applied sciences

MDPI

Article

Multi-Region and Multi-Band Electroencephalogram Emotion Recognition Based on Self-Attention and Capsule Network

Sheng Ke [1], Chaoran Ma [1], Wenjie Li [2], Jidong Lv [2] and Ling Zou [1,2,*]

[1] School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou 213159, China; s21150812043@smail.cczu.edu.cn (S.K.); s21150812004@smail.cczu.edu.cn (C.M.)

[2] School of Microelectronics and Control Engineering, Changzhou University, Changzhou 213159, China; lwj@cczu.edu.cn (W.L.); vveaglevv@163.com (J.L.)

* Correspondence: zouling@cczu.edu.cn

Abstract: Research on emotion recognition based on electroencephalogram (EEG) signals is important for human emotion detection and improvements in mental health. However, the importance of EEG signals from different brain regions and frequency bands for emotion recognition is different. For this problem, this paper proposes the Capsule–Transformer method for multi-region and multi-band EEG emotion recognition. First, the EEG features are extracted from different brain regions and frequency bands and combined into feature vectors which are input into the fully connected network for feature dimension alignment. Then, the feature vectors are inputted into the Transformer for calculating the self-attention of EEG features among different brain regions and frequency bands to obtain contextual information. Finally, utilizing capsule networks captures the intrinsic relationship between local and global features. It merges features from different brain regions and frequency bands, adaptively computing weights for each brain region and frequency band. Based on the DEAP dataset, experiments show that the Capsule–Transformer method achieves average classification accuracies of 96.75%, 96.88%, and 96.25% on the valence, arousal, and dominance dimensions, respectively. Furthermore, in emotion recognition experiments conducted on individual brain regions or frequency bands, it was observed that the frontal lobe exhibits the highest average classification accuracy, followed by the parietal, temporal, and occipital lobes. Additionally, emotion recognition performance is superior for high-frequency band EEG signals compared to low-frequency band signals.

Keywords: EEG; emotion recognition; Transformer; capsule network; brain region; frequency band

check for updates

Citation: Ke, S.; Ma, C.; Li, W.; Lv, J.; Zou, L. Multi-Region and Multi-Band Electroencephalogram Emotion Recognition Based on Self-Attention and Capsule Network. *Appl. Sci.* **2024**, *14*, 702. https://doi.org/10.3390/app14020702

Academic Editor: Andrea Prati

Received: 12 December 2023
Revised: 11 January 2024
Accepted: 12 January 2024
Published: 14 January 2024

1. Introduction

Emotion is a psychological state arising in response to internal and external stimuli [1]. It is a complex and subjective experience that significantly influences both work and life [2]. Human emotions can be assessed through dimensional affective models, such as the widely used valence–arousal–dominance model [3]. Valence refers to emotional positivity or negativity, Arousal indicates the intensity of the emotion, and dominance signifies the subjective sense of control [4]. Due to its excellent temporal resolution, resistance to deception, and independence from subjective consciousness [5], EEG signals have found widespread applications in emotional assessment for depression patients [6] and real-time emotional monitoring for drivers [7]. Emotion recognition from EEG signals primarily involves two steps: EEG feature extraction and emotional state classification [8]. EEG features are primarily extracted from the time domain and the frequency domain [8]. Time domain features typically describe the temporal characteristics of the signal and have intuitive physical meanings, such as mean, standard deviation, high-order crossover, kurtosis, etc. [9]. Frequency domain features characterize the frequency distribution properties of the signal, including power spectral density (PSD), differential entropy (DE), conditional entropy (CE), and so on [10]. For EEG emotional state classification, traditional

machine learning methods have been used a lot in this field. Liu et al. [11] constructed a dynamic functional brain network based on the SEED [12] dataset and used Support Vector Machines (SVM) to conduct emotion recognition studies. Veeramallu et al. [13] classified the extracted nonlinear EEG features based on the empirical mode decomposition (EMD) by feeding them into the Random Forest (RF) to classify the emotional states. However, machine learning methods are often deficient in acquiring deep features and have difficulty in capturing complex nonlinear relationships [14]. Deep learning also has more applications in the field of EEG emotion recognition due to its ability to handle complex tasks and large-scale data [15]. Gong et al. [16] integrated temporal, spatial, and spectral information of EEG signals to cascade convolutional neural networks (CNNs) and Transformer in a new way for emotion recognition tasks. Liu et al. [17] proposed a multilevel feature-guided capsule network (MLF-CapsNet) with multichannel emotion recognition based on whole-brain signals, but did not focus on the variability of different brain regions. Khamis et al. [18] utilized a multilayer perceptron (MLP) to perform emotion recognition experiments on different EEG bands, but the study did not address the application of combining multi-band EEG features for emotion recognition. Rupal et al. [19] divided the brain into eight regions and used a 1D convolutional LSTM network for emotion recognition, but they only spliced the signals from different brain regions for the experiments, ignoring the different roles and importance of the different brain regions for emotion recognition.

For the above problems, this paper proposes the Capsule–Transformer method for EEG signal emotion recognition research. Firstly, the EEG signal channels are categorized into four brain regions (frontal lobe, temporal lobe, parietal lobe, and occipital lobe). Within each brain region's EEG signal channels, emotion features are extracted from different frequency bands, and these features are combined into a feature sequence, serving as the input for the Transformer. The Transformer can capture global information from the entire feature vector. Subsequently, emotional features from different brain regions are combined into primary capsules. Through a dynamic routing mechanism [17] of the capsule network, these primary capsules are aggregated into emotion capsules. Dynamic routing mechanism means that when transferring information between different layers of the network, the network can dynamically adjust the weights between different primary capsules, so this process can better capture the potential relationship between local features and global features. In addition, after the dynamic routing mechanism, the mode lengths of different emotion capsules represent the probability of different emotion states, respectively, so it is able to categorize the emotion states through emotion capsules.

The main contributions of this paper can be summarized as follows:

The innovative combination of brain region self-attention, frequency band self-attention and dynamic routing mechanism enables the proposed model to both simultaneously capture information about the differences in emotional features across different brain regions and frequency bands, and further extract the intrinsic connections between signals from different brain regions in emotional activities. This paper proposed a new Capsule–Transformer method for emotion recognition of EEG signals, and explored the role and importance of EEG signals from different brain regions and frequency bands in emotion recognition. Based on this method, we conducted subject-dependent experiments on the DEAP public dataset.

