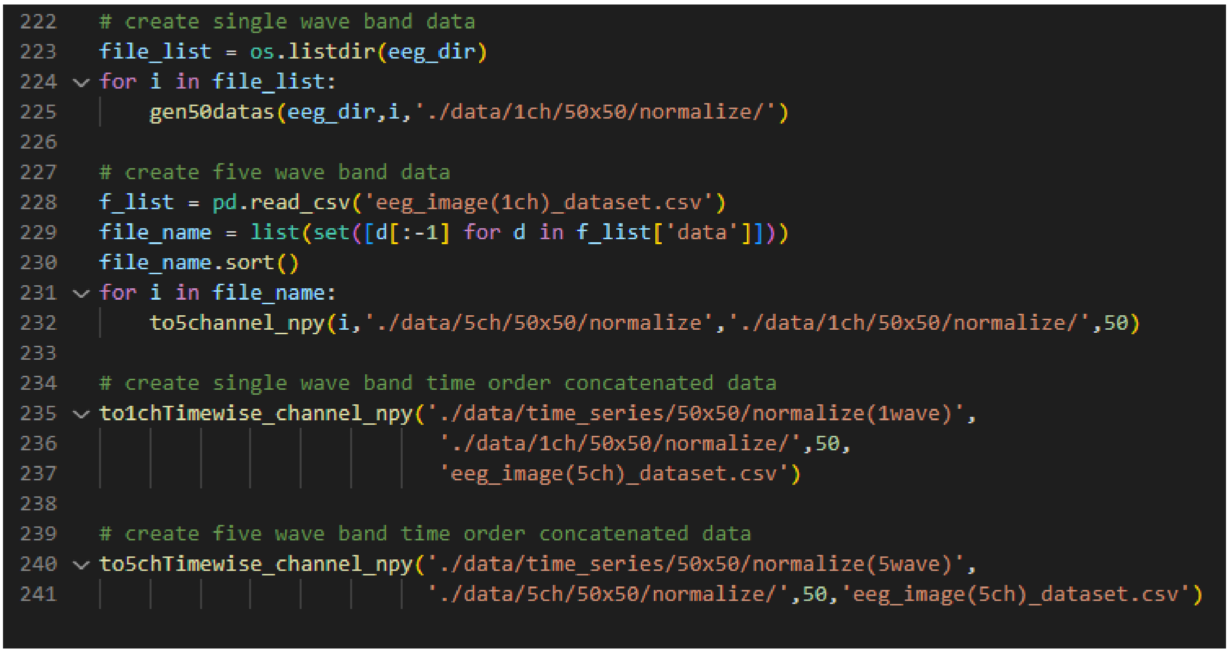
```
# Normalise all DE features
path = './SEED-V/EEG_DE_features/'
new_path = './SEED-V/Normalized_EEG'

normalise(path, new_path)
```

“DataGenerator.py” contains functions to generate EEG images. If the path of normalised data changed, change ‘eeg_dir’.



```
eeg_dir = './SEED-V/Normalized_EEG/'
```