2. Materials and Methods

2.1. EEG Dataset

DEAP [20] is a multimodal standard emotion dataset that collected physiological signals from 32 subjects (16 males, and 16 females), each of whom was physically and psychologically healthy and none of whom had a history of psychiatric disorders, addictive behaviors, or a history of psychotropic substance use. Subjects were instructed to watch 40 one-minute-long music videos, i.e., 40 trails per subject, while physiological signals were simultaneously recorded from 40 electrode channels (32 channels for EEG signals, 8 channels for peripheral physiological signals) following the international 10–20 system [21].

After watching each music video, subjects were asked to rate the 4 dimensions of valence, arousal, dominance, and liking ranging from 1 to 10 to assess the current emotional state. Valence refers to the positive or negative state of the emotion, arousal refers to the high or low intensity of the emotion, dominance refers to the degree of subjective control, and liking refers to the degree of preference for the stimulus or experience [20]. In this paper, 5 is taken as the threshold for a high or low emotional state on each dimension. It is worth noting that the ratings of the 27th subject on the dominance dimension for all experimental signals were consistently above 5. To avoid ineffective training, samples from this subject on the dominance dimension are not utilized. The data format for each subject is illustrated in Table 1. This paper utilizes the preprocessed version provided by the DEAP dataset. In this preprocessed version, the signals are downsampled to 128 Hz, bandpass filtering is applied with a range of 4.0–45.0 Hz, and independent component analysis (ICA) is employed to eliminate electrooculography (EOG) artifacts.

Table 1. The data format for each subject in the DEAP dataset.

Array Name	Array Shape	Array Contents
Data	40 × 40 × 8064	Video/trail × channel × data
Label	40 × 4	Video/trail × label (valence, arousal, dominance, liking)

2.2. Feature Extraction

Differential entropy (DE) is often used as a measure of the complexity of continuous random variables and has been demonstrated to be suitable for EEG emotion recognition studies [22]. The standard definition of DE is as Equation (1):

$$DE(X) = -\int_X f(x)\log(f(x))dx \tag{1}$$

where X represents the EEG sequence and $f(x)$ represents its probability density function. Given that X follows a Gaussian distribution $N(\mu, \sigma^2)$, the DE can be further represented as Equation (2):

$$DE(X) = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\frac{(x-\mu)^2}{2\sigma^2} \log(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\frac{(x-\mu)^2}{2\sigma^2})dx = \frac{1}{2}\log 2\pi e\sigma^2 \tag{2}$$

For EEG samples of T seconds in length, an $N \times C \times T$ dimensional DE feature matrix can be extracted using a 1 s non-overlapping time window, where N denotes the number of EEG channels and C is the number of EEG frequency bands.

Following the international 10–20 system, this paper divides the EEG signal channels of the DEAP dataset into 4 different brain regions. Specifically, the frontal lobe (Fp1, Fp2, F3, F4, Fz, AF3, AF4), temporal lobe (F7, T7, P7, FC5, CP5, F8, T8, P8, FC6, CP6), parietal lobe (P3, P4, Pz, C3, C4, Cz, CP1, CP2, FC1, FC2), and occipital lobe (O1, O2, Oz, PO3, PO4). For each subject, the EEG signal was initially decomposed into 4 frequency bands (θ: 4–7 Hz, α: 8–13 Hz, β: 14–30 Hz, γ: 31–45 Hz). Subsequently, using 1 s non-overlapping time windows on signal channels corresponding to each brain region, DE features are extracted for the four frequency bands. Therefore, after the brain region division and feature extraction, each subject consists of 4 feature matrices with shapes (40, k, 4, 60). Where 40 denotes the number of trails, k denotes the number of electrode channels contained in different brain regions (frontal lobe contains 7 electrodes, parietal lobe contains 10 electrodes, temporal lobe contains 10 electrodes, and occipital lobe contains 5 electrodes), 4 denotes the number of frequency bands, and 60 denotes the number of DE features on each frequency band. After data transformation, the shape of the feature matrix is converted to (2400, 4, k), which corresponds to 2400 labels in the valence, arousal, dominance, and liking dimensions, respectively. It is evident that for each specific brain region, there are 2400 samples, each composed of DE features from 4 frequency bands, and each frequency band contains k

DE features. Furthermore, if no brain region division is performed, each subject contains 1 feature matrix with a shape of (2400, 4, 32), where k represents the 32 EEG signal channels covering the entire brain.

2.3. Capsule–Transformer

The structure of the Capsule–Transformer consists of 3 main modules: Linear layer, Transformer, and Emotion capsules. The Linear layer is used for feature dimension alignment. The Transformer module captures contextual information about emotional features between different brain regions and between different frequency bands within each brain region through a self-attention mechanism. The Emotion capsules module integrates features from different brain regions and frequency bands into a single emotion capsule for emotional state classification. The details are illustrated in Figure 1.

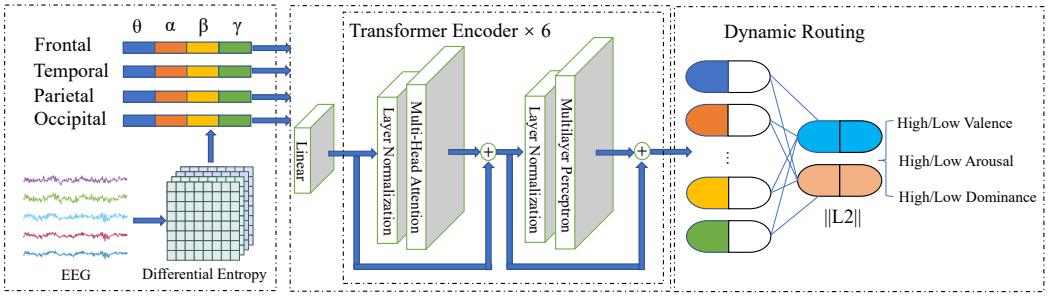

Figure 1. The structure of the proposed Capsule–Transformer method for EEG emotion recognition.

2.3.1. Transformer

The DE features $x_{11}, x_{12}, x_{13}, x_{14} \in R^{d1}$ for frontal multi-band EEG signals, $x_{21}, x_{22}, x_{23}, x_{24} \in R^{d2}$ for temporal multi-band EEG signals, $x_{31}, x_{32}, x_{33}, x_{34} \in R^{d3}$ for parietal multi-band EEG signals, and $x_{31}, x_{32}, x_{33}, x_{34} \in R^{d4}$ for occipital multi-band EEG signals, respectively, are inputted into the fully connected layer to obtain $f_{ij} \in R^{d}(1 \leq i \leq 4, 1 \leq j \leq 4)$, $f = concat(f_{11}, f_{12}, \ldots, f_{44}) \in R^{N \times d}$. Where $d1$, $d2$, $d3$, and $d4$ are the dimensions of the frequency band features within each brain region, and d is the dimension of the feature after dimensional alignment, totaling 16 frequency band features, i.e., $N = 16$. Then, the intrinsic contextual connections of features in different brain regions and frequency bands are learned based on the Transformer encoder structure in the following process.

The length of the feature vectors is first compressed to between 0 and 1 by a layer normalization (LN) structure, and the input feature vectors are mapped to the QKV space: $Q = W_q f$, $K = W_k f$, and $V = W_v f$, where $Q, K, V \in R^{N \times d}$, respectively. Then the scaled dot product is used to calculate the similarity between the frequency band features within the current brain region, between the current brain region and the features of other brain regions, and these similarities are normalized to the attention weights. Finally, the frequency band features of each brain region are weighted and summed with the corresponding attentional weights to obtain the output of the self-attention mechanism, as shown in Equation (3):

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{3}$$

To obtain richer contextual information in different brain regions and frequency bands, this paper uses the multi-head self-attention mechanism (MSA) to generate multiple QKV spaces, and then the attention outputs computed in each QKV space are spliced and linearly transformed thus obtaining richer feature representations, as shown in Equation (4):

$$MSA(Q, K, V) = Concat(Attention_1, \ldots, Attention_h)W^o \tag{4}$$

where W^o is the learnable parameter matrix and h is the number of self-attention heads, the detailed structure of MSA is shown in Figure 2.

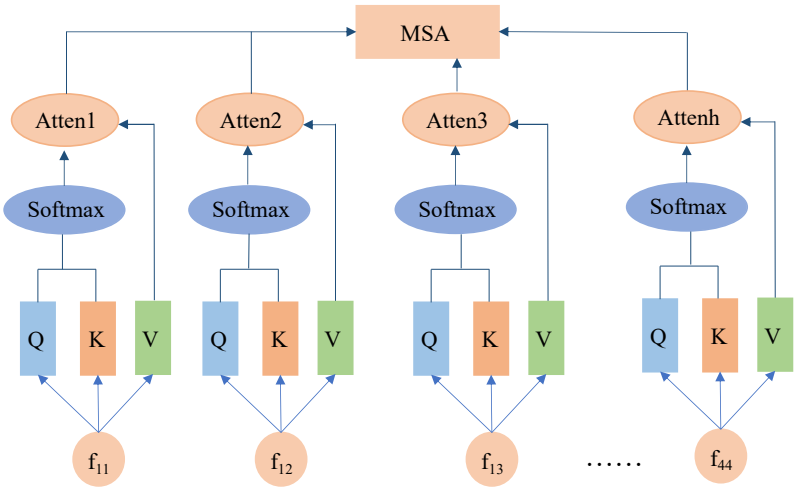

Figure 2. The structure of the multi-head self-attention (MSA).

Based on the above steps, a single Transformer encoder is calculated as shown in Equations (5) and (6):

$$f' = f + MSA(LN(f)) \tag{5}$$

$$f^* = f' + MLP(LN(f')) \tag{6}$$

where f , f^* denote the input and output of the encoder, respectively.

2.3.2. Emotion Capsules

This module, based on a dynamic routing mechanism, merges features from different brain regions and frequency bands into a single feature vector for emotion classification. The computational process of the dynamic routing mechanism is illustrated in Figure 3.

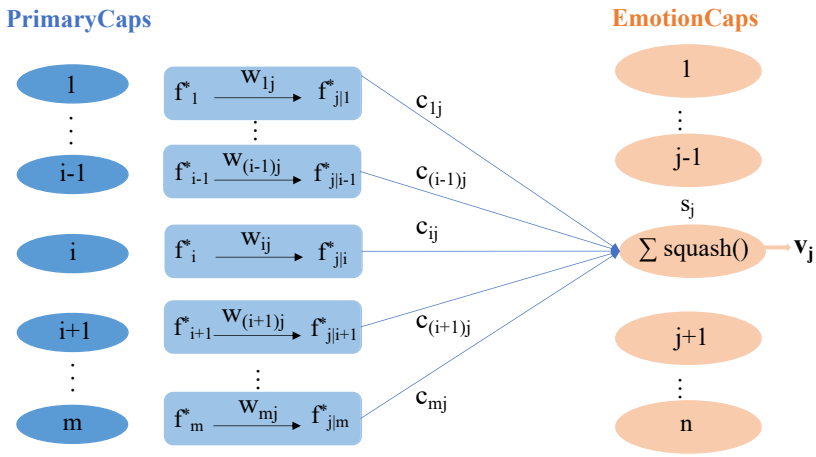

Figure 3. The architecture of dynamic routing mechanisms.

Firstly, the output of the above Transformer module is taken as the primary capsule $f^*{}_i \in R^{D_p}$. Since the primary capsules do not contain information about their relative positions to each other, the matrix $W_{ij} \in R^{Dc \times Dp}$ is needed to record the mapping relationship from the primary capsules to the emotion capsules, where D_c denotes the dimension of the emotion capsules, D_p denotes the dimension of the primary capsules, and $f^*{}_{j|i} \in R^{Dc}(1 \le i \le m, 1 \le j \le n, m = 16, n = 2)$, as shown in Equation (7):

$$f^*_{j|i} = W_{ij} f^*_i \tag{7}$$

Then, since each primary capsule contributes differently to the final emotional state classification, the weight b_{ij} assigned to each primary capsule has an initial value of 0. The sum of the weights of all the primary capsules is made to be 1 by the $softmax$ function. Each primary capsule is represented using feature vectors, and therefore vector addition operation is used to obtain the emotion capsules, as shown in Equation (8):

$$s_j = \sum_i softmax(b_{ij}) f^*_{j|i} \tag{8}$$

The emotion capsules use the mode length to represent the category probability, so it is necessary to compress the mode length to between 0 and 1, which is realized by using the activation function "squashing", and at the same time improves the nonlinear representation of the model, as shown in Equation (9):

$$v_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \frac{s_j}{\|s_j\|} \tag{9}$$

The weights of the primary capsules are updated by calculating the similarity of the inner product of the primary capsules to the emotion capsules, as shown in Equation (10):

$$b_{ij} = b_{ij} + v_j \cdot f^*_{ji} \tag{10}$$

Finally, the model is converged using an iterative approach, and the overall process is shown in Algorithm 1:

Algorithm 1 Capsule–Transformer

Input: Routing iterations r

$f^*_i = Transformer(x_{11}, x_{12}, ..., x_{14}), i = 1, ..., 16$
$f^*_{j|i} = W_{ij} \cdot f^*_i$
for all capsule i in layer l and capsule j in layer (l+1): $b_{ij} = 0$
for r iteration **do**
    for all capsule i in layer l: $c_j = softmax(b_i)$
    for all capsule j in layer (l + 1): $s_j = \sum_i c_{ij} f^*_{j|i}$
    for all capsule j in layer (l + 1): $v_j = squash(s_j)$
    for all capsule i in layer l and capsule j in layer (l + 1): $b_{ij} \leftarrow b_{ij} + f^*_{j|i} \cdot v_j$

Output: $||v_j||$ // L2 norm representing class probability

During the model training phase, margin loss is utilized as the loss function, as shown in Equation (11), where T_k represents the one-hot encoding of the labels. When the network predicts correctly, m^+ is applied to stretch the output vector. In case of a prediction error, m^- is employed to compress the output vector. λ serves as the balancing coefficient, and $||v_k||$ denotes the magnitude of the output vector. Additionally, in subsequent relevant experiments, cross-entropy is employed as the loss function for the Transformer model, as shown in Equation (12):

$$L_k = T_k \max(0, m^+ - \|v_k\|) + \lambda(1 - T_k) \max(0, \|v_k\| - m^-) \tag{11}$$

$$H(p, q) = -\sum_{i=1}^{n} p(x_i) \log(q(x_i)) \tag{12}$$

3. Experiments

3.1. Implementation Details

This paper relies on the valence–arousal–dominance dimensional emotion model and conducts emotion recognition experiments based on a subject-dependent approach. Taking

80% of the sample size of each subject for training and averaging the results across all subjects as a measure of the model metrics for the final results. The number of Transformer encoder layers is set to 6 and the number of self-attention heads h is 12. The dimension of the primary capsules is set to 768, the dimension of the emotion capsules to 32, and the number of routing iterations to 3. The deep learning framework used is PyTorch, and the model is trained in an environment based on CUDA 11.6 and Tesla V100, with a batch size of 64. During the training process, the Adam optimizer is used to update the trainable parameters, and the learning rate is set to 0.001. Experiments have found that the model can be adequately trained when the epoch is set to 50, as shown in Figure 4, which is a graph of the training loss value of subject 1 in the valence dimension with the number of iterations.

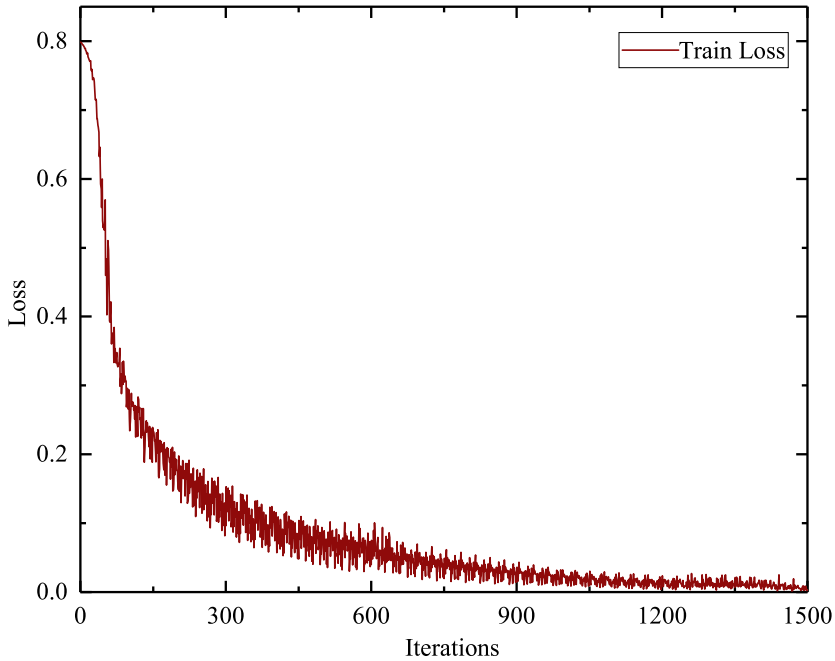

Figure 4. The training loss graph.

3.2. Evaluation Criteria

The model evaluation metrics used in this paper contain accuracy, precision, recall, and F1 score. TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative, respectively. The details are shown as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{13}$$

$$Precision = \frac{TP}{TP + FP} \tag{14}$$

$$Recall = \frac{TP}{TP + FN} \tag{15}$$

$$F1 = \frac{2Precision \times Recall}{Precision + Recall} \tag{16}$$

3.3. Experimental Results and Analysis

3.3.1. Brain Region Division Strategy

This paper validates the impact of brain region division strategy on EEG signal emotion recognition using the Capsule–Transformer method. Without brain region division, a single subject contains only a DE feature matrix of shape (2400, 4, 32), which is used as input

to the model. The results are shown in Table 2. Compared with no brain division, the classification accuracy of the Capsule–Transformer in the three dimensions is improved by 7.23%, 8.42%, and 6.61%, respectively, which indicates that the brain partitioning strategy has a facilitating effect on the EEG emotion recognition.

Table 2. The average recognition results with and without the brain region division strategy.

Label	Method	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
Valence	without division	89.52	89.44	88.79	89.46
	with division	**96.75**	**97.22**	**96.64**	**96.59**
Arousal	without division	88.46	88.59	88.25	89.26
	with division	**96.88**	**96.75**	**97.18**	**96.63**
Dominance	without division	89.64	89.27	88.53	89.29
	with division	**96.25**	**95.93**	**96.36**	**96.18**

3.3.2. Single Brain Region Emotion Recognition

To explore the importance of different brain regions for EEG emotion recognition, this paper explores and analyzes the effect of emotion recognition using signal features from a single brain region based on the Capsule–Transformer method. For each subject, the frontal, temporal, parietal, and occipital lobes were used as inputs to the model using DE feature matrices of shapes (2400, 4, 7), (2400, 4, 10), (2400, 4, 10), and (2400, 4, 5), respectively. The results, as shown in Table 3, showed that the frontal region EEG signals were the best for emotion recognition, followed by the parietal and temporal lobes, and finally the occipital lobe. The frontal EEG signal features achieved 91.88%, 91.25%, and 90.62% classification accuracy on the three dimensional classification tasks, respectively.

Table 3. The comparison of emotion recognition in different brain regions.