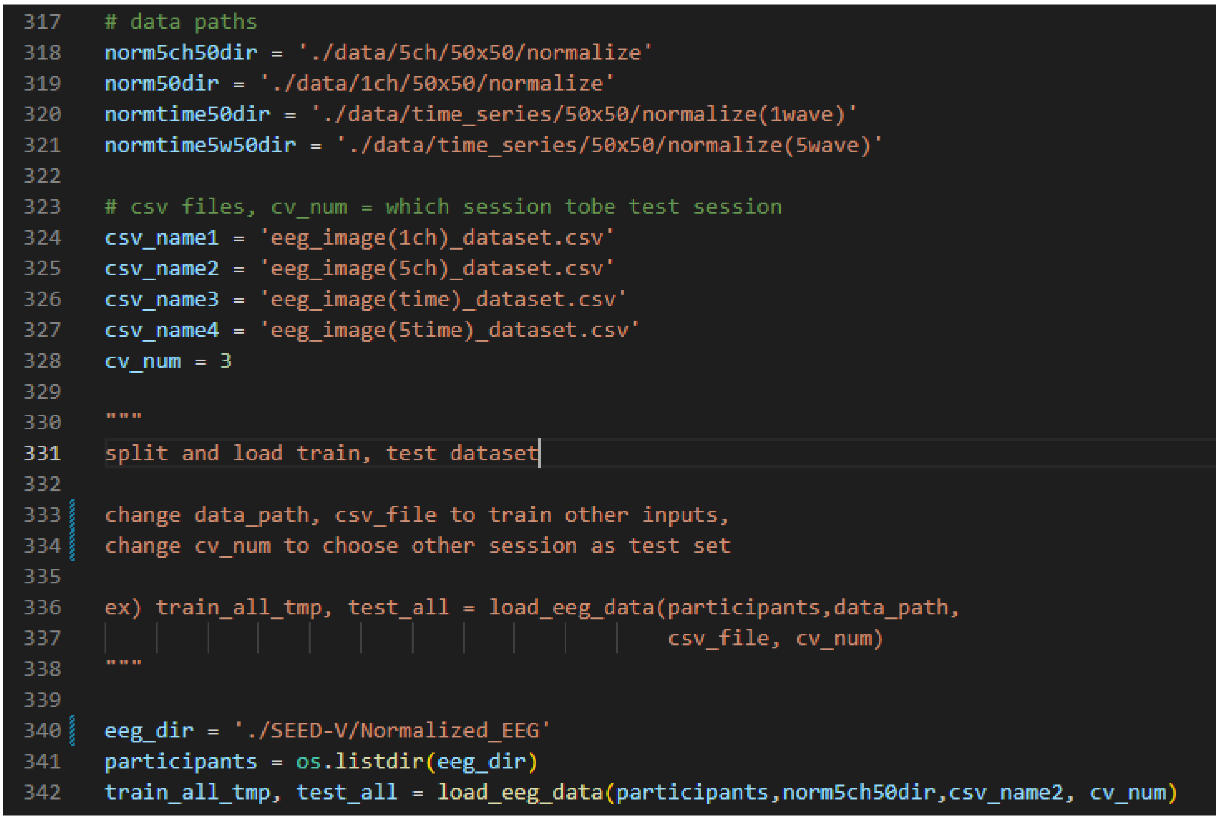
And then change the paths such as “./data/1ch/50x50/normalize/” to save images in other paths if necessary.



```
# create single wave band data
file_list = os.listdir(eeg_dir)
for i in file_list:
    gen50datas(eeg_dir,i,'./data/1ch/50x50/normalize/')

# create five wave band data
f_list = pd.read_csv('eeg_image(1ch)_dataset.csv')
file_name = list(set([d[:-1] for d in f_list['data']]))
file_name.sort()
for i in file_name:
    to5channel_npy(i,'./data/5ch/50x50/normalize','./data/1ch/50x50/normalize/',50)

# create single wave band time order concatenated data
to1chTimewise_channel_npy('./data/time_series/50x50/normalize(1wave)',
                           './data/1ch/50x50/normalize/',50,
                           'eeg_image(5ch)_dataset.csv')

# create five wave band time order concatenated data
to5chTimewise_channel_npy('./data/time_series/50x50/normalize(5wave)',
                          './data/5ch/50x50/normalize/',50,'eeg_image(5ch)_dataset.csv')
```

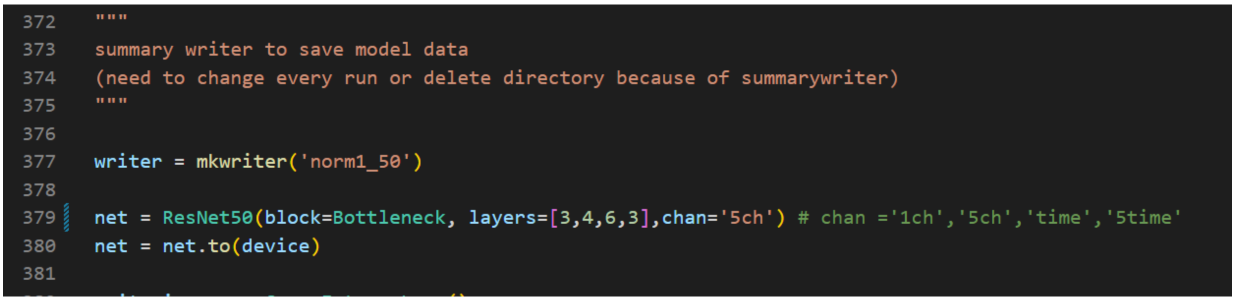
“ResNet50.py” contains the codes to run models.