Label	Method	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
Valence	Parietal	88.75	88.62	88.37	88.44
	Temporal	86.88	87.06	86.63	86.76
	Occipital	84.38	84.87	83.53	83.92
	Frontal	**91.88**	**91.76**	**91.59**	**91.67**
Arousal	Parietal	88.13	87.15	87.83	87.45
	Temporal	86.25	86.24	86.08	86.14
	Occipital	83.13	83.95	83.08	83.13
	Frontal	**91.25**	**91.17**	**90.52**	**90.65**
Dominance	Parietal	89.38	89.31	89.17	88.97
	Temporal	86.87	88.27	87.33	87.31
	Occipital	80.63	80.82	80.43	80.51
	Frontal	**90.62**	**89.73**	**90.42**	**90.23**

3.3.3. Single Frequency Band Emotion Recognition

In this paper, we further explore the effect of using single frequency band EEG signal features for emotion recognition based on the Capsule–Transformer method. For each subject, the frontal, temporal, parietal, and occipital lobes were used as inputs to the model using only a single band of DE features with the shapes (2400, 1, 7), (2400, 1, 10), (2400, 1, 10), (2400, 1, 5), respectively. The results, as shown in Figure 5, show that the emotion recognition effect of high-frequency EEG signal features is better than that of low-frequency EEG signal features, and the Gamma band reaches 90.59%, 89.79%, and 90.28% on the dimensions of valence, arousal, and dominance, respectively.

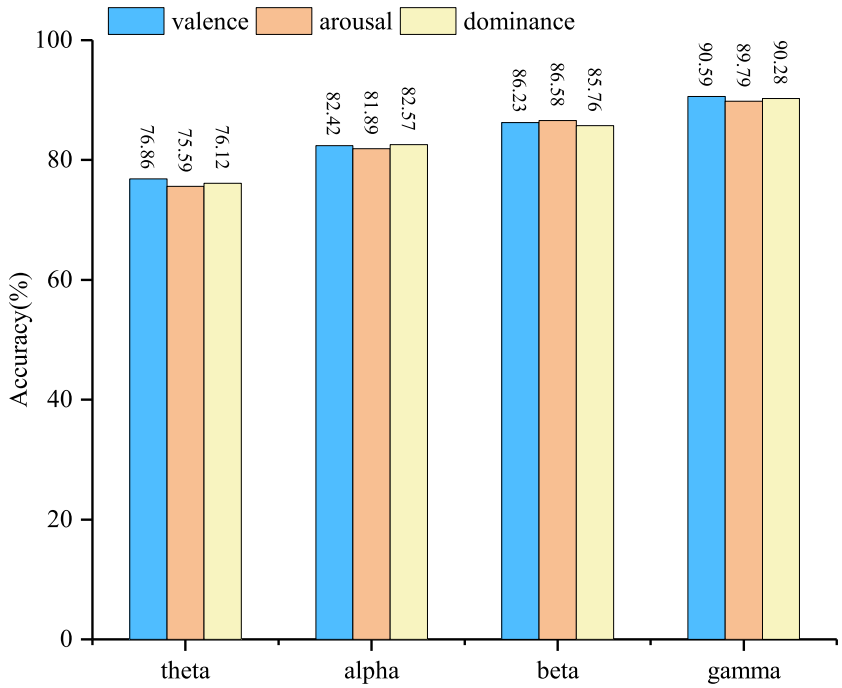

Figure 5. The comparison of emotion recognition in different frequency bands.

3.4. Ablation Experiment

To verify the effectiveness of the Capsule–Transformer method, based on the DEAP dataset, this paper compares the performance of this method with the Transformer and the original CapsNet for emotion recognition. In Transformer, the final emotion classification is based on class token, and cross-entropy is used as the loss function. In the original CapsNet, margin loss is used as the loss function, and the EEG signal features are directly used as the primary capsules after dimensional alignment, and then the emotion capsules are generated through the dynamic routing mechanism for emotion state classification. The results are shown in Table 4. The performance of the Capsule–Transformer method for emotion recognition in all three dimensions is significantly improved compared to the Transformer and capsule network.

Table 4. The comparison of Capsule–Transformer, Transformer, and CapsNet.

Label	Method	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
Valence	CapsNet	84.25	85.37	86.32	84.81
	Transformer	91.36	92.28	91.67	90.89
	Ours	**96.75**	**97.22**	**96.64**	**96.59**
Arousal	CapsNet	85.31	84.27	86.62	85.29
	Transformer	91.67	90.83	92.49	91.22
	Ours	**96.88**	**96.75**	**97.18**	**96.63**
Dominance	CapsNet	84.22	85.63	84.81	84.53
	Transformer	92.77	92.43	92.57	91.59
	Ours	**96.25**	**95.93**	**96.36**	**96.18**

3.5. Comparison with Existing Methods

In this paper, we compare the classification accuracy of the Capsule–Transformer method with seven representative EEG signal emotion recognition methods, including three traditional and four deep learning methods, and all of them are manually reproduced. The three traditional methods are KNN [23], SVM [24], and MLP [25]. The SVM is nonlinear and uses a Gaussian kernel function. The data inputs for KNN and SVM are DE features

extracted from four frequency bands (θ, α, β, γ), and the data inputs for MLP are the raw EEG signals after preprocessing. The four deep learning methods are parallel convolutional recurrent neural network (CNN-RNN) [26], continuous convolutional neural network (Conti-CNN) [27], depthwise convolutional Transformer encoder (DCoT) [28], and the spatial-frequency convolutional self-attention network (SFCSAN) [29]. The CNN-RNN combines the advantages of convolutional neural networks and recurrent neural networks to effectively extract spatial and temporal features of EEG signals. The Conti-CNN combines features of different frequency band signals to construct 3D EEG signal cubes as input and perform emotion recognition. The DCoT combines depthwise convolution and Transformer encoders to capture the dependence of emotion recognition on each EEG channel. The SFCSAN combines spatial and frequency band information to fully use the spatial and frequency domain information of EEG signals for emotion recognition. CNN-RNN takes the preprocessed raw EEG signals as the data input to the model, and the data inputs for Conti-CNN, DCoT, and SFCSAN are DE features extracted from the four frequency bands (θ, α, β, γ), respectively. In addition, all methods used the same data preprocessing method, training parameters, and sliding window length, which facilitated the fairness of the comparison experiments. As shown in Table 5, compared to the traditional KNN, SVM, and MLP methods, the Capsule–Transformer method has a large advantage, with a large increase in the average classification accuracy in all three dimensions. In addition, compared to the other four deep learning methods, the Capsule–Transformer method still shows some advantages. First, compared with the two CNN-based methods (CNN-RNN, Conti-CNN), the Capsule–Transformer method improves the average classification accuracy in each of the three dimensions by about 10%. Second, compared to DCoT using the channel attention mechanism, SFCSAN using convolutional self-attention, the Capsule–Transformer model still achieves better classification performance on the three classification tasks.

Table 5. Average accuracy (%) of different methods on the valence, arousal, and dominance dimensions of the DEAP dataset (mean ± std. dev.).

Method	Inputs	Valence	Arousal	Dominance
KNN	DE features	73.21 ± 4.88	73.28 ± 5.13	74.40 ± 5.29
SVM	DE features	81.66 ± 4.82	81.13 ± 4.18	81.98 ± 4.05
MLP	Raw EEG	83.75 ± 4.75	84.52 ± 5.09	84.49 ± 5.05
CNN-RNN	Raw EEG	84.66 ± 4.20	85.06 ± 4.11	85.03 ± 3.32
Conti-CNN	DE features	86.63 ± 3.20	87.22 ± 3.37	86.61 ± 3.16
DCoT	DE features	90.77 ± 3.09	90.19 ± 3.12	90.62 ± 3.12
SFCSAN	DE features	92.85 ± 2.23	92.68 ± 2.54	92.43 ± 2.15
Ours	DE features	96.75 ± 1.32	96.88 ± 1.46	96.25 ± 1.23

Figures 6–8 give the line graphs of the results of 32 subjects' emotion recognition based on the different methods mentioned above in the three classification tasks of valence, arousal, and dominance, where the horizontal axis is the number of subjects, and the vertical axis is the accuracy rate of emotion classification under each dimension. It is easy to see that Capsule–Transformer is characterized by better classification performance and high robustness compared to other methods.

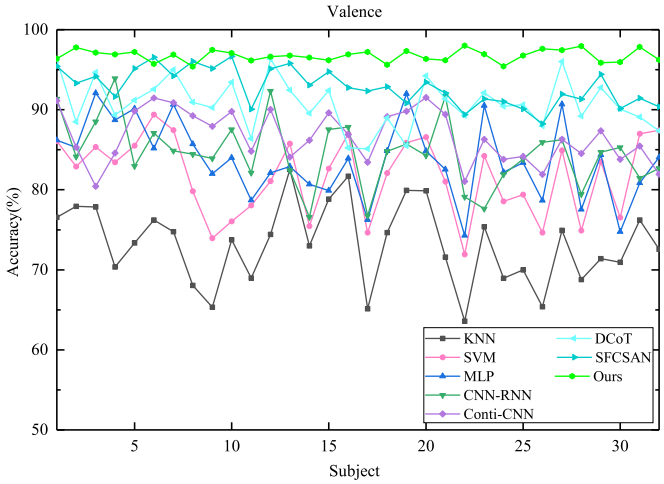

Figure 6. Comparison of the recognition accuracy of each subject using different methods on the valence dimension of the DEAP dataset.

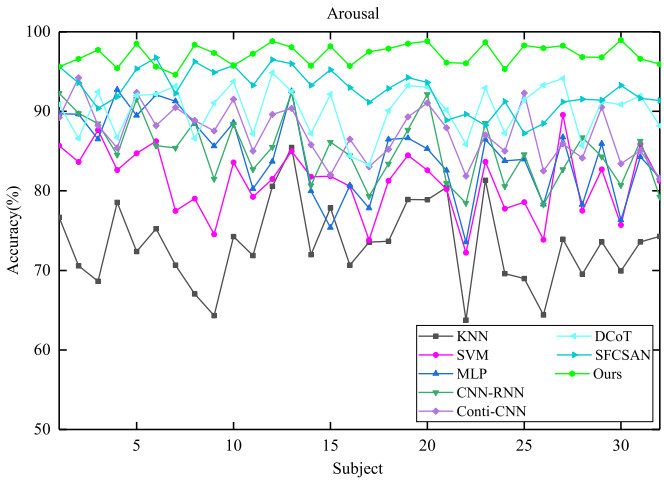

Figure 7. Comparison of the recognition accuracy of each subject using different methods on the arousal dimension of the DEAP dataset.

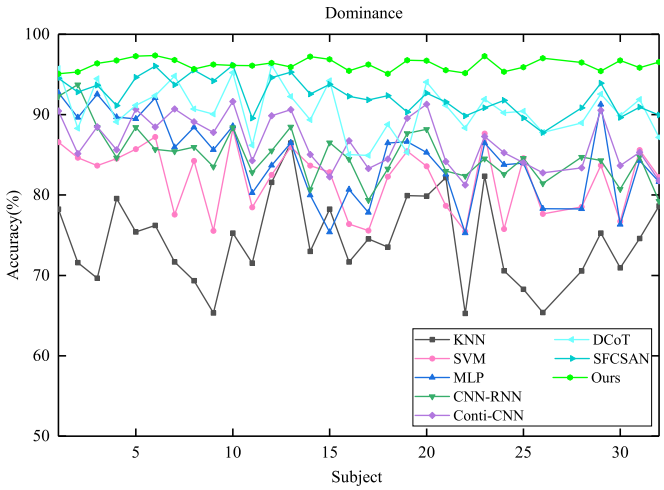

Figure 8. Comparison of the recognition accuracy of each subject using different methods on the dominance dimension of the DEAP dataset.

4. Discussion

Due to the complexity of human emotions, when expressing an emotion, EEG signals from different brain regions respond differently, and there are differences in EEG signals

from different frequency bands [30–32]. Most of the previous emotion recognition studies are based on EEG signals from whole regions [15,27,33]. In this paper, emotional EEG features are extracted from different brain regions and frequency bands, and a more advanced emotion classification performance is realized based on the Capsule–Transformer method. Then, emotion recognition experiments were conducted on each brain region separately, and the results showed that the accuracy of emotion classification of EEG signals in the frontal lobe region was higher than that of other brain regions. Zhong et al. [34] found that the frontal and parietal lobes were significantly more strongly activated compared to other brain regions in all frequency bands, suggesting that emotional processing is more closely correlated with these regions. Rupal et al. [19] similarly concluded that emotional activity varies in different regions of the brain, and conducted a more detailed regional division of EEG signals, which showed large differences in emotion recognition results in different brain regions. To explore the correlation between frequency band features and emotion recognition, this paper conducted emotion recognition experiments on four EEG signal bands, respectively, and the results showed that Gamma band EEG features have the best emotion classification effect; the higher the frequency band, the stronger the correlation of the EEG signals with human emotions. Gao et al. [35] concluded that high-frequency oscillations of EEG signals have a stronger ability to predict emotions compared to low-frequency oscillations. Zheng et al. [36] proposed a deep belief network-based method for detecting key frequency bands in EEG emotion recognition, and concluded that Beta and Gamma band EEG features contain more discriminative information conducive to emotion recognition. This is consistent with the viewpoint of this paper. Accurate and efficient EEG emotion recognition can help diagnose and assess patients with mood disorders such as depression and anxiety. By analyzing EEG signals in key brain regions and frequency bands, physicians can obtain objective information about patients' emotional states more quickly, which can help diagnose and differentiate different emotional disorders more accurately. The findings of this paper can provide information for clinical practice of emotion recognition.