Change the paths, cv_num, and CSV file names if necessary. cv_num is the test session number. If it is 3, then session 1,2 will be used to train and validate the model and session 3 will be the test set.



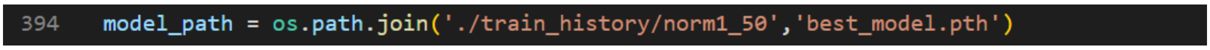
```
# data paths
norm5ch50dir = './data/5ch/50x50/normalize'
norm50dir = './data/1ch/50x50/normalize'
normtime50dir = './data/time_series/50x50/normalize(1wave)'
normtime5w50dir = './data/time_series/50x50/normalize(5wave)'

# csv files, cv_num = which session tobe test session
csv_name1 = 'eeg_image(1ch)_dataset.csv'
csv_name2 = 'eeg_image(5ch)_dataset.csv'
csv_name3 = 'eeg_image(time)_dataset.csv'
csv_name4 = 'eeg_image(5time)_dataset.csv'
cv_num = 3

"""
split and load train, test dataset

change data_path, csv_file to train other inputs,
change cv_num to choose other session as test set

ex) train_all_tmp, test_all = load_eeg_data(participants,data_path,
                                            csv_file, cv_num)
"""

eeg_dir = './SEED-V/Normalized_EEG'
participants = os.listdir(eeg_dir)
train_all_tmp, test_all = load_eeg_data(participants,norm5ch50dir,csv_name2, cv_num)
```

Change the summary writer path every attempt and change ‘chan’ which depends on our input (‘1ch’, ’5ch’, ‘time’, and ‘5time’).



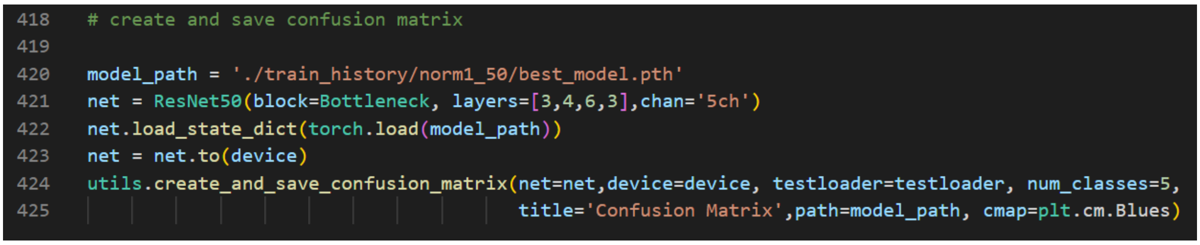
```
"""
summary writer to save model data
(need to change every run or delete directory because of summarywriter)
"""

writer = mkwriter('norm1_50')

net = ResNet50(block=Bottleneck, layers=[3,4,6,3],chan='5ch') # chan ='1ch','5ch','time','5time'
net = net.to(device)

```

Change ‘./train_history/norm1_50’ to ‘./train_history/(your summary writer path)’



```
model_path = os.path.join('./train_history/norm1_50','best_model.pth')
```

And finally, change ‘model_path’ where your model is saved and change ‘chan’ which depends on your input.



```
# create and save confusion matrix

model_path = './train_history/norm1_50/best_model.pth'
net = ResNet50(block=Bottleneck, layers=[3,4,6,3],chan='5ch')
net.load_state_dict(torch.load(model_path))
net = net.to(device)
utils.create_and_save_confusion_matrix(net=net,device=device, testloader=testloader, num_classes=5,
                                       title='Confusion Matrix',path=model_path, cmap=plt.cm.Blues)
```