The generation of human emotions is a complex physiological activity, and the responses of emotions to EEG signals often vary greatly from person to person [37]. Currently, the stability of the Capsule–Transformer method in cross-subject EEG emotion recognition studies needs to be improved. In addition, the method relies on EEG signal preprocessing and feature extraction work. Manual feature extraction for EEG signals from different brain regions and frequency bands and inputting them into the network can effectively improve the accuracy of emotion recognition, but this increases the workload and complexity of the experiment.

In the future, we will collect more emotional EEG data and, based on the method, explore the performance of emotion recognition in more complex scenarios, and explore the kinds of EEG features other than DE that are suitable for emotion recognition. In addition, we will also conduct end-to-end EEG emotion recognition research based on the method, which is conducive to reducing the workload and complexity of EEG emotion recognition experiments.

5. Conclusions

In this paper, a novel Capsule–Transformer method is proposed for multi-region, multi-band EEG signal emotion recognition research. The method utilizes the Transformer's ability to focus on the contextual information of EEG signal features in different brain regions and frequency bands, as well as the capsule network's ability to capture the intrinsic relationship between local and global features, to achieve a more advanced emotion recognition performance compared to most methods. Based on the DEAP dataset, we divided the EEG signal channels according to four different brain regions (frontal lobe, temporal lobe, parietal lobe, and occipital lobe) and extracted the EEG features from different signal bands for the experiments, and the results proved the effectiveness of the combination of brain region self-attention, frequency band self-attention, and the dynamic

routing mechanism for the recognition of emotions in EEG. In addition, the experimental results on single brain regions and single frequency bands show that the frontal lobe region has the highest accuracy in emotion recognition, followed by the parietal lobe, temporal lobe, and occipital lobe, and the Gamma frequency band has higher accuracy than Beta, Alpha, and Theta. it is clear that the research of EEG emotion recognition should pay more attention to the frontal lobe, parietal lobe region, and the high-frequency band EEG signals.

Author Contributions: Conceptualization, L.Z. and S.K.; methodology, S.K.; software, S.K.; validation, C.M., W.L. and J.L.; formal analysis, L.Z.; investigation, C.M.; resources, W.L.; data curation, S.K.; writing—original draft preparation, S.K.; writing—review and editing, S.K.; visualization, S.K.; supervision, W.L.; project administration, J.L.; funding acquisition, L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the project of Jiangsu Key Research and Development Plan (BE2021012-2 and BE2021012-5), Changzhou Science and Technology Bureau Plan (CE20225034), Key Laboratory of Brain Machine Collaborative Intelligence Foundation of Zhejiang Province (2020E10010-04), and Human–Machine Intelligence and Interaction International Joint Laboratory Project.

Institutional Review Board Statement: This study was approved by the administrators of the public dataset used in the article.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available dataset was analyzed in this study. This data can be found here: http://www.eecs.qmul.ac.uk/mmv/datasets/deap/, accessed on 28 December 2023.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Othman, M.; Wahab, A.; Karim, I.; Dzulkifli, M.A.; Alshaikli, I.F.T. EEG emotion recognition based on the dimensional models of emotions. *Procedia-Soc. Behav. Sci.* **2013**, *97*, 30–37. [CrossRef]
2. Suhaimi, N.S.; Mountstephens, J.; Teo, J. EEG-based emotion recognition: A state-of-the-art review of current trends and opportunities. *Comput. Intell. Neurosci.* **2020**, *2020*, 8875426. [CrossRef] [PubMed]
3. Trull, T.J.; Durrett, C.A. Categorical and dimensional models of personality disorder. *Annu. Rev. Clin. Psychol.* **2005**, *1*, 355–380. [CrossRef] [PubMed]
4. Wang, X.W.; Nie, D.; Lu, B.L. Emotional state classification from EEG data using machine learning approach. *Neurocomputing* **2014**, *129*, 94–106. [CrossRef]
5. Li, X.; Song, D.; Zhang, P.; Yu, G.; Hou, Y.; Hu, B. Emotion recognition from multi-channel EEG data through convolutional recurrent neural network. In Proceedings of the 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Shenzhen, China, 15–18 December 2016; pp. 352–359.
6. Duan, L.; Duan, H.; Qiao, Y.; Sha, S.; Qi, S.; Zhang, X.; Huang, J.; Huang, X.; Wang, C. Machine learning approaches for MDD detection and emotion decoding using EEG signals. *Front. Hum. Neurosci.* **2020**, *14*, 284. [CrossRef]
7. Fan, X.A.; Bi, L.Z.; Chen, Z.L. Using EEG to detect drivers' emotion with Bayesian Networks. In Proceedings of the 2010 International Conference on Machine Learning and Cybernetics, Qingdao, China, 11–14 July 2010; Volume 3, pp. 1177–1181.
8. Ackermann, P.; Kohlschein, C.; Bitsch, J.A.; Wehrle, K.; Jeschke, S. EEG-based automatic emotion recognition: Feature extraction, selection and classification methods. In Proceedings of the 2016 IEEE 18th International Conference on E-Health Networking, Applications and Services (Healthcom), Munich, Germany, 14–16 September 2016; pp. 1–6.
9. Al-Nafjan, A.; Hosny, M.; Al-Ohali, Y.; Al-Wabil, A. Review and classification of emotion recognition based on EEG brain–computer interface system research: A systematic review. *Appl. Sci.* **2017**, *7*, 1239. [CrossRef]
10. Zhang, Y.; Ji, X.; Zhang, S. An approach to EEG-based emotion recognition using combined feature extraction method. *Neurosci. Lett.* **2016**, *633*, 152–157. [CrossRef]
11. Liu, X.; Li, T.; Tang, C.; Xu, T.; Chen, P.; Bezerianos, A.; Wang, H. Emotion recognition and dynamic functional connectivity analysis based on EEG. *IEEE Access* **2019**, *7*, 143293–143302. [CrossRef]
12. Wagh, K.P.; Vasanth, K. Performance evaluation of multi-channel electroencephalogram signal (EEG) based time frequency analysis for human emotion recognition. *Biomed. Signal Process. Control* **2022**, *78*, 103966. [CrossRef]
13. Veeramallu, G.K.P.; Anupalli, Y.; kumar Jilumudi, S.; Bhattacharyya, A. EEG based automatic emotion recognition using EMD and random forest classifier. In Proceedings of the 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 6–8 July 2019; pp. 1–6.
14. Houssein, E.H.; Hammad, A.; Ali, A.A. Human emotion recognition from EEG-based brain–computer interface using machine learning: A comprehensive review. *Neural Comput. Appl.* **2022**, *34*, 12527–12557. [CrossRef]

15. Topic, A.; Russo, M. Emotion recognition based on EEG feature maps through deep learning network. *Eng. Sci. Technol. Int. J.* **2021**, *24*, 1442–1454. [CrossRef]
16. Gong, L.; Li, M.; Zhang, T.; Chen, W. EEG emotion recognition using attention-based convolutional Transformer neural network. *Biomed. Signal Process. Control* **2023**, *84*, 104835. [CrossRef]
17. Liu, Y.; Ding, Y.; Li, C.; Cheng, J.; Song, R.; Wan, F.; Chen, X. Multi-channel EEG-based emotion recognition via a multi-level features guided capsule network. *Comput. Biol. Med.* **2020**, *123*, 103927. [CrossRef] [PubMed]
18. Alarabi Aljribi, K. A comparative analysis of frequency bands in eeg based emotion recognition system. In Proceedings of the 7th International Conference on Engineering & MIS 2021, New York, NY, USA, 28 September 2021; pp. 1–7.
19. Agarwal, R.; Andujar, M.; Canavan, S. Classification of emotions using eeg activity associated with different areas of the brain. *Pattern Recognit. Lett.* **2022**, *162*, 71–80. [CrossRef]
20. Koelstra, S.; Muhl, C.; Soleymani, M.; Lee, J.S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. Deap: A database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput.* **2011**, *3*, 18–31. [CrossRef]
21. Herwig, U.; Satrapi, P.; Schönfeldt-Lecuona, C. Using the international 10-20 EEG system for positioning of transcranial magnetic stimulation. *Brain Topogr.* **2003**, *16*, 95–99. [CrossRef]
22. Duan, R.N.; Zhu, J.Y.; Lu, B.L. Differential entropy feature for EEG-based emotion classification. In Proceedings of the 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER), San Diego, CA, USA, 6–8 November 2013; pp. 81–84.
23. Li, M.; Xu, H.; Liu, X.; Lu, S. Emotion recognition from multichannel EEG signals using K-nearest neighbor classification. *Technol. Health Care* **2018**, *26*, 509–519. [CrossRef]
24. Mehmood, R.M.; Lee, H.J. Emotion classification of EEG brain signal using SVM and KNN. In Proceedings of the 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, Italy, 29 June–3 July 2015; pp. 1–5.
25. Yaacob, H.; Abdul, W.; Kamaruddin, N. Classification of EEG signals using MLP based on categorical and dimensional perceptions of emotions. In Proceedings of the 2013 5th International Conference on Information and Communication Technology for the Muslim World (ICT4M), Rabat, Morocco, 26–27 March 2013; pp. 1–6.
26. Yang, Y.; Wu, Q.; Qiu, M.; Wang, Y.; Chen, X. Emotion recognition from multi-channel EEG through parallel convolutional recurrent neural network. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–7.
27. Yang, Y.; Wu, Q.; Fu, Y.; Chen, X. Continuous convolutional neural network with 3D input for EEG-based emotion recognition. In Proceedings of the Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, 13–16 December 2018; Proceedings, Part VII 25; Springer: Berlin/Heidelberg, Germany, 2018; pp. 433–443.
28. Guo, J.Y.; Cai, Q.; An, J.P.; Chen, P.Y.; Ma, C.; Wan, J.H.; Gao, Z.K. A Transformer based neural network for emotion recognition and visualizations of crucial EEG channels. *Phys. Stat. Mech. Its Appl.* **2022**, *603*, 127700. [CrossRef]
29. Li, D.; Xie, L.; Chai, B.; Wang, Z.; Yang, H. Spatial-frequency convolutional self-attention network for EEG emotion recognition. *Appl. Soft Comput.* **2022**, *122*, 108740. [CrossRef]
30. Dimitrakopoulos, G.N.; Kakkos, I.; Dai, Z.; Lim, J.; deSouza, J.J.; Bezerianos, A.; Sun, Y. Task-independent mental workload classification based upon common multi-band EEG cortical connectivity. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2017**, *25*, 1940–1949. [CrossRef]
31. Wang, H.; Wu, X.; Yao, L. Identifying cortical brain directed connectivity networks from high-density EEG for emotion recognition. *IEEE Trans. Affect. Comput.* **2020**, *13*, 1489–1500. [CrossRef]
32. Syrjälä, J.; Basti, A.; Guidotti, R.; Marzetti, L.; Pizzella, V. Decoding working memory task condition using magnetoencephalography source level long-range phase coupling patterns. *J. Neural Eng.* **2021**, *18*, 016027. [CrossRef] [PubMed]
33. Cui, H.; Liu, A.; Zhang, X.; Chen, X.; Wang, K.; Chen, X. EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network. *Knowl.-Based Syst.* **2020**, *205*, 106243. [CrossRef]
34. Zhong, P.; Wang, D.; Miao, C. EEG-based emotion recognition using regularized graph neural networks. *IEEE Trans. Affect. Comput.* **2020**, *13*, 1290–1301. [CrossRef]
35. Gao, Z.; Cui, X.; Wan, W.; Gu, Z. Recognition of emotional states using multiscale information analysis of high frequency EEG oscillations. *Entropy* **2019**, *21*, 609. [CrossRef]
36. Zheng, W.L.; Guo, H.T.; Lu, B.L. Revealing critical channels and frequency bands for emotion recognition from EEG with deep belief network. In Proceedings of the 2015 7th International IEEE/EMBS Conference on Neural Engineering (NER), Montpellier, France, 22–24 April 2015; pp. 154–157.
37. Wang, Y.; Liu, J.; Ruan, Q.; Wang, S.; Wang, C. Cross-subject EEG emotion classification based on few-label adversarial domain adaption. *Expert Syst. Appl.* **2021**, *185*, 115581. [CrossRef]